Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, publishes research manuscripts out of all mathematical fields and is refereed in the traditional manner. It is indexed in Mathematical Reviews, Science Citation Index Expanded, Zentralblatt für Mathematik.

Manuscripts should be submitted as TeX-files by e-mail to one of the editors. Hints for manuscript preparation can be found under the following web address.

http://www.math.uni-bielefeld.de/documenta

Geschäftsführende Herausgeber / Managing Editors:

Ulf Rehmann (tech.), Bielefeld rehmann@math.uni-bielefeld.de
Stefan Teufel, Tübingen stefan.teufel@uni-tuebingen.de
Otmar Venjakob, Heidelberg venjakob@mathi.uni-heidelberg.de

Herausgeber / Editors:

Christian Bär, Potsdam bae@math.unipotsdam.de
Don Blasius, Los Angeles blasius@math.ucla.edu
Joachim Cuntz, Münster cuntz@math.uni-muenster.de
Patrick Delorme, Marseille delorme@iml.univ-mrs.fr
Gavril Farkas, Berlin (HU) farkas@math.hu-berlin.de
Edward Frenkel, Berkeley frenkel@math.berkeley.edu
Friedrich Götze, Bielefeld goetze@math.uni-bielefeld.de
Ursula Hamenstädt, Bonn ursula@math.uni-bonn.de
Lars Hesselholt, Cambridge, MA (MIT) larsh@math.nagoya-u.ac.jp
Max Karoubi, Paris karoubi@math.jussieu.fr
Stephen Lichtenbaum Stephen_Lichtenbaum@brown.edu
Alfred K. Louis, Saarbrücken louis@num.uni-sb.de
Eckhard Meinrenken, Toronto mein@math.toronto.edu
Alexander S. Merkurjev, Los Angeles merkurjev@math.ucla.edu
Anil Nerode, Ithaca anil@math.cornell.edu
Thomas Peternell, Bayreuth Thomas.Peternell@uni-bayreuth.de
Takeshi Saito, Tokyo t-saito@ms.u-tokyo.ac.jp
Peter Schneider, Münster pschnet@math.uni-muenster.de
Stefan Schwede, Bonn schwede@math.uni-bonn.de
Heinz Siedentop, München (LMU) h.s@hu.de
Wolfgang Soergel, Freiburg soergel@mathematik.uni-freiburg.de

ISSN 1431-0635 (Print), ISSN 1431-0643 (Internet)

Documenta Mathematica is a Leading Edge Partner of SPARC, the Scholarly Publishing and Academic Resource Coalition of the Association of Research Libraries (ARL), Washington DC, USA.
MATTEO LONGO, MARC-HUBERT NICOLE
The \(\Lambda\)-Adic Shimura-Shintani-Waldspurger Correspondence 1–21

HÉLÈNE ENSAULT, ADRIAN LANDER
On a Positive Equicharacteristic Variant of the \(p\)-Curvature Conjecture 23–50

PABLO PELAEZ
Birational Motivc Homotopy Theories and the Slice Filtration 51–70

ANDRÉ JOYAL AND JOACHIM KOCK
Coherence for Weak Units 71–110

INDRANIL BISWAS
A Criterion for Flatness of Sections of Adjoint Bundle of a Holomorphic Principal Bundle over a Riemann Surface 111–120

BRIAN OSSERMAN AND SAM PAYNE
Lifting Tropical Intersections 121–175

K. SATO
Cycle Classes for \(p\)-Adic Étale Tate Twists and the Image of \(p\)-Adic Regulators 177–247

CHRISTOF GEISS, BERNARD LECLERC, AND JAN SCHRÖER
Factorial Cluster Algebras 249–274

J. KAMPF AND M. KIDERLEN
Large Parallel Volumes of Finite and Compact Sets in \(d\)-Dimensional Euclidean Space 275–295

FENG-YU WANG
Transportation-Cost Inequalities on Path Space over Manifolds with Boundary 297–322

DRAGOS OPREA
Framed Sheaves over Treefolds and Symmetric Obstruction Theories 323–342

WALID ABOU SALEM, THOMAS CHEN, AND VITALI VOGALTER
On the Generalized Semi-Relativistic Schrödinger-Poisson System in \(\mathbb{R}^n\) 343–357

DANNY NEFTIN AND UZI VISHNE
Realizability and Admissibility under Extension of \(p\)-Adic and Number Fields 359–382
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eva Bayer–Fluckiger, Nivedita Bhaskhar, and Raman Parimala</td>
<td>383–392</td>
</tr>
<tr>
<td>HASSE PRINCIPLE FOR G-QUADRATIC FORMS</td>
<td></td>
</tr>
<tr>
<td>E. Bayer–Fluckiger and R. Parimala</td>
<td>393–394</td>
</tr>
<tr>
<td>ERRATUM FOR &quot;GALOIS ALGEBRAS, HASSE PRINCIPLE, AND INDUCTION–RESTRICTION METHODS&quot;</td>
<td></td>
</tr>
<tr>
<td>cf. Documenta Math. 16 (2011), 677–707</td>
<td></td>
</tr>
<tr>
<td>Francois Brunault</td>
<td>395–412</td>
</tr>
<tr>
<td>ON ZAGIER’S CONJECTURE FOR BASE CHANGES OF ELLIPTIC CURVES</td>
<td></td>
</tr>
<tr>
<td>Vladimir Chernousov, Alberto Elduque, Max-Albert Knus, and Jean-Pierre Tignol</td>
<td>413–468</td>
</tr>
<tr>
<td>ALGEBRAIC GROUPS OF TYPE D₄, TRIALITY, AND COMPOSITION ALGEBRAS</td>
<td></td>
</tr>
<tr>
<td>Kristian Ranestad and Frank-Olaf Schreyer</td>
<td>469–505</td>
</tr>
<tr>
<td>THE VARIETY OF POLAR SIMPLICES</td>
<td></td>
</tr>
<tr>
<td>D. Huybrechts and M. Kemeny</td>
<td>507–517</td>
</tr>
<tr>
<td>STABLE MAPS AND CHOW GROUPS</td>
<td></td>
</tr>
<tr>
<td>Stephane Garnier and Tilmann Wurzbacher</td>
<td>519–545</td>
</tr>
<tr>
<td>INTEGRATION OF VECTOR FIELDS ON SMOOTH AND HOLomorphic SUPERMANIFOLDS</td>
<td></td>
</tr>
<tr>
<td>Victor Abrashkin</td>
<td>547–619</td>
</tr>
<tr>
<td>PROJECTIVE VARIETIES WITH BAD SEMI-stABLE REDUCTION AT 3 ONLY</td>
<td></td>
</tr>
<tr>
<td>Meng Fai Lim and Romyar T. Sharifi</td>
<td>621–678</td>
</tr>
<tr>
<td>NEKOVÁŘ DUALITY OVER p-ADIC LIE EXTENSIONS OF GLOBAL FIELDS</td>
<td></td>
</tr>
<tr>
<td>B. Pasquier and N. Perrin</td>
<td>679–706</td>
</tr>
<tr>
<td>ELLIPTIC CURVES ON SOME HOMOGENEOUS SPACES</td>
<td></td>
</tr>
<tr>
<td>Vladimir Dotsenko and Anton Khoroshkin</td>
<td>707–747</td>
</tr>
<tr>
<td>QUILLEN HOMOLOGY FOR OPERADS VIA GRÖBNER BASES</td>
<td></td>
</tr>
<tr>
<td>Reza Taleb</td>
<td>749–783</td>
</tr>
<tr>
<td>AN EQUIVARIANT MAIN CONJECTURE IN IWASAWA THEORY AND THE COATES-SINNOTT CONJECTURE</td>
<td></td>
</tr>
<tr>
<td>Jan Möllers</td>
<td>785–855</td>
</tr>
<tr>
<td>A GEOMETRIC QUANTIZATION OF THE KOSTANT–SEKIGUCHI CORRESPONDENCE FOR SCALAR TYPE UNITARY HIGHEST WEIGHT REPRESENTATIONS</td>
<td></td>
</tr>
<tr>
<td>Tom Leinster</td>
<td>857–905</td>
</tr>
<tr>
<td>THE MAGNITUDE OF METRIC SPACES</td>
<td></td>
</tr>
</tbody>
</table>
Marc Levine
Convergence of Voevodsky’s Slice Tower 907–941

Alexandru Buium
p-Jets of Finite Algebras, I: p-Divisible Groups 943–969

Alexandru Buium
p-Jets of Finite Algebras, II: p-Typical Witt Rings 971–996

Shu Sasaki
On Artin Representations and Nearly Ordinary Hecke Algebras over Totally Real Fields 997–1038

Bjorn Poonen
Extending Self-Maps to Projective Space over Finite Fields 1039–1044

Ian Agol
The Virtual Haken Conjecture 1045–1087

I. Bauer and F. Catanese
Burniat Surfaces III: Deformations of Automorphisms and Extended Burniat Surfaces 1089–1136

P. P. Varjú
Random Walks in Compact Groups 1137–1175

Indranil Biswas, Tomás L. Gómez, and Norbert Hoffmann
Torelli Theorem for the Deligne–Hitchin Moduli Space, II 1177–1189

Nadir Matringe
Essential Whittaker Functions for $GL(n)$ 1191–1214

Lars Kindler
Evidence for a Generalization of Gieseker’s Conjecture on Stratified Bundles in Positive Characteristic 1215–1242

Kazunori Noguchi
The Zeta Function of a Finite Category 1243–1274

Masoud Kamgarpour and Travis Schedler
Ramified Satake Isomorphisms for Strongly Parabolic Characters 1275–1300

Estanislao Herscovici
On the Multi-Koszul Property for Connected Algebras 1301–1347

M. Brannan
Reduced Operator Algebras of Trace-Preserving Quantum Automorphism Groups 1349–1402
Siddarth Sankaran
Unitary Cycles on Shimura Curves and the Shimura Lift I

Indranil Biswas and S. Senthamarai Kannan
Stability of the Tangent Bundle of the Wonderful Compactification of an Adjoint Group

Indranil Biswas and Viktoria Heu
Holomorphic Connections on Filtered Bundles over Curves

Volker Bach, Sébastien Breteaux, Tim Tzaneteas
Minimization of the Energy of the Nonrelativistic One-Electron Pauli-Fierz Model over Quasifree States

Charles Vial
Algebraic Cycles and Fibrations

Olivier Haution
Invariants of Upper Motives

W. Młotkowski, K. A. Penson, K. Życzkowski
Densities of the Raney Distributions

Julian Kellerhals, Nicolas Monod, and Mikael Rørdam
Non-Supramenable Groups Acting on Locally Compact Spaces
The $\Lambda$-Adic Shimura-Shintani-Waldspurger Correspondence

Matteo Longo, Marc-Hubert Nicole

Received: October 25, 2012
Communicated by Takeshi Saito

Abstract. We generalize the $\Lambda$-adic Shintani lifting for $GL_2(\mathbb{Q})$ to indefinite quaternion algebras over $\mathbb{Q}$.

2010 Mathematics Subject Classification: Primary 11F37, 11F30, 11F85

1. Introduction

Langlands’s principle of functoriality predicts the existence of a staggering wealth of transfers (or lifts) between automorphic forms for different reductive groups. In recent years, attempts at the formulation of $p$-adic variants of Langlands’s functoriality have been articulated in various special cases. We prove the existence of the Shimura-Shintani-Waldspurger lift for $p$-adic families. More precisely, Stevens, building on the work of Hida and Greenberg-Stevens, showed in [21] the existence of a $\Lambda$-adic variant of the classical Shintani lifting of [20] for $GL_2(\mathbb{Q})$. This $\Lambda$-adic lifting can be seen as a formal power series with coefficients in a finite extension of the Iwasawa algebra $\Lambda := \mathbb{Z}_p[[X]]$ equipped with specialization maps interpolating classical Shintani lifts of classical modular forms appearing in a given Hida family.

Shimura in [19], resp. Waldspurger in [22] generalized the classical Shimura-Shintani correspondence to quaternion algebras over $\mathbb{Q}$, resp. over any number field. In the $p$-adic realm, Hida ([7]) constructed a $\Lambda$-adic Shimura lifting, while Ramsey ([17]) (resp. Park [12]) extended the Shimura (resp. Shintani) lifting to the overconvergent setting.

In this paper, motivated by ulterior applications to Shimura curves over $\mathbb{Q}$, we generalize Stevens’s result to any non-split rational indefinite quaternion algebra $B$, building on work of Shimura [19] and combining this with a result of Longo-Vigni [9]. Our main result, for which the reader is referred to Theorem 3.8 below, states the existence of a formal power series and specialization maps interpolating Shimura-Shintani-Waldspurger lifts of classical forms in a given
$p$-adic family of automorphic forms on the quaternion algebra $B$. The $Λ$-adic variant of Waldspurger’s result appears computationally challenging (see remark in [13 Intro.]), but it seems within reach for real quadratic fields (cf. [13]).

As an example of our main result, we consider the case of families with trivial character. Fix a prime number $p$ and a positive integer $N$ such that $p \nmid N$. Embed the set $\mathbb{Z}_{\geq 2}$ of integers greater or equal to 2 in $\text{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times)$ by sending $k \in \mathbb{Z}_{\geq 2}$ to the character $x \mapsto x^{k-2}$. Let $f_\infty$ be an Hida family of tame level $N$ passing through a form $f_0$ of level $\Gamma_0(Np)$ and weight $k_0$. There is a neighborhood $U$ of $k_0$ in $\text{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times)$ such that, for any $k \in \mathbb{Z}_{\geq 2} \cap U$, the weight $k$ specialization of $f_\infty$ gives rise to an element $f_k \in S_k(\Gamma_0(Np))$.

Fix a factorization $N = MD$ with $D > 1$ a square-free product of an even number of primes and $\gcd(M, D) = 1$ (we assume that such a factorization exists). Applying the Jacquet-Langlands correspondence we get for any $k \in \mathbb{Z}_{\geq 2} \cap U$ a modular form $f_k^{\text{JL}}$ on $\Gamma$, which is the group of norm-one elements in an Eichler order $R$ of level $Mp$ contained in the indefinite rational quaternion algebra $B$ of discriminant $D$. One can show that these modular forms can be $p$-adically interpolated, up to scaling, in a neighborhood of $k_0$. More precisely, let $\mathcal{O}$ be the ring of integers of a finite extension $F$ of $\mathbb{Q}_p$ and let $D$ denote the $\mathcal{O}$-module of $\mathcal{O}$-valued measures on $\mathbb{Z}_p^2$ which are supported on the set of primitive elements in $\mathbb{Z}_p^2$. Let $\Gamma_0$ be the group of norm-one elements in an Eichler order $R_0 \subseteq B$ containing $R$. There is a canonical action of $\Gamma_0$ on $D$ (see [9, §2.4] for its description). Denote by $F_k$ the extension of $F$ generated by the Fourier coefficients of $f_k$. Then there is an element $\Phi \in H^1(\Gamma_0, D)$ and maps $\rho_k : H^1(\Gamma_0, D) \rightarrow H^1(\Gamma, F_k)$ such that $\rho(k)(\Phi) = \phi_k$, the cohomology class associated to $f_k^{\text{JL}}$, with $k$ in a neighborhood of $k_0$ (for this we need a suitable normalization of the cohomology class associated to $f_k^{\text{JL}}$, which we do not touch for simplicity in this introduction). We view $\Phi$ as a quaternionic family of modular forms. To each $\phi_k$ we may apply the Shimura-Shintani-Waldspurger lifting ([13]) and obtain a modular form $h_k$ of weight $k + 1/2$, level $4Np$ and trivial character. We show that this collection of forms can be $p$-adically interpolated. For clarity’s sake, we present the liftings and their $Λ$-adic variants in a diagram, in which the horizontal maps are specialization maps of the $p$-adic family to weight $k$; $\text{JL}$ stands for the Jacquet-Langlands correspondence; $\text{SSW}$ stands for the Shimura-Shintani-Waldspurger lift; and the dotted arrows are constructed in this paper:
More precisely, as a particular case of our main result, Theorem 3.8, we get the following

**Theorem 1.1.** There exists a $p$-adic neighborhood $U_0$ of $k_0$ in $\text{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times)$, $p$-adic periods $\Omega_k$ for $k \in U_0 \cap \mathbb{Z}^{\geq 2}$ and a formal expansion

$$\Theta = \sum_{\xi \geq 1} a_\xi q^\xi$$

with coefficients $a_\xi$ in the ring of $\mathbb{C}_p$-valued functions on $U_0$, such that for all $k \in U_0 \cap \mathbb{Z}^{\geq 2}$ we have

$$\Theta(k) = \Omega_k \cdot h_k.$$

Further, $\Omega_{k_0} \neq 0$.

2. **Shintani integrals and Fourier coefficients of half-integral weight modular forms**

We express the Fourier coefficients of half-integral weight modular forms in terms of period integrals, thus allowing a cohomological interpretation which is key to the production of the $\Lambda$-adic version of the Shimura-Shintani-Waldspurger correspondence. For the quaternionic Shimura-Shintani-Waldspurger correspondence of interest to us (see [15], [22]), the period integrals expressing the values of the Fourier coefficients have been computed generally by Prasanna in [16].

2.1. **The Shimura-Shintani-Waldspurger lifting.** Let $4M$ be a positive integer, $2k$ an even non-negative integer and $\chi$ a Dirichlet character modulo $4M$ such that $\chi(-1) = 1$. Recall that the space of half-integral weight modular forms $\mathcal{S}_{k+1/2}(4M, \chi)$ consists of holomorphic cuspidal functions $h$ on the upper-half place $\mathcal{H}$ such that

$$h(\gamma(z)) = j^{1/2}(\gamma, z)^{2k+1}\chi(d)h(z),$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4M)$, where $j^{1/2}(\gamma, z)$ is the standard square root of the usual automorphy factor $j(\gamma, z)$ (cf. [15, 2.3]).

To any quaternionic integral weight modular form we may associate a half-integral weight modular form following Shimura’s work [19], as we will describe below.

Fix an odd square free integer $N$ and a factorization $N = M \cdot D$ into coprime integers such that $D > 1$ is a product of an even number of distinct primes. Fix a Dirichlet character $\psi$ modulo $M$ and a positive even integer $2k$. Suppose that

$$\psi(-1) = (-1)^k.$$ 

Define the Dirichlet character $\chi$ modulo $4N$ by

$$\chi(x) := \psi(x) \left( \frac{-1}{x} \right)^k .$$
Let $B$ be an indefinite quaternion algebra over $\mathbb{Q}$ of discriminant $D$. Fix a maximal order $\mathcal{O}_B$ of $B$. For every prime $\ell | M$, choose an isomorphism

$$i_\ell : B \otimes \mathbb{Q}_\ell \simeq M_2(\mathbb{Q}_\ell)$$

such that $i_\ell(\mathcal{O}_B \otimes \mathbb{Z}_\ell) = M_2(\mathbb{Z}_\ell)$. Let $R \subseteq \mathcal{O}_B$ be the Eichler order of $B$ of level $M$ defined by requiring that $i_\ell(R \otimes \mathbb{Z}_\ell)$ is the suborder of $M_2(\mathbb{Z}_\ell)$ of upper triangular matrices modulo $\ell$ for all $\ell | M$. Let $\Gamma$ denote the subgroup of the group $R_1^\ast$ of norm 1 elements in $R^\ast$ consisting of those $\gamma$ such that $i_\ell(\gamma) \equiv \left( \begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix} \right)$ mod $\ell$ for all $\ell | M$. We denote by $S_{2k}(\Gamma)$ the $\mathbb{C}$-vector space of weight $2k$ modular forms on $\Gamma$, and by $S_{2k}(\Gamma, \psi^2)$ the subspace of $S_{2k}(\Gamma)$ consisting of forms having character $\psi^2$ under the action of $R_1^\ast$. Fix a Hecke eigenform

$$f \in S_{2k}(\Gamma, \psi^2)$$

as in [19, Section 3].

Let $V$ denote the $\mathbb{Q}$-subspace of $B$ consisting of elements with trace equal to zero. For any $v \in V$, which we view as a trace zero matrix in $M_2(\mathbb{R})$ (after fixing an isomorphism $i_\infty : B \otimes \mathbb{R} \simeq M_2(\mathbb{R})$), set

$$G_v := \{ \gamma \in SL_2(\mathbb{R}) | \gamma^{-1} v \gamma = v \}$$

and put $\Gamma_v := G_v \cap \Gamma$. One can show that there exists an isomorphism

$$\omega : \mathbb{R}^\times \xrightarrow{\sim} G_v$$

defined by $\omega(s) = \beta^{-1} \left( \begin{smallmatrix} s & 0 \\ 0 & 1/s \end{smallmatrix} \right) \beta$, for some $\beta \in SL_2(\mathbb{R})$. Let $t_v$ be the order of $\Gamma_v \cap \{ \pm 1 \}$ and let $\gamma_v$ be an element of $\Gamma_v$ which generates $\Gamma_v \{ \pm 1 \} / \{ \pm 1 \}$. Changing $\gamma_v$ to $\gamma_v^{-1}$ if necessary, we may assume $\gamma_v = \omega(t)$ with $t > 0$. Define $V^\ast$ to be the $\mathbb{Q}$-subspace of $V$ consisting of elements with strictly negative norm. For any $\alpha = \left( \begin{smallmatrix} a & b \\ c & -a \end{smallmatrix} \right) \in V^\ast$ and $z \in \mathcal{H}$, define the quadratic form

$$Q_\alpha(z) := cz^2 - 2az - b.$$ 

Fix $\tau \in \mathcal{H}$ and set

$$P(f, \alpha, \Gamma) := -\left(2(\text{nr}(\alpha))^{1/2}/t_v\right) \int_{\tau}^{\gamma_v(\tau)} Q_\alpha(z)^{k-1} f(z) dz$$

where $\text{nr} : B \to \mathbb{Q}$ is the norm map. By [19, Lemma 2.1], the integral is independent on the choice $\tau$, which justifies the notation.

**Remark 2.1.** The definition of $P(f, \alpha, \Gamma)$ given in [19, (2.5)] looks different: the above expression can be derived as in [19, page 629] by means of [19, (2.20) and (2.22)].

Let $R(\Gamma)$ denote the set of equivalence classes of $V^\ast$ under the action of $\Gamma$ by conjugation. By [19, (2.6)], $P(f, \alpha, \Gamma)$ only depends on the conjugacy class of $\alpha$, and thus, for $C \in R(\Gamma)$, we may define $P(f, C, \Gamma) := P(f, \alpha, \Gamma)$ for any choice of $\alpha \in C$. Also, $q(C) := \text{nr}(\alpha)$ for any $\alpha \in C$.

Define $\mathcal{O}_B'$ to be the maximal order in $B$ such that $\mathcal{O}_B' \otimes \mathbb{Z}_\ell \simeq \mathcal{O}_B \otimes \mathbb{Z}_\ell$ for all $\ell \nmid M$ and $\mathcal{O}_B' \otimes \mathbb{Z}_\ell$ is equal to the local order of $B \otimes \mathbb{Q}_\ell$ consisting of
Define a locally constant function \( H \) is an Hecke pair, we let 

\[
\gamma \quad \text{define the action of the Hecke action on cohomology groups; for details, see [9, 2.2.].}
\]

Then, by [19, Theorem 3.1],

\[
\text{for any } C \in R(\Gamma), \text{ fix } \alpha C \in C. \text{ For any integer } \xi \geq 1, \text{ define}
\]

\[
a_\xi(h) := (2\mu(\Gamma \backslash \delta))^{-1} \cdot \sum_{C \in R(\Gamma) \mid q(C) = \xi} \eta_0(\alpha C) \xi^{-1/2} P(f, C, \Gamma).
\]

Then, by [19 Theorem 3.1],

\[
\tilde{h} := \sum_{\xi \geq 1} a_\xi(h) \xi^\ell \in S_{k+1/2}(\Gamma_1, \chi)
\]

is called the Shimura-Shintani-Waldspurger lifting of \( f \).

2.2. Cohomological interpretation. We introduce necessary notation to define the action of the Hecke algebra on cohomology groups; for details, see [9 §2.1]. If \( G \) is a subgroup of \( B^\times \) and \( S \) a subsemigroup of \( B^\times \) such that \((G, S)\) is an Hecke pair, we let \( \mathcal{H}(G, S) \) denote the Hecke algebra corresponding to \((G, S), \) whose elements are written as \( T(s) = GsG = \prod_i Gs_i \) for \( s, s_i \in S \) (finite disjoint union). For any \( s \in S, \) let \( s^* := \text{norm}(s)s^{-1} \) and denote by \( S^* \) the set of elements of the form \( s^* \) for \( s \in S. \) For any \( \mathbb{Z}[S^*]-\text{module } M \) we let \( T(s) \) act on \( H^1(G, M) \) at the level of cochains \( c \in Z^1(G, M) \) by the formula \( (c|T(s)) (\gamma) = \sum_i s_i^* c(t_i(\gamma)) \), where \( t_i(\gamma) \) are defined by the equations \( Gs_i \gamma = Gs_j \) and \( s_i \gamma = t_i(\gamma) s_j. \) In the following, we will consider the case of \( G = \Gamma \) and

\[
S = \{ s \in B^\times | \text{iq}(s) \text{ is congruent to } (1,1) \text{ mod } \ell \text{ for all } \ell | M \}.
\]

For any field \( L \) and any integer \( n \geq 0, \) let \( V_n(L) \) denote the \( L \)-dual of the \( L \)-vector space \( P_n(L) \) of homogeneous polynomials in 2 variables of degree \( n. \) We let \( \mathcal{H}_2(L) \) act from the right on \( P(x, y) \) as \( P|\gamma(x, y) := P(\gamma(x, y)) \), where for \( \gamma = \left( \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \) we put

\[
\gamma(x, y) := (ax + by, cx + dy).
\]

This also equips \( V_n(L) \) with a left action by \( \gamma : \varphi \mapsto \varphi(\gamma). \) To simplify the notation, we will write \( P(\gamma(z)) \) for \( P(\gamma, 1). \)

Let \( F \) denote the finite extension of \( \mathbb{Q} \) generated by the eigenvalues of the Hecke action on \( f. \) For any field \( K \) containing \( F, \) set

\[
\mathcal{W}_f(K) := H^1(\Gamma, V_{k-2}(K))^f
\]

where the superscript \( f \) denotes the subspace on which the Hecke algebra acts via the character associated with \( f. \) Also, for any sign \( \pm, \) let \( \mathcal{W}_f^\pm(K) \) denote the \( \pm \)-eigenspace for the action of the archimedean involution \( \iota. \) Remember that \( \iota \) is defined by choosing an element \( \omega_\infty \) of norm \(-1 \) in \( R^\times \) such that such
that $i_\ell (\omega_\infty) \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ mod $M$ for all primes $\ell | M$ and then setting $\iota := T(w_\infty)$ (see [9, §2.1]). Then $W_f^\infty(K)$ is one dimensional (see, e.g., [9] Proposition 2.2); fix a generator $\phi_f^\pm$ of $W_f^\infty(F)$.

To explicitly describe $\phi_f^\pm$, let us introduce some more notation. Define

$$f|_{\omega_\infty}(z) := (Cz + D)^{-k/2}f(\omega_\infty(z))$$

where $i_\infty(\omega_\infty) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. Then $f|_{\omega_\infty} \in S_{2k}(\Gamma)$ as well. If the eigenvalues of the Hecke action on $f$ are real, then we may assume, after multiplying $f$ by a scalar, that $f|_{\omega_\infty} = f$ (see [19, p. 627] or [10, Lemma 4.15]). In general, let $I(f)$ denote the class in $H^1(\Gamma, V_{k-2}(\mathbb{C}))$ represented by the cocycle

$$\gamma \mapsto \left[ P \mapsto I_\gamma(f)(P) := \int_\gamma \frac{f(z)}{P(z)} dz \right]$$

for any $\tau \in \mathcal{H}$ (the corresponding class is independent on the choice of $\tau$). With this notation,

$$P(f, \alpha, \Gamma) = -(2(-n(r(\alpha))^{1/2}/\kappa_n) \cdot I_{\gamma_n}(f)(Q_{\alpha_n}(z)^{k-1}).$$

Denote by $I^\pm(f) := (1/2) \cdot I(f) \pm (1/2) \cdot I(f)|\omega_\infty$, the projection of $I(f)$ to the eigenspaces for the action of $\omega_\infty$. Then $I(f) = I^+(f) + I^-(f)$ and $I_f^\pm = \Omega_f^\pm \cdot \phi_f^\pm$, for some $\Omega_f^\pm \in \mathbb{C}^\times$.

Given $\alpha \in V^\times$ of norm $-\xi$, put $\alpha' := \omega_\infty^{-1} \omega_\infty$. By [19, 4.19], we have

$$\eta(\alpha)\xi^{-1/2}P(f, \alpha, \Gamma) + \eta(\alpha')\xi^{-1/2}P(f, \alpha', \Gamma) = -\eta(\alpha) \cdot t_{\alpha_1}^{-1} \cdot I_{\gamma_1}^+(Q_{\alpha_1}(z)^{k-1}).$$

We then have

$$a_\xi(\tilde{h}) = \sum_{c \in R_{\Gamma}(\Gamma), q(c) = \xi} \frac{-\eta(\alpha)}{2\mu(\Gamma \setminus \mathcal{H}) \cdot t_{\alpha}} \cdot I_{\gamma_n}^+(Q_{\alpha_1}(z)^{k-1}).$$

We close this section by choosing a suitable multiple of $h$ which will be the object of the next section. Given $Q_\alpha(z) = cz^2 - 2az - b$ as above, with $\alpha$ in $V^\times$, define $Q_\alpha(z) := M \cdot Q_\alpha(z)$. Then, clearly, $I^{k}(f)(Q_{\alpha_1}(z)^{k-1})$ is equal to $M^{k-1}I^{k}(f)(Q_{\alpha_1}(z)^{k-1})$. We thus normalize the Fourier coefficients by setting

$$a_\xi(h) := \frac{a_\xi(\tilde{h}) \cdot M^{k-1} \cdot 2\mu(\Gamma \setminus \mathcal{H})}{\Omega_f^\pm} = \sum_{c \in R(\Gamma, q(c) = \xi)} \eta(\alpha) \cdot t_{\alpha} \cdot \phi_f^\pm (Q_{\alpha_1}(z)^{k-1}).$$

So

$$h := \sum_{\xi \geq 1} a_\xi(h) q^\xi$$

belongs to $S_{k+1/2}(4N, \chi)$ and is a non-zero multiple of $\tilde{h}$. 

Documenta Mathematica 18 (2013) 1–21.
3. THE $\Lambda$-ADIC SHIMURA-SHINTANI-WALDSPURGER CORRESPONDENCE

At the heart of Stevens’s proof lies the control theorem of Greenberg-Stevens, which has been worked out in the quaternionic setting by Longo–Vigni [9]. Recall that $N \geq 1$ is a square free integer and fix a decomposition $N = M \cdot D$ where $D$ is a square free product of an even number of primes and $M$ is coprime to $D$. Let $p \nmid N$ be a prime number and fix an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$.

3.1. THE HIDA HECKE ALGEBRA. Fix an ordinary $p$-stabilized newform

$$f_0 \in S_{k_0}(\Gamma_1(Mp^{\infty}) \cap \Gamma_0(D), \epsilon_0)$$

of level $\Gamma_1(Mp^{\infty}) \cap \Gamma_0(D)$, Dirichlet character $\epsilon_0$ and weight $k_0$, and write $\mathcal{O}$ for the ring of integers of the field generated over $\mathbb{Q}_p$ by the Fourier coefficients of $f_0$.

Let $\Lambda$ (respectively, $\mathcal{O}[[\mathbb{Z}_p^n]]$) denote the Iwasawa algebra of $W := 1 + p\mathbb{Z}_p$ (respectively, $\mathbb{Z}_p^n$) with coefficients in $\mathcal{O}$. We denote group-like elements in $\Lambda$ and $\mathcal{O}[[\mathbb{Z}_p^n]]$ as $[t]$. Let $\mathfrak{h}_{\infty}^\text{rd}$ denote the $p$-ordinary Hida Hecke algebra with coefficients in $\mathcal{O}$ of tame level $\Gamma_1(N)$. Denote by $\mathcal{L} := \text{Frac}(\Lambda)$ the fraction field of $\Lambda$. Let $\mathcal{R}$ denote the integral closure of $\Lambda$ in the primitive component $\mathcal{K}$ of $\mathfrak{h}_{\infty}^\text{rd} \otimes_\Lambda \mathcal{L}$ corresponding to $f_0$. It is well known that the $\Lambda$-algebra $\mathcal{R}$ is finitely generated as $\Lambda$-module.

Denote by $\mathcal{X}$ the $\mathcal{O}$-module $\text{Hom}^\text{cont}_{\mathcal{O}_\text{alg}}(\mathcal{R}, \overline{\mathbb{Q}}_p)$ of continuous homomorphisms of $\mathcal{O}$-algebras. Let $\mathcal{X}_\text{arithmetic}$ be the set of arithmetic homomorphisms in $\mathcal{X}$, defined in [9] [2.2] by requiring that the composition

$$W \longrightarrow \Lambda \longrightarrow \overline{\mathbb{Q}}_p$$

has the form $\gamma \mapsto \psi_\kappa(\gamma)\gamma^{n_\kappa}$ for an integer $k_\kappa \geq 2$ (called the weight of $\kappa$) and a finite order character $\psi_\kappa : W \rightarrow \overline{\mathbb{Q}}_p$ (called the wild character of $\kappa$). Denote by $r_\kappa$ the smallest among the positive integers $t$ such that $1 + p^t \mathbb{Z}_p \subseteq \ker(\psi_\kappa)$. For any $\kappa \in \mathcal{X}_\text{arithmetic}$, let $P_\kappa$ denote the kernel of $\kappa$ and $\mathcal{R}_{P_\kappa}$ the localization of $\mathcal{R}$ at $\kappa$. The field $F_\kappa := \mathcal{R}_{P_\kappa}/P_\kappa\mathcal{R}_{P_\kappa}$ is a finite extension of $\text{Frac}(\mathcal{O})$. Further, by duality, $\kappa$ corresponds to a normalized eigenform

$$f_\kappa \in S_{k_\kappa}(\Gamma_0(Np^{\infty}), \epsilon_{k_\kappa})$$

for a Dirichlet character $\epsilon_{k_\kappa} : (\mathbb{Z}/Np^{\infty}\mathbb{Z})^\times \rightarrow \overline{\mathbb{Q}}_p$. More precisely, if we write $\psi_\mathcal{R}$ for the character of $\mathcal{R}$, defined as in [9] Terminology p. 555], and we let $\omega$ denote the Teichmüller character, we have $\epsilon_{k_\kappa} := \psi_{k_\kappa} \cdot \psi_{\mathcal{R}} \cdot \omega^{-n_\kappa}$ (see [9] Cor. 1.6]). We call $(\epsilon_{k_\kappa}, k_\kappa)$ the signature of $\kappa$. We let $k_0$ denote the arithmetic character associated with $f_0$, so $f_0 = f_{k_0}$, $k_0 = k_{k_\kappa}$, $\epsilon_0 = \epsilon_{k_\kappa}$, and $r_0 = r_{k_0}$. The eigenvalues of $f_\kappa$ under the action of the Hecke operators $T_n$ ($n \geq 1$ an integer) belong to $F_\kappa$. Actually, one can show that $f_\kappa$ is a $p$-stabilized newform on $\Gamma_1(Mp^{\infty}) \cap \Gamma_0(D)$.

Let $\Lambda_N$ denote the Iwasawa algebra of $\mathbb{Z}_p^n \times (\mathbb{Z}/N\mathbb{Z})^\times$ with coefficients in $\mathcal{O}$. To simplify the notation, define $\Delta := (\mathbb{Z}/Np\mathbb{Z})^\times$. We have a canonical isomorphism of rings $\Lambda_N \simeq \Lambda[\Delta]$, which makes $\Lambda_N$ a $\Lambda$-algebra, finitely generated as
\(\Lambda\)-module. Define the tensor product of \(\Lambda\)-algebras
\[
\mathcal{R}_N := \mathcal{R} \otimes_{\Lambda} \Lambda_N,
\]
which is again a \(\Lambda\)-algebra (resp. \(\Lambda_N\)-algebra) finitely generated as a \(\Lambda\)-module, (resp. as a \(\Lambda_N\)-module). One easily checks that there is a canonical isomorphism of \(\Lambda\)-algebras
\[
\mathcal{R}_N \simeq \mathcal{R}[\Delta]
\]
(where \(\Lambda\) acts on \(\mathcal{R}\)); this is also an isomorphism of \(\Lambda_N\)-algebras, when we let \(\Lambda_N \simeq \Lambda[\Delta]\) act on \(\mathcal{R}[\Delta]\) in the obvious way.

We can extend any \(\kappa \in \mathcal{X}^{\text{arith}}\) to a continuous \(\mathcal{O}\)-algebra morphism
\[
\kappa_N : \mathcal{R}_N \longrightarrow \bar{\mathbb{Q}}_p
\]
setting
\[
\kappa_N \left( \sum_{i=1}^{n} r_i \cdot \delta_i \right) := \sum_{i=1}^{n} \kappa(r_i) \cdot \psi_R(\delta_i)
\]
for \(r_i \in \mathcal{R}\) and \(\delta_i \in \Delta\). Therefore, \(\kappa_N\) restricted to \(\mathbb{Z}_p^\times\) is the character \(t \mapsto \epsilon_\kappa(t) t^{n_\kappa}\). If we denote by \(\mathcal{X}_N\) the \(\mathcal{O}\)-module of continuous \(\mathcal{O}\)-algebra homomorphisms from \(\mathcal{R}_N\) to \(\bar{\mathbb{Q}}_p\), the above correspondence sets up an injective map \(\mathcal{X}^{\text{arith}} \hookrightarrow \mathcal{X}_N\). Let \(\mathcal{X}_N^{\text{arith}}\) denote the image of \(\mathcal{X}^{\text{arith}}\) under this map. For \(\kappa_N \in \mathcal{X}_N^{\text{arith}}\), we define the signature of \(\kappa_N\) to be that of the corresponding \(\kappa\).

3.2. The control theorem in the quaternionic setting. Recall that \(B/\mathbb{Q}\) is a quaternion algebra of discriminant \(D\). Fix an auxiliary real quadratic field \(F\) such that all primes dividing \(D\) are inert in \(F\) and all primes dividing \(Mp\) are split in \(F\), and an isomorphism \(i_F : B \otimes_{\mathbb{Q}} F \simeq M_2(F)\). Let \(\mathcal{O}_B\) denote the maximal order of \(B\) obtained by taking the intersection of \(B\) with \(M_2(\mathcal{O}_F)\), where \(\mathcal{O}_F\) is the ring of integers of \(F\). More precisely, define
\[
\mathcal{O}_B := \iota^{-1}(i_F^{-1}(i_F(B \otimes 1) \cap M_2(\mathcal{O}_F)))
\]
where \(\iota : B \hookrightarrow B \otimes_{\mathbb{Q}} F\) is the inclusion defined by \(b \mapsto b \otimes 1\). This is a maximal order in \(B\) because \(i_F(B \otimes 1) \cap M_2(\mathcal{O}_F)\) is a maximal order in \(i_F(B \otimes 1)\). In particular, \(i_F\) and our fixed embedding of \(\bar{\mathbb{Q}}\) into \(\bar{\mathbb{Q}}_p\) induce an isomorphism
\[
i_p : B \otimes_{\mathbb{Q}} \mathbb{Q}_p \simeq M_2(\mathbb{Q}_p)
\]
such that \(i_p(\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_p) = M_2(\mathbb{Z}_p)\). For any prime \(\ell|M\), also choose an embedding \(\mathbb{Q} \hookrightarrow \mathbb{Q}_\ell\) which, composed with \(i_F\), yields isomorphisms
\[
i_\ell : B \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \simeq M_2(\mathbb{Q}_\ell)
\]
such that \(i_p(\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_\ell) = M_2(\mathbb{Z}_\ell)\). Define an Eichler order \(R \subseteq \mathcal{O}_B\) of level \(M\) by requiring that for all primes \(\ell|M\) the image of \(R \otimes_{\mathbb{Z}} \mathbb{Z}_\ell\) via \(i_\ell\) consists of upper triangular matrices modulo \(\ell\). For any \(r \geq 0\), let \(\Gamma_r\) denote the subgroup of the group \(\mathbb{R}^*_+\) of norm-one elements in \(R\) consisting of those \(\gamma\) such that \(i_\ell(\gamma) = \left( \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right)\) with \(c \equiv 0 \mod Mp\) and \(a \equiv d \equiv 1 \mod Mp\).
for all primes $\ell | M_p$. To conclude this list of notation and definitions, fix an embedding $F \hookrightarrow \mathbb{R}$ and let $i_\infty : B \otimes_{\mathbb{Q}} \mathbb{R} \simeq M_2(\mathbb{R})$ be the induced isomorphism.

Let $Y := \mathbb{Z}_p^2$ and denote by $X$ the set of primitive vectors in $Y$. Let $D$ denote the $O$-module of $O$-valued measures on $Y$ which are supported on $X$. Note that $M_2(\mathbb{Z}_p)$ acts on $Y$ by left multiplication; this induces an action of $M_2(\mathbb{Z}_p)$ on the $O$-module of $O$-valued measures on $Y$, which induces an action on $D$. The group $R^\times$ acts on $D$ via $i_\infty$. In particular, we may define the group:

$$W := H^1(\Gamma_0, D).$$

Then $D$ has a canonical structure of $O[\mathbb{Z}_p^\times]$-module, as well as $K_{2d}$-action, as described in [9 §2.4]. In particular, let us recall that, for any $[t] \in O[\mathbb{Z}_p^\times]$, we have

$$\int_X \varphi(x, y) d([t] \cdot \nu) = \int_X \varphi(tx, ty) d\nu,$$

for any locally constant function $\varphi$ on $X$.

For any $\kappa \in X_{\text{arith}}$ and any sign $\pm \in \{-, +\}$, set

$$W^\pm_\kappa := W^\pm_{f_{\kappa}}(F_\kappa) = H^1(\Gamma_{\kappa}, V_{n_\kappa}(F_\kappa))^{f_{\kappa}, \pm}$$

where $f_{\kappa}$ is any Jacquet-Langlands lift of $f_\kappa$ to $\Gamma_{\kappa}$; recall that the superscript $f_\kappa$ denotes the subspace on which the Hecke algebra acts via the character associated with $f_\kappa$, and the superscript $\pm$ denotes the $\pm$-eigenspace for the action of the archimedean involution $\iota$. Also, recall that $W^\pm_\kappa$ is one dimensional and fix a generator $\phi^\pm_\kappa$ of it.

We may define specialization maps

$$\rho_\kappa : \mathbb{D} \rightarrow V_{n_\kappa}(F_\kappa)$$

by the formula

$$\rho_\kappa(\nu)(P) := \int_{\mathbb{Z}_p \times \mathbb{Z}_p^\times} e_\kappa(y) P(x, y) d\nu$$

which induces (see [9 §2.5]) a map:

$$\rho_\kappa : \mathbb{W} \rightarrow \mathbb{W}_\kappa.$$

Here $\mathbb{W}$ and $\mathbb{W}_\kappa$ denote the ordinary submodules of $\mathbb{W}$ and $\mathbb{W}_\kappa$, respectively, defined as in [3 Definition 2.2] (see also [9 §3.5]). We also let $\mathbb{W}_R := \mathbb{W} \otimes_{\Lambda} \mathbb{R}$, and extend the above map $\rho_\kappa$ to a map

$$\rho_\kappa : \mathbb{W}_R \rightarrow \mathbb{W}_\kappa$$

by setting $\rho_\kappa(x \otimes r) := \rho_\kappa(x) \cdot \kappa(r)$.

**Theorem 3.1.** There exists a $p$-adic neighborhood $U_0$ of $\kappa_0$ in $X$, elements $\Phi^\pm$ in $\mathbb{W}_R^\text{ord}$ and choices of $p$-adic periods $\Omega^\pm_\kappa \in F_\kappa$ for $\kappa \in U_0 \cap X_{\text{arith}}$ such that, for all $\kappa \in U_0 \cap X_{\text{arith}}$, we have

$$\rho_\kappa(\Phi^\pm) = \Omega^\pm_\kappa \cdot \phi^\pm_\kappa.$$
and $\Omega_{\kappa_0} \neq 0$.

**Proof.** This is an easy consequence of [9, Theorem 2.18] and follows along the lines of the proof of [21, Theorem 5.5], cf. [10, Proposition 3.2]. □

We now normalize our choices as follows. With $U_0$ as above, define

$$U_0^\text{arith} := U_0 \cap \mathcal{X}_0^\text{arith}.$$ 

Fix $\kappa \in U_0^\text{arith}$ and an embedding $\bar{Q}_p \hookrightarrow \mathbb{C}$. Let $f^\text{HL}_\kappa$ denote a modular form on $\Gamma_{r,\kappa}$, corresponding to $f_\kappa$ by the Jacquet-Langlands correspondence, which is well defined up to elements in $\Lambda \widehat{\otimes} \mathbb{C}$. View $\Phi_{\pm,\kappa}$ as an element in $H^1(\Gamma_{r,\kappa}, V^\kappa_{\text{fl}}(\mathbb{C}))_\pm$.

Choose a representative $\Phi_{\pm,\gamma}$ of $\Phi_{\pm,\kappa}$, by which we mean that if $\Phi_{\pm,\kappa} = \sum_i \Phi_{\pm,\gamma} \otimes r_i$, then we choose a representative $\Phi_{\pm,\gamma}$ for all $i$. Also, we will write $\rho_{\kappa}(\Phi)(P) := \int_{\mathbb{Z}_p \times \mathbb{Z}_p} \epsilon_\kappa(y) P(x, y) d\Phi_{\pm,\gamma}(x, y)$.

With this notation, we see that the two cohomology classes $\gamma \mapsto \int_{\mathbb{Z}_p \times \mathbb{Z}_p} \epsilon_\kappa(y) P(x, y) d\Phi_{\pm,\gamma}(x, y)$ and $\gamma \mapsto \Omega_{\kappa}^\pm \cdot \int_{\mathbb{Z}_p} \psi_{\kappa}(\gamma) \cdot \int_{\mathbb{Z}_p} P(z, 1) f^\text{HL}_{\kappa,\gamma}(z) dz$ are cohomologous in $H^1(\Gamma_{r,\kappa}, V_{\text{fl}}^\kappa(\mathbb{C}))$, for any choice of $\tau \in \mathcal{H}$.

### 3.3. Metaplectic Hida Hecke algebras

Let $\sigma : \Lambda_N \to \Lambda_N$ be the ring homomorphism associated to the group homomorphism $t \mapsto t^2$ on $\mathbb{Z}_p^\times \times (\mathbb{Z}/N\mathbb{Z})^\times$, and denote by the same symbol its restriction to $\Lambda$ and $\mathcal{O}[\mathbb{Z}_p^\times]$. We let $\Lambda_{\sigma}$, $\mathcal{O}[\mathbb{Z}_p^\times]$, and $\Lambda_{N,\sigma}$ denote, respectively, $\Lambda$, $\mathcal{O}[\mathbb{Z}_p^\times]$ and $\Lambda_N$ viewed as algebras over themselves via $\sigma$. The ordinary metaplectic $p$-adic Hida Hecke algebra we will consider is the $\Lambda$-algebra $\widetilde{R} := R \otimes_{\Lambda} \Lambda_{\sigma}$.

Define as above

$$\tilde{X} := \text{Hom}_{\mathcal{O}_{\text{alg}}}(\tilde{R}, \bar{Q}_p)$$

and let the set $\tilde{X}^\text{arith}$ of arithmetic points in $\tilde{X}$ to consist of those $\tilde{\kappa}$ such that the composition

$$W^\kappa \xrightarrow{\Lambda_{r,\kappa}} \Lambda_{r,\kappa+1,\kappa} \xrightarrow{\Lambda_{r,\kappa}} \tilde{R} \xrightarrow{\psi_{\kappa}} \bar{Q}_p$$

has the form $\gamma \mapsto \psi_{\kappa}(\gamma) \gamma^{n_{\kappa}}$ with $n_{\kappa} := k_{\bar{\kappa}} - 2$ for an integer $k_{\bar{\kappa}} \geq 2$ (called the weight of $\bar{\kappa}$) and a finite order character $\psi_{\kappa} : W \to \bar{Q}$ (called the wild character of $\bar{\kappa}$). Let $r_{\kappa}$ the smallest among the positive integers $t$ such that $1 + p^t \mathbb{Z}_p \subseteq \ker(\psi_{\kappa})$.

We have a map $p : \tilde{X} \to X$ induced by pull-back from the canonical map $R \to \tilde{R}$. The map $p$ restricts to arithmetic points.
As above, define the $\Lambda$-algebra (or $\Lambda_N$-algebra)

$$\tilde{\mathcal{R}}_N := \mathcal{R} \otimes_{\Lambda} \Lambda_N, \sigma$$

via $\lambda \mapsto 1 \otimes \lambda$.

We easily see that $\tilde{\mathcal{R}}_N \simeq \tilde{\mathcal{R}}[\Delta]$ as $\Lambda_N$-algebras, where we enhance $\tilde{\mathcal{R}}[\Delta]$ with the following structure of $\Lambda_N$-algebra: for $\sum_i \lambda_i \cdot \delta_i \in \tilde{\mathcal{R}}[\Delta]$ (with $\lambda_i \in \Lambda$ and $\delta_i \in \Delta$) and $\sum_j r_j \cdot \delta'_j \in \tilde{\mathcal{R}}[\Delta]$ (with $r_j = \sum_h r_{j,h} \otimes \lambda_{j,h} \in \tilde{\mathcal{R}}$, $r_{j,h} \in \mathcal{R}$, $\lambda_{j,h} \in \Lambda_\sigma$, and $\delta'_j \in \Delta$), we set

$$\left( \sum_i \lambda_i \cdot \delta_i \right) \cdot \left( \sum_j r_j \cdot \delta'_j \right) := \sum_{i,j,h} \left( r_{j,h} \otimes (\lambda_i \lambda_{j,h}) \right) \cdot (\delta_i \delta'_j).$$

As above, extend $\tilde{\kappa} \in \tilde{\mathcal{X}}_{\text{arith}}$ to a continuous $\mathcal{O}$-algebra morphism $\tilde{\kappa}_N : \tilde{\mathcal{R}}_N \to \bar{\mathbb{Q}}_p$ by setting

$$\tilde{\kappa}_N \left( \sum_{i=1}^n x_i \cdot \delta_i \right) := \sum_{i=1}^n \kappa(x_i) \cdot \psi_\mathcal{R}(\delta_i)$$

for $x_i \in \tilde{\mathcal{R}}$ and $\delta_i \in \Delta$, where $\psi_\mathcal{R}$ is the character of $\mathcal{R}$. If we denote by $\tilde{\mathcal{X}}_N$ the $\mathcal{O}$-module of continuous $\mathcal{O}$-linear homomorphisms from $\tilde{\mathcal{R}}_N$ to $\bar{\mathbb{Q}}_p$, the above correspondence sets up an injective map $\tilde{\mathcal{X}}_{\text{arith}} \hookrightarrow \tilde{\mathcal{X}}_N$ and we let $\tilde{\mathcal{X}}_{N, \text{arith}}$ denote the image of $\tilde{\mathcal{X}}_{\text{arith}}$. We also have a map $\rho_N : \tilde{\mathcal{X}}_N \to \mathcal{X}_N$ induced from the map $\mathcal{R}_N \to \tilde{\mathcal{R}}_N$ taking $r \mapsto r \otimes 1$ by pull-back. The map $\rho_N$ also restricts to arithmetic points. The maps $\rho$ and $\rho_N$ make the following diagram commute:

$$\begin{array}{ccc}
\tilde{\mathcal{X}}_{\text{arith}} & \longrightarrow & \tilde{\mathcal{X}}_N \\
\rho \downarrow & & \downarrow \rho_N \\
\mathcal{X}_{\text{arith}} & \longrightarrow & \mathcal{X}_N
\end{array}$$

where the projections take a signature $(\epsilon, k)$ to $(\epsilon^2, 2k)$.

### 3.4. The $\Lambda$-adic Correspondence

In this part, we combine the explicit integral formula of Shimura and the fact that the toric integrals can be $p$-adically interpolated to show the existence of a $\Lambda$-adic Shimura-Shintani-Waldspurger correspondence with the expected interpolation property. This follows very closely [21, §6].

Let $\tilde{\kappa}_N \in \mathcal{X}_{\text{arith}}$ of signature $(\epsilon_\mathcal{E}, k_\mathcal{E})$. Let $L_r$ denote the order of $\mathbb{M}_2(F)$ consisting of matrices $\begin{pmatrix} a & b \frac{1}{Mp^e} \\ c & d \end{pmatrix}$ with $a, b, c, d \in \mathcal{O}_F$. Define

$$\mathcal{O}_{B,r} := \mathcal{O}_r^{-1} \left( \mathcal{O}_r (F(B \otimes 1) \cap L_r) \right)$$

Documenta Mathematica 18 (2013) 1–21
Then $O_{B,r}$ is the maximal order introduced in [2.1] (and denoted $O_B^r$ there) defined in terms of the maximal order $O_B$ and the integer $Mp'$. Also,

$$S := O_B \cap O_{B,r}$$

is an Eichler order of $B$ of level $Mp$ containing the fixed Eichler order $R$ of level $M$. With $\alpha \in V^* \cap O_{B,1}$, we have

$$i_F(\alpha) = \begin{pmatrix} a & b/(Mp) \\ c & -a \end{pmatrix}$$

in $M_2(F)$ with $a, b, c \in \mathcal{O}_F$ and we can consider the quadratic forms

$$Q_\alpha(x, y) := cx^2 - 2axy - (b/(Mp))y^2,$$

and

$$Q_\alpha(x, y) := Mp \cdot Q_\alpha(x, y) = Mpcx^2 - 2Mpaxy - by^2.$$  

Then $\tilde{Q}_\alpha(x, y)$ coefficients in $\mathcal{O}_F$ and, composing with $F \hookrightarrow \mathbb{R}$ and letting $x = z$, $y = 1$, we recover $Q_\alpha(z)$ and $\tilde{Q}_\alpha(z)$ of [2.1] (defined by means of the isomorphism $i_\infty$). Since each prime $\ell | Mp$ is split in $F$, the elements $a, b, c$ can be viewed as elements in $\mathbb{Z}_\ell$ via our fixed embedding $\bar{Q} \hookrightarrow \mathbb{Q}_\ell$, for any prime $\ell | Mp$ (we will continue writing $a, b, c$ for these elements, with a slight abuse of notation). So, letting $b_\alpha \in \mathbb{Z}$ such that $\tilde{i}_\ell(\alpha) \equiv (\ast b_\alpha/(Mp)) \mod \ell$, for all $\ell | Mp$, we have $b \equiv b_\alpha$ modulo $Mp\mathbb{Z}_\ell$ as elements in $\mathbb{Z}_\ell$, for all $\ell | Mp$, and thus we get

$$\eta_{\alpha}(\alpha) = \varepsilon_{\alpha}(b_\alpha) = \varepsilon_{\alpha}(b)$$

for $b$ as in [7].

For any $\nu \in \mathbb{D}$, we may define an $\mathcal{O}$-valued measure $j_\alpha(\nu)$ on $\mathbb{Z}_p^{\times}$ by the formula:

$$\int_{\mathbb{Z}_p^{\times}} f(t)j_\alpha(\nu)(t) := \int_{\mathbb{Z}_p^{\times} \times \mathbb{Z}_p^{\times}} f(\tilde{Q}_\alpha(x, y))d
\nu(x, y).$$

for any continuous function $f : \mathbb{Z}_p^{\times} \to \mathbb{C}$. Recall that the group of $\mathcal{O}$-valued measures on $\mathbb{Z}_p^{\times}$ is isomorphic to the Iwasawa algebra $\mathcal{O}[\mathbb{Z}_p^{\times}]$, and thus we may view $j_\alpha(\nu)$ as an element in $\mathcal{O}[\mathbb{Z}_p^{\times}]$ (see, for example, [1, §3.2]). In particular, for any group-like element $[\lambda] \in \mathcal{O}[\mathbb{Z}_p^{\times}]$ we have:

$$\int_{\mathbb{Z}_p^{\times}} f(t)d([\lambda] \cdot j_\alpha(\nu))(t) = \int_{\mathbb{Z}_p^{\times}} \left( \int_{\mathbb{Z}_p^{\times}} f(ts)d[\lambda](s) \right) j_\alpha(\nu)(t) = \int_{\mathbb{Z}_p^{\times}} f(\lambda t)d_j(\nu)(t).$$

On the other hand,

$$\int_{\mathbb{Z}_p^{\times} \times \mathbb{Z}_p^{\times}} f(Q_\alpha(x, y))d(\lambda \cdot \nu) = \int_{\mathbb{Z}_p^{\times} \times \mathbb{Z}_p^{\times}} f(Q_\alpha(\lambda x, \lambda y))d\nu = \int_{\mathbb{Z}_p^{\times} \times \mathbb{Z}_p^{\times}} f(\lambda^2 Q_\alpha(x, y))d\nu$$

and we conclude that $j_\alpha(\lambda \cdot \nu) = [\lambda^2] \cdot j_\alpha(\nu)$. In other words, $j_\alpha$ is a $\mathcal{O}[\mathbb{Z}_p^{\times}]$-linear map

$$j_\alpha : \mathbb{D} \to \mathcal{O}[\mathbb{Z}_p^{\times}].$$

Before going ahead, let us introduce some notation. Let $\chi$ be a Dirichlet character modulo $Mp'$, for a positive integer $r$, which we decompose accordingly
with the isomorphism \((\mathbb{Z}/Np\mathbb{Z})^\times \simeq (\mathbb{Z}/N\mathbb{Z})^\times \times (\mathbb{Z}/p\mathbb{Z})^\times\) into the product \(\chi = \chi_N \cdot \chi_p\) with \(\chi_N : (\mathbb{Z}/N\mathbb{Z})^\times \rightarrow \mathbb{C}^\times\) and \(\chi_p : (\mathbb{Z}/p\mathbb{Z})^\times \rightarrow \mathbb{C}^\times\). Thus, we will write \(\chi(x) = \chi_N(xN) \cdot \chi_p(x_p)\), where \(x_N\) and \(x_p\) are the projections of \(x \in (\mathbb{Z}/Np\mathbb{Z})^\times\) to \((\mathbb{Z}/N\mathbb{Z})^\times\) and \((\mathbb{Z}/p\mathbb{Z})^\times\), respectively. To simplify the notation, we will suppress the \(N\) and \(p\) from the notation for \(x_N\) and \(x_p\), thus simply writing \(x\) for any of the two. Using the isomorphism \((\mathbb{Z}/N\mathbb{Z})^\times \simeq (\mathbb{Z}/M\mathbb{Z})^\times \times (\mathbb{Z}/D\mathbb{Z})^\times\), decompose \(\chi_N\) as \(\chi_N = \chi_M \cdot \chi_D\) with \(\chi_M\) and \(\chi_D\) characters on \((\mathbb{Z}/M\mathbb{Z})^\times\) and \((\mathbb{Z}/D\mathbb{Z})^\times\), respectively. In the following, we only need the case when \(\chi_D = 1\).

Using the above notation, we may define a \(\mathcal{O}[\mathbb{Z}^\times_p]\)-linear map \(J_\alpha : \mathbb{D} \rightarrow \mathcal{O}[\mathbb{Z}^\times_p]\) by

\[
J_\alpha(\nu) = \epsilon_{\bar{\nu},M}(b) \cdot \epsilon_{\bar{\nu},p}(-1) \cdot j_\alpha(\nu)
\]

with \(b\) as in \([6]\). Set \(\mathbb{D}_N := \mathbb{D} \otimes_{\mathcal{O}[\mathbb{Z}^\times_p]} \Lambda_N\), where the map \(\mathcal{O}[\mathbb{Z}^\times_p] \rightarrow \Lambda_N\) is induced from the map \(\mathbb{Z}^\times_p \rightarrow \mathbb{Z}^\times \times (\mathbb{Z}/N\mathbb{Z})^\times\) on group-like elements given by \(x \mapsto x \otimes 1\). Then \(J_\alpha\) can be extended to a \(\Lambda_N\)-linear map \(j_\alpha : \mathbb{D}_N \rightarrow \Lambda_N\). Setting \(\mathbb{R}_\mathbb{D} := \mathcal{R}_N \otimes_{\Lambda_N} \mathbb{D}_N\) and extending by \(\mathcal{R}_N\)-linearity over \(\Lambda_N\) we finally obtain a \(\mathcal{R}_N\)-linear map, again denoted by the same symbol,

\[
J_\alpha : \mathbb{D}_N \longrightarrow \mathbb{R}_\mathbb{D}.
\]

For \(\nu \in \mathbb{D}_N\) and \(r \in \mathcal{R}_N\) we thus have

\[
J_\alpha(r \otimes \nu) = \epsilon_{\bar{\nu},M}(b) \cdot \epsilon_{\bar{\nu},p}(-1) \cdot r \otimes j_\alpha(\nu).
\]

For the next result, for any arithmetic point \(\kappa_N \in \mathcal{X}^\text{arith}_N\) coming from \(\kappa \in \mathcal{X}^\text{arith}\), extend \(\rho_\kappa\) in \([5]\) by \(\mathcal{R}_N\)-linearity over \(\mathcal{O}[\mathbb{Z}^\times_p]\), to get a map

\[
\rho_\kappa : \mathbb{D}_N \longrightarrow V_{\eta_N}
\]

defined by \(\rho_\kappa(\nu \otimes r) := \rho_\kappa(\nu) \cdot \kappa_N(r)\), for \(\nu \in \mathbb{D}\) and \(r \in \mathcal{R}_N\). To simplify the notation, set

\[
(\nu, \alpha)_{\kappa_N} := \rho_\kappa(\nu)(\bar{Q}_\alpha^{n_N}/2).
\]

The following is essentially \([21]\) Lemma (6.1)].

**Lemma 3.2.** Let \(\bar{\kappa}_N \in \mathcal{X}^\text{arith}_N\) with signature \((\epsilon_{\bar{\kappa},k_\alpha})\) and define \(\kappa_N := p_N(\bar{\kappa}_N)\). Then for any \(\nu \in \mathbb{D}_\mathcal{R}_N\) we have:

\[
\bar{\kappa}_N(J_\alpha(\nu)) = \eta_{\epsilon_\alpha}(\alpha) \cdot (\nu, \alpha)_{\kappa_N}.
\]

**Proof.** For \(\nu \in \mathbb{D}_N\) and \(r \in \mathcal{R}_N\) we have

\[
\bar{\kappa}_N(J_\alpha(r \otimes \nu)) = \bar{\kappa}_N(\epsilon_{\bar{\nu},M}(b) \cdot \epsilon_{\bar{\nu},p}(-1) \cdot r \otimes j_\alpha(\nu))
\]

\[
= \epsilon_{\bar{\nu},M}(b) \cdot \epsilon_{\bar{\nu},p}(-1) \cdot \bar{\kappa}_N(r \otimes 1) \cdot \bar{\kappa}_N(1 \otimes j_\alpha(\nu))
\]

\[
= \epsilon_{\bar{\nu},M}(b) \cdot \epsilon_{\bar{\nu},p}(-1) \cdot \kappa_N(r) \cdot \int_{\mathbb{Z}_p^\times} \bar{\kappa}_N(t)d\rho_\alpha(\nu)
\]

and thus, noticing that \(\bar{\kappa}_N\) restricted to \(\mathbb{Z}_p^\times\) is \(\bar{\kappa}_N(t) = \epsilon_{\bar{\nu},p}(t)t^{n_N}\), we have

\[
\bar{\kappa}_N(J_\alpha(r \otimes \nu)) = \epsilon_{\bar{\nu},M}(b) \cdot \epsilon_{\bar{\nu},p}(-1) \cdot \kappa_N(r) \int_{\mathbb{Z}_p^\times \mathbb{Z}_p^\times} \epsilon_{\bar{\nu},p}(Q_\alpha(x, y))Q_\alpha(x, y)^{n_N}/2d\nu.
\]
Recalling \([8]\), and viewing \(a, b, c\) as elements in \(\mathbb{Z}_p\), we have, for \((x, y) \in \mathbb{Z}_p \times \mathbb{Z}_p^\times\),

\[
\varepsilon_{a, b}(\tilde{Q}_a(x, y)) = \varepsilon_{a, b}(-by^2) = \varepsilon_{a, b}(-b)e_{a, b}(y^2) = \varepsilon_{a, b}(-b)e_{a, b}(y).
\]

Thus, since \(\varepsilon_{\kappa}(-1)^2 = 1\), we get:

\[
\tilde{\kappa}_N(J_\nu(r \otimes \nu)) = \kappa_N(r) \cdot \varepsilon_{\kappa, M}(b) \cdot \varepsilon_{\kappa, p}(b) \cdot \rho_\nu(\nu)(\tilde{Q}_{\kappa, N}^{x, y}/2) = \eta_\kappa(\alpha) \cdot \langle \nu, \alpha \rangle_{\kappa, N}
\]

where for the last equality use \([9]\) and \([10]\). □

Define

\[
\mathbb{W}_{R_N} := \mathbb{W} \otimes_{\mathcal{O}[\mathbb{Z}_p^\times]} \mathcal{R}_N,
\]

the structure of \(\mathcal{O}[\mathbb{Z}_p^\times]\)-module of \(\mathcal{R}_N\) being that induced by the composition of the two maps \(\mathcal{O}[\mathbb{Z}_p^\times] \rightarrow \Lambda_N \rightarrow \mathcal{R}_N\) described above. There is a canonical map

\[
\vartheta : \mathbb{W}_{R_N} \rightarrow H^1(\Gamma_0, \mathcal{D}_{R_N})
\]

described as follows: if \(\nu_\gamma\) is a representative of an element \(\nu \in \mathbb{W}\) and \(r \in \mathcal{R}_N\), then \(\vartheta(\nu \otimes r)\) is represented by the cocycle \(\nu_\gamma \otimes r\).

For \(\nu \in \mathbb{W}_{R_N}\) represented by \(\nu_\gamma\) and \(\xi \geq 1\) an integer, define

\[
\theta_{\xi}(\nu) := \sum_{\gamma \in \Gamma_1 \backslash \Gamma, \varrho(\gamma) = \xi} J_{\nu_\gamma}(\nu_{\gamma_{\nu_\gamma}}) / t_{\nu_\gamma}.
\]

**Definition 3.3.** For \(\nu \in \mathbb{W}_{R_N}\), the formal Fourier expansion

\[
\Theta(\nu) := \sum_{\xi \geq 1} \theta_{\xi}(\nu) q^\xi
\]

in \(\mathcal{R}_N[q]\) is called the \(\Lambda\)-adic Shimura-Shintani-Waldspurger lift of \(\nu\). For any \(\tilde{\kappa} \in \tilde{X}_{\text{arith}}\), the formal power series expansion

\[
\Theta(\nu)(\tilde{\kappa}_N) := \sum_{\xi \geq 1} \tilde{\kappa}_N(\theta_{\xi}(\nu)) q^\xi
\]

is called the \(\tilde{\kappa}\)-specialization of \(\Theta(\nu)\).

There is a natural map

\[
\mathbb{W}_R \rightarrow \mathbb{W}_{R_N}
\]

taking \(\nu \otimes r\) to itself (use that \(\mathcal{R}\) has a canonical map to \(\mathcal{R}_N \simeq \mathcal{R}[\Delta]\), as described above). So, for any choice of sign, \(\Phi^\pm \in \mathbb{W}_R\) will be viewed as an element in \(\mathbb{W}_{R_N}\).

From now on we will use the following notation. Fix \(\tilde{\kappa}_0 \in \tilde{X}_{\text{arith}}\) and put \(\kappa_0 := p(\tilde{\kappa}_0) \in X_{\text{arith}}\). Recall the neighborhood \(U_0\) of \(\kappa_0\) in Theorem \([5, 1]\). Define \(U_{0}^\text{arith} := p^{-1}(U_0)\) and

\[
\tilde{U}_{0}^\text{arith} := \tilde{U}_0 \cap \tilde{X}_{\text{arith}}.
\]

For each \(\tilde{\kappa} \in \tilde{U}_{0}^\text{arith}\) put \(\kappa := p(\tilde{\kappa}) \in U_{0}^\text{arith}\). Recall that if \((\varepsilon_{\kappa, k_\kappa})\) is the signature of \(\tilde{\kappa}\), then \((\varepsilon_{\kappa, k_\kappa}) := (\varepsilon_{\kappa}^2, 2k_\kappa)\) is that of \(\kappa_0\). For any \(\kappa := p(\tilde{\kappa})\) as above, we may consider the modular form

\[
f_{\kappa}^{JL} \in \mathcal{S}_{\kappa, \gamma}(\Gamma_{\kappa}, \varepsilon_{\kappa})
\]
and its Shimura-Shintani-Waldspurger lift
\[ h_{\kappa} = \sum_{\xi} a_{\xi}(h_{\kappa})q^{\xi} \in S_{\kappa+1/2}(4Np^{\kappa}, \chi_{\kappa}), \quad \text{where } \chi_{\kappa}(x) := \epsilon_{\kappa}(x) \left( -\frac{1}{x} \right)^{\kappa}, \]
normalized as in \( \mathfrak{2} \) and \( \mathfrak{3} \). For our fixed \( \kappa_{0} \), recall the elements \( \Phi := \Phi^{+} \) chosen as in Theorem \( \mathfrak{5} \) and define \( \phi_{\kappa} := \phi_{\kappa}^{+} \) and \( \Omega_{\kappa} := \Omega_{\kappa}^{+} \) for \( \kappa \in \mathcal{U}_{0}^{\text{arith}} \).

**Proposition 3.4.** For all \( \bar{\kappa} \in \mathcal{U}_{0}^{\text{arith}} \) such that \( r_{\kappa} = 1 \) we have
\[ \tilde{\kappa}_{N}(\theta_{\kappa}(\Phi)) = \Omega_{\kappa} \cdot a_{\kappa}(h_{\kappa}) \quad \text{and} \quad \Theta(\Phi)(\tilde{\kappa}_{N}) = \Omega_{\kappa} \cdot h_{\kappa}. \]

**Proof.** By Lemma \( \mathfrak{3.2} \) we have
\[ \tilde{\kappa}_{N}(\theta_{\kappa}(\Phi)) = \sum_{C \in R(\Gamma_{1}, q(C)) = \xi} \eta_{\kappa}(\alpha_{C}) \frac{\eta_{\kappa}(\alpha_{C})}{t_{\alpha_{C}}} \rho_{\kappa,N}(\Phi)(\tilde{Q}_{\alpha_{C}}^{N/2}). \]
Using Theorem \( \mathfrak{3.4} \) we get
\[ \tilde{\kappa}_{N}(\theta_{\kappa}(\Phi)) = \sum_{C \in R(\Gamma_{1}, q(C)) = \xi} \eta_{\kappa}(\alpha_{C}) \frac{\eta_{\kappa}(\alpha_{C})}{t_{\alpha_{C}}} \phi_{\kappa}(\tilde{Q}_{\alpha_{C}}^{N/2}). \]
Now \( \mathfrak{2} \) shows the statement on \( \tilde{\kappa}_{N}(\theta_{2}(\Phi)) \), while that on \( \Theta(\Phi)(\tilde{\kappa}_{N}) \) is a formal consequence of the previous one. \( \square \)

**Corollary 3.5.** Let \( a_{p} \) denote the image of the Hecke operator \( T_{p} \) in \( \mathcal{R} \). Then \( \Theta(\Phi)|T_{p}^{2} = a_{p} \cdot \Theta(\Phi) \).

**Proof.** For any \( \kappa \in \mathcal{X}^{\text{arith}} \), let \( a_{p}(\kappa) := \kappa(T_{p}) \), which is a \( p \)-adic unit by the ordinarity assumption. For all \( \bar{\kappa} \in \mathcal{U}_{0}^{\text{arith}} \) with \( r_{\kappa} = 1 \), we have
\[ \Theta(\Phi)(\tilde{\kappa}_{N}|T_{p}^{2} = \Omega_{\kappa} \cdot h_{\kappa}|T_{p}^{2} = a_{p}(\kappa) \cdot \Omega_{\kappa} \cdot h_{\kappa} = a_{p}(\kappa) \cdot \Theta(\Phi)(\tilde{\kappa}_{N}). \]
Consequently,
\[ \tilde{\kappa}_{N}(\theta_{\ell^{2}}(\Phi)) = a_{p}(\kappa) \cdot \tilde{\kappa}_{N}(\theta_{2}(\Phi)) \]
for all \( \bar{\kappa} \) such that \( r_{\kappa} = 1 \). Since this subset is dense in \( \tilde{\mathcal{K}}_{N} \), we conclude that \( \theta_{\ell^{2}}(\Phi) = a_{p} \cdot \theta_{2}(\Phi) \) and so \( \Theta(\Phi)|T_{p}^{2} = a_{p} \cdot \Theta(\Phi) \). \( \square \)

For any integer \( n \geq 1 \) and any quadratic form \( Q \) with coefficients in \( F \), write \( \lfloor Q \rfloor_{n} \) for the class of \( Q \) modulo the action of \( \iota_{F}(\Gamma_{n}) \). Define \( \mathcal{F}_{n,\xi} \) to be the subset of the \( F \)-vector space of quadratic forms with coefficients in \( F \) consisting of quadratic forms \( Q_{\alpha} \) such that \( \alpha \in V^{*} \cap \mathcal{O}_{B,n} \) and \( -\nu(\alpha) = \xi \). Writing \( \delta_{Q_{\alpha}} \) for the discriminant of \( Q_{\alpha} \), the above set can be equivalently described as
\[ \mathcal{F}_{n,\xi} := \{ Q_{\alpha} | \alpha \in V^{*} \cap \mathcal{O}_{B,n}, \delta_{Q_{\alpha}} = Np^{\nu} \xi \}. \]
Define \( \mathcal{F}_{n,\xi}/\Gamma_{n} \) to be the set \( \{ [Q_{\alpha}]_{n} | Q_{\alpha} \in \mathcal{F}_{n,\xi} \} \) of equivalence classes of \( \mathcal{F}_{n,\xi} \) under the action of \( \iota_{F}(\Gamma_{n}) \). A simple computation shows that \( Q_{\alpha} - \gamma_{\alpha}g = Q_{\alpha} \) for all \( \alpha \in V^{*} \) and all \( g \in \Gamma_{n} \), and thus we find
\[ \mathcal{F}_{n,\xi}/\Gamma_{n} = \{ [Q_{\alpha}]_{n} | C \in R(\Gamma_{n}), \delta_{Q_{\alpha}} = Np^{\nu} \xi \}. \]
We also note that, in the notation of [21], if $f$ has weight character $\psi$, defined modulo $Np^m$, and level $\Gamma_n$, the Fourier coefficients $a_\xi(h)$ of the Shimura-Shintani-Waldspurger lift $h$ of $f$ are given by
\begin{equation}
 a_\xi(h) = \sum_{[Q] \in \mathcal{F}_m/\Gamma_n} \frac{\psi(Q)}{t_Q} \phi^+_Q(Q(z)^{k-1})
\end{equation}
and, if $Q = \bar{Q}_\alpha$, we put $\psi(Q) := \eta_\alpha(b_\alpha)$ and $t_Q := t_\alpha$. Also, if we let
\[ \mathcal{F}_n/\Gamma_n := \prod_\xi \mathcal{F}_{n,\xi}/\Gamma_n \]
we can write
\begin{equation}
 h = \sum_{[Q] \in \mathcal{F}_m/\Gamma_n} \frac{\psi(Q)}{t_Q} \phi^+_Q(Q(z)^{k-1})q^{\delta_Q/(Np^m)}.
\end{equation}
Fix now an integer $m \geq 1$ and let $n \in \{1, m\}$. For any $t \in (\mathbb{Z}/p^n\mathbb{Z})^\times$ and any integer $\xi \geq 1$, define $\mathcal{F}_{n,\xi,t}$ to be the subset of $\mathcal{F}_{n,\xi}$ consisting of forms such that $Np^m b_\alpha \equiv t \mod Np^m$. Also, define $\mathcal{F}_{n,\xi,t}/\Gamma_n$ to be the set of equivalence classes of $\mathcal{F}_{n,\xi,t}$ under the action of $i_F(\Gamma_n)$. If $\alpha \in V^* \cap O_{B,m}$ and
\[ i_F(\alpha) = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}, \]
then
\begin{equation}
 \bar{Q}_\alpha(\bar{x}, \bar{y}) = Np^m c x^2 - 2Np^m a xy - Np^m b y^2
\end{equation}
from which we see that there is an inclusion $\mathcal{F}_{m,\xi,t} \subseteq \mathcal{F}_{1,\xi,0^{-1},t}$. If $\bar{Q}_\alpha$ and $\bar{Q}_{\alpha'}$ belong to $\mathcal{F}_{m,\xi,t}$, and $\alpha' = gag^{-1}$ for some $g \in \Gamma_m$, then, since $\Gamma_m \subseteq \Gamma_1$, we see that $\bar{Q}_\alpha$ and $\bar{Q}_{\alpha'}$ represent the same class in $\mathcal{F}_{1,\xi,0^{-1},t}/\Gamma_1$. This shows that $[\bar{Q}_\alpha]_{m} \mapsto [\bar{Q}_{\alpha}]_{1}$ gives a well-defined map
\[ \pi_{m,\xi,t} : \mathcal{F}_{m,\xi,t}/\Gamma_m \rightarrow \mathcal{F}_{1,\xi,0^{-1},t}/\Gamma_1. \]

**Lemma 3.6.** The map $\pi_{m,\xi,t}$ is bijective.

**Proof.** We first show the injectivity. For this, suppose $\bar{Q}_\alpha$ and $\bar{Q}_{\alpha'}$ are in $\mathcal{F}_{m,\xi,t}$ and $[\bar{Q}_\alpha]_{m} = [\bar{Q}_{\alpha'}]_{1}$. So there exists $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in $i_F(\Gamma_1)$ such that such that $\bar{Q}_\alpha = \bar{Q}_{\alpha'}|g$. If $\bar{Q}_\alpha = cx^2 - 2axy - by^2$, and easy computation shows that $\bar{Q}_{\alpha'} = c'x^2 - 2a'xy - b'y^2$ with
\[ c' = ca^2 - 2ac\gamma - b\gamma^2 \]
\[ a' = -ca\beta + a\beta\gamma + a\alpha\delta + b\gamma\delta \]
\[ b' = -c\beta^2 + 2a\beta\delta + b\delta^2. \]
The first condition shows that $\gamma \equiv 0 \mod Np^m$. We have $b \equiv b' \equiv t \mod Np^m$, so $\delta^2 \equiv 1 \mod Np^m$. Since $\delta \equiv 1 \mod Np$, we see that $\delta \equiv 1 \mod Np^m$ too.
We now show the surjectivity. For this, fix \([\tilde{Q}_{\kappa_c}]_1\) in the target of \(\pi\), and choose a representative 

\[
\tilde{Q}_{\kappa_c} = cx^2 - 2axy - by^2
\]

(recall \(Mp^m\xi|\delta_{\tilde{Q}_{\kappa_c}}, Np[c, Np[a, and \(b \in \mathcal{O}_p^*\), the last condition due to \(\eta_\psi(\alpha_c) \neq 0\)). By the Strong Approximation Theorem, we can find \(\tilde{g} \in \Gamma_1\) such that

\[
i\ell(\tilde{g}) \equiv \begin{pmatrix} 1 & 0 \\ ab^{-1} & 1 \end{pmatrix} \text{ mod } Np^m
\]

for all \(\ell|Np\). Take \(g := i\ell(\tilde{g})\), and put \(\alpha := g^{-1}\alpha_cg\). An easy computation, using the expressions for \(a', b', c'\) in terms of \(a, b, c\) and \(g = \left( \begin{array}{cc} a' & b' \\ c' & d' \end{array} \right)\) as above, shows that \(\alpha \in V^* \cap \mathcal{O}_{B, m}, \eta_\psi(\alpha) = t\) and \(\delta_{\tilde{Q}_{\kappa}} = Np^m\xi\), and it follows that \(\tilde{Q}_\alpha \in \mathcal{F}_{m,t}\). Now

\[
\pi([\tilde{Q}_{\kappa_c}]_1) = [\tilde{Q}_\alpha]_1 = [\tilde{Q}_{\kappa} - t\alpha_c\bar{g}]_1 = [\tilde{Q}_{\kappa_c}]_1
\]

where the last equality follows because \(g \in \Gamma_1\).

**Proposition 3.7.** For all \(\tilde{k} \in \tilde{\mathcal{L}}^\text{arith}\) we have

\[
\Theta(\Phi)(\tilde{k}_{\kappa})|T_p^{r_\kappa - 1} = \Omega_{\kappa} \cdot h_{\kappa}.
\]

**Proof.** For \(r_\kappa = 1\), this is Proposition 3.4 above, so we may assume \(r_\kappa > 1\). As in the proof of Proposition 3.4, combining Lemma 3.2 and Theorem 3.1 we get

\[
\Theta(\Phi)(\tilde{k}_{\kappa}) = \sum_{\xi \geq 1} \left( \sum_{\ell \in R(1), \ell(\xi) = \xi} \frac{\eta_{\kappa}(\alpha_{\kappa}) \cdot \Omega_{\kappa}}{t_{\kappa}} \phi_{\kappa}(Q_{\kappa_c}^{1-1}) \right) q^\xi
\]

which, by Lemma 3.3 and 3.4 above we may rewrite as

\[
\Theta(\Phi)(\tilde{k}_{\kappa}) = \sum_{[Q] \in \mathcal{F}_1/\Gamma_1} \frac{\epsilon_{\kappa}(Q) \cdot \Omega_{\kappa}}{t_Q} \phi_{\kappa}(Q_{\kappa_c}^{1-1}) q^{\delta_Q/(Np)}
\]

By definition of the action of \(T_p\) on power series, we have

\[
\Theta(\Phi)(\tilde{k}_{\kappa})|T_p^{r_\kappa - 1} = \sum_{[Q] \in \mathcal{F}_1/\Gamma_1} \frac{\epsilon_{\kappa}(Q) \cdot \Omega_{\kappa}}{t_Q} \phi_{\kappa}(Q_{\kappa_c}^{1-1}) q^{\delta_Q/(Np^{r_\kappa})}.
\]

Setting \(\mathcal{F}_{n,t}/\Gamma_n := \prod_{\ell \geq 1} \mathcal{F}_{n, t, \ell}/\Gamma_n\) for \(n \in \{1, r_\kappa\}\), Lemma 3.6 shows that \(\mathcal{F}_{1, t}^* := \{[Q] \in \mathcal{F}_{1,t}/\Gamma_1, t\} such \ text{ that } p^r|\delta_Q\ is equal to \mathcal{F}_{r_\kappa, t}^*\). Therefore, splitting the above sum over \(t \in (\mathbb{Z}/Np^{r_\kappa}\mathbb{Z})^\times\), we get

\[
\Theta(\Phi)(\tilde{k}_{\kappa})|T_p^{r_\kappa - 1} = \sum_{t \in (\mathbb{Z}/p^{r_\kappa-1}\mathbb{Z})} \sum_{[Q] \in \mathcal{F}_{1,t}} \frac{\epsilon_{\kappa}(Q) \cdot \Omega_{\kappa}}{t_Q} \phi_{\kappa}(Q_{\kappa_c}^{1-1}) q^{\delta_Q/(Np^{r_\kappa})}
\]

\[
= \sum_{t \in (\mathbb{Z}/p^{r_\kappa-1}\mathbb{Z})} \sum_{[Q] \in \mathcal{F}_{m,t}/\Gamma_m} \frac{\epsilon_{\kappa}(Q) \cdot \Omega_{\kappa}}{t_Q} \phi_{\kappa}(Q_{\kappa_c}^{1-1}) q^{\delta_Q/(Np^{r_\kappa})}
\]

\[
= \sum_{[Q] \in \mathcal{F}_{m,t}/\Gamma_m} \frac{\epsilon_{\kappa}(Q) \cdot \Omega_{\kappa}}{t_Q} \phi_{\kappa}(Q_{\kappa_c}^{1-1}) q^{\delta_Q/(Np^{r_\kappa})}.
\]
Comparing this expression with \[\{12\}\] gives the result. □

We are now ready to state the analogue of \[\{21\}\] Thm. 3.3], which is our main result. For the reader’s convenience, we briefly recall the notation appearing below. We denote by $\mathcal{X}$ the points of the ordinary Hida Hecke algebra, and by $\mathcal{X}^{\text{arith}}$ its arithmetic points. For $\kappa_0 \in \mathcal{X}^{\text{arith}}$, we denote by $U_0$ the $p$-adic neighborhood of $\kappa_0$ appearing in the statement of Theorem 3.1 and put $U_0^{\text{arith}} := U_0 \cap \mathcal{X}^{\text{arith}}$. We also denote by $\Phi := \Phi^+ \in W_\text{ord}$ the cohomology class appearing in Theorem 3.1. The points $\tilde{\mathcal{X}}$ of the metaplectic Hida Hecke algebra defined in \[\{3.3\}\] are equipped with a canonical map $p : \mathcal{X}^{\text{arith}} \to \mathcal{X}^{\text{arith}}$ on arithmetic points. Let $\tilde{U}_0^{\text{arith}} := \tilde{U}_0 \cap \tilde{\mathcal{X}}^{\text{arith}}$. For each $\tilde{\kappa} \in \tilde{U}_0^{\text{arith}}$, put $\kappa = p(\tilde{\kappa}) \in U_0^{\text{arith}}$. Recall that if $(\epsilon_{\tilde{\kappa}}, k_{\tilde{\kappa}})$ is the signature of $\tilde{\kappa}$, then the signature of $\kappa$ is $(\epsilon_{\kappa}, k_{\kappa}) := (\epsilon_{\tilde{\kappa}}^2, 2k_{\tilde{\kappa}})$. For any $\kappa := p(\tilde{\kappa})$ as above, we may consider the modular form

$$f_{\kappa}^{\text{JL}} \in S_{\kappa}(\Gamma_{\kappa}, \epsilon_{\kappa})$$

and its Shimura-Shintani-Waldspurger lift

$$h_{\kappa} = \sum_{\xi} a_{\xi}(h_{\kappa}) q^\xi \in S_{\kappa + 1/2}(4Np^{r_{\kappa}}; \chi_{\kappa}),$$

where $\chi_{\kappa}(x) := \epsilon_{\kappa}(x) \left(\frac{-1}{x}\right)^{k_{\kappa}}$, normalized as in \[\{2\}\] and \[\{3\}\]. Finally, for $\tilde{\kappa} \in \tilde{\mathcal{X}}^{\text{arith}}$, we denote by $\tilde{\kappa}_N$ its extension to the metaplectic Hecke algebra $\tilde{\mathcal{R}}_N$ defined in \[\{3.3\}\].

**Theorem 3.8.** Let $\kappa_0 \in \mathcal{X}^{\text{arith}}$. Then there exists a choice of $p$-adic periods $\Omega_{\kappa}$ for $\kappa \in U_0$ such that the $\Lambda$-adic Shimura-Shintani-Waldspurger lift of $\Phi$

$$\Theta(\Phi) := \sum_{\xi \geq 1} \theta_{\xi}(\Phi) q^\xi$$

in $\mathcal{R}_N[q]$ has the following properties:

1. $\Omega_{\kappa_0} \neq 0$.
2. For any $\tilde{\kappa} \in \tilde{U}_0^{\text{arith}}$, the $\kappa$-specialization of $\Theta(\Phi)$

$$\Theta(\nu)(\tilde{\kappa}_N) := \sum_{\xi \geq 1} \tilde{\kappa}(\theta_{\xi}(\Phi)) q^\xi \text{ belongs to } S_{\kappa + 1/2}(4Np^{r_{\kappa}}; \chi_{\kappa}'),$$

where $\chi_{\kappa}'(x) := \chi_{\kappa}(x) \cdot (\chi_{\kappa}(x))^{k_{\kappa}}$, and satisfies

$$\Theta(\Phi)(\tilde{\kappa}_N) = \Omega_{\kappa} \cdot h_{\kappa}|T_p^{1-r_{\kappa}}.$$ 

**Proof.** The elements $\Omega_{\kappa}$ are those $\Omega_\kappa^+$ appearing in Theorem 3.1 which we used in Propositions 3.3 and 3.7 above, so (1) is clear. Applying $T_p^{r_{\kappa}-1}$ to the formula of Proposition 3.7 using Corollary 3.5 and applying $a_p(\kappa)^{1-r_{\kappa}}$ on both sides gives

$$\Theta(\Phi)(\tilde{\kappa}_N) = a_p(\kappa)^{1-r_{\kappa}} \cdot \Omega_{\kappa} \cdot h_{\kappa}|T_p^{r_{\kappa}-1}.$$ 

By \[\{18\}\] Prop. 1.9], each application of $T_p$ has the effect of multiplying the character by $(\chi_{\kappa}(x))^{k_{\kappa}}$, hence

$$h_{\kappa}' := h_{\kappa}|T_p^{r_{\kappa}-1} \in S_{\kappa + 1/2}(4Np^{r_{\kappa}}; \chi_{\kappa}').$$

Documenta Mathematica 18 (2013) 1–21
with \( \chi'_k \) as in the statement. This gives the first part of (2), while the last formula follows immediately from Proposition \ref{prop:2}. \qed

Remark 3.9. Theorem \ref{thm:1} is a direct consequence of Theorem \ref{thm:3.8}, as we briefly show below.

Recall the embedding \( \mathbb{Z}^{\geq 2} \hookrightarrow \text{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times) \) which sends \( k \in \mathbb{Z}^{\geq 2} \) to the character \( x \mapsto x^{k-2} \). Extending characters by \( \mathcal{O} \)-linearity gives a map

\[
\mathbb{Z}^{\geq 2} \rightarrow \mathcal{X}(\Lambda) := \text{Hom}_{\text{cont}}(\Lambda, \mathbb{Q}_p).
\]

We denote by \( k^{(\Lambda)} \) the image of \( k \in \mathbb{Z}^{\geq 2} \) in \( \mathcal{X}(\Lambda) \) via this embedding. We also denote by \( \varpi : \mathcal{X} \rightarrow \mathcal{X}(\Lambda) \) the finite-to-one map obtained by restriction of homomorphisms to \( \Lambda \). Let \( k^{(\mathcal{R})} \) be a point in \( \mathcal{X} \) of signature \( (k, 1) \) such that \( \varpi(k^{(\mathcal{R})}) = k^{(\Lambda)} \). A well-known result by Hida (see [6] Cor. 1.4]) shows that \( \mathcal{R}/\Lambda \) is unramified at \( k^{(\Lambda)} \). As shown in [21] §1, this implies that there is a section \( s_{k^{(\Lambda)}}(\varpi) \) which is defined on a neighborhood \( \mathcal{U}_{k^{(\Lambda)}} \) of \( k^{(\Lambda)} \) in \( \mathcal{X}(\Lambda) \) and sends \( k^{(\Lambda)} \) to \( k^{(\mathcal{R})} \).

Fix now \( k_0 \) as in the statement of Theorem \ref{thm:1} corresponding to a cuspidal \( f_0 \) of weight \( k_0 \) with trivial character. The form \( f_0 \) corresponds to an arithmetic character \( k_0^{(\mathcal{R})} \) of signature \( (1, k_0) \) belonging to \( \mathcal{X} \). Let \( \mathcal{U}_0 \) be the inverse image of \( \mathcal{U}_0 \) under the section \( s_{k_0}^{-1} \) of \( \mathcal{X} \), where \( \mathcal{U}_0 \subseteq \mathcal{X} \) is the neighborhood of \( k_0^{(\mathcal{R})} \) in Theorem \ref{thm:3.8}. Extending scalars by \( \mathcal{O} \) gives, as above, an injective continuous map \( \text{Hom}(\mathbb{Z}_p^\times, \mathbb{Z}_p^\times) \hookrightarrow \mathcal{X}(\Lambda) \), and we let \( \mathcal{U}_0' \) be any neighborhood of the character \( x \mapsto x^{k_0-2} \) which maps to \( \mathcal{U}_0 \) and is contained in the residue class of \( k_0 \) modulo \( p-1 \). Composing this map with the section \( \mathcal{U}_0 \hookrightarrow \mathcal{U}_0' \) gives a continuous injective map

\[
\varsigma : \mathcal{U}_0 \hookrightarrow \mathcal{U}_0' \hookrightarrow \mathcal{U}_0
\]

which takes \( k_0 \) to \( k_0^{(\mathcal{R})} \), since by construction the image of \( k_0 \) by the first map is \( k_0^{(\Lambda)} \). We also note that, more generally, \( \varsigma(k) = k^{(\mathcal{R})} \) because by construction \( \varsigma(k) \) restricts to \( k^{(\Lambda)} \) and its signature is \( (1, k) \), since the character of \( \varsigma(k) \) is trivial. To show the last assertion, recall that the character of \( \varsigma(k) \) is \( \psi_k \cdot \psi_\mathcal{R} \cdot \omega^{-k} \), and note that \( \psi_k \) is trivial because \( k^{(\Lambda)}(x) = x^{k_0-1} \), and \( \psi_\mathcal{R} \cdot \omega^{-k} \) is trivial because the same is true for \( k_0 \) and \( k \equiv k_0 \) modulo \( p-1 \). In other words, arithmetic points in \( \varsigma(\mathcal{U}_0) \) correspond to cusps with trivial character. This is the Hida family of forms with trivial character that we considered in the Introduction.

We can now prove Theorem \ref{thm:1}. For all \( k \in \mathcal{U}_0 \cap \mathbb{Z}^{\geq 2} \), put \( \Omega_k := \Omega_{k^{(\mathcal{R})}} \) and define \( \Theta := \Theta(\Phi) \circ \varsigma \) with \( \Phi \) as in Theorem \ref{thm:3.8} for \( \kappa_0 = k_0^{(\mathcal{R})} \). Applying Theorem \ref{thm:3.8} to \( k_0^{(\mathcal{R})} \), and restricting to \( \varsigma(\mathcal{U}_0) \), shows that \( U_0, \Omega_k \) and \( \Theta \) satisfy the conclusion of Theorem \ref{thm:1}.

Remark 3.10. For \( \kappa \in \mathcal{U}_0^{\text{cont}} \) of signature \( (\epsilon_\kappa, k_\kappa) \) with \( r_\kappa = 1 \) as in the above theorem, \( h_{\kappa} \) is trivial if \( (-1)^{k_\kappa} \neq \epsilon_\kappa(-1) \). However, since \( \phi_{\kappa_0} \neq 0 \), it follows that \( h_{\kappa_0} \) is not trivial as long as the necessary condition \( (-1)^{k_0} = \epsilon_0(-1) \) is verified.
Remark 3.11. This result can be used to produce a quaternionic \( \Lambda \)-adic version of the Saito-Kurokawa lifting, following closely the arguments in \cite{LongoNicole12}. 

References


Matteo Longo
Dipartimento di Matematica
Università di Padova
Via Trieste 63
35121 Padova
Italy
mlongo@math.unipd.it

Marc-Hubert Nicole
Institut de mathématiques de Luminy
Université d’Aix-Marseille
campus de Luminy, case 907
13288 Marseille cedex 9
France
nicole@iml.univ-mrs.fr
ON A POSITIVE EQUICHARACTERISTIC
VARIANT OF THE $p$-CURVATURE CONJECTURE

HÉLÈNE ESNault, ADRIAN LANGER

Received: October 26, 2012
Communicated by Takeshi Saito

Abstract. Our aim is to formulate and prove a weak form in equal characteristic $p > 0$ of the $p$-curvature conjecture. We also show the existence of a counterexample to a strong form of it.

2010 Mathematics Subject Classification: 14D05, 14E20, 14F35, 11G10, 11G99
Keywords and Phrases: varieties in positive characteristic, stratified bundles, étale trivializable bundles, monodromy group, abelian varieties

Introduction

If $(E, \nabla)$ is a vector bundle with an algebraic integrable connection over a smooth complex variety $X$, then it is defined over a smooth scheme $S$ over $\text{Spec} \mathbb{Z}[\frac{1}{N}]$ for some positive integer $N$, so $(E, \nabla) = (E_S, \nabla_S) \otimes_S \mathbb{C}$ over $X = X_S \otimes_S \mathbb{C}$ for a geometric generic point $\mathbb{Q}(S) \subset \mathbb{C}$. Grothendieck-Katz’s $p$-curvature conjecture predicts that if for all closed points $s$ of some non-trivial open $U \subset S$, the $p$-curvature of $(E_S, \nabla_S) \times_S s$ is zero, then $(E, \nabla)$ is trivialized by a finite étale cover of $X$ (see e.g. [An, Conj.3.3.3]). Little is known about it. N. Katz proved it for Gauss-Manin connections [Ka], for $S$ finite over $\text{Spec} \mathbb{Z}[\frac{1}{M}]$ (i.e., if $X$ can be defined over a number field), D. V. Chudnovsky and G. V. Chudnovsky in [CC] proved it in the rank 1 case and Y. André in [An] proved it in case the Galois differential Lie algebra of $(E, \nabla)$ at the generic point of $S$ is solvable (and for extensions of connections satisfying the conjecture). More recently, B. Farb and M. Kisin [FK] proved it for certain locally symmetric varieties $X$. In general, one is lacking methods to think of the problem.

The first author is supported by the SFB/TR45 and the ERC Advanced Grant 226257. The second author is supported by the Bessel Award of the Humboldt Foundation and a Polish MNiSW grant (contract number N N201 420639).
Y. André in [An II] and E. Hrushovskiy in [Hr V] formulated the following equal characteristic 0 analog of the conjecture: if $X \to S$ is a smooth morphism of smooth connected varieties defined over a characteristic 0 field $k$, then if $(E_S, \nabla_S)$ is a relative integrable connection such that for all closed points $s$ of some non-trivial open $U \subset S$, $(E_S, \nabla_S) \times_S s$ is trivialized by a finite étale cover of $X \times_S s$, then $(E, \nabla)|_{X_{\eta}}$ should be trivialized by a finite étale cover of $X_{\eta}$, where $\eta$ is a geometric generic point and $X_{\eta} = X \times_S \eta$. So the characteristic 0 analogy to integrable connections is simply integrable connections, and to the $p$-curvature condition is the trivialization of the connection by a finite étale cover. André proved it [An Prop. 7.1.1], using Jordan’s theorem and Simpson’s moduli of flat connections, while Hrushovsky [Hr p.116] suggested a proof using model theory.

It is tempting to formulate an equal characteristic $p > 0$ analog of Y. André’s theorem. A main feature of integrable connections over a field $k$ of characteristic 0 is that they form an abelian, rigid, $k$-linear tensor category. In characteristic $p > 0$, the category of bundles with an integrable connection is only $\mathcal{O}_{X^{(1)}}$-linear, where $X^{(1)}$ is the relative Frobenius twist of $X$, and the notion is too weak. On the other hand, in characteristic 0, the category of bundles with a flat connection is the same as the category of $\mathcal{O}_X$-coherent $\mathcal{D}_X$-modules. In characteristic $p > 0$, $\mathcal{O}_X$-coherent $\mathcal{D}_X$-modules over a smooth variety $X$ defined over a field $k$ form an abelian, rigid, $k$-linear tensor category (see [Gi]). It is equivalent to the category of stratified bundles. It bears strong analogies with the category of bundles with an integrable connection in characteristic 0. For example, if $X$ is projective smooth over an algebraically closed field, the triviality of the étale fundamental group forces all such $\mathcal{O}_X$-coherent $\mathcal{D}_X$-modules to be trivial (\cite{EM}).

So we raise the question 1: let $f : X \to S$ be a smooth projective morphism of smooth connected varieties, defined over an algebraically closed characteristic $p > 0$ field, let $(E, \nabla)$ be a stratified bundle relative to $S$, such that for all closed point $s$ of some non-trivial open $U \subset S$, the stratified bundle $(E, \nabla)|_{X_s}$ is trivialized by a finite étale cover of $X_s := X \times_S s$. Is it the case that the stratified bundle $(E, \nabla)|_{X_{\eta}}$ is trivialized by a finite étale cover of $X_{\eta}$?

In this form, this is not true. Y. Laszlo [L3] constructed a one dimensional non-trivial family of bundles over a curve over $\bar{F}_2$ which is fixed by the square of Frobenius, as a (negative) answer to a question of J. de Jong concerning the behavior of representations of the étale fundamental group over a finite field $\bar{F}_q$, $q = p^n$, with values in $GL(r, \bar{F}(i))$, where $\bar{F} \supset \bar{F}_2$ is a finite extension. In fact, Laszlo’s example yields also a counter-example to the question as stated above. We explain this in Sections[1]and[4] (see Corollary 4.3). We remark that if $E$ is a bundle on $X$, such that the bundle $E|_{X_s}$ is stable, numerically flat (see Definition 3.2) and moves in the moduli, then $E_\eta$ cannot be trivialized by a finite étale cover (see Proposition 4.4). In contrast, we show that if the family $X \to S$ is trivial (as it is in Laszlo’s example), thus $X = Y \times_k S$, if $k$ is algebraically closed, and if $(F^p_\mathcal{E} \times \text{id}_{Y})^*(E)|_{Y \times_S \mathbb{A}^1} \cong E|_{Y \times_S \mathbb{A}^1}$ for all closed points $s$ of some non-trivial open in $S$ and some fixed natural number $n$, then the moduli points of $E|_{Y \times_S \mathbb{A}^1}$ are constant (see Proposition 4.4). Here $F_\mathcal{E} : Y \to Y$ is the absolute Frobenius of $Y$. In Laszlo’s example, one does have $(F^p_\mathcal{E} \times \text{id}_{Y})^*(E)|_{Y \times_S \mathbb{A}^1} \cong E|_{Y \times_S \mathbb{A}^1}$ but
only over $k = \mathbb{F}_2$ (i.e., $S$ is also defined over $\mathbb{F}_2$). When one extends the family to the algebraic closure of $\mathbb{F}_2$, to go from the absolute Frobenius over $\mathbb{F}_2$, that is the relative Frobenius over $k$, to the absolute one, one needs to replace the power 2 with a higher power $n(s)$, which depends on the field of definition of $s$, and is not bounded.

So we modify question 1 in question 2: let $f : X \to S$ be a smooth projective morphism of smooth connected varieties, defined over an algebraically closed characteristic field $k$ of characteristic $p > 0$, let $E$ be a bundle such that for all closed points $s$ of some non-trivial open $U \subset S$, the bundle $E|_{X_s}$ is trivialized by a finite Galois étale cover of $X_s := X \times_S s$ of order prime to $p$. Is it the case that the bundle $E|_{X_{\bar{\eta}}}$ is trivialized by a finite étale cover of $X_{\bar{\eta}}$?

The answer is nearly yes: it is the case if $k$ is not algebraic over its prime field (Theorem 5.1). If $k = \mathbb{F}_p$, it might be wrong (Remarks 5.4), but what remains true is that there exists a finite étale cover of $X_{\bar{\eta}}$ over which the pull-back of $E$ is a direct sum of line bundles (Theorem 5.1). The assumption on the degrees of the Galois covers of $X_s$ trivializing $E|_{X_s}$ is necessary (as follows from Laszlo’s example) and it allows us to apply Brauer-Feit’s theorem in place of Jordan’s theorem used by André. However, there is no direct substitute for Simpson’s moduli spaces of flat bundles. Instead, we use the moduli spaces constructed in [La1] and we carefully analyze subloci containing the points of interest, that is the numerically flat bundles. The necessary material needed on moduli is gathered in Section 3.

Finally we raise the general question 3: let $f : X \to S$ be a smooth projective morphism of smooth connected varieties, defined over an algebraically closed characteristic $p > 0$ field, let $(E, \nabla)$ be a stratified bundle relative to $S$, such that for all closed points $s$ of some non-trivial open $U \subset S$, the stratified bundle $(E, \nabla)|_{X_s}$ is trivialized by a finite Galois étale cover of $X_s := X \times_S s$ of order prime to $p$. Is it the case that the bundle $(E, \nabla)|_{X_{\bar{\eta}}}$ is trivialized by a finite étale cover of $X_{\bar{\eta}}$?

We give the following not quite complete answer. If the rank of $E$ is 1, (in which case the assumption on the degrees of the Galois covers is automatically fulfilled), then the answer is yes provided $S$ is projective, and for any $s \in U$, Pic$(X_s)$ is reduced (see Theorem 7.1). The proof relies on (a variant of) an idea of M. Raynaud [Ra], using the height function associated to a symmetric line bundle (that is the reason for our assumption on $S$) on the abelian scheme and its dual, to show that an infinite Verschiebung-divisible point has height equal to 0 (Theorem 6.2). If $E$ has any rank, then the answer is yes if $k$ is not $\mathbb{F}_p$ (Theorem 7.2). In general, there is a prime to $p$-order Galois cover of $X_{\bar{\eta}}$ such that the pull-back of $E$ becomes a sum of stratified line bundles (Theorem 7.2).

Acknowledgements: The first author thanks Michel Raynaud for the fruitful discussions in November 2009, which are reflected in [Ra] and in Section 6. The first author thanks Johan de Jong for a beautiful discussion in November 2010 on the content of [EM], where she suggested question 1 to him, and where he replied that Laszlo’s example should contradict this, and that this should be better understood. The second author would like to thank Stefan Schröer for destroying his naive hopes concerning
Néron models of Frobenius twists of an abelian variety. We thank Damian Rössler for discussions on $p$-torsion on abelian schemes over functions fields. We thank the referee of a first version of the article. He/she explained to us that the dichotomy in Theorem 5.12) and in Theorem 7.22) should be $\mathbb{F}_p$ or not rather that countable or not, thereby improving our result.

1 Preliminaries on relative stratified sheaves

Let $S$ be a scheme of characteristic $p$ (i.e., $\mathcal{O}_S$ is an $\mathbb{F}_p$-algebra). By $F^r_S : S \to S$ we denote the $r$-th absolute Frobenius morphism of $S$ which corresponds to the $p^r$-th power mapping on $\mathcal{O}_S$.

If $X$ is an $S$-scheme, we denote by $X^{(r)}_S$ the fiber product of $X$ and $S$ over the $r$-th Frobenius morphism of $S$. If it is clear with respect to which structure $X$ is considered, we simplify the notation to $X^{(r)}$. Then the $r$-th absolute Frobenius morphism of $X$ induces the relative Frobenius morphism $F^r_{X/S} : X \to X^{(r)}$. In particular, we have the following commutative diagram:

\[
\begin{array}{ccc}
X & \xrightarrow{F^r_X} & X^{(r)} \\
\downarrow{X^{(r)}_S} & & \downarrow{W^{(r)}_X} \\
S & \xrightarrow{F^r_S} & S \\
\end{array}
\]

which defines $W^r_{X/S} : X^{(r)} \to X$.

Making $r = 1$ and replacing $X$ by $X^{(i)}$, this induces the similar diagram

\[
\begin{array}{ccc}
X^{(i)} & \xrightarrow{F^r_{X^{(i)}}} & X^{(i+1)} \\
\downarrow{X^{(i+1)}_S} & & \downarrow{W^{(i+1)}_{X^{(i)}}} \\
S & \xrightarrow{F^r_{X^{(i)}}} & S \\
\end{array}
\]

We assume that $X/S$ is smooth. A relative stratified sheaf on $X/S$ is a sequence $\{E_i, \sigma_i\}_{i \in \mathbb{N}}$ of locally free coherent $\mathcal{O}_{X^{(i)}}$-modules $E_i$ on $X^{(i)}$ and isomorphisms $\sigma_i : F^r_{X^{(i)}}E_i \to E_i$ of $\mathcal{O}_{X^{(i)}}$-modules. A morphism of relative stratified sheaves $\{\alpha_i : E_i \rightarrow E'_i\}$ is a sequence of $\mathcal{O}_{X^{(i)}}$-linear maps $\alpha_i : E_i \to E'_i$ compatible with the $\sigma_i$, that is such that $\sigma'_i \circ F^r_{X^{(i)}} \alpha_{i+1} = \alpha_i \circ \sigma_i$.

This forms a category Strat$(X/S)$, which is contravariant for morphisms $\phi : T \to S$: to $\{E_i, \alpha_i\} \in \text{Start}(X/S)$ one assigns $\phi^*\{E_i, \alpha_i\} \in \text{Strat}(X \times_ST/T)$ in the obvious way: $\phi$ induces $1_{X^{(i)}_X} \times \phi : X^{(i)}_X \times_ST \to X^{(i)}_X$ and $(\phi^*\{E_i, \alpha_i\})_i = \{(1_{X^{(i)}_X} \times \phi)^*E_i, (1_{X^{(i)}_X} \times \phi)^*\sigma_i\}$.
If \( S = \text{Spec} \ k \) where \( k \) is a field, \( \text{Strat}(X/k) \) is an abelian, rigid, tensor category. Giving a rational point \( x \in X(k) \) defines a fiber functor via \( \omega_x : \text{Strat}(X/k) \to \text{Vec}_k \), \( \omega_x(\{E, \sigma_i\}) = (E_0) \), in the category of finite dimensional vector spaces over \( k \), thus a \( k \)-group scheme \( \pi_*(\text{Strat}(X/k), \omega_x) = \text{Aut}^\circ(\omega_x) \). Tannaka duality implies that \( \text{Strat}(X/k) \) is equivalent via \( \omega_x \) to the representation category of \( \pi_*(\text{Strat}(X/k), \omega_x) \) with values in \( \text{Vec}_k \). For any object \( E := \{E_i, \sigma_i\} \in \text{Strat}(X/k) \), we define its monodromy group to be the \( k \)-affine group scheme \( \pi(E, \omega_x) \), where \( \langle E \rangle \subset \text{Strat}(X/k) \) is the full subcategory spanned by \( E \). This is the image of \( \pi(\text{Strat}(X/k), \omega_x) \) in \( GL(\omega_x(\langle E \rangle)) \) ([DM Proposition 2.21 a]). We denote by \( I_{X/k} \in \text{Strat}(X/k) \) the trivial object, with \( E' = \mathcal{O}_{X^{(i)}} \) and \( \sigma_i = \text{Identity} \).

**Lemma 1.1.** With the notation above

1) If \( h : Y \to X \) is a finite étale cover such that \( h^*E \) is trivial, then \( h_! \mathcal{L}_{Y/k} \)

has finite monodromy group and one has a faithfully flat homomorphism

\[
\pi(h_! \mathcal{L}_{Y/k}, \omega_x) \to \pi(\langle E \rangle, \omega_x).
\]

Thus in particular, \( E \) has finite monodromy group as well.

2) If \( E \in \text{Strat}(X/k) \) has finite monodromy group, then there exists a \( \pi((E), \omega_x) \)-torsor \( h : Y \to X \) such that \( h^*E \) is trivial in \( \text{Strat}(Y/k) \). Moreover, one has an isomorphism

\[
\pi(h_! \mathcal{L}_{Y/k}, \omega_x) \xrightarrow{\cong} \pi((E), \omega_x).
\]

**Proof.** We first prove 2). Assume \( \pi((E), \omega_x) = : G \) is a finite group scheme over \( k \). One applies Nori’s method [No, Chapter I, II]: the regular representation of \( G \) on the affine \( k \)-algebra \( k[G] \) of regular function defines the Artin \( k \)-algebra \( k[G] \) as a \( k \)-algebra object of the representation category of \( G \) on finite dimensional \( k \)-vector spaces, (such that \( k \subset k[G] \) is the maximal trivial subobject). Thus by Tannaka duality, there is an object \( \mathcal{A} = (A^i, \tau_i) \in \text{Strat}(X/k) \), which is an \( I_{X/k} \)-algebra object, (such that \( I_{X/k} \subset \mathcal{A} \) is the maximal trivial subobject). We define \( h_i : Y_i = X \to X_{(i)} \). Then the isomorphism \( \tau_i \) yields an \( \mathcal{O}_{X^{(i)}} \)-isomorphism between \( Y_{(i)} \xrightarrow{h_{(i)}} X^{(i)} \) and \( Y_i \xrightarrow{h_i} X_{(i)} \), (see, e.g., [SGAS Exposé XV, §1, Proposition 2]), and via this isomorphism, \( \mathcal{A} \) is isomorphic to \( h_* \mathcal{L}_{Y/k} \). On the other hand, \( \omega_x(\mathcal{A}) \) is a sub \( G \)-representation of \( k[G] \) for some \( n \in \mathbb{N} \), thus \( E \subset \mathcal{A}^{\otimes n} \in \text{Strat}(X/k) \), thus there is an inclusion \( E \subset (h_! \mathcal{L}_{Y/k})^{\otimes n} \in \text{Strat}(X/k) \), thus \( h^*E \subset (h^*h_! \mathcal{L}_{Y/k})^{\otimes n} \in \text{Strat}(Y/k) \). Since \( h^*h_! \mathcal{L}_{Y/k} \) is isomorphic to \( k[G] \) in \( \text{Strat}(Y/k) \) (recall that by [DM Proposition 13], \( G \) is an étale group scheme), then \( h^*E \) is isomorphic to \( \otimes^r \mathcal{L}_{Y/k} \), where \( r \) is the rank of \( E \). This shows the first part of the statement, and shows the second part as well: indeed, \( E \) is then a subobject of \( h_* \mathcal{L}_{Y/k} \) thus \( \mathcal{A} \subset \langle h_* \mathcal{L}_{Y/k} \rangle \) is a full subcategory. One applies [DM Proposition 2.21 a]) to show that the induced homomorphism \( \pi(h_* \mathcal{L}_{Y/k}, \omega_x) \to \pi((E), \omega_x) = G \) is faithfully flat. So \( \pi((h_* \mathcal{L}_{Y/k}, \omega_x) \) acts on \( \omega_x(h_* \mathcal{L}_{Y/k}) = k[G] \) via its quotient \( G \) and the regular representation \( G \subset GL(k[G]) \). Thus the homomorphism is an isomorphism.

We show 1). Assume that there is a finite étale cover \( h : Y \to X \) such that \( h^*E \) is isomorphic in \( \text{Strat}(Y/k) \) to \( \otimes^r \mathcal{L}_{Y/k} \) where \( r \) is the rank of \( E \). Then \( E \subset \otimes^r h_* \mathcal{L}_{Y/k} \), thus \( \pi(h_* \mathcal{L}_{Y/k}, \omega_x) \to \pi((E), \omega_x) \) is faithfully flat [DM loc. cit.,] so we are reduced to showing that \( h_* \mathcal{L}_{Y/k} \) has finite monodromy. But, by the same argument as on \( E \),
any of its objects of rank $r'$ lies in $\oplus_r h^r \mathbb{L}^r_{F/k}$. So we apply [DM Proposition 2.20 a)] to conclude that the monodromy of $h_{\mathbb{L}}^r_{F/k}$ is finite.

**Corollary 1.2.** With the notations as in [7] if $E \in \text{Strat}(X/k)$ has finite monodromy group, then for any field extension $K \supset k$, $E \otimes K \in \text{Strat}(X \otimes K/K)$ has finite monodromy group.

Let $E$ be an $\mathcal{O}_X$-module. We say that $E$ has a stratification relative to $S$ if there exists a relative stratified sheaf $\{E_i, \sigma_i\}$ such that $E_0 = E$.

Let us consider the special case $S = \text{Spec} k$, where $k$ is a perfect field, and $X/k$ is smooth. An (absolute) stratified sheaf on $X$ is a sequence $\{E_i, \sigma_i\}_{i \in \mathbb{N}}$ of coherent $\mathcal{O}_X$-modules $E_i$ on $X$ and isomorphisms $\sigma_i : F_q^i E_{i+1} \to E_i$ of $\mathcal{O}_X$-modules.

As $k$ is perfect, the $W_{X,0}$ are isomorphisms, thus giving an absolute stratified sheaf is equivalent to giving a stratified sheaf relative to $\text{Spec} k$.

We now go back to the general case and we assume that $S$ is an integral $k$-scheme, where $k$ is a field. Let us set $K = k(S)$ and let $\eta : \text{Spec} K \to S$ be the generic point of $S$. Let us fix an algebraic closure $\overline{k}$ of $k$ and let $\overline{\eta}$ be the corresponding generic geometric point of $S$.

By contravariance, a relative stratified sheaf $\{E_i, \sigma_i\}$ on $X/S$ restricts to a relative stratified sheaf $\{E_i, \sigma_i\}_{|X_s}$ in fibers $X_s$ for $s$ a point of $S$. We are interested in the relation between $\{E_i, \sigma_i\}_{|X_s}$ and $\{E_i, \sigma_i\}_{|X_s}$ for closed points $s \in |S|$. More precisely, we want to understand under which assumptions the finiteness of $\{E_i, \sigma_i\}_{|X_s}$ for all closed points $s \in |S|$ implies the finiteness of $\{E_i, \sigma_i\}_{|X_\eta}$. Recall that finiteness of $E \subset \text{Strat}(X)$ means that all objects of $\langle E \rangle$ are subquotients in $\text{Strat}(X)$ of direct sums of a single object, which is equivalent to saying that after the choice of a rational point, the monodromy group of $E$ is finite ([DM Proposition 2.20 (a)]).

Let $X$ be a smooth variety defined over $\mathbb{F}_q$ with $q = p^r$. For all $n \in \mathbb{N} \setminus \{0\}$, one has the commutative diagram

\[
\begin{array}{ccc}
X & \xrightarrow{\phi^m} & X^{(m)} \\
\downarrow \text{Spec } \mathbb{F}_q & & \downarrow \text{Spec } \mathbb{F}_q \\
\text{Spec } \mathbb{F}_q & \xrightarrow{\phi^m_{|\mathbb{F}_q}} & \text{Spec } \mathbb{F}_q
\end{array}
\]

which allows us to identify $X^{(m)}$ with $X$ (as an $\mathbb{F}_q$-scheme).

Let $S$ be an $\mathbb{F}_q$ connected scheme, with field of constants $k$, i.e. $k$ is the normal closure of $\mathbb{F}_q$ in $H^0(S, \mathcal{O}_X)$. We define $X_S := X \times_{\mathbb{F}_q} S$.

**Proposition 1.3.** Let $E$ be a vector bundle on $X_S$. Assume that there exists a positive integer $n$ such that we have an isomorphism

\[
\tau : ((F^r \times_{\mathbb{F}_q} \text{id})^n)^* E \simeq E.
\]

Then $E$ has a natural stratification $E\tau = \{E_i, \sigma_i\}$, $E_0 = E$ relative to $S$. 

\[
(1)
\]
Proof. We define
\[ E_{rn} = (W_{X/F_q}^m \times \mathbb{F}_q^{\text{id}_S})^* E. \]
Then we use the factorization
\[ \begin{array}{cccccc}
X & \xrightarrow{F_X/F_q} & X(1) & \xrightarrow{F_X/(1)/F_q} & X(rn-1) & \xrightarrow{F_X/(rn)/F_q} & X(rn) \\
& & & & \downarrow & \downarrow & \\
& & & & \text{Spec } F_q & \downarrow & \\
\end{array} \]
(4)
of \( F_{X/F_q}^m \) and we define
\[ E_{nr-1} = (F_{X/(rn-1)/F_q}^m \times \mathbb{F}_q^{\text{id}_S})^* E_{rn}, \ldots, E_1 = (F_{X/(1)/F_q}^m \times \mathbb{F}_q^{\text{id}_S})^* E_2 \]
(5)
with identity isomorphisms \( \sigma_{nr-1}, \ldots, \sigma_1 \). Then we use the isomorphism \( \tau \) to define
\[ \sigma_0 : E \simeq (F_{X/F_q}^m \times \mathbb{F}_q^{\text{id}_S})^* E_1. \]
(6)
Assume we constructed the bundles \( E_i \) on \( X^{(i)} \) for all \( i \leq arn \) for some integer \( a \geq 1 \).
We now replace the diagram (1) by the diagram
\[ \begin{array}{cccccc}
X^{(arn)} & \xrightarrow{F_{X^{(arn)}/F_q}^m} & X^{((a+1)rn)} & \xrightarrow{W_{X^{(arn)}/F_q}^m} & X^{(arn)} \\
& & & & \downarrow & \downarrow & \\
& & & & \text{Spec } F_q & \downarrow & \\
\end{array} \]
(7)
We then define
\[ E_{(a+1)rn} = (W_{X^{(arn)}/F_q}^m \times \mathbb{F}_q^{\text{id}_S})^* E_{arn} \]
(8)
(which is equal to \( E \) under identification of \( X^{(arn)} \) with \( X \)). Then we use the factorization
\[ \begin{array}{cccccc}
X^{(arn)} & \xrightarrow{F_{X^{(arn)}/F_q}^m} & X^{(arn+1)} & \xrightarrow{F_{X^{(arn+1)}/F_q}^m} & \ldots & \xrightarrow{F_{X^{((a+1)rn-1)/F_q}^m}} & X^{((a+1)rn-1)} & \xrightarrow{F_{X^{((a+1)rn)/F_q}^m}} & X^{((a+1)rn)} \\
& & & & \downarrow & \downarrow & \downarrow & \downarrow & \\
& & & & \text{Spec } F_q & \downarrow & \\
\end{array} \]
(9)
of \( F_{X^{(arn)}/F_q}^m \) to define
\[ E_{(a+1)rn-1} = (F_{X^{((a+1)rn-1)/F_q}^m} \times \mathbb{F}_q^{\text{id}_S})^* E_{(a+1)rn}, \ldots, \]
\[ E_{arn+1} = (F_{X^{(arn+1)}/F_q}^m \times \mathbb{F}_q^{\text{id}_S})^* E_{arn+2} \]
(10)
with identity isomorphisms \( \sigma_{(a+1)nr-1} \), \ldots, \( \sigma_{arn+1} \). Then we again use \( \tau \) to define

\[
\sigma_{arn} : E_{arn} \cong (F_{\omega}^{\infty})^{*}E_{arn+1}. 
\]

(11)

The above construction and [Gi] Proposition 1.7 imply

**Proposition 1.4.** Assume in addition to (2) that \( X \) is proper and \( \mathbb{F}_q \subset k \subset \overline{\mathbb{F}}_q \). Fix a rational point \( x \in X_0(k) \). Then for any closed point \( s \in |S| \), the Tannaka group scheme 
\[
\pi(\mathbb{E}_S, \omega_S(k,s)) \text{ of } \mathbb{E}_S := \mathbb{E}_X\big|_S \text{ over the residue field } k(s) \text{ of } s \text{ is finite.}
\]

**Proof.** The bundle \( E \) is base changed of a bundle \( E^0 \) defined over \( X \times_{\mathbb{F}_q} S_0 \) for some form \( S_0 \) of \( S \) defined over a finite extension \( \mathbb{F}_q' \) of \( \mathbb{F}_q \) such that \( x \) is base change of an \( \mathbb{F}_q' \)-rational point \( x_0 \) of \( X \times_{\mathbb{F}_q} S_0 \). We can also assume that \( \tau \) comes by base change from \( \tau_0 : ((F^r \times_{\mathbb{F}_q} id_{S_0})^0)^0 \cong E^0 \). Proposition 1.3 yields then a relative stratification \( \mathbb{E}_0 \circlearrowleft = (E^0_0, \alpha_0^0) \) of \( E^0 \) defined over \( \mathbb{F}_q' \), with \( E_i = E^0_0 \circlearrowleft F_0^i \). A closed point \( s \) of \( S = S_0 \times_{\mathbb{F}_q} k \) is a base change of some closed point \( s_0 \) of \( S_0 \) of degree \( b \) say over \( \mathbb{F}_q'^b \). By Corollary 1.2 we just have to show that \( \pi(\mathbb{E}_S, \omega_S(k_0)) \) is finite. So we assume that \( k = \mathbb{F}_q \), \( S = S_0 \), \( s = s_0 \). The underlying bundles of \( \mathbb{E}_S \) and \( \mathbb{E}_S' \) are by construction all isomorphic for \( m = ab \). Thus by [Gi] Proposition 1.7, \( \mathbb{E}_\tau \cong \mathbb{E}_{\tau^m} \) in \( \text{Strat}(X/k) \). But this implies that \( F_{\mathbb{F}_q}'(\mathbb{E}_\tau) \cong \mathbb{E}_{\tau^m} \). Thus \( E \) is algebraically trivializable on the Lang torsor \( h : Y \to X \times_{\mathbb{F}_q} \mathbb{F}_q' \) and the bundles \( E_i \) are trivializable on \( Y \times_{X \times_{\mathbb{F}_q} \mathbb{F}_q'} X^{(i)} = Y^{(i)}/\mathbb{F}_q' \). Thus the stratified bundle \( h^*\mathbb{E}_\tau \) on \( Y \) relative to \( \mathbb{F}_q' \) is trivial. We apply Lemma 1.4 to finish the proof.

2 Étale trivializable bundles

Let \( X \) be a smooth projective variety over an algebraically closed field \( k \). Let \( F_X : X \to X \) be the absolute Frobenius morphism.

A locally free sheaf on \( X \) is called étale trivializable if there exists a finite étale covering of \( X \) on which \( E \) becomes trivial. Note that if \( E \) is étale trivializable then it is numerically flat (see Definition 2.2 and the subsequent discussion). In particular, stability and semistability for such bundles are independent of a polarization (and Gieseker and slope stability and semistability are equivalent). More precisely, such \( E \) is stable if and only if it does not contain any locally free subsheaves of smaller rank and degree 0 (with respect to some or equivalently to any polarization).

**Proposition 2.1.** (see [S]) If there exists a positive integer \( n \) such that \( (F_X^n)^*E \cong E \) then \( E \) is étale trivializable. Moreover, if \( k = \overline{\mathbb{F}}_p \) then \( E \) is étale trivializable if and only if there exists a positive integer \( n \) and an isomorphism \( (F_X^n)^*E \cong E \).

**Proposition 2.2.** (see [BD]) If there exists a finite degree \( d \) étale Galois covering \( f : Y \to X \) such that \( f^*E \) is trivial and \( E \) is stable, then one has an isomorphism \( \alpha : (F_X^d)^*E \cong E \).
As a corollary we see that a line bundle on $X/k$ is étale trivializable if and only if it is torsion of order prime to $p$. One implication follows from the above proposition. The other one follows from the fact that $(F^i_k)^*L \simeq L$ is equivalent to $L^{0(p^i-1)} \simeq \mathcal{O}_X$ and for any integer $n$ prime to $p$ we can find $d$ such that $p^d - 1$ is divisible by $n$.

We recall that if $E$ is any vector bundle on $X$ such that there is a $d \in \mathbb{N} \setminus \{0\}$ and an isomorphism $\alpha : (F^i_k)^*E \cong E$, then $E$ carries an absolute stratified structure $E_{\alpha}$, i.e. a stratified structure relative to $\mathbb{F}_p$ by the procedure of Proposition 1.3. On the other hand, any stratified stratified structure $\{E_i, \sigma_i\}$ relative to $\mathbb{F}_p$ induces in an obvious way a stratified structure relative to $k$: the absolute Frobenius $F^q : X \to X$ factors through $W^q_{X/k} : X^{(n)} \to X$, so $\{W^q_{X/k}\}^*E_n, (W^q_{X/k})^* \sigma_n\}$ is the relative stratified structure, denoted by $E_{\alpha/k}$. Proposition 2.2 together with Lemma 1.1 show

**Corollary 2.3.** Under the assumptions of Proposition 2.2 we can take $d = \text{length}_k(\pi((E_{\alpha/k}, \sigma_h))).$

Let us also recall that there exist examples of étale trivializable bundles such that $(F^i_k)^*E \not\cong E$ for every positive integer $n$ (see Laszlo’s example in [BD]).

**Proposition 2.4.** (Deligne; see [Ls] 3.2) Let $X$ be an $\mathbb{F}_p$-scheme. If $G$ is a connected linear algebraic group defined over a finite field $\mathbb{F}_p$ then the embedding $G(\mathbb{F}_p) \hookrightarrow G$ induces an equivalence of categories between the category of $G(\mathbb{F}_p)$-torsors on $X$ and $G$-torsors $P$ over $X$ with an isomorphism $(F^q_k)^*P \simeq P$. In particular, if $G$ is a connected reductive algebraic group defined over an algebraically closed field $k$ and $P$ is a principal $G$-bundle on $X/k$ such that there exists an isomorphism $(F^q_k)^*P \simeq P$ for some natural number $n > 0$, then there exists a Galois étale cover $f : Y \to X$ with Galois group $G(\mathbb{F}_p)$ such that $f^*P$ is trivial. Indeed, every reductive group has a $\mathbb{Z}$-form so we can use the above proposition.

## 3 Preliminaries on Relative Moduli Spaces of Sheaves

Let $S$ be a scheme of finite type over a universally Japanese ring $R$. Let $f : X \to S$ be a projective morphism of $R$-schemes of finite type with geometrically connected fibers and let $\mathcal{O}_X(1)$ be an $f$-very ample line bundle.

A family of pure Gieseker semistable sheaves on the fibres of $X_T = X \times_S T \to T$ is a $T$-flat coherent $\mathcal{O}_{X_T}$-module $E$ such that for every geometric point $t$ of $T$ the restriction of $E$ to the fibre $X_t$ is pure (i.e., all its associated points have the same dimension) and Gieseker semistable (which is semistability with respect to the growth of the Hilbert polynomial of subsheaves defined by $\mathcal{O}_{X_T}(1)$ (see [HL] 1.2]). We introduce an equivalence relation $\sim$ on such families in the following way. $E \sim E'$ if and only if there exist filtrations $0 = E_0 \subset E_1 \subset \cdots \subset E_m = E$ and $0 = E'_0 \subset E'_1 \subset \cdots \subset E'_m = E'$ by coherent $\mathcal{O}_{X_T}$-modules such that $\oplus_{i=0}^m E_i / E_{i-1}$ is a family of pure Gieseker semistable sheaves on the fibres of $X_T$ and there exists an invertible sheaf $L$ on $T$ such that $\oplus_{i=1}^m E'_i / E'_{i-1} \simeq (\oplus_{i=1}^m E_i / E_{i-1}) \otimes \mathcal{O}_T L$.

Let us define the moduli functor

$$\mathcal{M}_p(X/S) : \text{Sch}/S^{op} \to \text{Sets}$$
from the category of locally noetherian schemes over $S$ to the category of sets by

$$\mathcal{M}_P(X/S)(T) = \left\{ \begin{array}{l}
\sim \text{ equivalence classes of families of pure Gieseker} \\
\text{semistable sheaves on the fibres of } T \times_S X \to T, \\
\text{which have Hilbert polynomial } P.
\end{array} \right\}.$$ 

Then we have the following theorem (see [La1 Theorem 0.2]).

**Theorem 3.1.** Let us fix a polynomial $P$. Then there exists a projective $S$-scheme $M_P(X/S)$ of finite type over $S$ and a natural transformation of functors

$$\theta : \mathcal{M}_P(X/S) \to \text{Hom}_S(\cdot, M_P(X/S)),$$

which uniformly corepresents the functor $\mathcal{M}_P(X/S)$. For every geometric point $s \in S$ the induced map $\theta(s)$ is a bijection. Moreover, there is an open scheme $M^s_k(P) \subset M_P(X/S)$ that universally corepresents the subfunctor of families of geometrically Gieseker stable sheaves.

Let us recall that $M_P(X/S)$ uniformly corepresents $\mathcal{M}_P(X/S)$ means that for every flat base change $T \to S$ the fiber product $M_P(X/S) \times_S T$ corepresents the fiber product functor $\text{Hom}_S(\cdot, T) \times_{\text{Hom}_S(\cdot, S)} \mathcal{M}_P(X/S)$. For the notion of corepresentability, we refer to [HL] Definition 2.2.1. In general, for every $S$-scheme $T$ we have a well defined morphism $M_P(X/S) \times_S T \to M_P(X_T/T)$ which for a geometric point $T = \text{Spec } k(s) \to S$ is bijection on points.

The moduli space $M_P(X/S)$ in general depends on the choice of polarization $\mathcal{O}_X(1)$.

**Definition 3.2.** Let $k$ be a field and let $Y$ be a projective $k$-variety. A coherent $\mathcal{O}_Y$-module $E$ is called numerically flat, if it is locally free and both $E$ and its dual $E^+ = \mathcal{H}^0(E, \mathcal{O}_Y)$ are numerically effective on $Y \otimes k$, where $k$ is an algebraic closure of $k$.

Assume that $Y$ is smooth. Then a numerically flat sheaf is strongly slope semistable of degree 0 with respect to any polarization (see [La2 Proposition 5.1]). But such a sheaf has a filtration with quotients which are numerically flat and slope stable (see [La2 Theorem 4.1]). Let us recall that a slope stable sheaf is Gieseker stable and any extension of Gieseker semistable sheaves with the same Hilbert polynomial is Gieseker semistable. Thus a numerically flat sheaf is Gieseker semistable with respect to any polarization.

Let $P$ be the Hilbert polynomial of the trivial sheaf of rank $r$. In case $S$ is a spectrum of a field we write $M_X(r)$ to denote the subscheme of the moduli space $M_P(X/k)$ corresponding to locally free sheaves. For a smooth projective morphism $X \to S$ we also define the moduli subscheme $M(X/S, r) \to S$ of the relative moduli space $M_P(X/S)$ as a union of connected components which contains points corresponding to numerically flat sheaves of rank $r$. Note that in positive characteristic numerical flatness is not an open condition. More precisely, on a smooth projective variety $Y$ with an ample divisor $H$, a locally free sheaf with numerically trivial Chern classes, that is with Chern classes $c_i$ in the Chow group of codimension $i$ cycles intersecting trivially $H^{\dim(Y) - i}$
for all \(i \geq 1\), is numerically flat if and only if it is strongly slope semistable (see [La2, Proposition 5.1]).

By definition for every family \(E\) of pure Gieseker semistable sheaves on the fibres of \(X_T\) we have a well defined morphism \(\varphi_E = \theta([E]) : T \to M_p(X/S)\), which we call a classifying morphism.

**Proposition 3.3.** Let \(X\) be a smooth projective variety defined over an algebraically closed field \(k\) of positive characteristic. Let \(S\) be a \(k\)-variety and let \(E\) be a rank \(r\) locally free sheaf on \(X \times_k S\) such that for every \(s \in S(k)\) the restriction \(E_s\) is Gieseker semistable with numerically trivial Chern classes. Assume that the classifying morphism \(\varphi_E : S \to M_k(r)\) is constant and for a dense subset \(S' \subset S(k)\) the bundle \(E_s\) is étale trivializable for \(s \in S'\). Then \(E_{\bar{s}}\) is étale trivializable.

**Proof.** If \(E_s\) is stable for some \(k\)-point \(s \in S\) then there exists an open neighbourhood \(U\) of \(\varphi_E(s)\), a finite étale morphism \(U' \to U\) and a locally free sheaf \(\mathcal{V}\) on \(X \times_k U'\) such that the pull backs of \(E\) and \(\mathcal{V}\) to \(X \times_k (\varphi_E^{-1}(U) \times U')\) are isomorphic (this is called existence of a universal bundle on the moduli space in the étale topology). But \(\varphi_E(S)\) is a point, so this proves that there exists a vector bundle on \(X\) such that \(E\) is its pull back by the projection \(X \times_k S \to X\). In this case the assertion is obvious.

Now let us assume that \(E_s\) is not stable for all \(s \in S(k)\). If \(0 = E_0^s \subset E_1^s \subset \ldots \subset E_m^s = E_s\) is a Jordan–Hölder filtration (in the category of slope semistable torsion free sheaves), then by assumption the isomorphism classes of semi-simplifications \(\oplus_{i=1}^{m} E_i^s / E_{i-1}^s\) do not depend on \(s \in S(k)\). Let \((r_1, \ldots, r_m)\) denote the sequence of ranks of the components \(E_i^s / E_{i-1}^s\) for some \(s \in S(k)\). Since there is only finitely many such sequences (they differ only by permutation), we choose some permutation that appears for a dense subset \(S' \subset S\).

Now let us consider the scheme of relative flags \(f : \text{Flag}(E/S; P_1, \ldots, P_m) \to S\), where \(P_i\) is the Hilbert polynomial of \(E_i^s\). By our assumption the image of \(f\) contains \(S'\). Therefore by Chevalley’s theorem it contains an open subscheme \(U\) of \(S\). Let us recall that the scheme of relative flags \(\text{Flag}(E[1\times_k U]; P_1, \ldots, P_m) \to U\) is projective. In particular, using Bertini’s theorem (\(k\) is algebraically closed) we can find a generically finite morphism \(W \to U\) factoring through this flag scheme. Let us consider pull back of the universal filtration \(0 = F_0 \subset F_1 \subset \ldots \subset F_m = E_W\) to \(X \times_k W\). Note that the quotients \(F_i = F_i / F_{i-1}\) are \(W\)-flat and by shrinking \(W\) we can assume that they are families of Gieseker stable locally free sheaves (since by assumption \(F_i\) is Gieseker stable and locally free for some points \(s \in W(k) \cap S'\)). This and the first part of the proof implies that \(E_{\bar{s}}\) has a filtration by subbundles such that the associated graded sheaf is étale trivializable. By Lemma 5.2 this implies that \(E_{\bar{s}}\) is étale trivializable. \(\square\)

4 Laszlo’s example

Let us describe Laszlo’s example of a line in the moduli space of bundles on a curve fixed by the second Verschiebung morphism (see [La, Section 3]).

Let us consider a smooth projective genus 2 curve \(X\) over \(\mathbb{F}_2\) with affine equation

\[y^2 + x(x+1)y = x^3 + x^2 + x.\]
In this case the moduli space $M_X(2, \mathcal{O}_X)$ of rank 2 vector bundles on $X$ with trivial determinant is an $\mathbb{P}^2$-scheme isomorphic to $\mathbb{P}^3$. The pull back of bundles by the relative Frobenius morphism defines the Verschiebung map

$$V : M_{X^{(1)}}(2, \mathcal{O}_{X^{(1)}}) \simeq \mathbb{P}^3 \dashrightarrow M_X(2, \mathcal{O}_X) \simeq \mathbb{P}^3$$

which in appropriate coordinates can be described as

$$[a : b : c : d] \mapsto [a^2 + b^2 + c^2 + d^2 : ab + cd : ac + bd : ad + bc].$$

The restriction of $V$ to the line $\Delta \cong \mathbb{P}^1$ given by $b = c = d$ is an involution and it can be described as $[a : b] \mapsto [a + b : b]$.

Using a universal bundle on the moduli space (which exists locally in the étale topology around points corresponding to stable bundles) and taking a finite covering $\Delta \rightarrow X$ we obtain the following theorem:

**Theorem 4.1.** ([Ls Corollary 3.2]) There exist a smooth quasi-projective curve $S$ defined over some finite extension of $\mathbb{F}_2$ and a locally free sheaf $E$ of rank 2 on $X \times S$ such that $(F^2 \times \text{id}_S)^*E \simeq E$, $\det E \simeq \mathcal{O}_{X \times S}$ and the classifying morphism $\varphi_E : S \rightarrow M_X(2, \mathcal{O}_X)$ is not constant. Moreover, one can choose $S$ so that $E_s$ is stable for every closed point $s$ in $S$.

Now note that the map $(F_2)^* : M_X(2, \mathcal{O}_X) \rightarrow M_X(2, \mathcal{O}_X)$ defined by pulling back bundles by the absolute Frobenius morphism can be described on $\Delta$ as $[a : b] \mapsto [a^2 + b^2 : b^2]$. In particular, the map $(F_2^\infty)^*|_{\Delta}$ is described as $[a : b] \mapsto [a^{2^n}, b^{2^n}]$. It follows that if a stable bundle $E$ corresponds to a modular point of $\Delta(\mathbb{P}^2_2) \setminus \Delta(\mathbb{P}^2_{2-1})$ (or, equivalently, $E$ is defined over $\mathbb{F}_2$) then $(F_2^m)^*E \simeq E$ and $(F_2^m)^*E \not\simeq E$ for $0 < m < 2n$.

This implies that for $k = \mathbb{F}_2$ and for every $s \in S(k)$, the bundle $E_s$ which is the restriction to $X \times \mathbb{F}_2$ of the bundle $E$ from Theorem 4.1 is étale trivializable.

Let $X, S$ be varieties defined over an algebraically closed field $k$ of positive characteristic. Assume that $X$ is projective. Let us set $K = k(S)$. Let $\bar{\eta}$ be a generic geometric point of $S$.

**Proposition 4.2.** Let $E$ be a bundle on $X_S = X \times_k S \rightarrow S$ which is numerically flat on the closed fibres of $X_S = X \times_k S \rightarrow S$. Assume that for some $s \in S$ the bundle $E_s$ is stable and the classifying morphism $\varphi_E : S \rightarrow M_X(r)$ is not constant. Then $E_{\bar{\eta}} = E|_{X_{\bar{\eta}}}$ is not étale trivializable.

**Proof.** Assume that there exists a finite étale cover $\pi' : Y' \rightarrow X_{\bar{\eta}}$ such that $(\pi')^*E_{\bar{\eta}} \simeq \mathcal{O}_{Y'}^{\oplus r}$. As $k$ is algebraically closed, one has the base change $\pi_1(X) \rightarrow \pi_1(X_{\bar{k}})$ for the étale fundamental group ([SGA1 Exp. X, Cor.1.8]), so there exists a finite étale cover $\pi : Y \rightarrow X$ such that $\pi = \pi \otimes \bar{k}$. Hence there exists a finite morphism $T \rightarrow U$ over some open subset $U$ of $S$, such that $\pi_T^*E_T$ is trivial where $\pi_T = \pi \times_k \text{id}_T : Y \times_k T \rightarrow X \times_k T$ and $E_T =$ pull back by $X \times_k T \rightarrow X \times_k U$ of $E|_{X \times_k U}$.

So for any $k$-rational point $t \in T$, one has $\pi_*E_t \subset \mathcal{O}_T^{\oplus r}$, where $r$ is the rank of $E$. Hence $E_t \subset \pi_\ast \mathcal{O}_T^{\oplus r}$. Hence $E_t \subset \pi_\ast \mathcal{O}_T^{\oplus r}$, i.e., all the bundles $E_t$ lie in one fixed bundle $\pi_\ast \mathcal{O}_T^{\oplus r}$. 
Since \( \pi \) is étale, the diagram

\[
\begin{array}{ccc}
Y & \xrightarrow{F_Y} & Y \\
\downarrow{\pi} & & \downarrow{\pi} \\
X & \xrightarrow{F_X} & X
\end{array}
\]

is cartesian (see, e.g., \cite{SGA5} Exp. XIV, §1, Prop. 2). Since \( X \) is smooth, \( F_X \) is flat. By flat base change we have isomorphisms \( F_Y^*(\pi_*\mathcal{O}_Y) \cong \pi_*(F_Y^*\mathcal{O}_Y) \cong \pi_*\mathcal{O}_Y \). In particular, this implies that \( \pi_*\mathcal{O}_Y \) is strongly semistable of degree 0. Therefore if \( E_i \) is stable then it appears as one of the factors in a Jordan–Hölder filtration of \( \pi_*\mathcal{O}_Y \). Since the direct sum of factors in a Jordan–Hölder filtration of a semistable sheaf does not depend on the choice of the filtration, there are only finitely many possibilities for the isomorphism classes of stable sheaves \( E_i \) for \( i \in T(k) \).

It follows that in \( U \subset S \) there is an infinite sequence of \( k \)-rational points \( s_i \) with the property that \( E_{s_i} \) is stable (since stability is an open property) and \( E_{s_i} \cong E_{s_{i+1}} \). This contradicts our assumption that the classifying morphism \( \varphi_E \) is not constant.

**Corollary 4.3.** There exist smooth curves \( X \) and \( S \) defined over an algebraic closure \( k \) of \( \mathbb{P}^2 \) such that \( X \) is projective and there exists a locally free sheaf \( E \) on \( X \times_k S \to S \) such that for every \( s \in S(k) \), the bundle \( E_s \) is étale trivializable but \( E_{\bar{s}} \) is not étale trivializable. Moreover, on \( E \) there exists a structure of a relatively stratified sheaf \( \mathbb{E} \) such that for every \( s \in S(k) \), the bundle \( \mathbb{E}_s \) has finite monodromy but the monodromy group of \( \mathbb{E}_{\bar{s}} \) is infinite.

The second part of the corollary follows from Proposition 4.3. The above corollary should be compared to the following fact:

**Proposition 4.4.** Let \( X \) be a projective variety defined over an algebraically closed field \( k \) of positive characteristic. Let \( S \) be a \( k \)-variety and let \( E \) be a rank \( r \) locally free sheaf on \( X \times_k S \rightarrow S \). Assume that there exists a positive integer \( n \) such that for every \( s \in S(k) \) we have \( (F_X^n)^*E_s \cong E_s \), where \( F_X \) denotes the absolute Frobenius morphism. Then the classifying morphism \( \varphi_E : S \to M_X(r) \) is constant and \( E_{\bar{s}} \) is étale trivializable.

**Proof.** By Proposition 2.1 if \( (F_X^n)^*E_s \cong E_s \) then there exists a finite étale Galois cover \( \pi_i : Y_i \rightarrow X \) with Galois group \( G = \text{GL}_r(\mathbb{F}_p) \) such that \( \pi_i^*E_s \) is trivial (in this case it is essentially due to Lange and Stuhler; see \cite{LS}). This implies that \( E_s \subset \langle \pi_i \rangle, \pi_i^*E_s \cong (\langle \pi_i \rangle, \mathcal{O}_Y)^{\otimes r} \) and hence \( \text{gr}_{JH}E_s \subset (\text{gr}_{JH}(\pi_i, \mathcal{O}_Y))^{\otimes r} \).

Since \( X \) is proper, the étale fundamental group of \( X \) is topologically finitely generated and hence there exists only finitely many finite étale coverings of \( X \) of fixed degree (up to an isomorphism). This theorem is known as the Lang–Serre theorem (see \cite{LS} Théorème 4)). Let \( \mathcal{S} \) be the set of all Galois coverings of \( X \) with Galois group \( G \). Then for every closed \( k \)-point \( s \) of \( S \) the semi-simplification of \( E_s \) is contained in \( (\text{gr}_{JH} \alpha_* \mathcal{O}_Y)^{\otimes r} \) for some \( \alpha \in \mathcal{S} \). Therefore there are only finitely many possibilities for images of \( k \)-points \( s \) in \( M_X(r) \). Since \( S \) is connected, it follows that \( \varphi_E : S \to M_X(r) \) is constant.
The remaining part of the proposition follows from Proposition 3.3.

Note that by Proposition 4.2 together with Corollary 2.3, the monodromy groups of $E_s$ in Theorem 4.1 for $s \in S(k)$ are not uniformly bounded. In fact, only if $k$ is an algebraic closure of a finite field do we know that the monodromy groups of $E_s$ are finite because then $E_s$ can be defined over some finite subfield of $k$ and the isomorphism $(F^2)^* E_s \simeq E_s$ implies that for some $n$ we have $(F^n)^* E_s \simeq E_s$ (see the paragraph following Theorem 4.1).

Moreover, the above proposition shows that in Theorem 4.1, we cannot hope to replace $F$ with the absolute Frobenius morphism $F_X$.

5 Analogue of the Grothendieck-Katz conjecture in positive equicharacteristic

As Corollary 4.3 shows, the positive equicharacteristic version of the Grothendieck–Katz conjecture which requests a relatively stratified bundle to have finite monodromy group on the geometric generic fiber once it does on all closed fibers, does not hold in general. But one can still hope that it holds for a family of bundles coming from representations of the prime-to-$p$ quotient of the étale fundamental group. In this section we follow André’s approach [An, Théorème 7.2.2] in the equicharacteristic zero case to show that this is indeed the case.

Let $k$ be an algebraically closed field of positive characteristic $p$. Let $f : X \to S$ be a smooth projective morphism of $k$-varieties (in particular, integral $k$-schemes). Let $\eta$ be the generic point of $S$. In particular, $X_{\eta}$ is smooth (see [SGA1, Defn 1.1]).

**Theorem 5.1.** Let $E$ be a locally free sheaf of rank $r$ on $X$. Let us assume that there exists a dense subset $U \subset S(k)$ such that for every $s \in U$, there is a finite étale covering $\pi_s : Y_s \to X_s$ of Galois group of order prime-to-$p$ such that $\pi_s^* (E_s)$ is trivial.

1) Then there exists a finite Galois étale covering $\pi_\eta : Y_\eta \to X_\eta$ of order prime-to-$p$ such that $\pi_\eta^* E_\eta$ is a direct sum of line bundles.

2) If $k$ is not algebraic over its prime field and $U$ is open in $S$, then $E_\eta$ is étale trivializable on a finite étale cover $Z_\eta \to X_\eta$ which factors as a Kummer (thus finite abelian of order prime to $p$) cover $Z_\eta \to Y_\eta$ and a Galois cover $Y_\eta \to X_\eta$ of order prime to $p$.

**Proof.** Without loss of generality, shrinking $S$ if necessary, we may assume that $S$ is smooth. Moreover, by passing to a finite cover of $S$ and replacing $U$ by its inverse image, we can assume that $f$ has a section $\sigma : S \to X$.

By assumption for every $s \in U$ there exists a finite étale Galois covering $\pi_s : Y_s \to X_s$ with Galois group $\Gamma_s$ of order prime-to-$p$ and such that $\pi_s^* E_s$ is trivial. To these data one can associate a representation $\rho_s : \pi_s^*(\text{trace of } \sigma(s)) \to \Gamma_s \subset \text{GL}_r(k)$ of the prime-to-$p$ quotient of the étale fundamental group.
By the Brauer–Feit version of Jordan’s theorem (see [BF, Theorem]) there exist a constant $j(r)$ such that $\Gamma_s$ contains an abelian normal subgroup $A_s$ of index $\leq j(r)$ (here we use assumption that the $p$-Sylow subgroup of $\Gamma$ is trivial).

For a $k$-point $s$ of $S$ we have a homomorphism of specialization
\[ \alpha_s : \pi_1(X, \sigma(\bar{\eta})) \to \pi_1(X, \sigma(s)), \]
which induces an isomorphism of the prime-to-$p$ quotients of the étale fundamental groups.

So for every $s \in U$ we can define the composite morphism
\[ \tilde{\rho}_s : \pi'_1(X, \sigma(\bar{\eta})) \xrightarrow{\alpha_s} \pi'_1(X, \sigma(s)) \xrightarrow{\tilde{\rho}} \Gamma_x \to \Gamma_x/A_s. \]

Let $K$ be the kernel of the canonical homomorphism $\pi_s : \pi_1(X, \sigma(\bar{\eta})) \to \pi_1(S, \bar{\eta})$, let $K'$ be its maximal pro-$p'$-quotient. Then by [SGA1, Exp. XIII, Proposition 4.3 and Exemples 4.4], one has $K' = \pi'_1(X, \sigma(\bar{\eta}))$, the maximal pro-$p'$-quotient of $\pi_1(X, \sigma(\bar{\eta}))$, and one has a short exact sequence
\[ \{1\} \to \pi'_1(X, \sigma(\bar{\eta})) \to \pi'_1(X, \sigma(\bar{\eta})) \xrightarrow{\pi} \pi_1(S, \bar{\eta}) \to \{1\}, \]
where $\pi'_1(X, \sigma(\bar{\eta}))$ is defined as the push-out of $\pi_1(X, \sigma(\bar{\eta}))$ by $K \to K'$.

Since $X_{\bar{\eta}}$ is proper, $\pi_1(X_{\bar{\eta}}, \sigma(\bar{\eta}))$ is topologically finitely generated. Therefore $\pi'_1(X, \sigma(\bar{\eta}))$ is also topologically finitely generated and hence it contains only finitely many subgroups of indices $\leq j(r)$. Let $G$ be the intersection of all such subgroups in $\pi'_1(X, \sigma(\bar{\eta}))$. It is a normal subgroup of finite index. Since $\ker(\tilde{\rho}_s)$ is a normal subgroup of index $\leq j(r)$ in $\pi'_1(X, \sigma(\bar{\eta}))$ we have
\[ G \subset \bigcap_{s \in U} \ker(\tilde{\rho}_s). \]

Now let us consider the commutative diagram
\[
\begin{array}{cccccc}
\pi_1(X_{\bar{\eta}}, \sigma(\bar{\eta})) & \longrightarrow & \pi_1(X, \sigma(\bar{\eta})) & \longrightarrow & \pi_1(S, \bar{\eta}) & \longrightarrow & \{1\} \\
\downarrow & & \downarrow & & \downarrow & & \\
\{1\} & \longrightarrow & \pi'_1(X_{\bar{\eta}}, \sigma(\bar{\eta})) & \longrightarrow & \pi'_1(X, \sigma(\bar{\eta})) & \longrightarrow & \pi_1(S, \bar{\eta}) & \longrightarrow & \{1\}
\end{array}
\]

Then $G \cdot \sigma(\pi_1(S, \bar{\eta})) \subset \pi'_1(X, \sigma(\bar{\eta}))$ is a subgroup of finite index. It is open by the Nikolov–Segal theorem [NS, Theorem 1.1]. So the pre-image $H$ of this subgroup under the quotient homomorphism $\pi_1(X, \sigma(\bar{\eta})) \to \pi_1(X, \sigma(\bar{\eta}))$ defines a finite étale covering $h : X' \to X$.

Let us take $s \in S(k)$. Since the composition
\[ H \subset \pi_1(X, \sigma(\bar{\eta})) \to \pi_1(X, \sigma(s)) \to \pi_1(S, s) \]
is surjective, the geometric fibres of $X' \to S$ are connected. Let us choose a $k$-point in $X'$ lying over $\sigma(s)$. By abuse of notation we call it $\sigma'(s)$. Similarly, let us choose a geometric point $\sigma'(\bar{\eta})$ of $X'_n$ lying over $\sigma(\bar{\eta})$. Then for any $s \in U$ we have the following commutative diagram:

\[
\begin{array}{cccc}
\pi'_1(X'_n, \sigma'(\bar{\eta})) & \xrightarrow{h_s} & \pi'_1(X_n, \sigma(\bar{\eta})) & \xrightarrow{G} \pi'_1(X_n, \sigma(\bar{\eta}))/G \\
\Downarrow \simeq & & \Downarrow \simeq & \\
\pi''_1(X'_n, \sigma'(s)) & \xrightarrow{h_s} & \pi''_1(X_n, \sigma(s)) & \xrightarrow{\Gamma_s/A_s} \Gamma_s/A_s
\end{array}
\]

This diagram shows that $\pi''_1(X'_n, \sigma'(s)) \to \Gamma_s$ factors through $A_s$ and hence $E'_s = (h^*E)_s$ is trivialized by a finite étale Galois covering. $\pi'_1(Y'_s \to X'_s)$ with an abelian Galois group of order prime to $p$, which is a subgroup of $A_s$. Since $E'_s \subset (\pi'_1)^*E'_s \simeq ((\pi'_1)^*O_{Y'_s})^{\oplus r}$, and $(\pi'_1)^*O_{Y'_s}$ is a direct sum of torsion line bundles of orders prime to $p$, it follows that for every $s \in U$ the bundle $E'_s$ is also a direct sum of torsion line bundles of order prime to $p$.

We consider the union $M(X'/S, r)$ of the components of $M_{\eta}(X'/S)$ containing moduli points of numerically flat bundles, as defined in Section 3. Let us consider the $S$-morphism $\psi : M(X'/S, 1)^{\times r} \to M(X'/S)$ given by $([L_1], \ldots, [L_r]) \mapsto [\square L]$ (in fact we give it by this formula on the level of functors; existence of the morphism follows from the fact that moduli schemes corepresent these functors). The bundle $E'$ gives us a section $\tau : S \to M(X'/S, r)$, and by the above for every $k$-rational point $s$ of $U$, the point $\tau(s)$ is contained in the image of $\psi$. Therefore $\tau(S)$ is contained in the image of $\psi$ as $\psi$ is projective (thus proper).

Let us consider the fibre product

\[
\begin{array}{ccc}
M(X'/S, 1)^{\times r} \times_{M(X'/S, r)} S & \to & S \\
\downarrow \tau & & \downarrow \tau \\
M(X'/S, 1)^{\times r} & \to & M(X'/S, r)
\end{array}
\]

Let us recall that in positive characteristic the canonical map $M(X' \times SS' \to S, r) \to M(X'/S, r) \times SS'$ need not be an isomorphism (although it is an isomorphism for $r = 1$). Anyway we can find an étale morphism $S' \to S$ over some non-empty open subset of $S$, such that there exists a map $\nu : S' \to M(X' \times SS' \to SS', 1)^{\times r}$ which composed with $M(X' \times SS' \to SS', 1)^{\times r} \to M(X' \times SS' \to SS, r) \to M(X'/S, r)$ gives the composition of $S' \to S$ with $\tau$. This shows that the pull back $E''_s$ of $E'$ to $X' \times SS'$ has a filtration whose quotients are line bundles which are of degree 0 on the fibres of $X' \times SS' \to SS'$.

Now let us note the following lemma:
Lemma 5.2. Let \( f : X \to S \) be a projective morphism of \( k \)-varieties. Let \( 0 \to G_1 \to G \to G_2 \to 0 \) be a sequence of locally free sheaves on \( X \). Assume that there exists a dense subset \( U \subset S(k) \) such that for each \( s \in U \) this sequence splits after restricting to \( X_s \). Then it splits on the fibre \( X_\eta \) over the generic point \( \eta \) of \( S \).

Proof. By shrinking \( S \) if necessary, we may assume that \( S \) is affine and the relative cohomology sheaf \( \mathcal{R}^1 p_* \mathcal{H}om(G_2, G_1) \) is locally free. The above short exact sequence defines a class \( \lambda \in \text{Ext}^1(G_2, G_1) \cong H^0(S, \mathcal{R}^1 f_* \mathcal{H}om(G_2, G_1)) \), such that \( \lambda(s) = 0 \) for every \( k \)-rational point \( s \) of \( U \). It follows that \( \lambda = 0 \) and hence the sequence is split over the generic point of \( S \).

Now let us note that on a smooth projective variety every short exact sequence of the form \( 0 \to G_1 \to G \to G_2 \to 0 \) in which \( G \) is a direct sum of line bundles of degree 0 and \( G_2 \) is a line bundle of degree 0 splits. So the filtration of \( E^n \) restricted to the closed fibers splits. Therefore the above lemma and easy induction show that \( E^n \eta \) is a direct sum of line bundles, where \( \eta' \) is the generic point of \( S' \). This shows the first part of the theorem.

To prove the second part of the theorem, we may assume that \( U = S \). Let us take a line bundle \( L \) on \( X \) such that for every \( k \)-rational point \( s \) the line bundle \( L_s \) is étale trivializable. We need to prove that there exists a positive integer \( n \) prime to \( p \) and such that \( L^{\otimes n}_\eta \cong \mathcal{O}_{X_\eta} \).

We thank the referee for showing us the following lemma.

Lemma 5.3. Let \( g : A \to S \) be an abelian scheme and let \( \sigma \) be a section of \( g \) such that for all \( s \in S(k) \), \( \sigma(s) \) is torsion of order prime to \( p \). Then \( \sigma \) is torsion of order prime to \( p \).

Proof. We may assume that \( S \) is normal and affine. Let us choose a subfield \( k' \subset k \) that is finitely generated and transcendental over \( F_p \), and such that \( A \to S \) and \( \sigma \) come by base change \( \text{Spec} k \to \text{Spec} k' \) from an abelian scheme \( g' : A' \to S' \) and a section \( \sigma' \) defined over \( k' \). Let \( m > 1 \) be prime to \( p \) and let \( \Gamma \) be the subgroup \( A'(S') \cap [m]^{-1}(\mathbb{Z}, \sigma') \) of \( A'(S') \). Then \( \Gamma \) is a finitely generated group. Note that assumptions of Néron’s specialization theorem [Chapter 9, Theorem 6.2] are satisfied and therefore there exists a Hilbert subset \( \Sigma \) of points \( s' \in S' \) for which the specialization map \( \mathcal{A}'(S') \to \mathcal{A}'_s(k(s')) \) is injective on \( \Gamma \). Since the Hilbert subset \( \Sigma \subset S' \) contains infinitely many closed points (see [Chapter 9, Theorems 5.1, 5.2 and 4.2]), there is a closed point \( s \in S \) the image of which in \( S' \) lies in \( \Sigma \). The specialization of \( \mathbb{Z}, \sigma \) at \( s \) is injective and hence \( \sigma \) is torsion of order dividing the order of \( \sigma(s) \), which is prime to \( p \).

Let us first assume that \( X \to S \) is of relative dimension 1. By passing to a finite cover of \( S \) we can assume that \( f \) has a section. The relative Picard scheme \( A = \text{Pic}^0(X/S) \to S \) is smooth. Using the above lemma to the section corresponding to the line bundle \( L \) we see that there exists some positive integer \( n \) prime to \( p \) and a line bundle \( M \) on \( S \) such that \( L^{\otimes n} \cong f^* M \). In particular, \( L^{\otimes n}_\eta \cong \mathcal{O}_{X_\eta} \).
Let us keep the notation from the beginning of the section, i.e., we see that the map $p$ closed field of positive characteristic $k$ of $X$. Therefore the family $(L^n_{\eta})_{n\in\mathbb{Z}}$ is bounded. Thus for any sufficiently ample divisor $H$ on $X_{\eta}$ we have $H^1(X_{\eta}, L^n_{\eta}(\eta)) = 0$ for all integers $n$. We consider such an $H$ which is defined over $\eta$.

Using Bertini’s theorem we can find a very ample divisor $Y \subset X$ in the linear system $|H|$ such that $f|_Y : Y \rightarrow S$ is smooth (possibly after shrinking $S$) and such that for every positive integer $n$ we have $H^1(X_{\eta}, L^{\otimes n}(-Y)|_{\eta}) = 0$. Indeed, shrinking $S$ and using semicontinuity of cohomology, we may assume that $H$ is defined over $S$, that the function $\dim H^0(X_{s}, \mathcal{O}_{X_{s}}(H))$ is constant and $S$ is affine. Let us choose a $k$-rational point $s$ in $S$. Then by Grauert’s theorem (see [Ha, Chapter III, Corollary 12.9]) the restriction map

$$H^0(X_{s}, \mathcal{O}_{X_{s}}(H)) \rightarrow H^0(X_{s}, \mathcal{O}_{X_{s}}(H))$$

is surjective. By Bertini’s theorem in the linear system $|\mathcal{O}_{X_{s}}(H)|$ there exists a smooth divisor. By the above we can lift it to a divisor $Y \subset X$, which after shrinking $S$ is the required divisor.

Applying our induction assumption to $L|_Y$ on $Y \rightarrow X$ we see that there exists a positive integer $n$ prime to $p$ such that $(L|_Y)^{\otimes n}_{\eta} = \mathcal{O}_{Y_{\eta}}$. Using the short exact sequence

$$0 \rightarrow L^{\otimes n}_{\eta}(-Y_{\eta}) \rightarrow L^{\otimes n}_{\eta} \rightarrow (L^{\otimes n}_{Y_{\eta}})_{\eta} \rightarrow 0$$

we see that the map

$$H^0(X_{\eta}, L^{\otimes n}_{\eta}) \rightarrow H^0(Y_{\eta}, (L^{\otimes n}_{Y_{\eta}})_{\eta})$$

is surjective. In particular, $L^{\otimes n}_{\eta}$ has a section and hence it is trivial.

Remarks 5.4. 1. Laszlo’s example shows that the first part of the theorem is false if one does not assume that orders of the monodromy groups of $E_s$ are prime to $p$ (in this example $E_{\eta}$ is a stable rank 2 vector bundle). Note that in this example, $E$ has even the richer structure of a relatively stratified bundle (see Proposition 1.3).

2. Let $E$ be a supersingular elliptic curve defined over $k = \overline{\mathbb{F}}_p$. Let $M$ be a line bundle of degree 0 and of infinite order on $E_{\overline{\mathbb{F}}_p(\ell)}$. Then one can find a smooth curve $S$ defined over $k$ such that there exists a line bundle $L$ on $X = S \times_k E \rightarrow S$ such that $L_{\eta} \simeq M$. In this example the line bundle $L_s$ is torsion for every $k$-rational point $s$ of $S$ as it is defined over a finite field. Since $E$ is a supersingular elliptic curve, there are no torsion line bundles of order divisible by $p$. In this case all line bundles $L_s$ for $s \in S(k)$ are étale trivializable (and the monodromy group has order prime to $p$).

This shows that the second part of Theorem 5.1 is no longer true if $k$ is an algebraic closure of a finite field.

Let us keep the notation from the beginning of the section, i.e., $k$ is an algebraically closed field of positive characteristic $p$ and $f : X \rightarrow S$ is a smooth projective morphism of $k$-varieties (in particular connected) with geometrically connected fibers. For simplicity, we also assume that $f$ has a section $\sigma : S \rightarrow X$. 

Documenta Mathematica 18 (2013) 23–50
Lemma 5.5. Let $E$ be a locally free sheaf on $X$. If there exists a point $s_0 \in S(k)$ such that $E_{s_0}$ is numerically flat then $E_\eta$ is also numerically flat. In particular, if there exists a point $s_0 \in S(k)$ such that there is a finite covering $\pi_{s_0} : X_{s_0} \to X_0$ such that $\pi_{s_0}^*(E_{s_0})$ is trivial, then $E_\eta$ is also numerically flat.

Proof. Let us fix a relatively ample line bundle. If $E_{s_0}$ is numerically flat then it is strongly semistable with numerically trivial Chern classes (see [La2, Proposition 5.1]). Since $E$ is $S$-flat, the restriction of $E$ to any fiber has numerically trivial Chern classes (as intersection numbers remain constant on fibers). Now note that for any $n$ the sheaf $(F^n_{X_0/k})^*E_{s_0}$ is slope semistable. Since slope semistability is an open property, it follows that $(F^n_{X_0/k})^*E_\eta$ is also slope semistable. By [HL, Corollary 1.3.8] it follows that $(F^n_{X_0/k})^*E_\eta$ is also slope semistable. Thus $E_\eta$ is strongly semistable with vanishing Chern classes and hence it is numerically flat by [La2, Proposition 5.1].

Let us recall that numerically flat sheaves on a proper $k$-variety $Y$ form a Tannakian category. A rational point $y \in Y(k)$ neutralizes it. Thus we can define $S$-fundamental group scheme of $Y$ at the point $y$ (see [La2, Definition 6.1]). For a numerically flat sheaf $E$ on $Y$, we consider the Tannaka $k$-group $\pi_S((E), y) := Aut^\otimes((E), y) \subset GL(E_y)$, where now $(E)$ is the full tensor subcategory of numerically flat bundles spanned by $E$. We call it the $S$-monodromy group scheme. Using this language we can reformulate Theorem 5.5 in the following way (for simplicity we reformulate only the second part of the theorem).

Theorem 5.6. Let $E$ be an $S$-flat family of numerically flat sheaves on the fibers of $X \to S$. Let us assume that $k$ is not algebraic over its prime field and there exists a non-empty open subset $U \subset S(k)$ such that for every $s \in U$, the $S$-monodromy group scheme $\pi_S((E_s), \sigma(s))$ is finite étale of order prime-to-$p$. Then $\pi_S((E_\eta), \sigma(\eta))$ is also finite étale.

6 Verschiebung divisible points on abelian varieties: on the theorem by M. Raynaud

Let $K$ be an arbitrary field of positive characteristic $p$ and let $A$ be an abelian variety defined over $K$. The multiplication by $p^n$ map $[p^n] : A \to A$ factors through the relative Frobenius morphism $F_{A/K}^n : A \to A^{(n)}$ and hence defines the Verschiebung morphism $V^n : A^{(n)} \to A$ such that $V^n F_{A/K}^n = [p^n]$.

Definition 6.1. A $K$-point $P$ of $A$ is said to be $V$-divisible if for every positive integer $n$ there exists a $K$-point $P_n$ in $A^{(n)}$ such that $V^n(P_n) = P$.

Let $T$ be an integral noetherian separated scheme of dimension 1 with field of rational functions $K$. Let us recall that a smooth, separated group scheme of finite type $\mathcal{G} \to T$ is called a Néron model of $A$ if the general fiber of $\mathcal{G} \to T$ is isomorphic to $A$ and for every smooth morphism $X \to T$, a morphism $X_K \to \mathcal{G}_K$ extends (then uniquely) to a $T$-morphism $X \to \mathcal{G}$.
Assume that the base field $K$ is the function field of a normal projective variety $S$ defined over a field $k$ of positive characteristic $p$.

We say that $A$ has a good reduction at a codimension 1 point $s \in S$ if the Néron model of $A$ over $\text{Spec} \mathcal{O}_{S,s}$ is an abelian scheme (the usual definition is slightly different as it assumes that the identity component of the special fibre of the Néron model is an abelian variety; it is equivalent to the above one by [BLR, 7.4, Theorem 5]). We say that $A$ has potential good reduction at a codimension 1 point $s \in S$ if there exists a finite Galois extension $K'$ over $K$ such that if $S'$ is the normalization of $S$ in $K'$ then $A_{K'}$ has good reduction at every codimension 1 point $s' \in S'$ lying over $s$.

We say that $A$ has (potential) good reduction if it has (potential) good reduction at every codimension 1 point of $S$. Assume that $A$ has good reduction at every codimension 1 point of $S$. Then there exists a big open subset $U \subset S$ (i.e., the codimension of the complement of $U$ in $S$ is $\geq 2$) and an abelian $U$-scheme $\mathfrak{A} \to U$. Note that the group $A(K)$ of $K$-points of $A$ is isomorphic via the restriction map to the group of rational sections $U \dashrightarrow \mathfrak{A}$ of $\mathfrak{A} \to U$ defined over some big open subset of $U$. The section corresponding to $P \in A(K)$ will be denoted by $\hat{P}: U \dashrightarrow \mathfrak{A}$.

Let $c \in \text{Pic} A$ be a class of a line bundle $L$. By the theorem of the cube $c$, satisfies the following equality:

$$m_1^{12}c - m_1^{12}c - m_2^{12}c + m_1^c + m_2^c + m_3^c = 0,$$

where $m_I$ for $I \subset \{1,2,3\}$ is the map $A \times_K A \times_K A \to A$ defined by addition over the factors in $I$. (In particular, $m_i$ is the $i$-th projection). Combining [MB, Chapter III, 3.1] (relying on the Proposition 1.2.1), the line bundle $L \in \text{Pic}(A)$ extends uniquely (at least if we fix a rigidification) to a line bundle $\hat{L}$ over $\mathfrak{A}$ such that the class $\hat{c} = [\hat{L}] \in \text{Pic}(\mathfrak{A})$ is cubical, i.e., satisfies the relation

$$\hat{m}_1^{12}\hat{c} - \hat{m}_1^{12}\hat{c} - \hat{m}_2^{12}\hat{c} + \hat{m}_1^c + \hat{m}_2^c + \hat{m}_3^c = 0,$$

where $V \subset U$ is a big open subset and where $\hat{m}_I$ for $I \subset \{1,2,3\}$ is the map $\mathfrak{A} \times_S \mathfrak{A} \times_S \mathfrak{A} \to \mathfrak{A}$ defined by addition over the factors in $I$.

Now let us choose an ample line bundle $H$ on $S$. Then the map $\hat{h}_c: A(K) \to \mathbb{Z}$ given by

$$\hat{h}_c(P) = \text{deg}_H(\hat{P} - \hat{0})^*\hat{c}$$

is well defined as $\hat{P}$ is defined on a big open subset of $S$ and $\hat{P}^*\hat{L}$ extends to a rank 1 reflexive sheaf on $S$. This map is the canonical (Néron–Tate) height of $A$ corresponding to $c$ (see [MB, Chapter III, Section 3]).

The following theorem was suggested to the authors by M. Raynaud (in the good reduction case over a curve $S$, and with a somewhat different proof).

**Theorem 6.2.** Assume that $A$ has potential good reduction. If $P \in A(K)$ is $V$-divisible and $c$ is symmetric then $\hat{h}_c(P) = 0$.

**Proof.** Let us first assume that $A$ has good reduction. By assumption there exists a $K$-point $P_n$ of $A^{(n)}$ such that $V^n(P_n) = P$. Since $\mathfrak{A} \to U$ is an abelian scheme, so is $\mathfrak{A}^{(n)} \to U$, thus $P_n$ is the restriction to $\text{Spec} K$ of $\hat{P}_n \in \mathfrak{A}^{(n)}(U)$. 

Let us factor the absolute Frobenius morphism $F^n_A$ into the composition of the relative Frobenius morphism $F^n_{A/K} : A \to A^{(n)}$ and $W_n : A^{(n)} \to A$. Let us set $c_n = W^n_n c$. Its cubical extension $\tilde{c}_n \in \text{Pic}(\omega^{(n)}_A)$, for some big open $V_n \subset U$, together with $H$ allows one to define $\tilde{h}_{c_n}(P_n)$ by the corresponding formula. Since $(F^n_A)^*c = p^n c$, we have $(F^n_{A/K})^*c_n = p^n c$. On the other hand, since $c$ is symmetric, we have $[p^n]^*c = p^{2n}c$ and hence $(F^n_{A/K})^*((V^n)^*c) = p^{2n}c$. Therefore

$$(F^n_{A/K})^*((V^n)^*c - p^n c_n) = 0.$$  

Since $F^n_{A/K}$ is an isogeny this implies that the class $d = (V^n)^*c - p^n c_n$ is torsion. By additivity and functoriality of the canonical height (see [Se, Theorem, p. 35]) we have

$$\tilde{h}_c(P) = \tilde{h}_{(V^n)^*c}(P_n) = \tilde{h}_{p^n c_n}(P_n) + \tilde{h}_d(P_n) = p^n \cdot \tilde{h}_{c_n}(P_n)$$

(note that additivity implies that $\tilde{h}_{md} = m\tilde{h}_d$, so since $md = 0$ for some $m$, we get $\tilde{h}_d = 0$). Therefore if $\tilde{h}_c(P) \neq 0$ then $|\tilde{h}_c(P)| \geq p^n$ and we get a contradiction if $n$ is sufficiently large.

Now let us consider the general case. Since there exist only finitely many codimension 1 points $s \in S$ at which $A$ has bad reduction, one can find a finite Galois extension $K'$ of $K$ such that if $S'$ is the normalization of $S$ in $K'$ then $A_{K'}$ has good reduction at every codimension 1 point $s' \in S'$. On the other hand, if $P \in A(K)$ is $V$-divisible on $A$, $P \otimes K' \subset A(K')$ is $V$ divisible on $A_{K'}$. Then by the above we have $h_{V^n}(P') = 0$ and functoriality of the canonical height implies that $\tilde{h}_c(P) = 0$.

**Remark 6.3.** It is an interesting problem whether Theorem 6.2 holds for an arbitrary abelian variety $A/K$. Its proof shows that one can use the semiabelian reduction theorem to reduce the general statement to the case when $A$ has semiabelian reduction (see [BLR, 7.4, Theorem 1]).

Now assume that $S$ is geometrically connected. Then the extension $k \subset K$ is regular (i.e., $K/k$ is separable and $k$ is algebraically closed in $K$). Let $(B, \tau)$ be the $K/k$-trace of the abelian $K$-variety $A$, where $B$ is an abelian $k$-variety and $\tau : B_k \to A$ is a homomorphism of abelian $K$-varieties (it exists by [Co] Theorem 6.2). Let us recall that by definition $(B, \tau)$ is a final object in the category of pairs consisting of an abelian $k$-variety and a $K$-map from the scalar $K$-extension of this variety to $A$.

Since the extension $k \subset K$ is regular, the kernel $K$-group scheme of $\tau$ is connected (with connected dual) ([Co] Theorem 6.12). Therefore $\tau$ is injective on $K$-points and in particular we can treat $B(k)$ as a subgroup of $A(K)$.

**Corollary 6.4.** Assume that $A$ has potential good reduction. If $P \in A(K)$ is $V$-divisible then $[P] \in (A(K)/B(k))_{\text{tors}}$. In particular, if $k$ is algebraically closed then $P \in B(k) + A(K)_{\text{tors}} \subset A(K)$.

**Proof.** We can choose the class $c \in \text{Pic}(A)$ so that it is ample and symmetric. Then the first part of the corollary follows from Theorem 6.2 and [Co, Theorem 9.15] (which is true for regular extensions $K/k$).
In this section we use the height estimate of the previous section and the fact that torsion stratified line bundles on a perfect field have order prime to $p$ (apply Proposition \ref{prop:height-estimation} together with Lemma \ref{lem:height-bound}).

7 Stratified bundles

To prove the second part take positive integer $m$ such that $mP = Q \in B(k)$. Since $k$ is algebraically closed, the set $B(k)$ is divisible and there exists $Q' \in B(k)$ such that $mQ' = Q$. Then $P = Q' + (P - Q')$, where $m(P - Q') = 0$.

Let us assume that the field $k$ is algebraically closed. It is an interesting question whether a $V$-divisible $K$-point $P$ of $A$ can be written as a sum of $Q + R$, where $Q \in B(k)$ and $R \in A(K)_{\text{tors}}$ is torsion of order prime-to-$p$.

By the Lang–Néron theorem (\cite[Theorem 2.1]{Co}), the groups $A^{(i)}(K)/B^{(i)}(k)$ are finitely generated. It follows that the groups $G_i = (A^{(i)}(K)/B^{(i)}(k))_{\text{tors}}$ are finite.

Note that the homomorphism $B(k) \to B^{(i)}(k)$ induced by $F_{i/k}$ is a bijection. One has a factorization $F_{i/k} : A(K^{1/p^i}) \to A^{(i)}(K) \to A^{(i)}(K^{1/p^i})$, inducing a bijection $A(K^{1/p^i}) \to A^{(i)}(K)$. Thus in particular,

$$F_i : A(K)/B(k) \to A^{(i)}(K)/B^{(i)}(k)$$

is injective.

Moreover, the Verschiebung morphism induces the homomorphisms

$$V_i : A^{(i)}(K)/B^{(i)}(k) \to A(K)/B(k)$$

such that $V_i F_i = p^i$ and $F_i V_i = p^i$. This shows that prime-to-$p$ torsion subgroups of groups $G_i$ are isomorphic and in particular have the same order $m$.

Now let us assume that orders of the $p$-primary torsion subgroups of the abelian groups $G_i$ are uniformly bounded by some $p^e$. Then for all $i \geq e$

$$F_i([mP]) = F_i(V_i([mP_0])) = p^i m P_0 = 0.$$  

This implies that $mP_0 = 0$, so $mP \in B(k)$. Now $B(k)$ is a divisible group so there exists some $Q' \in B(k)$ such that $mQ' = mQ$. Then $R = P - Q \in A(K)$ is torsion of order prime to $p$. So we conclude

\textbf{Lemma 6.5.} \textit{If the order of the $G_i$ is bounded as $i$ goes to infinity, under the assumption the Theorem \ref{prop:height-estimation} there exists a positive integer $m$, prime to $p$ and such that $m \cdot P_i \in B(k)$ for every integer $i$.}

Note that the above assumption on $G_i$ is satisfied, e.g., if $A$ is an elliptic curve over the function field $K$ of a smooth curve over $k = \overline{k}$. If $A$ is isotrivial then the assertion is clear. If $A$ is not isotrivial then the $j$-invariant of $A$ is transcendental over $k$. In this case $A(K^{1/p^i})_{\text{tors}}$ is finite (see \cite{Le}) so orders of the groups $G_i = A^{(i)}(K)_{\text{tors}}$ are uniformly bounded.

Documenta Mathematica 18 (2013) 23–50
Let \( k \) be an algebraically closed field of positive characteristic \( p \). Let \( f : X \to S \) be a smooth projective morphism of \( k \)-varieties with geometrically connected fibres. Assume that \( S \) is projective, which surely is a very strong assumption. Indeed, if \( k \neq \mathbb{F}_p \), and in the statement of Theorem 7.1 \( S' \) is open, then one obtains the stronger Theorem 7.2. For simplicity, let us also assume that \( f \) has a section \( \sigma : S \to X \).

Consider the torsion component \( \text{Pic}^\tau(X/S) \to \text{Pic}(X/S) \) of identity of \( \text{Pic}(X/S) \to S \). Let \( \phi_n : \text{Pic}(X/S) \to \text{Pic}(X/S) \) be the multiplication by \( n \) map. Then there exists an open subgroup scheme \( \text{Pic}^\tau(X/S) \) of \( \text{Pic}(X/S) \) such that every geometric point \( s \) of \( S \) the fibre of \( \text{Pic}^\tau(X/S) \) over \( s \) is the union

\[
\bigcup_{n > 0} \phi_n^{-1}(\text{Pic}^0(X_i)),
\]

where \( \text{Pic}^0(X_i) \) is the connected component of the identity of \( \text{Pic}(X_i/s) \). It is well known that \( \text{Pic}^\tau(X/S) \to S \) is also a closed subgroup scheme of \( \text{Pic}(X/S) \). Moreover, the morphism \( \text{Pic}^\tau(X/S) \to \text{Pic}(X/S) \) is projective and the formation of \( \text{Pic}^\tau(X/S) \to S \) commutes with a base change of \( S \) (see, e.g., [K], Theorem 6.16 and Exercise 6.18). We assume that \( \text{Pic}^0(X_i) \) is reduced for every point \( s \in S \).

**Theorem 7.1.** Let \( L = \{L_i, \sigma_i\} \) be a relatively stratified line bundle on \( X/S \). Assume that there exists a dense subset \( S' \subset S(k) \) such that for every \( s \in S', \) the stratified bundle \( L_s = L_s|_{X_s} \) has finite monodromy. Then \( L_{\eta} \) has finite monodromy.

**Proof.** Replacing \( L \) by a power \( L \otimes N \), where \( N \) is sufficiently large, we may assume that \( L_{\eta} \in \text{Pic}^0(X_i) \) for all closed points \( s \) in \( S \) (see [K], Corollary 6.17)). By assumption \( \tilde{\pi} : \mathcal{A} = \text{Pic}^0(X/S) \to S \) is an abelian scheme. Let us consider the dual abelian scheme \( \mathcal{A}' \to S \). We have a well defined Albanese morphism \( g : (X, \sigma) \to (\mathcal{A}', e) \) (see [FGA], Exposé VI, Théorème 3.3)). Moreover, the map \( g^*: \text{Pic}^0(\mathcal{A}/S) \to \mathcal{A}' = \text{Pic}^0(X/S) \) is an isomorphism of \( S \)-schemes. Let us set \( \hat{A} = \mathcal{A}'_{\eta} \).

Let \( P_i \) be the \( K \)-point of \( \hat{A}^{(i)} \) corresponding to \((L_i)_{\eta}\). Note that the \( K \)-point \( P_0 \in \hat{A} \) is \( V \)-divisible. Indeed, by the definition of a relative stratification we have \( V^P(P_0) = P_0 \) for all integers \( n \). Similarly, we see that all the points \( P_i \in \hat{A}^{(i)}(K) \) are \( V \)-divisible. By Corollary 6.4 it follows that \( P_i \in \hat{B}^{(i)}(k) + \hat{A}^{(i)}(K)_{\text{crys}} \), where \( \hat{B}/k, \tilde{\pi} : \hat{B} \to \hat{A} \) is the \( K/k \)-trace of \( \hat{A} \) (note that \( \hat{B}^{(i)}/k, \tilde{\pi}^{(i)} \) is the \( K/k \)-trace of \( \hat{A}^{(i)} \)). So for every \( i \geq 0 \) we can write \( P_i = Q_i + R_i \) for some \( Q_i \in \hat{B}^{(i)}(k) \) and \( R_i \in \hat{A}^{(i)}(K)_{\text{crys}} \).

Now we transpose the above by duality. Let \( A \) be the dual abelian \( K \)-variety of \( \hat{A} \) and \( B \) the dual abelian \( k \)-variety of \( \hat{B} \). We have the \( K/k \)-images \( \tau^{(i)} : A^{(i)}_{\eta} \to B^{(i)}_k \) and an \( S \)-morphism \( \tau : \mathcal{A} \to B \times_k S \) (possibly after shrinking \( S \)). By abuse of notation we can treat \( L_i \) as line bundles on \( \mathcal{A} \) because \( g^*: \text{Pic}^0(\mathcal{A}/S) \to \text{Pic}^0(X/S) \) is an isomorphism. Let \( M_i \) be the line bundle on \( B^{(i)} \) corresponding to \( Q_i \) and let \( \pi_i : B^{(i)} \times_k S \to B^{(i)} \) denote the projection. Let us fix a non-negative integer \( i \) and take a positive integer \( n_i \), such that \( n_i R_i = 0 \). Then the line bundle \( L^{(i)n_i} \otimes \tau^* \pi_i^* M^{(i)n_i} \) has degree 0 on every fiber of \( \mathcal{A} \to S \). Thus it is trivial after restriction to \( \mathcal{A}_{\eta} \). Hence after shrinking \( S \) we can assume that \( L^{(i)n_i} \simeq \tau^* \pi_i^* M^{(i)n_i} \).

Let us fix a point \( s \in S(k) \) and consider the morphism

\[
\pi_i' = (\tau^{(i)} \pi_i)_{\mathcal{A}^{(i)}_{\eta}} : \mathcal{A}^{(i)}_{\eta} \to B^{(i)}.
\]
Note that $\tau^{(i)}$ has connected fibres and hence $(\pi')_i^*\mathcal{O}_{\mathcal{X}^{(i)}} = \mathcal{O}_{\mathcal{Y}^{(i)}}$. By assumption there exists a positive integer $a_i$, such that for every $i$ the order of the line bundle $(L_i)_s$ divides $a_i$. The important point is that $a_i$ is prime to $p$.

Therefore $(\pi')_1^*M_i^{[a_i]} \simeq \mathcal{O}_{A_i}$ and by the projection formula

$$M_i^{[a_i]} \simeq (\pi')_1^* \mathfrak{M}_i^{[a_i]} \simeq (\pi')_1^* \mathcal{O}_{A_i} \simeq \mathcal{O}_B.$$ 

This implies that $M_i$ is a torsion line bundle and hence $Q_i \in \mathcal{A}^{(i)}(K)_{\text{tors}}$. Therefore

$$P_i = Q_i + R_i \in \mathcal{A}^{(i)}(K)_{\text{tors}}.$$ 

Let us recall that the set of $p$-torsion points of $\mathcal{A}(K)$ is finite. Assuming it is not empty, we can therefore find a non-empty open subset $U \subset S$ such that for every $s \in U(k)$ and every $p$-torsion point $\mathfrak{T} \in \mathcal{A}(K)$ the section $\mathfrak{T}$ is defined on $U$ and the point $\mathfrak{T}(s)$ is non-zero.

Let us write the order of $P_i$ as $m_i p^{e_i}$, where $m_i$ is not divisible by $p$. If $e_0 \geq 1$ then the point $m_0 p^{e_0 - 1} P_0$ is $p$-torsion in $\mathcal{A}(K)$. If we take $s \in S \cap U(k)$, then $a_i m_i p^{e_i - 1} P_0(s) = [L_i^{[a_i]}]_{\mathfrak{X}_0} = 0$, a contradiction. It follows that $m_0 P_0 = 0$. Similarly, the order of all $P_i$ is prime to $p$.

As already mentioned in the last section, the homomorphism $\mathcal{A}(K^{1/p}) \to \mathcal{A}^{(i)}(K)$ induced by $F_i^{1/p}_{A/K}$ is a bijection. So we have an induced injection

$$F_i: \mathcal{A}(K) \to \mathcal{A}^{(i)}(K).$$ 

On the other hand, the Verschiebung morphism induces homomorphisms

$$V_i: \mathcal{A}^{(i)}(K) \to \mathcal{A}(K)$$

such that $V_i F_i(P) = p^i P$ and $F_i V_i(Q) = p^i Q$ for all $P \in \mathcal{A}(K)$ and $Q \in \mathcal{A}^{(i)}(K)$. Hence

$$p^i m_i P_i = F_i V_i(m_i P_i) = F_i(m_i P_0) = 0$$

and since the order of $P_i$ is prime to $p$ we have $m_0 P_i = 0$ for all $i \geq 0$. Therefore $(L_i)^{[m_0]} \simeq \mathcal{O}_{\mathfrak{X}_0}$ for all $i$ and the stratified line bundle $\mathcal{L}_\eta$ has finite monodromy.

Now we fix the following notation: $k$ is an algebraically closed field of positive characteristic $p$ and $f: X \to S$ is a smooth projective morphism of $k$-varieties with geometrically connected fibres.

**Theorem 7.2.** Let $\mathcal{E} = \{E_i, \sigma_i\}$ be a relatively stratified bundle on $X/S$. Assume that there exists a dense subset $U \subset S(k)$ such that for every $s \in U$ the stratified bundle $\mathcal{E}_s = \mathcal{E}|_X$, has finite monodromy of order prime to $p$.

1) Then there exists a finite Galois étale covering $\pi_\mathfrak{q}: \mathcal{Y}_\mathfrak{q} \to \mathcal{X}_\mathfrak{q}$ of order prime-to-$p$ such that $\pi_\mathfrak{q}^* \mathcal{E}_\mathfrak{q}$ is a direct sum of stratified line bundles.
2) If \( k \neq \bar{F} \) and \( U \) is open in \( S(k) \), then the monodromy group of \( \mathbb{E}_\eta \) is finite, and \( \mathbb{E}_\eta \) trivializes on a finite étale cover \( Z_\eta \rightarrow X_\eta \) which factors as a Kummer (thus finite abelian of order prime to \( p \)) cover \( Z_\eta \rightarrow Y_\eta \) and a Galois cover \( Y_\eta \rightarrow X_\eta \) of order prime to \( p \).

**Proof.** We prove 1). Let us first remark that the schemes \( X_{i \eta} \), \( i \geq 0 \), are all isomorphic (as schemes, not as \( k \)-schemes). Therefore the relative Frobenius induces an isomorphism on fundamental groups.

By the first part of Theorem 5.1 we know that there exists a finite Galois étale covering \( \pi_i: Y_{i \eta} \rightarrow X_{i \eta} \) of degree prime to \( p \) such that \( \pi_i^*(E_i) \) is a direct sum of line bundles \( \bigoplus_{j=1}^r L_{i j} \). Note that from the proof of Theorem 5.1 the degree of \( \pi_i \) depends only on \( \pi_{\sigma_i}^*(X_{i \eta}, \sigma_i^*(\bar{\eta})) \) and the Brauer-Feit constant \( j(r) \), and therefore it can be bounded independently of \( i \). Using the Lang–Serre theorem (see [LS, Théorème 4]) we can therefore assume that \( Y_{i \eta} = Y_{\eta} \), where \( Y_\eta = Y_{0 \eta} \). Now we know that

\[
\bigoplus_{j=1}^r L_{i j} \simeq (F_i^{j})^{*}(\bigoplus_{j'=1}^{r'} L_{i+1, j'}) .
\]

By the Krull-Schmidt theorem, the set of isomorphism classes of line bundles \( \{L_{i j}\}_j \) is the same as the set of isomorphism classes of lines bundles which come by pull-back \( \{(F_i^{j})^{*}(L_{i+1, j'})\}_j \). So we can reorder the indices \( j' \) so that

\[
(F_i^{j})^{*}(L_{i+1, j}) \simeq L_{i, j} .
\]

This finishes the proof of 1).

To prove 2), we do the proof 1) replacing \( Y_\eta \rightarrow X_\eta \) by \( Z_\eta \rightarrow X_\eta \) of Theorem 5.1 2). This finishes the proof of 2).

\[\square\]

**Remarks 7.3.**

1) Case 2) of Theorem 7.2 applied to a line bundle extends Theorem 7.1 where \( S \) was assumed to be projective, \( \text{Pic}^0(X_s) \) reduced for all \( s \in S \) closed, \( S' \subset S(k) \) dense, to the case when \( S \) is not necessarily projective and \( S' \subset S(k) \) is open and dense, but we have to assume that \( k \) is not algebraic over its prime field.

2) If \( Y_\eta \) has a good projective model satisfying assumptions of Theorem 7.1 then it follows that \( \mathbb{E}_\eta \) has finite monodromy.

**References**


BIRATIONAL MOTIVIC HOMOTOPY THEORIES
AND THE SLICE FILTRATION

PABLO PELAEZ

Received: May 21, 2012
Revised: October 5, 2012
Communicated by Lars Hesselholt

ABSTRACT. We show that there is an equivalence of categories between the orthogonal components for the slice filtration and the birational motivic stable homotopy categories which are constructed in this paper. Relying on this equivalence, we are able to describe the slices for projective spaces (including $\mathbb{P}^\infty$), Thom spaces and blow ups.

2000 Mathematics Subject Classification: 14F42
Keywords and Phrases: Birational Invariants, Motivic Homotopy Theory, Motivic Spectral Sequence, Slice Filtration.

1 Definitions and Notation

Our main result, theorem 3.6, shows that there is an equivalence of categories between the orthogonal components for the slice filtration (see definition 1.1) and the weakly birational motivic stable homotopy categories which are constructed in this paper (see definition 2.9). Relying on this equivalence; we are able to describe over an arbitrary base scheme (see theorems 4.2, 4.4 and 4.6) the slices for projective spaces (including $\mathbb{P}^\infty$), Thom spaces and blow ups. We also construct the birational motivic stable homotopy categories (see definition 2.4), which are a natural generalization of the weakly birational motivic stable homotopy categories, and show (see proposition 2.12) that there exists a Quillen equivalence between them when the base scheme is a perfect field. Our approach was inspired by the work of Kahn-Sujatha [1] on birational motives, where the existence of a connection between the layers of the slice filtration and birational invariants is explicitly suggested. Furthermore, this approach...
allows to obtain analogues for the slice filtration in the unstable setting (see remark 3.8).

In this paper, $X$ will denote a Noetherian separated base scheme of finite Krull dimension, $\text{Sch}_X$ the category of schemes of finite type over $X$ and $\text{Sm}_X$ the full subcategory of $\text{Sch}_X$ consisting of smooth schemes over $X$ regarded as a site with the Nisnevich topology. All the maps between schemes will be considered over the base $X$. Given $Y \in \text{Sch}_X$, all the closed subsets $Z$ of $Y$ will be considered as closed subschemes with the reduced structure.

Let $\mathcal{M}$ be the category of pointed simplicial presheaves in $\text{Sm}_X$ equipped with the motivic Quillen model structure [14] constructed by Morel-Voevodsky [8, p. 86 Thm. 3.2], taking the affine line $\mathbb{A}_X^1$ as interval. Given a map $f : Y \to W$ in $\text{Sm}_X$, we will abuse notation and denote by $f$ the induced map $f : Y_+ \to W_+$ in $\mathcal{M}$ between the corresponding pointed simplicial presheaves represented by $Y$ and $W$ respectively.

We define $T$ in $\mathcal{M}$ to be the pointed simplicial presheaf represented by $S^1 \wedge \mathbb{G}_m$, where $\mathbb{G}_m$ is the multiplicative group $\mathbb{A}_X^1 - \{0\}$ pointed by 1, and $S^1$ denotes the simplicial circle. Given an arbitrary integer $r \geq 1$, $S^r$ (respectively $\mathbb{G}_m^r$) will denote the iterated smash product $S^1 \wedge \cdots \wedge S^1$ (respectively $\mathbb{G}_m^r$) with $r$-factors; $S^0 = \mathbb{G}_m^0$ will be by definition equal to the pointed simplicial presheaf $X_+$ represented by the base scheme $X$.

Let $\text{Spt}(\mathcal{M})$ denote Jardine’s category of symmetric $T$-spectra on $\mathcal{M}$ equipped with the motivic model structure defined in [6, Thm. 4.15] and let $\mathcal{SH}$ denote its homotopy category, which is triangulated. We will follow Jardine’s notation [6, p. 506-507] where $F_n$ denotes the left adjoint to the $n$-evaluation functor

$$
\text{Spt}(\mathcal{M}) \xrightarrow{ev_n} \mathcal{M} \\
(X^m)_{m \geq 0} \xrightarrow{X^n} \mathcal{M}
$$

Notice that $F_0(A)$ is just the usual infinite suspension spectrum $\Sigma^\infty A$.

For every integer $q \in \mathbb{Z}$, we consider the following family of symmetric $T$-spectra

$$
C^q_{\text{eff}} = \{ F_n(S^r \wedge \mathbb{G}_m^s \wedge U_+) \mid n, r, s \geq 0; s - n \geq q; U \in \text{Sm}_X \} \tag{1.1}
$$

where $U_+$ denotes the simplicial presheaf represented by $U$ with a disjoint base point.

Let $\Sigma^q_\mathcal{SH}^{\text{eff}}$ denote the smallest full triangulated subcategory of $\mathcal{SH}$ which contains $C^q_{\text{eff}}$ and is closed under arbitrary coproducts. Voevodsky [16] defines the slice filtration in $\mathcal{SH}$ to be the following family of triangulated subcategories

$$
\cdots \subseteq \Sigma^{q+1}_\mathcal{SH}^{\text{eff}} \subseteq \Sigma^q_\mathcal{SH}^{\text{eff}} \subseteq \Sigma^{q-1}_\mathcal{SH}^{\text{eff}} \subseteq \cdots
$$

It follows from the work of Neeman [9], [10] that the inclusion

$$
i_q : \Sigma^q_\mathcal{SH}^{\text{eff}} \to \mathcal{SH}$$

Documenta Mathematica 18 (2013) 51–70
Birational Motivic Homotopy Theories

has a right adjoint \( r_q : \mathcal{SH} \rightarrow \Sigma_q^T \mathcal{SH}^{eff} \), and that the following functors

\[
\begin{align*}
    f_q &: \mathcal{SH} \rightarrow \mathcal{SH} \\
    s_{<q} &: \mathcal{SH} \rightarrow \mathcal{SH} \\
    s_q &: \mathcal{SH} \rightarrow \mathcal{SH}
\end{align*}
\]

are triangulated, where \( f_q \) is defined as the composition \( i_q \circ r_q \); and \( s_{<q}, s_q \) are characterized by the fact that for every \( E \in \mathcal{SH} \), we have distinguished triangles in \( \mathcal{SH} \):

\[
\begin{align*}
    f_q E \to^\theta E \to s_{<q} E \to S^1 \wedge f_q E \\
    f_{q+1} E \to^\theta f_q E \to s_q E \to S^1 \wedge f_{q+1} E
\end{align*}
\]

We will refer to \( f_q E \) as the \((q - 1)\)-connective cover of \( E \), to \( s_{<q} E \) as the \( q \)-orthogonal component of \( E \), and to \( s_q E \) as the \( q \)-slice of \( E \). It follows directly from the definition that \( s_{<q+1} E, s_q E \) satisfy that for every symmetric \( T \)-spectrum \( K \) in \( \Sigma_{q+1}^T \mathcal{SH}^{eff} \):

\[
\text{Hom}_{\mathcal{SH}}(K, s_{<q+1} E) = \text{Hom}_{\mathcal{SH}}(K, s_q E) = 0
\]

**Definition 1.1.** Let \( E \in Spt(\mathcal{M}) \) be a symmetric \( T \)-spectrum. We will say that \( E \) is \( n \)-orthogonal, if for all \( K \in \Sigma_n^T \mathcal{SH}^{eff} \)

\[
\text{Hom}_{\mathcal{SH}}(K, E) = 0
\]

Let \( \mathcal{SH}^{\perp}(n) \) denote the full subcategory of \( \mathcal{SH} \) consisting of the \( n \)-orthogonal objects.

The slice filtration admits an alternative definition in terms of (left and right) Bousfield localization of \( Spt(\mathcal{M}) \) [11, 12]. The Bousfield localizations are constructed following Hirschhorn’s approach [2]. In order to be able to apply Hirschhorn’s techniques, it is necessary to know that \( Spt(\mathcal{M}) \) is cellular [2, Def. 12.1.1] and proper [2, Def. 13.1.1].

**Theorem 1.2.** The Quillen model category \( Spt(\mathcal{M}) \) is:

1. cellular (see [5], [3, Cor. 1.6] or [12, Thm. 2.7.4]).
2. proper (see [6, Thm. 4.15]).

For details and definitions about Bousfield localization we refer the reader to Hirschhorn’s book [2]. Let us just mention the following theorem of Hirschhorn, which guarantees the existence of left and right Bousfield localizations.
Let $A$ be a Quillen model category which is cellular and proper. Let $L$ be a set of maps in $A$ and let $K$ be a set of objects in $A$. Then:

1. The left Bousfield localization of $A$ with respect to $L$ exists.
2. The right Bousfield localization of $A$ with respect to the class of $K$-colocal equivalences exists.

Now, we can describe the slice filtration in terms of suitable Bousfield localizations of $Spt(M)$.

**Theorem 1.4** (see [12]). 1. Let $R_{C_{q}^{\text{eff}}}Spt(M)$ be the right Bousfield localization of $Spt(M)$ with respect to the set of objects $C_{q}^{\text{eff}}$ (see Eqn. (1.1)). Then its homotopy category $R_{C_{q}^{\text{eff}}}S_{q}H$ is triangulated and naturally equivalent to $\Sigma_{q}S_{q}H^{\text{eff}}$. Moreover, the functor $f_{q}$ is canonically isomorphic to the following composition of triangulated functors:

$$S_{q}H \xrightarrow{R} R_{C_{q}^{\text{eff}}}S_{q}H \xrightarrow{C_{q}} S_{q}H$$

where $R$ is a fibrant replacement functor in $Spt(M)$, and $C_{q}$ a cofibrant replacement functor in $R_{C_{q}^{\text{eff}}}Spt(M)$.

2. Let $L_{<q}Spt(M)$ be the left Bousfield localization of $Spt(M)$ with respect to the set of maps

$$\{ F_{n}(S^{r} \wedge G_{m}^{s} \wedge U_{+}) \to * \mid F_{n}(S^{r} \wedge G_{m}^{s} \wedge U_{+}) \in C_{q}^{\text{eff}} \}$$

Then its homotopy category $L_{<q}S_{q}H$ is triangulated and naturally equivalent to $S_{q}H^{\perp}(q)$. Moreover, the functor $s_{<q}$ is canonically isomorphic to the following composition of triangulated functors:

$$S_{q}H \xrightarrow{Q} L_{<q}S_{q}H \xrightarrow{W_{q}} S_{q}H$$

where $Q$ is a cofibrant replacement functor in $Spt(M)$, and $W_{q}$ a fibrant replacement functor in $L_{<q}Spt(M)$.

3. Let $S^qSpt(M)$ be the right Bousfield localization of $L_{<q+1}Spt(M)$ with respect to the set of objects

$$\{ F_{n}(S^{r} \wedge G_{m}^{s} \wedge U_{+}) \mid n, r, s \geq 0; s - n = q; U \in Sm_{X} \}$$

Then its homotopy category $S^qS_{q}H$ is triangulated and the identity functor $id : R_{C_{q}^{\text{eff}}}Spt(M) \to S^qSpt(M)$

is a left Quillen functor. Moreover, the functor $s_{q}$ is canonically isomorphic to the following composition of triangulated functors:

$$S_{q}H \xrightarrow{R} R_{C_{q}^{\text{eff}}}S_{q}H \xrightarrow{C_{q}} S^{q}S_{q}H \xrightarrow{W_{q+1}} R_{C_{q}^{\text{eff}}}S_{q}H \xrightarrow{C_{q}} S_{q}H$$
Proof. (1) and (3) follow directly from \([12, \text{Thms. 3.3.9, 3.3.25, 3.3.50, 3.3.68}]\). On the other hand, (2) follows from proposition 3.2.27(3) together with theorem 3.3.26; proposition 3.3.30 and theorem 3.3.45 in \([12]\).

2 Birational and Weakly Birational Cohomology Theories

In this section, we construct the birational and weakly birational motivic stable homotopy categories. These are defined as left Bousfield localizations of \(\text{Spt}(\mathcal{M})\) with respect to maps which are induced by open immersions with a numerical condition in the codimension of the closed complement (which is assumed to be smooth in the weakly birational case). The existence of the left Bousfield localizations considered in this section follows immediately from theorems 1.2 and 1.3.

Lemma 2.1. Let \(a,a',b,b',p,p' \geq 0\) be integers such that \(a - p = a' - p'\) and \(b - p = b' - p'\). Assume that \(p \geq p'\), then for every \(Y \in \text{Sm}_X\), there is a weak equivalence in \(\text{Spt}(\mathcal{M})\), which is natural with respect to \(Y\):

\[ g_{p,p'}^{a,b}(Y) : F_p(S^a \wedge \mathbb{G}_m^b \wedge Y_+) \to F_{p'}(S^{a'} \wedge \mathbb{G}_m^{b'} \wedge Y_+) \]

Proof. We have the following adjunction (see \([12, \text{Def. 2.6.8}]\))

\[(F_p, ev_p, \varphi) : \mathcal{M} \to \text{Spt}(\mathcal{M})\]

Using this adjunction, we define \(g_{p,p'}^{a,b}(Y)\) as adjoint to the identity map:

\[ S^a \wedge \mathbb{G}_m^b \wedge Y_+ \xrightarrow{id} ev_p(F_p(S^a \wedge \mathbb{G}_m^b \wedge Y_+)) \cong S^{p-p'} \wedge \mathbb{G}_m^{p-p'} \wedge S^{a'} \wedge \mathbb{G}_m^{b'} \wedge Y_+ \]

\[ \cong S^a \wedge \mathbb{G}_m^b \wedge Y_+ \]

Thus, it is clear that \(g_{p,p'}^{a,b}(Y)\) is natural in \(Y\), and it follows from \([12, \text{Prop. 2.4.26}]\) that it is a weak equivalence in \(\text{Spt}(\mathcal{M})\).

Definition 2.2 (see \([13, \text{section 7.5}]\)). Let \(Y \in \text{Sch}_X\), and \(Z\) a closed subscheme of \(Y\). The codimension of \(Z\) in \(Y\), \(\text{codim}_Y Z\) is the infimum (over the generic points \(z_i\) of \(Z\)) of the dimensions of the local rings \(\mathcal{O}_{Y,z_i}\).

Since \(X\) is Noetherian of finite Krull dimension and \(Y\) is of finite type over \(X\), \(\text{codim}_Y Z\) is always finite.

Definition 2.3. We fix an arbitrary integer \(n \geq 0\), and consider the following set of open immersions which have a closed complement of codimension at least \(n + 1\):

\[ B_n = \{\iota_{U,Y} : U \to Y \text{ open immersion} \mid Y \in \text{Sm}_X; Y \text{ irreducible}; (\text{codim}_Y Y \setminus U) \geq n + 1\} \]

The letter \(B\) stands for birational.
Now we consider the left Bousfield localization of \( \text{Spt}(\mathcal{M}) \) with respect to a suitable set of maps induced by the families of open immersions \( B_n \) described above.

**Definition 2.4.** Let \( n \in \mathbb{Z} \) be an arbitrary integer.

1. Let \( B_n \text{Spt}(\mathcal{M}) \) denote the left Bousfield localization of \( \text{Spt}(\mathcal{M}) \) with respect to the set of maps \( sB_n = \{ F_p(\mathbb{G}^b_m \wedge \iota_{U,Y}) : b, p, r \geq 0, b - p \geq n - r, \iota_{U,Y} \in B_r \} \).

2. Let \( b^{(n)} \) denote its fibrant replacement functor and \( \mathcal{S}H(B_n) \) its associated homotopy category.

For \( n \neq 0 \) we will call \( \mathcal{S}H(B_n) \) the codimension \( n + 1 \)-birational motivic stable homotopy category, and for \( n = 0 \) we will call it the birational motivic stable homotopy category.

**Lemma 2.5.** Let \( n \in \mathbb{Z} \) be an arbitrary integer. Then for every \( a \geq 0 \), the maps \( S^a \wedge sB_n = \{ F_p(S^a \wedge \mathbb{G}^b_m \wedge \iota_{U,Y}) : b, p, r \geq 0, b - p \geq n - r, \iota_{U,Y} \in B_r \} \) are weak equivalences in \( B_n \text{Spt}(\mathcal{M}) \).

**Proof.** Let \( F_p(\mathbb{G}^b_m \wedge \iota_{U,Y}) \in sB_n \) with \( \iota_{U,Y} \in B_r \). Both \( F_p(\mathbb{G}^b_m \wedge U_\ast) \) and \( F_p(\mathbb{G}^b_m \wedge Y_\ast) \) are cofibrant in \( \text{Spt}(\mathcal{M}) \) (see [12, Props. 2.4.17, 2.6.18 and Thm. 2.6.30]) and hence also in \( B_n \text{Spt}(\mathcal{M}) \). By construction, \( F_p(\mathbb{G}^b_m \wedge \iota_{U,Y}) \) is a weak equivalence in \( B_n \text{Spt}(\mathcal{M}) \); and [2, Thm. 4.1.1.(4)] implies that \( B_n \text{Spt}(\mathcal{M}) \) is a simplicial model category. Thus, it follows from Ken Brown’s lemma (see [4, lemma 1.1.12]) that \( F_p(S^a \wedge \mathbb{G}^b_m \wedge \iota_{U,Y}) \) is also a weak equivalence in \( B_n \text{Spt}(\mathcal{M}) \) for every \( a \geq 0 \).

**Proposition 2.6.** Let \( E \) be an arbitrary symmetric \( T \)-spectrum. Then \( E \) is fibrant in \( B_n \text{Spt}(\mathcal{M}) \) if and only if the following conditions hold:

1. \( E \) is fibrant in \( \text{Spt}(\mathcal{M}) \).

2. For every \( a, b, p, r \geq 0 \) such that \( b - p \geq n - r \); and every \( \iota_{U,Y} \in B_r \), the induced map

\[
\text{Hom}_{\mathcal{S}H}(F_p(S^a \wedge \mathbb{G}^b_m \wedge Y_\ast), E) \xrightarrow{\iota_{U,Y}^*} \text{Hom}_{\mathcal{S}H}(F_p(S^a \wedge \mathbb{G}^b_m \wedge U_\ast), E)
\]

is an isomorphism.

**Proof.** (\( \Rightarrow \)): Since the identity functor

\[
id : \text{Spt}(\mathcal{M}) \to B_n \text{Spt}(\mathcal{M})
\]

Documenta Mathematica 18 (2013) 51–70
is a left Quillen functor, the conclusion follows from the derived adjunction
\[(Q, b^{(n)}, \varphi) : \mathcal{SH} \to \mathcal{SH}(B_n)\]
together with lemma 2.5.
\((\Leftarrow)\): Assume that \(E\) satisfies (1) and (2). Let \(\omega_0, \eta_0\) denote the base points of the pointed simplicial sets \(\text{Map}_*(F_p(G^b_m \wedge Y_+), E)\) and \(\text{Map}_*(F_p(G^b_m \wedge U_+), E)\) respectively. Since \(F_p(G^b_m \wedge Y_+)\) and \(F_p(G^b_m \wedge U_+)\) are always cofibrant, by [2, Def. 3.1.4(1)(a) and Thm. 4.1.1(2)] it is enough to show that every map in \(sB_n\) induces a weak equivalence of simplicial sets:
\[
\text{Map}_*(F_p(G^b_m \wedge Y_+), E) \xrightarrow{\iota_{Y,Y}} \text{Map}_*(F_p(G^b_m \wedge U_+), E)
\]
Since \(Spt(M)\) is a pointed simplicial model category, we observe that lemma 6.1.2 in [4] and remark 2.4.3(2) in [12] imply that the following diagram is commutative for \(a \geq 0\) and all the vertical arrows are isomorphisms
\[
\begin{array}{ccc}
\pi_{a,\omega_0} \text{Map}_*(F_p(G^b_m \wedge Y_+), E) & \xrightarrow{=} & \pi_{a,\eta_0} \text{Map}_*(F_p(G^b_m \wedge U_+), E) \\
\downarrow & & \downarrow \\
\text{Hom}_{\mathcal{SH}}(F_p(S^a \wedge G^b_m \wedge Y_+), E) & \xrightarrow{\iota_{Y,Y}} & \text{Hom}_{\mathcal{SH}}(F_p(S^a \wedge G^b_m \wedge U_+), E)
\end{array}
\]
by hypothesis, the bottom row is an isomorphism, hence the top row is also an isomorphism. This implies that for every map in \(sB_n\), the induced map
\[
\text{Map}_*(F_p(G^b_m \wedge Y_+), E) \xrightarrow{\iota_{Y,Y}} \text{Map}_*(F_p(G^b_m \wedge U_+), E)
\]
is a weak equivalence when it is restricted to the path component of \(\text{Map}_*(F_p(G^b_m \wedge Y_+), E)\) containing \(\omega_0\). This holds in particular for
\[
\text{Map}_*(F_{p+1}(G^b_{m+1} \wedge Y_+), E) \xrightarrow{\iota_{Y,Y}} \text{Map}_*(F_{p+1}(G^b_{m+1} \wedge U_+), E)
\]
Therefore, the following map is a weak equivalence of pointed simplicial sets, since taking \(S^1\)-loops kills the path components that do not contain the base point
\[
\begin{array}{ccc}
\text{Map}_*(S^1, \text{Map}_*(F_{p+1}(G^b_{m+1} \wedge Y_+), E)) & \xrightarrow{=} & \text{Map}_*(S^1, \text{Map}_*(F_{p+1}(G^b_{m+1} \wedge U_+), E))
\end{array}
\]
Now, since $Spt(M)$ is a simplicial model category we deduce that the rows in the following commutative diagram are isomorphisms

$$\text{Map}_* (S^1, \text{Map}_* (F_{p+1} (G_{m}^{b+1} \wedge Y_+), E))$$

$$\cong$$

$$\text{Map}_* (F_{p+1} (S^1 \wedge G_{m}^{b+1} \wedge Y_+), E)$$

$$\text{Map}_* (S^1, \text{Map}_* (F_{p+1} (G_{m}^{b+1} \wedge U_+), E))$$

$$\cong$$

$$\text{Map}_* (F_{p+1} (S^1 \wedge G_{m}^{b+1} \wedge U_+), E)$$

Thus, by the three out of two property for weak equivalences, we conclude that

$$\text{Map}_* (F_{p+1} (S^1 \wedge G_{m}^{b+1} \wedge Y_+), E) \xrightarrow{i_{U,Y}^*} \text{Map}_* (F_{p+1} (S^1 \wedge G_{m}^{b+1} \wedge U_+), E)$$

is also a weak equivalence of pointed simplicial sets. Finally, lemma 2.1 implies that the following diagram is commutative and the vertical arrows are weak equivalences in $Spt(M)$

$$\text{Map}_* (F_{p+1} (S^1 \wedge G_{m}^{b+1} \wedge Y_+), E) \xrightarrow{i_{U,Y}^*} \text{Map}_* (F_{p+1} (S^1 \wedge G_{m}^{b+1} \wedge U_+), E)$$

$$\text{Map}_* (F_{p} (G_{m}^{b} \wedge Y_+), E) \xrightarrow{i_{U,Y}^*} \text{Map}_* (F_{p} (G_{m}^{b} \wedge U_+), E)$$

Thus, we conclude by the two out of three property for weak equivalences that the bottom arrow is also a weak equivalence in $Spt(M)$. 

**Proposition 2.7.** The homotopy category $SH(B_n)$ is a compactly generated triangulated category in the sense of Neeman [9, Def. 1.7].

**Proof.** We will prove first that $SH(B_n)$ is a triangulated category. For this, it is enough to show that the smash product with the simplicial circle induces a Quillen equivalence (see [14, sections 1.2, 1.3])

$$(- \wedge S^1, \Omega_{S^1} \varphi) : B_n Spt(M) \to B_n Spt(M)$$

It follows from [2, Thm. 4.1.1.(4)] that this adjunction is a Quillen adjunction, and the same argument as in [12, Cor. 3.2.38] (replacing [12, Prop. 3.2.32] with proposition 2.6) allows us to conclude that it is a Quillen equivalence. Finally, since $SH$ is a compactly generated triangulated category (see [12, Prop. 3.1.5]) and the identity functor is a left Quillen functor

$$id : Spt(M) \to B_n Spt(M)$$
it follows from the derived adjunction
\[(Q, b^n, \varphi) : \mathcal{SH} \to \mathcal{SH}(B_n)\]
that \(\mathcal{SH}(B_n)\) is also compactly generated, having exactly the same set of generators as \(\mathcal{SH}\).

**Definition 2.8.** We fix an arbitrary integer \(n \geq 0\), and consider the following set of open immersions with smooth closed complement of codimension at least \(n + 1\)
\[WB_n = \{\iota_{U, Y} : U \to Y \text{ open immersion} \mid Y, Z = Y \setminus U \in Sm_X; Y \text{ irreducible}; (\text{codim}_Y Z) \geq n + 1\}\]
Notice that every map in \(WB_n\) is also in \(B_n\), but the converse doesn’t hold.

The reason to consider maps \(\iota_{U, Y}\) in \(WB_n\) is that if the closed complement is smooth, then the Morel-Voevodsky homotopy purity theorem (see [8, Thm. 2.23]) characterizes the homotopy cofibre of \(\iota_{U, Y}\) in terms of the Thom space of the normal bundle for the closed immersion \(Y \setminus U \to Y\).

**Definition 2.9.** Let \(n \in \mathbb{Z}\) be an arbitrary integer.

1. Let \(WB_n Spt(M)\) denote the left Bousfield localization of \(Spt(M)\) with respect to the set of maps
\[sWB_n = \{F_p(S^a \wedge \mathbb{G}_m^b \wedge Y, E) \to \iota_{U, Y}^* F_p(S^a \wedge \mathbb{G}_m^b \wedge U, E)\}\]
where \(\iota_{U, Y} : U \to Y\) is open immersion.

2. Let \(wb^{(n)}\) denote its fibrant replacement functor and \(\mathcal{SH}(WB_n)\) its associated homotopy category.

For \(n \neq 0\) we will call \(\mathcal{SH}(WB_n)\) the \(\text{codimension} n + 1\)-weakly birational motivic stable homotopy category, and for \(n = 0\) we will call it the \(\text{weakly birational motivic stable homotopy category}\).

**Proposition 2.10.** Let \(E\) be an arbitrary symmetric \(T\)-spectrum. Then \(E\) is fibrant in \(WB_n Spt(M)\) if and only if the following conditions hold:

1. \(E\) is fibrant in \(Spt(M)\).
2. For every \(a, b, p, r \geq 0\) such that \(b - p \geq n - r\); and every \(\iota_{U, Y} \in WB_n\), the induced map
\[\text{Hom}_{\mathcal{SH}}(F_p(S^a \wedge \mathbb{G}_m^b \wedge Y, E) \to \iota_{U, Y}^* F_p(S^a \wedge \mathbb{G}_m^b \wedge U, E)\]
is an isomorphism.

**Proof.** The proof is exactly the same as in proposition 2.6. \(\square\)
Proposition 2.11. The homotopy category $\mathcal{SH}(WB_n)$ is a compactly generated triangulated category in the sense of Neeman.

Proof. The proof is exactly the same as in proposition 2.7.

Proposition 2.12. Assume that the base scheme $X = \text{Spec } k$, with $k$ a perfect field, then the Quillen adjunction:

$$(\text{id}, \text{id}, \varphi) : WB_n \text{Spt}(\mathcal{M}) \to B_n \text{Spt}(\mathcal{M})$$

is a Quillen equivalence.

Proof. Consider the following commutative diagram

$$\begin{array}{ccc}
\text{Spt}(\mathcal{M}) & \xrightarrow{\text{id}} & \text{Spt}(\mathcal{M}) \\
\downarrow \text{id} & & \downarrow \text{id} \\
WB_n \text{Spt}(\mathcal{M}) & \xrightarrow{-} & B_n \text{Spt}(\mathcal{M})
\end{array}$$

where the solid arrows are left Quillen functors. Clearly, $WB_r \subseteq B_r$ for every $r \geq 0$, so $sWB_n \subseteq sB_n$, and we conclude that every $sWB_n$-local equivalence is a $sB_n$-local equivalence. Therefore, the universal property of left Bousfield localizations implies that the horizontal arrow is also a left Quillen functor. The universal property for left Bousfield localizations also implies that it is enough to show that all the maps in

$$sB_n = \{ F_p(G_m^b \wedge \iota_{U,Y}) : b, p, r \geq 0, b - p \geq n - r; \iota_{U,Y} \in B_r \}$$

become weak equivalences in $WB_n \text{Spt}(\mathcal{M})$. Given $F_p(G_m^b \wedge \iota_{U,Y}) \in sB_n$ with $\iota_{U,Y} \in B_r$, we proceed by induction on the dimension of $Z = Y \setminus U$. If $\dim Z = 0$, then $Z \in Sm_X$ since $k$ is a perfect field (and we are considering $Z$ with the reduced scheme structure), hence $F_p(G_m^b \wedge \iota_{U,Y}) \in sWB_n$ and then a weak equivalence in $WB_n \text{Spt}(\mathcal{M})$.

If $\dim Z > 0$, then we consider the singular locus $Z_s$ of $Z$ over $X$. We have that $\dim Z_s < \dim Z$ since $k$ is a perfect field. Therefore, by induction on the dimension $F_p(G_m^b \wedge \iota_{V,Y})$ is a weak equivalence in $WB_n \text{Spt}(\mathcal{M})$, where $V = Y \setminus Z_s$. On the other hand, $F_p(G_m^b \wedge \iota_{U,V})$ is also a weak equivalence in $WB_n \text{Spt}(\mathcal{M})$ since $\iota_{U,V}$ is also in $B_r$ and its closed complement $V \setminus U = Z \setminus Z_s$ is smooth over $X$, by construction of $Z_s$.

But $F_p(G_m^b \wedge \iota_{U,Y}) = F_p(G_m^b \wedge \iota_{U,V}) \circ F_p(G_m^b \wedge \iota_{V,Y})$, so by the two out of three property for weak equivalences we conclude that $F_p(G_m^b \wedge \iota_{U,Y})$ is a weak equivalence in $WB_n \text{Spt}(\mathcal{M})$.

3 A Characterization of the Slices

This section contains our main results. We give a characterization of the slices in terms of effectivity and birational conditions (in the sense of definition 3.1), and we also show that there is an equivalence between the notion of orthogonality (see definition 1.1) and weak birationality (see definition 3.1).
Definition 3.1. Let $E \in Spt(\mathcal{M})$ be a symmetric $T$-spectrum and $n \in \mathbb{Z}$.

1. We will say that $E$ is $n+1$-birational (respectively weakly $n+1$-birational), if $E$ is fibrant in $B_n Spt(\mathcal{M})$ (respectively $WB_n Spt(\mathcal{M})$). If $n = 0$, we will simply say that $E$ is birational (respectively weakly birational).

2. We will say that $E$ is an $n$-slice if $E$ is isomorphic in $\mathcal{S}H$ to $s_n(E')$ for some symmetric $T$-spectrum $E'$.

Definition 3.2. 1. Let $\iota_{U,Y}$ be an open immersion in $Sm_X$. Let $Y/U$ denote the pushout of the following diagram in $\mathcal{M}$ (i.e. the homotopy cofibre of $\iota_{U,Y}$ in $\mathcal{M}$)

\[ \begin{array}{ccc}
U & \xrightarrow{\iota_{U,Y}} & Y \\
\downarrow & & \downarrow \\
X & \xrightarrow{=} & Y/U
\end{array} \]

2. Given a vector bundle $\pi : V \to Y$ with $Y \in Sm_X$, let $Th(V)$ denote the Thom space of $V$, i.e. $V/(V \setminus \sigma_0(Y))$, where $\sigma_0 : Y \to V$ denotes the zero section of $V$.

Lemma 3.3. Let $\iota_{U,Y} \in WB_r$ for some $r \geq 0$, and let $a, b, p \geq 0$ be arbitrary integers such that $b - p \geq n - r$.

Then

\[ F_p(S^a \wedge \mathbb{G}_m^b \wedge Y/U) \in \Sigma_{T}^{n+1} \mathcal{S}H^{eff} \]

Proof. Since $\Sigma_{T}^{n+1} \mathcal{S}H^{eff}$ is a triangulated category, it is enough to consider the case $a = 0$. It is also clear that it suffices to show that $F_0(Y/U) \in \Sigma_{T}^{r+1} \mathcal{S}H^{eff}$. Now, it follows from the Morel-Voevodsky homotopy purity theorem (see [8, Thm. 2.23]) that there is an isomorphism in $\mathcal{S}H$

\[ F_0(Y/U) \to F_0(Th(N)) \]

where $Th(N)$ is the Thom space of the normal bundle $N$ of the (smooth) complement $Z$ of $U$ in $Y$:

\[ e : Y \setminus U = Z \to Y \]

But, $\iota_{U,Y} \in WB_r$; so $e$ is a regular embedding of codimension $c$ at least $r + 1$, hence $N$ is a vector bundle of rank at least $r + 1$. Therefore, if $N$ is a trivial vector bundle we conclude from [8, Prop. 2.17(2)] that

\[ F_0(Th(N)) \cong F_0(S^c \wedge \mathbb{G}_m^c \wedge Z_+) \in \Sigma_{T}^{r+1} \mathcal{S}H^{eff} \subseteq \Sigma_{T}^{n+1} \mathcal{S}H^{eff} \]

Finally, we conclude in the general case by choosing a Zariski cover of $Z$ which trivializes $N$ and using the Mayer-Vietoris property for Zariski covers.  

Documenta Mathematica 18 (2013) 51–70
Lemma 3.4. Let $U \in Sm_X$. Consider the open immersion in $Sm_X$

$$m_U : \mathbb{A}^1_U \setminus U \to \mathbb{A}^1_U$$

given by the complement of the zero section. Then $m_U \in WB_0$, and there exists a weak equivalence in $Spt(M)$ between its homotopy cofibre in $M$, $\mathbb{A}^1_U/(\mathbb{A}^1_U \setminus U)$ and $S^1 \wedge G_m \wedge U_+$

$$t_U : \mathbb{A}^1_U/(\mathbb{A}^1_U \setminus U) \to S^1 \wedge G_m \wedge U_+$$

Proof. Since the zero section $i_0 : U \to \mathbb{A}^1_U$ is a closed embedding of codimension 1 between smooth schemes over $X$, it follows from the definition of $WB_0$ that $m_U \in WB_0$. Finally, [8, Prop. 2.17(2)] implies the existence of the weak equivalence $t_U$.

Proposition 3.5. Let $E \in Spt(M)$ be a symmetric $T$-spectrum and $n \in \mathbb{Z}$. Consider the following conditions:

1. $E$ is fibrant in $L_{<n+1}Spt(M)$.
2. $E$ is weakly $n+1$-birational (see definition 3.1(1)).
3. $E$ is $n+1$-birational (see definition 3.1(1)).

Then (1) and (2) are equivalent. In addition, if the base scheme $X = \text{Spec } k$, with $k$ a perfect field, then (1), (2) and (3) are equivalent.

Proof. (1)⇒(2): Assume that $E$ is fibrant in $L_{<n+1}Spt(M)$. By proposition 2.10 it suffices to show that for every $a, b, p, r \geq 0$ with $b - p \geq n - r$, and every $\iota_U, \iota_Y \in WB_r$; the induced map

$$\text{Hom}_{SH}(F_p(S^a \wedge G_m \wedge U_+), E) \xrightarrow{\iota_U, \iota_Y} \text{Hom}_{SH}(F_p(S^a \wedge G_m \wedge U_+), E)$$

is an isomorphism. We observe that

$$F_p(S^a \wedge G_m \wedge -) : \mathcal{M} \to Spt(M)$$

is a left Quillen functor, therefore the following

$$F_p(S^a \wedge G_m \wedge U_+) \xrightarrow{F_p(S^a \wedge G_m \wedge Y_+)} F_p(S^a \wedge G_m \wedge Y_+) \longrightarrow F_p(S^a \wedge G_m \wedge Y/U)$$

is a cofibre sequence in $Spt(M)$. However, $\mathcal{SH}$ is a triangulated category and lemma 2.1 implies that

$$F_{p+1}(S^a \wedge G_m^{b+1} \wedge Y/U) \cong \Omega_{S^1} \circ R \circ F_p(S^a \wedge G_m^{b} \wedge Y/U)$$

are isomorphic in $\mathcal{SH}$, where $R$ denotes a fibrant replacement functor in $Spt(M)$. Hence it suffices to show that

$$\text{Hom}_{SH}(F_{p+1}(S^a \wedge G_m^{b+1} \wedge Y/U), E) = \text{Hom}_{SH}(F_p(S^a \wedge G_m^{b} \wedge Y/U), E) = 0$$
Finally, it follows from lemma 3.4 that the following groups are isomorphic
of weak equivalences.

\[ \text{Hom}_{\text{SH}}(F_p(S^n \wedge G^b_m \wedge U_+), E) = 0 \]

for every \( F_p(S^n \wedge G^b_m \wedge U_+ \in C^{n+1}_{\text{eff}} \).

The same argument as in lemma 2.5 implies that it is enough to consider the
case when \( b, p \geq 1 \) and \( F_p(S^1 \wedge G^b_m \wedge U_+) \in C^{n+1}_{\text{eff}} \). In effect, if \( F_p(G^b_m \wedge U_+) \in C^{n+1}_{\text{eff}} \),
then lemma 2.1 implies that the natural map
\[ g_{p+1}^{1,b+1}(U) : F_{p+1}(S^1 \wedge G^b_m \wedge U_+) \rightarrow F_p(G^b_m \wedge U_+) \]
is a weak equivalence in \( \text{Spt}(\mathcal{M}) \).

Now, it follows from lemma 3.4, that if \( b \geq 1 \), and \( 0 - p + (b - 1) \geq n \) (i.e.
\( b - p \geq n + 1 \); then \( F_p(G^{b-1}_m \wedge mU) \in sW B_n \), i.e. a weak equivalence in \( WB_n \text{Spt}(\mathcal{M}) \).

Since \( \text{SH}(WB_n) \) is a triangulated category, \( id : \text{Spt}(\mathcal{M}) \rightarrow WB_n \text{Spt}(\mathcal{M}) \) is a
left Quillen functor, and \( F_p(G^{b-1}_m \wedge (A^1_{\text{eff}}/(A^1_{\text{eff}} \wedge U_+))) \) is the homotopy cofibre of
\( F_p(G^{b-1}_m \wedge mU) \); we deduce that \( E \) being \( n + 1 \)-weakly birational implies that
\[ \text{Hom}_{\text{SH}}(F_p(G^{b-1}_m \wedge (A^1_{\text{eff}}/(A^1_{\text{eff}} \wedge U_+))), E) = 0 \]

Finally, it follows from lemma 3.4 that the following groups are isomorphic
\[ 0 = \text{Hom}_{\text{SH}}(F_p(G^{b-1}_m \wedge (A^1_{\text{eff}}/(A^1_{\text{eff}} \wedge U_+))), E) \]
\[ \cong \text{Hom}_{\text{SH}}(F_p(S^1 \wedge G^b_m \wedge U_+), E) \]

(2)⇒(1): Assume that \( E \) is \( n + 1 \)-weakly birational. Then, proposition 3.3.30
in [12] implies that it suffices to show that
\[ \text{Hom}_{\text{SH}}(F_p(S^n \wedge G^b_m \wedge U_+), E) = 0 \]

But this follows from lemma 3.3 together with [12, Prop. 3.3.30], since we are
assuming that \( E \) is fibrant in \( L_{<n+1} \text{Spt}(\mathcal{M}) \).

\[ \text{The Quillen adjunction} \]
\[ (id, id, \varphi) : WB_n \text{Spt}(\mathcal{M}) \rightarrow L_{<n+1} \text{Spt}(\mathcal{M}) \]
is a Quillen equivalence. In addition, if the base scheme \( X = \text{Spec } k \), with \( k \) a
perfect field, then the Quillen adjunction
\[ (id, id, \varphi) : B_n \text{Spt}(\mathcal{M}) \rightarrow L_{<n+1} \text{Spt}(\mathcal{M}) \]
is also a Quillen equivalence.

Proof. We show first that \( WB_n \text{Spt}(\mathcal{M}) \) and \( L_{<n+1} \text{Spt}(\mathcal{M}) \) are Quillen equivalent.
Since \( WB_n \text{Spt}(\mathcal{M}) \), \( L_{<n+1} \text{Spt}(\mathcal{M}) \) are both left Bousfield localizations of \( \text{Spt}(\mathcal{M}) \), we deduce that they are simplicial model categories with the same
cofibrant replacement functor \( Q \). Thus, it suffices to show that they have the
same class of weak equivalences.

Documenta Mathematica 18 (2013) 51–70
However, proposition 3.5 implies that $WB_n Spt(M)$, and $L_{<n+1} Spt(M)$ also have the same class of fibrant objects. Therefore, it follows from [2, Thm. 9.7.4] that they have exactly the same class of weak equivalences.

Finally, if the base scheme is a perfect field, by proposition 2.12 we conclude that $WB_n Spt(M)$ and $B_n Spt(M)$ are Quillen equivalent.

**Theorem 3.7.** Let $E$ be fibrant in $Spt(M)$. Then $E$ is an $n$-slice (see definition 3.1(2)) if and only if the following conditions hold:

S1 $E$ is $n$-effective, i.e. $E \in \Sigma^n TSH^{eff}$.

S2 $E$ is $n+1$-weakly birational.

In addition, if the base scheme $X = \text{Spec } k$, with $k$ a perfect field, then $E$ is an $n$-slice if and only if the following conditions hold:

GSS1 $E$ is $n$-effective, i.e. $E \in \Sigma^n_T SH^{eff}$.

GSS2 $E$ is $n+1$-birational.

**Proof.** Assume that $E$ is an $n$-slice. Then theorems 1.4(1) and 1.4(3) imply that $E$ is $n$-effective and fibrant in $L_{<n+1} Spt(M)$.

Hence, proposition 3.5 implies that $E$ is also $n+1$-weakly birational.

Now we assume that $E$ satisfies the conditions S1 and S2 above. Then, proposition 3.5 implies that $E$ is fibrant in $L_{<n+1} Spt(M)$. Therefore, theorem 1.4(3) implies that $E$ is isomorphic in $SH$ to its own $n$-slice $s_n(E)$.

Finally, if the base scheme is a perfect field, then by proposition 3.5 the conditions S2 and GSS2 are equivalent; hence we can conclude applying the same argument as above.

**Remark 3.8.** Notice that theorem 3.6 implies that it is possible to construct the slice filtration directly from the Quillen model categories $WB_n Spt(M)$ described in definition 2.9 without making any reference to the effective categories $\Sigma^n TSH^{eff}$. One of the interesting consequences of this fact is that it is possible to obtain analogues of the slice filtration in the unstable setting, since the suspension with respect to $T$ or $S^1$ does not play an essential role in the construction of $WB_n Spt(M)$, i.e. we could consider the left Bousfield localization of the motivic unstable homotopy category $M$ with respect to the maps in definition 2.8. We will study the details of this construction in a future work.

4 Some Computations

In this section we use the characterization of the slices obtained in theorem 3.7 to describe the slices of projective spaces, Thom spaces and blow ups.

To simplify the notation, given a simplicial presheaf $K \in M$ or a map $f \in M$; let $s_j(K)$, $s_j(f)$ (respectively $s_{<j}(K)$, $s_{<j}(f)$) denote $s_j(F_0(K))$, $s_j(F_0(f))$ (respectively $s_{<j}(F_0(K))$, $s_{<j}(F_0(f))$).
Lemma 4.1. Let \( g : E \to F \) be a map in \( SH \) such that \( s_{<n}(g) \) and \( s_{<n+1}(g) \) are both isomorphisms in \( SH \). Then the \( n \)-slice of \( g \), \( s_n(g) \) is also an isomorphism in \( SH \).

Proof. It follows from [12, Prop. 3.1.19] that the rows in the following commutative diagram are distinguished triangles in \( SH \)

\[
\begin{array}{cccc}
s_n(E) & s_{<n+1}(E) & s_{<n}(E) & S^1 \land s_n(E) \\
\downarrow s_n(g) & \downarrow s_{<n+1}(g) & \downarrow s_{<n}(g) & \downarrow S^1 \land s_n(g) \\
s_n(F) & s_{<n+1}(F) & s_{<n}(F) & S^1 \land s_n(F)
\end{array}
\]

Thus, we conclude that \( s_n(g) \) is also an isomorphism in \( SH \).

Consider \( Y \in Sm_X \). Let \( \mathbb{P}^n(Y) \) denote the trivial projective bundle of rank \( n \) over \( Y \), and let \( \mathbb{P}^\infty(Y) \) denote the colimit in \( M \) of the following filtered diagram

\[
\mathbb{P}^0(Y) \to \mathbb{P}^1(Y) \to \cdots \to \mathbb{P}^n(Y) \to \cdots
\]

given by the inclusions of the respective hyperplanes at infinity.

Theorem 4.2. Let \( Y \in Sm_X \). Then for any integer \( j \leq n \), the diagram 4.1 induces the following isomorphisms in \( SH \)

\[
s_j(\mathbb{P}^n(Y)_+) \cong s_j(\mathbb{P}^{n+1}(Y)_+) \cong \cdots \cong s_j(\mathbb{P}^\infty(Y)_+)
\]

Proof. Let \( k > n \), and consider the closed embedding induced by the diagram (4.1) \( \lambda^k_n : \mathbb{P}^n(Y) \to \mathbb{P}^k(Y) \). It is possible to choose a linear embedding \( \mathbb{P}^{k-n-1}(Y) \to \mathbb{P}^k(Y) \) such that its open complement \( U_{k,n} \) contains \( \mathbb{P}^n(Y) \) and has the structure of a vector bundle over \( \mathbb{P}^n(Y) \), with zero section \( \sigma^k_n \):

\[
\begin{array}{ccc}
U_{k,n} & \xrightarrow{s_n^k} & \mathbb{P}^k(Y) \\
\downarrow \sigma^k_n & & \downarrow \lambda^k_n \\
\mathbb{P}^n(Y) & & \mathbb{P}^{k-n-1}(Y)
\end{array}
\]

By homotopy invariance \( s_{<j}(\sigma^k_n) \) is an isomorphism in \( SH \) for every integer \( j \). On the other hand, if \( j \leq n \), then \( F_0(v^k_n) \) is a weak equivalence in \( WB_j Spt(M) \) since the codimension of its closed complement is \( n+1 \). Thus, theorems 1.4(2) and 3.6 imply that if \( j \leq n + 1 \), then \( s_{<j}(v^k_n) \) is also an isomorphism in \( SH \).

Therefore, \( s_{<j}(\lambda^k_n) = s_{<j}(v^k_n) \circ s_{<j}(\sigma^k_n) \) is an isomorphism in \( SH \) for \( j \leq n+1 \); and using lemma 4.1 we conclude that the induced map on the slices \( s_j(\lambda^k_n) \) is also an isomorphism for \( j \leq n \).

Finally, the result for \( \mathbb{P}^\infty(Y) \) follows directly from the fact that the slices commute with filtered homotopy colimits.

\[
Documenta Mathematica 18 (2013) 51–70
\]
Let $HZ$ denote Voevodsky’s Eilenberg-MacLane spectrum (see [15, section 6.1]) representing motivic cohomology in $\mathcal{SH}$.

**Corollary 4.3.** Assume that the base scheme $X = \text{Spec } k$, with $k$ a perfect field. Then, in the following diagram all the symmetric $T$-spectra are isomorphic to $HZ$:

$$
HZ \xrightarrow{\cong} s_0(\mathbb{P}^0(k)_+) \xrightarrow{\cong} s_0(\mathbb{P}^1(k)_+) \xrightarrow{\cong} \cdots
$$

$$
\cdots \xrightarrow{\cong} s_0(\mathbb{P}^n(k)_+) \xrightarrow{\cong} \cdots \xrightarrow{\cong} s_0(\mathbb{P}^{\infty}(k)_+)
$$

**Proof.** This follows immediately from theorem 4.2 together with the computation of Levine [7, Thm. 10.5.1] and Voevodsky [17] for the zero slice of the sphere spectrum.

**Theorem 4.4.** Let $\omega_{U,Y} \in WB_n$, $\pi : V \to Y$ a vector bundle of rank $r$ together with a trivialization $t : \pi^{-1}(U) \to \mathbb{A}^r_U$ of its restriction to $U$. Then for every integer $j \leq n$, there exists an isomorphism in $\mathcal{SH}$ (see definition 3.2(2))

$$s_j(Th(V)) \cong S^r \wedge \mathbb{G}_m^r \wedge s_{j-r}(Y_+)$$

**Proof.** Let $Z \in Sm_X$ be the closed complement of $\omega_{U,Y}$. Consider the following diagram in $Sm_X$, where all the squares are cartesian

$$
\begin{array}{ccc}
\pi^{-1}(Z) \cap (V \setminus \sigma_0(Y)) & \xrightarrow{$\beta$} & \pi^{-1}(U) \cap (V \setminus \sigma_0(Y)) \\
\downarrow & & \downarrow \\
\pi^{-1}(Z) & \xrightarrow{$\alpha$} & \pi^{-1}(U) \\
\downarrow & & \downarrow \\
Z & \xrightarrow{$\pi$} & Y \xrightarrow{$\omega_{U,Y}$} U
\end{array}
$$

and let $\gamma : Th(\pi^{-1}(U)) \to Th(V)$ be the induced map between the corresponding Thom spaces. We observe that $\alpha, \beta$ also belong to $WB_n$; thus, if $j \leq n$ we conclude that $F_0(\omega_{U,Y}), F_0(\alpha), F_0(\beta)$ are all weak equivalences in $WB_j \text{spt}(\mathcal{M})$. Therefore, theorems 1.4(2) and 3.6 imply that if $j \leq n+1$, then $s_{<j}(\omega_{U,Y}), s_{<j}(\alpha), s_{<j}(\beta)$ are isomorphisms in $\mathcal{SH}$. We claim that if $j \leq n+1$, then

$$s_{<j}(\gamma) : s_{<j}(Th(\pi^{-1}(U))) \to s_{<j}(Th(V))$$

is also an isomorphism in $\mathcal{SH}$. In effect, by construction of the Thom spaces, we deduce that for any integer $j \in \mathbb{Z}$, the rows in the following commutative diagram in $\mathcal{SH}$ are in fact distinguished triangles

$$
\begin{array}{ccc}
s_{<j}(\pi^{-1}(U) \cap (V \setminus \sigma_0(Y)))_+ & \xrightarrow{s_{<j}(\pi^{-1}(U)_+)} & s_{<j}(Th(\pi^{-1}(U))) \\
\downarrow & & \downarrow \\
s_{<j}(\sigma_0(Y)_+)_+ & \xrightarrow{s_{<j}(\sigma_0(Y)_+)} & s_{<j}(Th(V)) \\
\downarrow & & \downarrow \\
s_{<j}(V)_+ & \xrightarrow{s_{<j}(\gamma)} & s_{<j}(Th(V))
\end{array}
$$
Since \( s_{<j}(\alpha), s_{<j}(\beta) \) are isomorphisms in \( \mathcal{SH} \) for \( j \leq n + 1 \), we conclude that for \( j \leq n + 1 \), \( s_{<j}(\gamma) \) is also an isomorphism in \( \mathcal{SH} \). Thus, lemma 4.1 implies that for \( j \leq n \), \( s_j(U,Y) \), \( s_j(\gamma) \) are isomorphisms in \( \mathcal{SH} \). Now, we use the trivialization \( t \) to obtain the following commutative diagram in \( Sm_X \) where the rows are isomorphisms

\[
\begin{array}{c}
K_U \setminus U & \overset{\approx}{\rightarrow} & \pi^{-1}(U) \cap (V \setminus \sigma_0(Y)) \\
\downarrow & & \downarrow \\
K_U & \overset{\approx}{\rightarrow} & \pi^{-1}(U) \\
\downarrow & & \downarrow \\
U & \overset{\pi_U}{\rightarrow} & \pi^{-1}(U)
\end{array}
\]

The same argument as above, shows that for every integer \( j \in \mathbb{Z} \), there is an isomorphism in \( \mathcal{SH} \)

\[
s_j(t) : s_j(Th(\pi^{-1}(U))) \rightarrow s_j(Th(K_U))
\]

On the other hand, [8, Prop. 2.17(2)] implies that there is a weak equivalence \( w : F_0(Th(K_U)) \rightarrow S^r \wedge G^r_m \wedge F_0(U_+) \) in \( Spt(M) \). Thus, for \( j \leq n \) there exist isomorphisms in \( \mathcal{SH} \)

\[
s_j(Th(\pi^{-1}(U))) \overset{s_j(t)}{\rightarrow} s_j(Th(K_U)) \overset{s_j(\gamma)}{\downarrow} s_j(Th(V)) \overset{s_j(w)}{\rightarrow} s_j(S^r \wedge G^r_m \wedge U_+) \]

However, there exists a canonical isomorphism in \( \mathcal{SH} \)

\[
s_j(S^r \wedge G^r_m \wedge U_+) \rightarrow S^r \wedge G^r_m \wedge s_j-r(U_+)
\]

Finally, we conclude by using the isomorphism \( s_{j-r}(t_{U,Y}) \) (notice that if \( j \leq n \) then certainly \( j - r \leq n \), since \( r \geq 0 \)).

**Corollary 4.5.** Assume that the base scheme \( X = \text{Spec} \ k \), with \( k \) a perfect field. Let \( t_{U,Y} \in B_n, \pi : V \rightarrow Y \) a vector bundle of rank \( r \) together with a trivialization \( t : \pi^{-1}(U) \rightarrow K_U \) of its restriction to \( U \). Then for every integer \( j \leq n \), there exists an isomorphism in \( \mathcal{SH} \)

\[
s_j(Th(V)) \cong S^r \wedge G^r_m \wedge s_j-r(Y_+).
\]

**Proof.** Proposition 2.12 implies that \( F_0(t_{U,Y}) \) is a weak equivalence in \( WB_j Spt(M) \) for \( j \leq n \). Hence, the result follows using exactly the same argument as in theorem 4.4. \( \square \)
Given a closed embedding $Z \to Y$ of smooth schemes over $X$, let $\mathcal{B}ZY$ denote the blowup of $Y$ with center in $Z$.

**Theorem 4.6.** Let $i_{U,Y} \in WB_n$ with closed complement $Z$, and $j \in \mathbb{Z}$ an arbitrary integer. Consider the following cartesian square in $\text{Sm}_X$

$$
\begin{array}{ccc}
D & \xrightarrow{d} & \mathcal{B}ZY \\
\downarrow q & & \downarrow u \\
Z & \xrightarrow{i} & Y
\end{array}
$$

(4.2)

and let $q_j, d_j, p_j, i_j$ denote $s_j(q), s_j(d), s_j(p), s_j(i)$ respectively. Then the cartesian square (4.2) induces the following distinguished triangle in $\mathcal{SH}$

$$
s_j(D_{+}) \xrightarrow{\left(\frac{d}{q_j}\right)} s_j(\mathcal{B}ZY_{+}) \oplus s_j(Z_{+}) \xrightarrow{\left(\frac{p_j}{i_j}\right)} s_j(Y_{+})
$$

(4.3)

If $j \leq n$, then $s_j(i_{U,Y})$ is an isomorphism in $\mathcal{SH}$, and the following distinguished triangles in $\mathcal{SH}$ split

$$
s_j(D_{+}) \xrightarrow{\left(\frac{d}{q_j}\right)} s_j(\mathcal{B}ZY_{+}) \oplus s_j(Z_{+}) \xrightarrow{\left(\frac{p_j}{i_j}\right)} s_j(Y_{+})
$$

(4.4)

$$
s_j(Y_{+}) \xrightarrow{r_j} s_j(\mathcal{B}ZY_{+}) \xrightarrow{\left(\frac{d}{q_j}\right)} s_j(Y_{+})
$$

(4.5)

where $r_j = s_j(u) \circ (s_j(i_{U,Y}))^{-1}$, and $O_D(1)$ denotes the canonical line bundle of the projective bundle $q : D \to Z$.

**Proof.** It follows from [8, Prop. 2.29 and Rmk. 2.30] that the following square is homotopy cocartesian in $\mathcal{M}$

$$
\begin{array}{ccc}
S^1 \wedge D_{+} & \xrightarrow{id \wedge d} & S^1 \wedge \mathcal{B}ZY_{+} \\
\downarrow id \wedge q & & \downarrow id \wedge p \\
S^1 \wedge Z_{+} & \xrightarrow{id \wedge i} & S^1 \wedge Y_{+}
\end{array}
$$

Thus, we deduce that the following diagram is a distinguished triangle in $\mathcal{SH}$

$$
F_0(D_{+}) \xrightarrow{\left(\frac{d}{q_j}\right)} F_0(\mathcal{B}ZY_{+}) \oplus F_0(Z_{+}) \xrightarrow{\left(\frac{p_j}{i_j}\right)} F_0(Y_{+})
$$

Since the slices $s_j$ are triangulated functors, it follows that diagram (4.3) is a distinguished triangle in $\mathcal{SH}$.
Now, we prove that $s_j(\iota U, Y)$ is an isomorphism for $j \leq n$. By lemma 4.1, it suffices to show that $s_{<j}(\iota U, Y)$ is an isomorphism in $\mathcal{SH}$ for $j \leq n + 1$. But this follows directly from theorems 3.6 and 1.4(2) since $F_0(\iota U, Y)$ is clearly a weak equivalence in $WB_j Spt(M)$ for $j \leq n$.

Thus, $r_j$ is well defined for $j \leq n$, and the following diagram shows that it gives a splitting for the distinguished triangle (4.4)

\[
\begin{array}{ccc}
    s_j(U_+) & \xrightarrow{s_j(u)} & s_j(B\ell Z Y) \\
    \downarrow & & \downarrow p_j \\
    s_j(U_+) & \xrightarrow{s_j(\iota U, Y)} & s_j(Y) \\
\end{array}
\] (4.6)

Finally, since the normal bundle of the closed embedding $d : D \to B\ell Z Y$ is given by $O_D(1)$, we deduce from the Morel-Voevodsky homotopy purity theorem (see [8, Thm. 2.23]) that the following diagram is a distinguished triangle in $\mathcal{SH}$

\[
s_j(U_+) \xrightarrow{s_j(u)} s_j(B\ell Z Y) \rightarrow s_j(Th(O_D(1)))
\]

Combining this distinguished triangle with diagram (4.6) above, we conclude that diagram (4.5) is a split distinguished triangle in $\mathcal{SH}$ for $j \leq n$.

**Acknowledgements**

The author would like to thank Marc Levine for bringing to his attention the connection between slices and birational cohomology theories, as well as for all his stimulating comments and questions, and also thank Chuck Weibel for his interest in this work and several suggestions which helped to improve the exposition.

**References**


Pablo Pelaez
Department of Mathematics
Rutgers University, U.S.A.
pablo.pelaez@rutgers.edu
COHERENCE FOR WEAK UNITS

ANDRÉ JOYAL AND JOACHIM KOCK

Received: September 21, 2011
Communicated by Stefan Schwede

Abstract. We define weak units in a semi-monoidal 2-category $C$ as cancellable pseudo-idempotents: they are pairs $(I, \alpha)$ where $I$ is an object such that tensoring with $I$ from either side constitutes a biequivalence of $C$, and $\alpha : I \otimes I \to I$ is an equivalence in $C$. We show that this notion of weak unit has coherence built in: Theorem A: $\alpha$ has a canonical associator 2-cell, which automatically satisfies the pentagon equation. Theorem B: every morphism of weak units is automatically compatible with those associators. Theorem C: the 2-category of weak units is contractible if non-empty. Finally we show (Theorem E) that the notion of weak unit is equivalent to the notion obtained from the definition of tricategory: $\alpha$ alone induces the whole family of left and right maps (indexed by the objects), as well as the whole family of Kelly 2-cells (one for each pair of objects), satisfying the relevant coherence axioms.

2010 Mathematics Subject Classification: 18D05; 18D10
Keywords and Phrases: Monoidal 2-categories, units, coherence.

Introduction

The notion of tricategory, introduced by Gordon, Power, and Street [2] in 1995, seems still to represent the highest-dimensional explicit weak categorical structure that can be manipulated by hand (i.e. without methods of homotopy theory), and is therefore an important test bed for higher-categorical ideas. In this work we investigate the nature of weak units at this level. While coherence for weak associativity is rather well understood, thanks to the geometrical insight provided by the Stasheff associahedra [12], coherence for unit structures is more mysterious, and so far there seems to be no clear geometric pattern for the coherence laws for units in higher dimensions. Specific interest in weak units
stems from Simpson’s conjecture [11], according to which strict \( n \)-groupoids with weak units should model all homotopy \( n \)-types.

In the present paper, working in the setting of a strict 2-category \( \mathscr{C} \) with a strict tensor product, we define a notion of weak unit by simple axioms that involve only the notion of equivalence, and hence in principle make sense in all dimensions. Briefly, a weak unit is a cancellable pseudo-idempotent. We work out the basic theory of such units, and compare with the notion extracted from the definition of tricategory. In the companion paper Weak units and homotopy 3-types [4] we employ this notion of unit to prove a version of Simpson’s conjecture for 1-connected homotopy 3-types, which is the first nontrivial case. The strictness assumptions of the present paper should be justified by that result.

By cancellable pseudo-idempotent we mean a pair \((I, \alpha)\) where \(I\) is an object in \( \mathscr{C} \) such that tensoring with \(I\) from either side is an equivalence of 2-categories, and \( \alpha : I \otimes I \cong I \) is an equi-arrow (i.e. an arrow admitting a pseudo-inverse).

The remarkable fact about this definition is that \(\alpha\), viewed as a multiplication map, comes with canonical higher order data built in: it possesses a canonical associator \(A\) which automatically satisfies the pentagon equation. This is our Theorem A. The point is that the arrow \(\alpha\) alone, thanks to the cancellability of \(I\), induces all the usual structure of left and right constraints with all the 2-cell data that goes into them and the axioms they must satisfy.

As a warm-up to the various constructions and ideas, we start out in Section 1 by briefly running through the corresponding theory for cancellable-idempotent units in monoidal 1-categories. This theory has been treated in detail in [8].

The rest of the paper is dedicated to the case of monoidal 2-categories. In Section 2 we give the definitions and state the main results: Theorem A says that there is a canonical associator 2-cell for \(\alpha\), and that this 2-cell automatically satisfies the pentagon equation. Theorem B states that unit morphisms automatically are compatible with the associators of Theorem A. Theorem C states that the 2-category of units is contractible if non-empty. Hence, ‘being unital’ is, up to homotopy, a property rather than a structure.

Next follow three sections dedicated to proofs of each of these three theorems. In Section 3 we show how the map \(\alpha : II \cong I\) alone induces left and right constraints, which in turn are used to construct the associator and establish the pentagon equation. The left and right constraints are not canonical, but surprisingly the associator does not depend on the choice of them. In Section 4 we prove Theorem B by interpreting it as a statement about units in the 2-category of arrows, where it is possible to derive it from Theorem A. In Section 5 we prove Theorem C. The key ingredient is to use the left and right constraints to link up all the units, and to show that the unit morphisms are precisely those compatible with the left and right constraints; this makes them ‘essentially unique’ in the required sense.
In Section 6 we go through the basic theory of classical units (i.e. as extracted from the definition of tricategory [2]). Finally, in Section 7 we show that the two notions of unit are equivalent. This is our Theorem E. A curiosity implied by the arguments in this section is that the left and right axioms for the 2-cell data in the Gordon-Power-Street definition (denoted TA2 and TA3 in [2]) imply each other.

(We have no Theorem D.)

This notion of weak units as cancellable idempotents is precisely what can be extracted from the more abstract, Tamsamani-style, theory of fair $n$-categories [7] by making an arbitrary choice of a fixed weak unit. In the theory of fair categories, the key object is a contractible space of all weak units, rather than any particular point in that space, and handling this space as a whole bypasses coherence issues. However, for the sake of understanding what the theory entails, and for the sake of concrete computations, it is interesting to make a choice and study the ensuing coherence issues, as we do in this paper. The resulting approach is very much in the spirit of the classical theory of monoidal categories, bicategories, and tricategories, and provides some new insight to these theories. To stress this fact we have chosen to formulate everything from scratch in such classical terms, without reference to the theory of fair categories.

In the case of monoidal 1-categories, the cancellable-idempotent viewpoint on units goes back to Saavedra [10]. The importance of this viewpoint in higher categories was first suggested by Simpson [11], in connection with his weak-unit conjecture. He gave an ad hoc definition in this style, as a mere indication of what needed to be done, and raised the question of whether higher homotopical data would have to be specified. The surprising answer is, at least here in dimension 3, that specifying $\alpha$ is enough, then the higher homotopical data is automatically built in.

ACKNOWLEDGEMENTS. We thank Georges Maltsiniotis for pointing out to us that the cancellable-idempotent notion of unit in dimension 1 goes back to Saavedra [10], and we thank Josep Elgueta for catching an error in an earlier version of our comparison with tricategories. The first-named author was supported by the NSERC. The second-named author was very happy to be a CIRGET postdoc at the UQAM in 2004, and currently holds support from grants MTM2009-10359 and MTM2010-20692 of Spain.

1 Units in monoidal categories

It is helpful first briefly to recall the relevant results for monoidal categories, referring the reader to [8] for further details of this case.

1.1. Semi-monoidal categories. A semi-monoidal category is a category $\mathcal{C}$ equipped with a tensor product (which we denote by plain juxtaposition),
i.e. an associative functor

\[ C \times C \rightarrow C \]

\((X,Y) \mapsto XY.\)

For simplicity we assume strict associativity, \(X(YZ) = (XY)Z.\)

1.2. **Monoidal categories.** (Mac Lane [9].) A semi-monoidal category \(C\) is a **monoidal category** when it is furthermore equipped with a distinguished object \(I\) and natural isomorphisms

\[ IX \overset{\lambda_X}{\longrightarrow} X \overset{\rho_X}{\longleftarrow} XI \]

obeying the following rules (cf. [9]):

\[ \lambda_I = \rho_I \quad (1) \]
\[ \lambda_{XY} = \lambda_X Y \quad (2) \]
\[ \rho_{XY} = X \rho_Y \quad (3) \]
\[ X \lambda_Y = \rho_X Y \quad (4) \]

Naturality of \(\lambda\) and \(\rho\) implies

\[ \lambda_{IX} = I \lambda_X, \quad \rho_{XI} = \rho_X I, \quad (5) \]

independently of Axioms (1)–(4).

1.3 **Remark.** Tensoring with \(I\) from either side is an equivalence of categories.

1.4 **Lemma.** (Kelly [5].) Axiom (4) implies axioms (1), (2), and (3).

**Proof.** (4) implies (2): Since tensoring with \(I\) on the left is an equivalence, it is enough to prove \(I\lambda_{XY} = I\lambda_X Y\). But this follows from Axiom (4) applied twice (swap \(\lambda\) out for a \(\rho\) and swap back again only on the nearest factor):

\[ I\lambda_{XY} = \rho_I XY = I\lambda_X Y. \]

Similarly for \(\rho\), establishing (3).

(4) and (2) implies (1): Since tensoring with \(I\) on the right is an equivalence, it is enough to prove \(\lambda_I I = \rho_I I\). But this follows from (2), (5), and (4):

\[ \lambda_I I = \lambda_{II} = I\lambda_I = \rho_I I. \]

The following alternative notion of unit object goes back to Saavedra [10]. A thorough treatment of the notion was given in [8].
1.5. Units as cancellable pseudo-idempotents. An object $I$ in a semi-monoidal category $\mathcal{C}$ is called **cancellable** if the two functors $\mathcal{C} \to \mathcal{C}$

\[
\begin{align*}
X & \longrightarrow IX \\
X & \longrightarrow XI
\end{align*}
\]

are fully faithful. By definition, a **pseudo-idempotent** is an object $I$ equipped with an isomorphism $\alpha : II \cong I$. Finally we define a **unit object** in $\mathcal{C}$ to be a cancellable pseudo-idempotent.

1.6 Lemma. [8] Given a unit object $(I, \alpha)$ in a semi-monoidal category $\mathcal{C}$, for each object $X$ there are unique arrows $\lambda_X : IX \to XI$ and $\rho_X : XI \to IX$ such that

\[
\begin{align*}
(L) & \quad I\lambda_X = \alpha X \\
(R) & \quad \rho_X I = X\alpha.
\end{align*}
\]

The $\lambda_X$ and $\rho_X$ are isomorphisms and natural in $X$.

**Proof.** Let $L : \mathcal{C} \to \mathcal{C}$ denote the functor defined by tensoring with $I$ on the left. Since $L$ is fully faithful, we have a bijection

\[
\text{Hom}(IX, X) \to \text{Hom}(II X, IX).
\]

Now take $\lambda_X$ to be the inverse image of $\alpha X$; it is an isomorphism since $\alpha X$ is. Naturality follows by considering more generally the bijection

\[
\text{Nat}(L, \text{id}_\mathcal{C}) \to \text{Nat}(L \circ L, L);
\]

let $\lambda$ be the inverse image of the natural transformation whose components are $\alpha X$. Similarly on the right. $\square$

1.7 Lemma. [8] For $\lambda$ and $\rho$ as above, the Kelly axiom (4) holds:

\[
X\lambda_Y = \rho_X Y.
\]

Therefore, by Lemma 1.6 a semi-monoidal category with a unit object is a monoidal category in the classical sense.

**Proof.** In the commutative square

\[
\begin{array}{ccc}
XIIY & \xrightarrow{XI\lambda_Y} & XIY \\
\rho_X Y \downarrow & & \rho_X Y \\
XIY & \xrightarrow{X\lambda_Y} & XY
\end{array}
\]
the top arrow is equal to $X\alpha Y$, by $X$ tensor $(L)$, and the left-hand arrow is also equal to $X\alpha Y$, by $(R)$ tensor $Y$. Since $X\alpha Y$ is an isomorphism, it follows that $X\lambda Y = \rho_X Y$.

1.8 Lemma. For a unit object $(I, \alpha)$ we have: (i) The map $\alpha : II \to I$ is associative. (ii) The two functors $X \mapsto IX$ and $X \mapsto XI$ are equivalences.

Proof. Since $\alpha$ is invertible, associativity amounts to the equation $I\alpha = \alpha I$, which follows from the previous proof by setting $X = Y = I$ and applying $L$ and $R$ once again. To see that $L$ is an equivalence, just note that it is isomorphic to the identity via $\lambda$.

1.9. Uniqueness of units. Just as in a semi-monoid a unit element is unique if it exists, one can show [8, 2.20] that in a semi-monoidal category, between any two units there is a unique isomorphism of units. This statement does not involve $\lambda$ and $\rho$, but the proof does: the canonical isomorphism $I \cong J$ is the composite $I \sim IJ \lambda_J \to J$.

2 Units in monoidal 2-categories: definition and main results

In this section we set up the necessary terminology and notation, give the main definition, and state the main results.

2.1. 2-categories. We work in a strict 2-category $\mathscr{C}$. We use the symbol $#$ to denote composition of arrows and horizontal composition of 2-cells in $\mathscr{C}$, always written from the left to the right, and occasionally decorating the symbol $#$ by the name of the object where the two arrows or 2-cells are composed. By an equi-arrow in $\mathscr{C}$ we understand an arrow $f$ admitting an (unspecified) pseudo-inverse, i.e. an arrow $g$ in the opposite direction such that $f#g$ and $g#f$ are isomorphic to the respective identity arrows, and such that the comparison 2-cells satisfy the usual triangle equations for adjunctions. (The usual word for ‘equi-arrow’ is of course ‘equivalence’; we reserve the latter word for equivalence of categories and 2-categories. We find it useful to have a different word for the equivalences inside a 2-category.) It is worth pointing out that it is not necessary to insist on the triangle equations. If the 2-cells exist but do not satisfy the triangle equation, they can always be replaced by 2-cells that do. We shall make extensive use of arguments with pasting diagrams [6]. Our drawings of 2-cells should be read from bottom to top, so that for example

\[
\begin{array}{ccc}
X & \xrightarrow{h} & Z \\
\downarrow{f} & & \downarrow{g} \\
U & \xrightarrow{U} & Y \\
\end{array}
\]

denotes $U : f \# g \Rightarrow h$. The symbol $\circ$ will denote identity 2-cells.
The few 2-functors we need all happen to be strict. By *natural transformation* we always mean pseudo-natural transformation. Hence a natural transformation $u : F \Rightarrow G$ between two 2-functors from $\mathcal{D}$ to $\mathcal{C}$ is given by an arrow $u_X : FX \to GX$ for each object $X \in \mathcal{D}$, and an invertible 2-cell

\[
\begin{array}{ccc}
FX & \xrightarrow{u_X} & GX \\
\downarrow{F(x)} & & \downarrow{G(x)} \\
FX' & \xrightarrow{u_{X'}} & GX'
\end{array}
\]

for each arrow $x : X \to X'$ in $\mathcal{D}$, subject to the usual compatibility conditions [6]. The modifications we shall need will happen to be invertible.

2.2. Semi-monoidal 2-categories. By *semi-monoidal 2-category* we mean a 2-category $\mathcal{C}$ equipped with a tensor product, i.e. an associative 2-functor

\[
\otimes : \mathcal{C} \times \mathcal{C} \longrightarrow \mathcal{C}
\]

\[
\begin{array}{ccc}
(X,Y) & \mapsto & XY,
\end{array}
\]

denoted by plain juxtaposition. We already assumed $\mathcal{C}$ to be a strict 2-category, and we also require $\otimes$ to be a strict 2-functor and to be strictly associative: $(XY)Z = X(YZ)$. This is mainly for convenience, to keep the focus on unit issues.

Note that the tensor product of two equi-arrows is again an equi-arrow, since its pseudo-inverse can be taken to be the tensor product of the pseudo-inverses.

2.3. Semi-monoids. A *semi-monoid* in $\mathcal{C}$ is a triple $(X, \alpha, \tilde{\alpha})$ consisting of an object $X$, a multiplication map $\alpha : XX \to X$, and an invertible 2-cell $\tilde{\alpha}$ called the *associator*,

\[
\begin{array}{ccc}
XXX & \xrightarrow{\alpha X} & XX \\
\downarrow{x_\alpha} & & \downarrow{\alpha} \\
XX & \xrightarrow{\alpha} & X
\end{array}
\]

required to satisfy the ‘pentagon equation’:

\[
\begin{array}{ccc}
XXX & \xrightarrow{\alpha X} & XXX \\
\downarrow{x_\alpha} & & \downarrow{\alpha} \\
XX & \xrightarrow{\alpha} & X
\end{array}
= 
\begin{array}{ccc}
XXX & \xrightarrow{\alpha X} & XXX \\
\downarrow{x_\alpha} & & \downarrow{\alpha} \\
XX & \xrightarrow{\alpha} & X
\end{array}
\]
In the applications, $\alpha$ will be an equi-arrow, and hence we will have

$$\tilde{A} = A \#_{XX} \alpha$$

for a some unique invertible

$$A : X \alpha \Rightarrow \alpha X,$$

which it will more convenient to work with. In this case, the pentagon equation is equivalent to the more compact equation

$$XXX \alpha_{XX} \tilde{A} \alpha_{XX} = XXX \alpha_{XX} (\circlearrowright) X \alpha_{XX}$$

which we shall also make use of.

2.4. **Semi-monoid maps.** A *semi-monoid map* $f : (X, \alpha, \tilde{A}) \to (Y, \beta, \tilde{B})$ is the data of an arrow $f : X \to Y$ in $\mathcal{C}$ together with an invertible 2-cell

$$\begin{array}{ccc}
XX & \xrightarrow{ff} & YY \\
\downarrow \alpha & & \downarrow \beta \\
X & \xrightarrow{f} & Y
\end{array}$$

such that this cube commutes:

$$XXX \alpha_{XX} \tilde{A} \alpha_{XX} = XXX \beta_{YY} \tilde{B} \beta_{YY}$$

When $\beta$ is an equi-arrow, the cube equation is equivalent to the simpler equa-
tion:

\[
\begin{array}{c}
\begin{array}{c}
\begin{array}{c}
XX \xrightarrow{\alpha} \alpha X \xrightarrow{f f} Y Y
\end{array}
\end{array}
\end{array}
\end{array}
\]

which will be useful.

2.5. SEMI-MONOID TRANSFORMATIONS. A semi-monoid transformation between two parallel semi-monoid maps \((f, F)\) and \((g, G)\) is a 2-cell \(T : f \Rightarrow g\) in \(\mathcal{C}\) such that this cylinder commutes:

\[
\begin{array}{c}
\begin{array}{c}
\begin{array}{c}
XX \xrightarrow{\alpha} \alpha X \xrightarrow{f f} Y Y
\end{array}
\end{array}
\end{array}
\end{array}
\]

2.6 LEMMA. Let \(f : X \to Y\) be a semi-monoid map. If \(f\) is an equi-arrow (as an arrow in \(\mathcal{C}\)) with quasi-inverse \(g : Y \to X\), then there is a canonical 2-cell \(G\) such that \((g, G)\) is a semi-monoid map.

Proof. The 2-cell \(G\) is defined as the mate [6] of the 2-cell \(F^{-1}\). It is routine to check the cube equation in 2.4. \(\square\)

2.7. PSEUDO-IDEMPOTENTS. A pseudo-idempotent is a pair \((I, \alpha)\) where \(\alpha : II \to I\) is an equi-arrow. A morphism of pseudo-idempotents from \((I, \alpha)\) to \((J, \beta)\) is a pair \((u, U)\) consisting of an arrow \(u : I \to J\) in \(\mathcal{C}\) and an invertible 2-cell

\[
\begin{array}{c}
\begin{array}{c}
\begin{array}{c}
II \xrightarrow{u} JJ
\end{array}
\end{array}
\end{array}
\end{array}
\]

If \((u, U)\) and \((v, V)\) are morphisms of pseudo-idempotents from \((I, \alpha)\) to \((J, \beta)\), a 2-morphism of pseudo-idempotents from \((u, U)\) to \((v, V)\) is a 2-cell \(T : u \Rightarrow v\) satisfying the cylinder equation of 2.5.
2.8. Cancellable objects. An object \( I \) in \( \mathcal{C} \) is called cancellable if the two 2-functors \( \mathcal{C} \to \mathcal{C} \)

\[
X \mapsto IX \\
X \mapsto XI
\]

are fully faithful. (Fully faithful means that the induced functors on hom categories are equivalences.) A cancellable morphism between cancellable objects \( I \) and \( J \) is an equi-arrow \( u : I \to J \). (Equivalently it is an arrow such that the functors on hom cats defined by tensoring with \( u \) on either side are equivalences, cf. 5.1.) A cancellable 2-morphism between cancellable arrows is any invertible 2-cell.

We are now ready for the main definition and the main results.

2.9. Units. A unit object is by definition a cancellable pseudo-idempotent. Hence it is a pair \( (I, \alpha) \) consisting of an object \( I \) and an equi-arrow \( \alpha : II \to I \), with the property that tensoring with \( I \) from either side define fully faithful 2-functors \( \mathcal{C} \to \mathcal{C} \).

A morphism of units is a cancellable morphism of pseudo-idempotents. In other words, a unit morphism from \( (I, \alpha) \) to \( (J, \beta) \) is a pair \( (u, U) \) where \( u : I \to J \) is an equi-arrow and \( U \) is an invertible 2-cell

\[
\begin{array}{ccc}
II & \xrightarrow{uu} & JJ \\
\downarrow \alpha & & \downarrow \beta \\
I & \xrightarrow{u} & J.
\end{array}
\]

A 2-morphism of units is a cancellable 2-morphism of pseudo-idempotents. Hence a 2-morphism from \( (u, U) \) to \( (v, V) \) is an invertible 2-cell \( T : u \Rightarrow v \) such that

\[
\begin{array}{ccc}
II & \xrightarrow{uv} & JJ \\
\downarrow \alpha & & \downarrow \beta \\
I & \xrightarrow{u} & J
\end{array} = \begin{array}{ccc}
II & \xrightarrow{uv} & JJ \\
\downarrow \alpha & & \downarrow \beta \\
I & \xrightarrow{v} & J.
\end{array}
\]

This defines the 2-category of units.

In the next section we’ll see how the notion of unit object induces left and right constraints familiar from standard notions of monoidal 2-category. It will then turn out (Lemmas 5.1 and 5.2) that unit morphisms and 2-morphisms can be characterised as those morphisms and 2-morphisms compatible with the left and right constraints.
Theorem A (Associativity). Given a unit object \((I, \alpha)\), there is a canonical invertible 2-cell

\[
\begin{array}{c}
III \\
\downarrow_{I_{\alpha}} \\
II \\
\downarrow_{\alpha} \\
I \\
\end{array}
\]

which satisfies the pentagon equation

\[
\begin{array}{c}
\begin{array}{c}
III \\
\downarrow_{I_{\alpha}} \\
II \\
\downarrow_{\alpha} \\
I \\
\end{array}
\end{array}
\sim

\begin{array}{c}
\begin{array}{c}
III \\
\downarrow_{I_{\alpha}} \\
II \\
\downarrow_{\alpha} \\
I \\
\end{array}
\end{array}
\]

In other words, a unit object is automatically a semi-monoid. The 2-cell \(A\) is characterised uniquely in 3.7.

Theorem B. A unit morphism \((u, U) : (I, \alpha) \to (J, \beta)\) is automatically a semi-monoid map, when \(I\) and \(J\) are considered semi-monoids in virtue of Theorem A.

Theorem C (Contractibility). The 2-category of units in \(\mathcal{C}\) is contractible, if non-empty.

In other words, between any two units there exists a unit morphism, and between any two parallel unit morphisms there is a unique unit 2-morphism. Theorem C shows that units objects are unique up to homotopy, so in this sense ‘being unital’ is a property not a structure.

The proofs of these three theorems rely on the auxiliary structure of left and right constraints which we develop in the next section, and which also displays the relation with the classical notion of monoidal 2-category: in Section 7 we show that the cancellable-idempotent notion of unit is equivalent to the notion extracted from the definition of tricategory of Gordon, Power, and Street [2]. This is our Theorem E.

3 Left and right actions, and associativity of the unit (Theorem A)

Throughout this section we fix a unit object \((I, \alpha)\).
3.1 Lemma. For each object $X$ there exists pairs $(\lambda_X, L_X)$ and $(\rho_X, R_X)$,

$$\begin{align*}
\lambda_X : IX \to X, & \quad L_X : I\lambda_X \Rightarrow \alpha X \\
\rho_X : XI \to X, & \quad R_X : X\alpha \Rightarrow \rho_X I
\end{align*}$$

where $\lambda_X$ and $\rho_X$ are equi-arrows, and $L_X$ are $R_X$ are invertible 2-cells.

For every such family, there is a unique way to assemble the $\lambda_X$ into a natural transformation (this involves defining 2-cells $\lambda_f$ for every arrow $f$ in $\mathcal{C}$) in such a way that $L$ is a natural modification. Similarly for the $\rho_X$ and $R_X$.

The $\lambda_X$ is an action of $I$ on each $X$, and the 2-cell $L_X$ expresses an associativity constraint on this action. Using these structures we will construct the associator for $\alpha$, and show it satisfies the pentagon equation. Once that is in place we will see that the actions $\lambda$ and $\rho$ are coherent too (satisfying the appropriate pentagon equations).

We shall treat the left action. The right action is of course equivalent to establish.

3.2. Construction of the left action. Since tensoring with $I$ is a fully faithful 2-functor, the functor

$$\text{Hom}(IX, X) \to \text{Hom}(IIX, IX)$$

is an equivalence of categories. In the second category there is the canonical object $\alpha X$. Hence there is a pseudo pre-image which we denote $\lambda_X : IX \to X$, together with an invertible 2-cell $L_X : I\lambda_X \Rightarrow \alpha X$:

$$\begin{array}{ccc}
IIX & \xrightarrow{I\lambda_X} & IX \\
\alpha X & \downarrow{L_X} & \\
IX & \xrightarrow{\lambda_X} & IX
\end{array}$$

Since $\alpha$ is an equi-arrow, also $\alpha X$ is equi, and since $L_X$ is invertible, we conclude that also $I\lambda_X$ is an equi-arrow. Finally since the 2-functor ‘tensoring with $I$’ is fully faithful, it reflects equi-arrows, so already $\lambda_X$ is an equi-arrow.

3.3. Naturality. A slight variation in the formulation of the construction gives directly a natural transformation $\Lambda$ and a modification $L$: Let $L : \mathcal{C} \to \mathcal{C}$ denote the 2-functor ‘tensoring with $I$ on the left’. Since $L$ is fully faithful, there is an equivalence of categories

$$\text{Nat}(L, \text{Id}_{\mathcal{C}}) \to \text{Nat}(L \circ L, L).$$

Now in the second category we have the canonical natural transformation whose $X$-component is $\alpha X$ (and with trivial components on arrows). Hence there is a pseudo pre-image natural transformation $\lambda : L \to \text{Id}_{\mathcal{C}}$, together with a modification $L$ whose $X$-component is $L_X : I\lambda_X \Rightarrow \alpha X$. 

Documenta Mathematica 18 (2013) 71–110
Coherence for Weak Units

However, we wish to stress the fact that the construction is completely object-wise. This fact is of course due to the presence of the isomorphism \( L_X \): something isomorphic to a natural transformation is again natural. More precisely, to provide the 2-cell data \( \lambda_f \) needed to make \( \lambda \) into a natural transformation, just pull back the 2-cell data from the natural transformation \( \alpha X \). In detail, we need invertible 2-cells

\[
\begin{array}{ccc}
IX & \xrightarrow{\lambda_X} & X \\
\downarrow \scriptstyle{f} & & \downarrow \scriptstyle{f} \\
IY & \xrightarrow{\lambda_Y} & Y
\end{array}
\]

(Here the commutative cell is actually the 2-cell part of the natural transformation \( \alpha X \).) Now the point is that each 2-cell \( \lambda_f \) is uniquely defined by this compatibility: indeed, since the other three 2-cells in the diagram are invertible, there is a unique 2-cell that can fill the place of \( I\lambda_f \), and since \( I \) is cancellable this 2-cell comes from a unique 2-cell \( \lambda_f \). The required compatibilities of \( \lambda_f \) with composition, identities, and 2-cells now follows from its construction: \( \lambda_f \) is just the translation via \( L \) of the identity 2-cell \( \alpha X \).

3.4. Uniqueness of the left constraints. There may be many choices for \( \lambda_X \), and even for a fixed \( \lambda_X \), there may be many choices for \( L_X \). However, between any two pairs \( (\lambda_X, L_X) \) and \( (\lambda'_X, L'_X) \) there is a unique invertible 2-cell \( U_X^{\alpha} : \lambda_X \Rightarrow \lambda'_X \) such that this compatibility holds:

\[
\begin{array}{ccc}
I\lambda_X & \xrightarrow{U_X^{\alpha}} & I\lambda'_X \\
\downarrow \scriptstyle{I} & & \downarrow \scriptstyle{I} \\
L_X & \xrightarrow{\alpha X} & L'_X
\end{array}
\]

(Here the commutative cell is actually the 2-cell part of the natural transformation \( \alpha X \).) Now the point is that each 2-cell \( \lambda_f \) is uniquely defined by this compatibility: indeed, since the other three 2-cells in the diagram are invertible, there is a unique 2-cell that can fill the place of \( I\lambda_f \), and since \( I \) is cancellable this 2-cell comes from a unique 2-cell \( \lambda_f \). The required compatibilities of \( \lambda_f \) with composition, identities, and 2-cells now follows from its construction: \( \lambda_f \) is just the translation via \( L \) of the identity 2-cell \( \alpha X \).

...
There is of course a completely analogous statement for right constraints.

3.5. Construction of the associator. We define $A : I\alpha \Rightarrow \alpha I$ as the unique 2-cell satisfying the equation

$$ III = RI^{II-1}_1 III $$

with $I\alpha I$. This definition is meaningful: since $I\alpha I$ is an equi-arrow, pre-composing with $I\alpha I$ is a 2-equivalence, hence gives a bijection on the level of 2-cells, so $A$ is determined by the right-hand side of the equation. Note that $A$ is invertible since all the 2-cells in the construction are.

The associator $\tilde{A}$ is defined as $A$-followed-by-$\alpha$:

$$ \tilde{A} := A \#_I \alpha, $$

but it will be more convenient to work with $A$.

3.6 Proposition. The definition of $A$ does not depend on the choices of left constraint $(\lambda, L)$ and right constraint $(\rho, R)$.

Proof. Write down the right-hand side of (9) in terms of different left and right constraints. Express these cells in terms of the original $L_I$ and $R_I$, using the comparison 2-cells $U^\text{left}_I$ and $U^\text{right}_I$ of 3.4. Finally observe that these comparison cells can be moved across the commutative square to cancel each other pairwise. $\square$
3.7. Uniqueness of $A$. Equation (9) may not appear familiar, but it is equivalent to the following ‘pentagon’ equation (after post-whiskering with $\alpha$): 

$$III \xrightarrow{\rho I} III \xrightarrow{\lambda \# \alpha} II = III \xrightarrow{\rho I} III \xrightarrow{\lambda \# \alpha} II$$

From this pentagon equation we shall derive the pentagon equation for $A$, asserted in Theorem A. To this end we need comparison between $\alpha$, $\lambda_I$, and $\rho_I$, which we now establish, in analogy with Axiom (1) of monoidal category: the left and right constraints coincide on the unit object, up to a canonical 2-cell:

3.8 Lemma. There are unique invertible 2-cells

$$\rho_I \xRightarrow{E} \alpha \xRightarrow{D} \lambda_I,$$

such that

$$III \xrightarrow{I\lambda} II = III \xrightarrow{\alpha I} II = III \xrightarrow{E I} II$$

Proof. The left-hand equation defines uniquely a 2-cell $I\alpha \Rightarrow I\lambda_I$, and since $I$ is cancellable, this cell comes from a unique 2-cell $\alpha \Rightarrow \lambda_I$ which we then call D. Same argument for E. \hfill $\square$

**Theorem A (Associativity).** Given a unit object $(I, \alpha)$, there is a canonical invertible 2-cell

$$III \xrightarrow{\alpha I} II$$

which satisfies the pentagon Equation (8).
Proof. On each side of the cube equation (10), paste the cell $EII$ on the top, and the cell $IID$ on the left. On the left-hand side of the equation we can use Equations (11) directly, while on the right-hand side we first need to move those cells across the commutative square before applying (11). The result is precisely the pentagon cube for $\tilde{A} = A \# \alpha$. 

3.9. Coherence of the actions. We have now established that $(I, \alpha, \tilde{A})$ is a semi-monoid, and may observe that the left and right constraints are coherent actions, i.e. that their ‘associators’ $L$ and $R$ satisfy the appropriate pentagon equations. For the left action this equation is:

Establishing this (and the analogous equation for the right action) is a routine calculation which we omit since we will not actually need the result. We also note that the two actions are compatible—i.e. constitute a two-sided action. Precisely this means that there is a canonical invertible 2-cell

\[
IXI \xrightarrow{\lambda \# I} XI
\]

This 2-cell satisfies two pentagon equations, one for $IIXI$ and one for $IXII$.

4 Units in the 2-category of arrows in $\mathscr{C}$, and Theorem B

In this section we prove Theorem B, which asserts that a morphism of units $(u, U) : (I, \alpha) \to (J, \beta)$ is automatically a semi-monoid map (with respect to the canonical associators $A$ and $B$ of the two units). We have to establish the cube equation of 2.4, or in fact the reduced version (7). The strategy to establish Equation (7) is to interpret everything in the 2-category of arrows of $\mathscr{C}$. The key point is to prove that a morphism of units is itself a unit in the 2-category of arrows. Then we invoke Theorem A to get an associator for this unit, and a pentagon equation, whose short form (6) will be the sought equation.
4.1. The 2-category of arrows. The 2-category of arrows in $\mathcal{C}$, denoted $\mathcal{C}^2$, is the 2-category described as follows. The objects of $\mathcal{C}^2$ are the arrows of $\mathcal{C}$,

$$X_0 \xrightarrow{x} X_1.$$ 

The arrows from $(X_0, X_1, x)$ to $(Y_0, Y_1, y)$ are triples $(f_0, f_1, F)$ where $f_0 : X_0 \to Y_0$ and $f_1 : X_1 \to Y_1$ are arrows in $\mathcal{C}$ and $F$ is a 2-cell

$$
\begin{array}{c}
X_0 \xrightarrow{f_0} Y_0 \\
X_1 \xrightarrow{f_1} Y_1.
\end{array}
$$

If $(g_0, g_1, G)$ is another arrow from $(X_0, X_1, x)$ to $(Y_0, Y_1, y)$, a 2-cell from $(f_0, f_1, F)$ to $(g_0, g_1, G)$ is given by a pair $(m_0, m_1)$ where $m_0 : f_0 \Rightarrow g_0$ and $m_1 : f_1 \Rightarrow g_1$ are 2-cells in $\mathcal{C}$ compatible with $F$ and $G$ in the sense that this cylinder commutes:

$$
\begin{array}{c}
X_0 \xleftarrow{m_0} Y_0 \\
X_1 \xleftarrow{f_1} Y_1.
\end{array}
$$

Composition of arrows in $\mathcal{C}^2$ is just pasting of squares. Vertical composition of 2-cells is just vertical composition of the components (the compatibility is guaranteed by pasting of cylinders along squares), and horizontal composition of 2-cells is horizontal composition of the components (compatibility guaranteed by pasting along the straight sides of the cylinders). Note that $\mathcal{C}^2$ inherits a tensor product from $\mathcal{C}$: this follows from functoriality of the tensor product on $\mathcal{C}$.

4.2 Lemma. If $I_0$ and $I_1$ are cancellable objects in $\mathcal{C}$ and $i : I_0 \to I_1$ is an equi-arrow, then $i$ is cancellable in $\mathcal{C}^2$.

Proof. We have to show that for given arrows $x : X_0 \to X_1$ and $y : Y_0 \to Y_1$, the functor

$$\text{Hom}_{\mathcal{C}^2}(x, y) \to \text{Hom}_{\mathcal{C}^2}(ix, iy)$$

defined by tensoring with $i$ on the left is an equivalence of categories (the check for tensoring on the right is analogous).
Let us first show that this functor is essentially surjective. Let

\[
\begin{array}{ccc}
I_0 X_0 & \xrightarrow{s_0} & I_0 Y_0 \\
ix \downarrow & S & iy \\
I_1 X_1 & \xrightarrow{s_1} & I_1 Y_1
\end{array}
\]

be an object in \(\text{Hom}_{\mathcal{C}_2}(ix, iy)\). We need to find a square

\[
\begin{array}{ccc}
X_0 & \xrightarrow{k_0} & Y_0 \\
x \downarrow & K & y \\
X_1 & \xrightarrow{k_1} & Y_1
\end{array}
\]

and an isomorphism \((m_0, m_1)\) from \((s_0, s_1, S)\) to \((I_0 k_0, I_1 k_1, iK)\), i.e. a cylinder

\[
\begin{array}{ccc}
I_0 X_0 & \xrightarrow{m_0} & I_0 Y_0 \\
ix \downarrow & S & iy \\
I_1 X_1 & \xrightarrow{s_1} & I_1 Y_1
\end{array} \quad \begin{array}{ccc}
I_0 X_0 & \xrightarrow{m_1} & I_1 Y_1 \\
ix \downarrow & S & iy \\
I_1 X_1 & \xrightarrow{s_1} & I_1 Y_1
\end{array}
\]

Since \(I_0\) is a cancellable object, the arrow \(s_0\) is isomorphic to \(I_0 k_0\) for some \(k_0 : X_0 \rightarrow Y_0\). Let the connecting invertible 2-cell be denoted \(m_0 : s_0 \Rightarrow I_0 k_0\). Similarly we find \(k_1\) and \(m_1 : s_1 \Rightarrow I_1 k_1\). Since \(m_0\) and \(m_1\) are invertible, there is a unique 2-cell

\[
\begin{array}{ccc}
I_0 X_0 & \xrightarrow{i k_0} & I_0 Y_0 \\
ix \downarrow & T & iy \\
I_1 X_1 & \xrightarrow{i k_1} & I_1 Y_1
\end{array}
\]

that can take the place of \(iK\) in the cylinder equation; it remains to see that \(T\) is of the form \(iK\) for some \(K\). But this follows since the map

\[
2\text{Cell}_{\mathcal{C}_2}(k_0 \# y, x \# k_1) \rightarrow 2\text{Cell}_{\mathcal{C}_2}(i(k_0 \# y), i(x \# k_1))
\]

\[
K \mapsto iK
\]

is a bijection. Indeed, the map factors as ‘tensoring with \(I_0\) on the left’ followed by ‘post-composing with \(iY_1\’; the first is a bijection since \(I_0\) is cancellable, the second is a bijection since \(i\) (and hence \(iY_1\) is an equi-arrow).
Now for the fully faithfulness of \( \text{Hom}_{\mathcal{C}^2}(x, y) \to \text{Hom}_{\mathcal{C}^2}(ix, iy) \). Fix two objects in the left-hand category, \( P \) and \( Q \):

\[
\begin{array}{ccc}
X_0 & \xrightarrow{p_0} & Y_0 \\
\downarrow x & & \downarrow y \\
X_1 & \xrightarrow{p_1} & Y_1 \\
\end{array}
\quad
\quad
\begin{array}{ccc}
X_0 & \xrightarrow{q_0} & Y_0 \\
\downarrow x & & \downarrow y \\
X_1 & \xrightarrow{q_1} & Y_1 \\
\end{array}
\]

The arrows from \( P \) to \( Q \) are pairs \((m_0, m_1)\) consisting of

\[
m_0 : p_0 \Rightarrow q_0 \quad m_1 : p_1 \Rightarrow q_1
\]

cylinder-compatible with the 2-cells \( P \) and \( Q \). The image of these two objects are

\[
\begin{array}{ccc}
I_0X_0 & \xrightarrow{I_0p_0} & I_0Y_0 \\
\downarrow ix & & \downarrow iy \\
I_1X_1 & \xrightarrow{I_1p_1} & I_1Y_1 \\
\end{array}
\quad
\begin{array}{ccc}
I_0X_0 & \xrightarrow{I_0q_0} & I_0Y_0 \\
\downarrow ix & & \downarrow iy \\
I_1X_1 & \xrightarrow{I_1q_1} & I_1Y_1 \\
\end{array}
\]

The possible 2-cells from \( iP \) to \( iQ \) are pairs \((n_0, n_1)\) consisting of

\[
n_0 : I_0p_0 \Rightarrow I_0q_0 \quad n_1 : I_1p_1 \Rightarrow I_1q_1
\]

cylinder-compatible with the 2-cells \( iP \) and \( iQ \). Now since \( I_0 \) is cancellable, every 2-cell \( n_0 \) like this is uniquely of the form \( I_0n_0 \) for some \( n_0 \). Hence there is a bijection between the possible \( m_0 \) and the possible \( n_0 \). Similarly for \( m_1 \) and \( n_1 \). So there is a bijection between pairs \((m_0, m_1)\) and pairs \((n_0, n_1)\).

Now by functoriality of tensoring with \( i \), all images of compatible \((m_0, m_1)\) are again compatible. It remains to rule out the possibility that some \((n_0, n_1)\) pair could be compatible without \((m_0, m_1)\) being so, but this follows again from the argument that ‘tensoring with \( i \) on the left’ is a bijection on hom sets, just like argued for (12).

4.3 LEMMA. An arrow in \( \mathcal{C}^2 \),

\[
\begin{array}{ccc}
X_0 & \xrightarrow{f_0} & Y_0 \\
\downarrow x & & \downarrow y \\
X_1 & \xrightarrow{f_1} & Y_1 \\
\end{array}
\]

is an equi-arrow in \( \mathcal{C}^2 \) if the components \( f_0 \) and \( f_1 \) are equi-arrows in \( \mathcal{C} \) and \( F \) is invertible.
Proof. We can construct an explicit quasi-inverse by choosing quasi-inverses to the components.

4.4 Corollary. If \((I_0, \alpha_0)\) and \((I_1, \alpha_1)\) are units in \(\mathcal{C}\), and \((u, U) : I_0 \to I_1\) is a unit map between them, then

\[ u : I_0 \to I_1 \]

is a unit object in \(\mathcal{C}^2\) with structure map

\[
\begin{array}{ccc}
I_0 I_0 & \xrightarrow{\alpha_0} & I_0 \\
\downarrow{uu} & & \downarrow{u} \\
I_1 I_1 & \xrightarrow{\alpha_1} & I_1 \\
\end{array}
\]

Proof. The object \(u\) is cancellable by Lemma 4.2, and the morphism \((\alpha_0, \alpha_1, U^{-1})\) from \(uu\) to \(u\) is an equi-arrow by Lemma 4.3.

Theorem B. Let \((I_0, \alpha_0)\) and \((I_1, \alpha_1)\) be units, with canonical associators \(A_0\) and \(A_1\), respectively. If \((u, U)\) is a unit map from \(I_0\) to \(I_1\) then it is automatically a semi-monoid map. That is,

\[
\begin{array}{ccc}
I_0 I_0 I_0 & \xrightarrow{uuu} & I_1 I_1 I_1 \\
\downarrow{I_0 \alpha_0} & & \downarrow{\alpha_1 I_1} \\
I_0 I_0 & \xrightarrow{uu} & I_1 I_1 \\
\end{array} = \begin{array}{ccc}
I_0 I_0 I_0 & \xrightarrow{uuu} & I_1 I_1 I_1 \\
\downarrow{uU} & & \downarrow{\alpha_1 I_1} \\
I_0 I_0 & \xrightarrow{uu} & I_1 I_1 \\
\end{array}
\]

Proof. By the previous Corollary, \((u, U^{-1})\) is a unit object in \(\mathcal{C}^2\). Hence there is a canonical associator

\[ B : uU^{-1} \leftrightarrow U^{-1}u. \]

By definition of 2-cells in \(\mathcal{C}^2\), this is a pair of 2-cells in \(\mathcal{C}\)

\[ B_0 : I_0 \alpha_0 \Rightarrow \alpha_0 I_0 \quad B_1 : I_1 \alpha_1 \Rightarrow \alpha_1 I_1, \]

fitting the cylinder equation

\[
\begin{array}{ccc}
I_0 I_0 I_0 & \xrightarrow{\alpha_0 I_0} & I_0 I_0 \\
\downarrow{uu} & & \downarrow{uu} \\
I_1 I_1 I_1 & \xrightarrow{I_0 \alpha_0} & I_0 I_0 \\
\end{array} = \begin{array}{ccc}
I_0 I_0 I_0 & \xrightarrow{\alpha_0 I_0} & I_0 I_0 \\
\downarrow{uU^{-1}} & & \downarrow{uu} \\
I_1 I_1 I_1 & \xrightarrow{\alpha_1 I_1} & I_1 I_1 \\
\end{array}
\]
This is precisely the cylinder diagram we are looking for—provided we can show that $B_0 = A_0$ and $B_1 = A_1$. But this is a consequence of the characterising property of the associator of a unit: first note that as a unit object in $\mathcal{C}^2$, $u$ induces left and right constraints: for each object $x : X_0 \to X_1$ in $\mathcal{C}^2$ there is a left action of the unit $u$, and this left action will induce a left action of $(I_0, \alpha_0)$ on $X_0$ and a left action of $(I_1, \alpha_1)$ on $X_1$ (the ends of the cylinders). Similarly there is a right action of $u$ which induces right actions at the ends of the cylinder. Now the unique $B$ that exists as associator for the unit object $u$ compatible with the left and right constraints induces $B_0$ and $B_1$ at the ends of the cylinder, and these will of course be compatible with the induced left and right constraints. Hence, by uniqueness of associators compatible with left and right constraints, these induced associators $B_0$ and $B_1$ must coincide with $A_0$ and $A_1$. Note that this does not dependent on choice of left and right constraints, cf. Proposition 3.6.

\[ \square \]

5 Contractibility of the space of weak units (Theorem C)

The goal of this section is to prove Theorem C, which asserts that the 2-category of units in $\mathcal{C}$ is contractible if non-empty. First we describe the unit morphisms and unit 2-morphisms in terms of compatibility with left and right constraints. This will show that there are not too many 2-cells. Second we use the left and right constraints to connect any two units.

The following lemma shows that just as the single arrow $\alpha$ induces all the $\lambda_X$ and $\rho_X$, the single 2-cell $U$ of a unit map induces families $U^\text{left}_X$ and $U^\text{right}_X$ expressing compatibility with $\lambda_X$ and $\rho_X$.

5.1 Lemma. Let $(I, \alpha)$ and $(J, \beta)$ be units, and let $(u, U)$ be a morphism of pseudo-idempotents from $(I, \alpha)$ to $(J, \beta)$. The following are equivalent.

(i) $u$ is an equi-arrow (i.e. $u$ is a morphism of units).

(ii) $u$ is left cancellable, i.e. tensoring with $u$ on the left is an equivalence of categories $\text{Hom}(X, Y) \to \text{Hom}(IX, JY)$.

(ii') $u$ is right cancellable, i.e. tensoring with $u$ on the right is an equivalence of categories $\text{Hom}(X, Y) \to \text{Hom}(XI, YJ)$.

(iii) For fixed left actions $(\lambda_X, L_X)$ for the unit $(I, \alpha)$ and $(\ell_X, L'_X)$ for the unit $(J, \beta)$, there is a unique invertible 2-cell $U^\text{left}_X$, natural in $X$:

\[
\begin{array}{ccc}
IX & \xrightarrow{u_X} & JX \\
\lambda_X & & \ell_X \\
X & \xrightarrow{U^\text{left}_X} & X
\end{array}
\]
such that this compatibility holds:

\[
\begin{array}{cccccc}
II X & \xrightarrow{\ u u X \ } & JJ X \\
I \lambda_X & \circ & \alpha_X & \circ & U_X & \beta_X & = & I \lambda_X & \circ & u U_X^{\text{left}} & \circ & J \ell_X & \circ & L_X' & \beta_X \\
IX & \xrightarrow{u_X} & JX \\
\end{array}
\]

(iii') For fixed right actions \((\rho_X, R_X)\) for the unit \((I, \alpha)\) and \((r_X, R'_X)\) for the unit \((J, \beta)\), there is a unique invertible 2-cell \(U_X^{\text{right}}\), natural in \(X\):

\[
\begin{array}{cccc}
XI & \xrightarrow{X u} & X J \\
\rho_X & \circ & U_X^{\text{right}} & \circ & r_X \\
X & \xrightarrow{X u} & X J \\
\end{array}
\]

such that this compatibility holds:

\[
\begin{array}{cccccc}
XII & \xrightarrow{X u u} & X J J \\
X \alpha & \circ & R_X & \circ & U_X^{\text{right}} & \circ & r_X J \\
XI & \xrightarrow{X u} & X J \\
\end{array}
\]

\[
\begin{array}{cccccc}
XII & \xrightarrow{X u u} & X J J \\
X \beta & \circ & R'_X & \circ & r_X J \\
XI & \xrightarrow{X u} & X J \\
\end{array}
\]

Proof. (i) implies (ii): ‘tensoring with \(u\)’ can be done in two steps: given an arrow \(X \rightarrow Y\), first tensor with \(I\) to get \(IX \rightarrow IY\), and then post-compose with \(uY\) to get \(IX \rightarrow JY\). The first step is an equivalence because \(I\) is a unit, and the second step is an equivalence because \(u\) is an equi-arrow.

(ii) implies (iii): In Equation (13), the 2-cell labelled \(u U_X^{\text{left}}\) is uniquely defined by the three other cells, and it is invertible since the three other cells are. Since tensoring with \(u\) on the left is an equivalence, this cell comes from a unique invertible cell \(U_X^{\text{right}}\), justifying the label \(u U_X^{\text{right}}\).

(iii) implies (i): The invertible 2-cell \(U_X^{\text{right}}\) shows that \(uX\) is isomorphic to an equi-arrow, and hence is an equi-arrow itself. Now take \(X\) to be a right cancellable object (like for example \(I\)) and conclude that already \(u\) is an equi-arrow.

Finally, the equivalence \((i) \Rightarrow (ii') \Rightarrow (iii') \Rightarrow (i)\) is completely analogous. \(\square\)

Note that for \((u, U)\) the identity morphism on \((I, \alpha)\), we recover Observation 3.4.
5.2 Lemma. Let \((I, \alpha)\) and \((J, \beta)\) be units; let \((u, U)\) and \((v, V)\) be morphisms of pseudo-idempotents from \(I\) to \(J\); and consider a 2-cell \(T : u \Rightarrow v\). Then the following are equivalent.

(i) \(T\) is an invertible 2-morphism of pseudo-idempotents.

(ii) \(T\) is a left cancellable 2-morphism of pseudo-idempotents (i.e., induces a bijection on hom sets (of hom cats) by tensoring with \(T\) from the left).

(ii') \(T\) is a right cancellable 2-morphism of pseudo-idempotents (i.e., induces a bijection on hom sets (of hom cats) by tensoring with \(T\) from the right).

(iii) For fixed left actions \((\lambda_X, L_X)\) for \((I, \alpha)\) and \((\ell_X, L'_X)\) for \((J, \beta)\), with induced canonical 2-cells \(U_X^{\text{left}}\) and \(V_X^{\text{left}}\) as in 5.1, we have:

\[
\begin{align*}
IX & \xrightarrow{\lambda_X} X \\
TX & \xrightarrow{\ell_X} X \\
IX & \xrightarrow{v_X} JX \\
U_X^{\text{left}} & \xrightarrow{X} X \\
X & \xrightarrow{\circ} X
\end{align*}
\]

\[
(15)
\]

(iii') For fixed right actions \((\rho_X, R_X)\) for \((I, \alpha)\) and \((r_X, R'_X)\) for \((J, \beta)\), with induced canonical 2-cells \(U_X^{\text{right}}\) and \(V_X^{\text{right}}\) as in 5.1, we have:

\[
\begin{align*}
XI & \xrightarrow{\rho_X} X \\
XT & \xrightarrow{Xu} X \\
XI & \xrightarrow{xv} X \\
U_X^{\text{right}} & \xrightarrow{X} X \\
X & \xrightarrow{\circ} X
\end{align*}
\]

\[
(16)
\]

Proof. It is obvious that (i) implies (ii). Let us prove that (iii) implies (iii'), so assume that tensoring with \(T\) on the left defines a bijection on the level of 2-cells. Start with the cylinder diagram for compatibility of tensor 2-cells.
(cf. 2.5). Tensor this diagram with $X$ on the right to get

\[
\begin{align*}
&\begin{array}{ccc}
&I \quad II \quad X^v \\
\downarrow \alpha X & & \downarrow \beta X \\
IX & & JX
\end{array}
\begin{array}{ccc}
&V \quad \downarrow uX \\
\downarrow \alpha X & & \downarrow \beta X \\
IX & & JX
\end{array}
= \\
\begin{array}{ccc}
&\downarrow \alpha X \\
\downarrow \alpha X & & \downarrow \beta X \\
IX & & JX
\end{array}
\end{align*}
\]

On each side of this equation, paste an $L_X$ along $\alpha X$, apply Equation (13) on each side, and cancel the $L'_X$ that appear on the other side of the square. The resulting diagram

\[
\begin{align*}
&\begin{array}{ccc}
&I \quad II \quad X^v \\
\downarrow \alpha X & & \downarrow \beta X \\
IX & & JX
\end{array}
\begin{array}{ccc}
&\downarrow \alpha X \\
\downarrow \alpha X & & \downarrow \beta X \\
IX & & JX
\end{array} = \\
\begin{array}{ccc}
&\downarrow \alpha X \\
\downarrow \alpha X & & \downarrow \beta X \\
IX & & JX
\end{array}
\end{align*}
\]

is the tensor product of $T$ with the promised equation (15). Since $T$ is cancellable, we can cancel it away to finish.

(iii) implies (i): the arguments in (ii)$\Rightarrow$(iii) can be reversed: start with (15), tensor with $T$ on the left, and apply (13) to arrive at the axiom for being a 2-morphism of pseudo-idempotents. Since both $U^\text{left}_X$ and $V^\text{right}_X$ are invertible, so is $TX$. Now take $X$ to be a right cancellable object, and cancel it away to conclude that already $T$ is invertible.

Finally, the equivalence (i)$\Rightarrow$(ii)$'\Rightarrow$(iii)$'$ implies (i) is completely analogous. □

5.3 Corollary. Given two parallel morphisms of units, there is a unique unit 2-morphism between them.

Proof. Choose left actions for $(I, \alpha)$ and $(J, \beta)$ as in Lemma 5.2 (iii), and take $X$ to be a right cancellable object. For given morphisms of units $u$ and $v$ as in Lemma 5.2, Equation (15) defines the 2-cell $T$ uniquely, since $\lambda_X$ is an equi-arrow and $X$ is right cancellable. □

Next we aim at proving that there is a unit morphism between any two units. The strategy is to use the left and right constraints to produce a unit morphism

\[
I \longrightarrow IJ \longrightarrow J.
\]

As a first step towards this goal we have:
5.4 Lemma. Let $I$ and $J$ be units, and pick a left constraint $\lambda$ for $I$ and a right constraint $r$ for $J$. Put

$$\gamma := r_I \lambda_J : IJJ \to IJ$$

Then $(IJ, \gamma)$ is a unit.

Proof. Since $I$ and $J$ are cancellable, clearly $IJ$ is cancellable too. Since $\lambda_J$ and $r_I$ are equi-arrows, $\gamma$ is too. \hfill $\square$

5.5 Lemma. There is an invertible 2-cell

$$\begin{array}{c}
IJJIJ \\
\downarrow \gamma \\
IJJ \\
\downarrow \lambda_J \\
IJ \\
\end{array} \xrightarrow{Z} \begin{array}{c}
JJ \\
\downarrow \beta \\
J. \\
\end{array}$$

Hence $(\lambda_J, Z)$ is a unit map. (And there is another 2-cell making $r_I$ a unit map.)

Proof. The 2-cell $Z$ is defined like this:

where the 2-cell $K^\lambda$ is constructed in Lemma 7.2. \hfill $\square$

5.6 Corollary. Given two units, there exists a unit morphism between them.

Proof. Continuing the notation from above, by Lemma 5.4, $(IJ, \gamma)$ is a unit, and by Lemma 5.5, $\lambda : IJ \to J$ is a morphism of units. Similarly, $r : IJ \to I$ is a unit morphism, and by Lemma 2.6 any chosen pseudo-inverse $r^{-1} : I \to IJ$ is again a unit morphism. Finally we take

$$I \xrightarrow{r^{-1}} IJ \xrightarrow{\lambda} J.$$ 

\hfill $\square$
Theorem C (Contractibility). The 2-category of units in \( \mathcal{C} \) is contractible, if non-empty. In other words, between any two units there exists a unit morphism, and between any two parallel unit morphisms there is a unique unit 2-morphism.

Proof. By Lemma 5.6 there is a unit morphism between any two units (an equi-arrow by definition), and by Corollary 5.3 there is a unique unit 2-morphism between any two parallel unit morphisms.

\[ \square \]

6 Classical units

In this section we review the classical theory of units in a monoidal 2-category, as extracted from the definition of tricategory of Gordon, Power, and Street [2]. In the next section we compare this notion with the cancellable-idempotent approach of this work. The equivalence is stated explicitly in Theorem E.

6.1. Tricategories. The notion of tricategory introduced by Gordon, Power, and Street [2] is roughly a weak category structure enriched over bicategories: this means that the structure maps (composition and unit) are weak 2-functors satisfying weak versions of associativity and unit constraints. For the associativity, the pentagon equation is replaced by a specified pentagon 3-cell (TD7), required to satisfy an equation corresponding to the 3-dimensional associahedron. This equation (TA1) is called the nonabelian 4-cocycle condition. For the unit structure, three families of 3-cells are specified (TD8): one corresponding to the Kelly axiom, one left variant, and one right variant (those two being the higher-dimensional analogues of Axioms (2) and (3) of monoidal category). Two axioms are imposed on these three families of 3-cells: one (TA2) relating the left family with the middle family, and one (TA3) relating the right family with the middle family. These are called left and right normalisation. (These two axioms are the higher-dimensional analogues of the first argument in Kelly’s lemma 1.6.) It is pointed out in [2] that the middle family together with the axioms (TA2) and (TA3) completely determine the left and right families if they exist.

6.2. Monoidal 2-categories. By specialising the definition of tricategory to the one-object case, and requiring everything strict except the units, we arrive at the following notion of monoidal 2-category: a monoidal 2-category is a semi-monoidal 2-category (cf. 2.2) equipped with an object \( I \), two natural transformations \( \lambda \) and \( \rho \) with equi-arrow components

\[ \lambda_X : IX \to X \]

\[ \rho_X : XI \to X \]
and (invertible) 2-cell data

$$
\begin{array}{c}
IX \xrightarrow{\lambda_X} X \\
\downarrow{\lambda_f} \quad \downarrow{f} \\
IY \xrightarrow{\lambda_Y} Y
\end{array}
\quad
\begin{array}{c}
XI \xrightarrow{\rho_X} X \\
\downarrow{\rho_f} \quad \downarrow{f} \\
YI \xrightarrow{\rho_Y} Y,
\end{array}
$$

together with three natural modifications $K$, $K^\lambda$, and $K^\rho$, with invertible components

$$
K : X \lambda_Y \Rightarrow \rho_X Y \\
K^\lambda : \lambda_X Y \Rightarrow \lambda_X Y \\
K^\rho : X \rho_Y \Rightarrow \rho_{XY}.
$$

We call $K$ the Kelly cell.

These three families are subject to the following two equations:

$$
\begin{array}{c}
X\lambda_Y Z \xrightarrow{XK^\lambda_{Y,Z}} X\lambda_Y Z \\
\downarrow{\kappa_{X,Y,Z}} \quad \downarrow{\kappa_{X,Y,Z}} \\
\rho_X Y Z
\end{array}
$$

(17)

$$
\begin{array}{c}
X\rho_Y Z \xrightarrow{XK^\rho_{Y,Z}} \rho_{XY} Z \\
\downarrow{\kappa_{X,Y,Z}} \quad \downarrow{\kappa_{X,Y,Z}} \\
XY\lambda_Y Z
\end{array}
$$

(18)

6.3 Remark. We have made one change compared to [2], namely the direction of the arrow $\rho_X$: from the viewpoint of $\alpha$ it seems more practical to work with $\rho_X : XI \rightarrow X$ rather than with the convention of $\rho_X : X \rightarrow XI$ chosen in [2]. Since in any case it is an equi-arrow, the difference is not essential. (Gurski in his thesis [3] has studied a version of tricategory where all the equi-arrows in the definition are equipped with specified pseudo-inverses. This has the advantage that the definition becomes completely algebraic, in a technical sense.)

6.4 Lemma. The object $I$ is cancellable (independently of the existence of $K$, $K^\lambda$, and $K^\rho$.)

Proof. We need to establish that ‘tensoring with $I$ on the left’,

$$
L : \text{Hom}(X,Y) \rightarrow \text{Hom}(IX, IY),
$$
is an equivalence of categories. But this follows since the diagram

\[
\begin{array}{c}
\text{Hom}(X,Y) \xrightarrow{L} \text{Hom}(IX, IY) \\
\downarrow \quad \text{Id} \quad \downarrow \text{Id} \\
\text{Hom}(X,Y) \xrightarrow{\lambda_Y \#} \text{Hom}(IX, Y)
\end{array}
\]

is commutative up to isomorphism: the component at \( f : X \to Y \) of this isomorphism is just the naturality square \( \lambda_f \). Since the functors \( \lambda_X \# \) and \( \# \lambda_Y \) are equivalences, it follows from this isomorphism that \( L \) is too. \( \square \)

6.5. **Coherence of the Kelly cell.** As remarked in [2], if the \( K^\lambda \) and \( K^\rho \) exist, they are determined uniquely from \( K \) and the two axioms. Indeed, the two equations

\[
\begin{align*}
I\lambda_Y Z \xrightarrow{\rho_Y I} & I\lambda_Y Z \\
\kappa_{I,Y,Z} \xrightarrow{\rho_I Y Z} & \kappa_{I,Y,Z} \\
\rho_{I,Y Z} \xrightarrow{\rho_{I,Y Z}} & \rho_{I,Y Z}
\end{align*}
\]

\[
\begin{align*}
X\rho_Y I \xrightarrow{\rho_{XY I}} & \rho_{XY I} \\
X\kappa_{Y,I} \xrightarrow{X\kappa_{Y,I}} & X\kappa_{Y,I} \\
\rho_{XY I} \xrightarrow{X\rho_Y I} & \rho_{XY I}
\end{align*}
\]

which are just special cases of (17) and (18) uniquely determine \( K^\lambda \) and \( K^\rho \), by cancellability of \( I \). But these two special cases of the axioms do not imply the general case.

We shall take the Kelly cell \( K \) as the main structure, and say that \( K \) is *coherent on the left* (resp. *on the right*) if Axiom (17) (resp. (18)) holds for the induced cell \( K^\lambda \) (resp. \( K^\rho \)). We just say *coherent* if both hold. We shall see (7.8) that in fact coherence on the left implies coherence on the right and vice versa.

6.6. **Naturality.** The Kelly cell is a modification. For future reference we spell out the naturality condition satisfied: given arrows \( f : X \to X' \) and \( g : Y \to Y' \), we have

\[
\begin{align*}
X\lambda_Y \xrightarrow{f} & X\lambda_Y \\
\rho_{XY} \xrightarrow{f g} & \rho_{XY} \\
X'\lambda_Y \xrightarrow{f g} & X'\lambda_Y
\end{align*}
\]

\[
\begin{align*}
\rho_{XY} \xrightarrow{f g} & \rho_{XY} \\
\rho_{XY} \xrightarrow{f g} & \rho_{XY}
\end{align*}
\]
6.7 Remark. Particularly useful is naturality of $\lambda$ with respect to $\lambda_X$ and naturality of $\rho$ with respect to $\rho_X$. In these cases, since $\lambda_X$ and $\rho_X$ are equi-arrows, we can cancel them and find the following invertible 2-cells:

$$N^\lambda : I\lambda_X \Rightarrow \lambda_I$$

$$N^\rho : \rho_X I \Rightarrow X\rho_I,$$

in analogy with Observation (5) of monoidal categories.

The following lemma holds for $\mathbf{K}$ independently of Axioms (17) and (18):

6.8 Lemma. The Kelly cell $\mathbf{K}$ satisfies the equation

$$XIIY \xrightarrow{\rho_X IY} X\lambda_Y \Rightarrow X\lambda_Y = XIIY \xrightarrow{N^\rho IY} X\rho_Y \Rightarrow X\rho_Y \xrightarrow{N^\rho} \mathbf{K}$$

Proof. It is enough to establish this equation after post-whiskering with $X\lambda_Y$. The rest is a routine calculation, using on one side the definition of the cell $N^\lambda$, then naturality of $\mathbf{K}$ with respect to $f = X$ and $g = \lambda_Y$. On the other side, use the definition of $N^\rho$ and then naturality of $\mathbf{K}$ with respect to $f = \rho_X$ and $g = Y$. In the end, two $\mathbf{K}$-cells cancel.

Combining the 2-cells described so far we get

$$\rho_I I \xrightarrow{\mathbf{K}^{-1}} I\lambda_I \xrightarrow{\mathbf{N}^\lambda} \lambda_I I \Rightarrow \lambda_I I$$

and hence, by cancelling $I$ on the right, an invertible 2-cell

$$P : \rho_I \Rightarrow \lambda_I.$$

Now we could also define $Q : \rho_I \Rightarrow \lambda_I$ in terms of

$$I\rho_I \xrightarrow{\mathbf{K}^e} \rho_I I \xrightarrow{\mathbf{N}^e} \rho_I I \Rightarrow I\lambda_I.$$

Finally, in analogy with Axiom (1) for monoidal categories:

6.9 Lemma. We have $P = Q$. (This is true independently of Axioms (17) and (18).)

Proof. Since $I$ is cancellable, it is enough to show $IP = IQ$. To establish this equation, use the constructions of $P$ and $Q$, then substitute the characterising Equations (19) for the auxiliary cells $\mathbf{K}^\lambda$ and $\mathbf{K}^\rho$, and finally use Lemma 6.8. □
6.10. The 2-category of GPS units. For short, we shall say GPS unit for
the notion of unit just introduced. In summary, a GPS unit is a quadruple
\((I, \lambda, \rho, K)\) where \(I\) is an object, \(\lambda_X\) and \(\rho_X\) are natural
transformations with equi-arrow components, and \(K : X\lambda_Y \Rightarrow \rho_XY\) is
a coherent Kelly cell (natural in \(X\) and \(Y\), of course).

A morphism of GPS units from \((I, \lambda, \rho, K)\) to \((J, \ell, r, H)\) is an arrow \(u : I \rightarrow J\)
equipped with natural families of invertible 2-cells

\[
\begin{align*}
IX \xrightarrow{u_X} JX \\
\lambda_X & \downarrow \quad \Upsilon_{\lambda X}^\leftarrow \quad \ell_X \\
X & \xrightarrow{u} X
\end{align*}
\begin{align*}
XI \xrightarrow{\rho_X} XJ \\
\rho & \downarrow \quad \Upsilon_{\rho X}^\rightarrow \quad r_X \\
X & \xrightarrow{} X
\end{align*}
\]

satisfying the equation

\[
\begin{align*}
XIY \xrightarrow{XuY} XJY \\
\lambda_{XY} K \quad \rho_X \quad \Upsilon_{\rho X Y}^\rightarrow \quad r_{XY} Y \\
XY \xrightarrow{} XY
\end{align*} = \begin{align*}
XIY \xrightarrow{XuY} XJY \\
\lambda_{XY} \quad \Upsilon_{\lambda Y X}^\leftarrow \quad \ell_X Y \quad H \quad r_{XY} Y \\
XY \xrightarrow{} XY
\end{align*}
\]

Finally, a 2-morphism of GPS unit maps is a 2-cell \(T : u \Rightarrow v\) satisfying the
compatibility conditions (15) and (16) of Lemma 5.2.

6.11. Remarks. Note first that \(u\) is automatically an equi-arrow. Observe
also that \(\Upsilon_{u X}^\leftarrow\) and \(\Upsilon_{\rho X Y}^\rightarrow\) completely determine each other by Equation (20), as
is easily seen by taking on the one hand \(X\) to be a left cancellable object and
on the other hand \(Y\) to be a right cancellable object. Finally note that there are
two further equations, expressing compatibility with \(K^\lambda\) and \(K^\rho\), but
they can be deduced from Equation (20), independently of the coherence Axioms (17)
and (18). Here is the one for \(K^\lambda\) for future reference:

\[
\begin{align*}
IXY \xrightarrow{u_{XY}} JXY \\
\lambda_{XY} K^\lambda \quad \lambda_{XY} \quad \Upsilon_{\lambda Y X}^\leftarrow \quad \ell_{XY} Y \\
XY \xrightarrow{} XY
\end{align*} = \begin{align*}
IXY \xrightarrow{u_{XY}} JXY \\
\lambda_{XY} \quad \Upsilon_{\lambda Y X}^\leftarrow \quad \ell_{XY} Y \quad H^\ell \quad \ell_{XY} Y \\
XY \xrightarrow{} XY
\end{align*}
\]

7 Comparison with classical theory (Theorem E)

In this section we prove the equivalence between the two notions of unit.
7.1. From cancellable-idempotent units to GPS units. We fix a unit object \((I, \alpha)\). We also assume chosen a left constraint \(\lambda_X : IX \to X\) with \(L_X : I\lambda_X \equiv \alpha X\), and a right constraint \(\rho_X : XI \to X\) with \(R_X : X\alpha \equiv \rho_X I\). First of all, in analogy with Axioms (2) and (3) of monoidal categories we have:

7.2 Lemma. In the situation of 7.1, there are unique natural invertible 2-cells

\[
\begin{align*}
K^\lambda : \lambda_{XY} &\Rightarrow \lambda_X Y \\
K^\rho : X\rho_Y &\Rightarrow \rho_{XY}
\end{align*}
\]

satisfying

\[
IIXY \xrightarrow{IK^\lambda} IXY = IIXY \xrightarrow{L^{-1}Y} L^{-1}Y \xrightarrow{\lambda_{XY}} IXY \tag{22}
\]

\[
XYII \xrightarrow{K^\rho I} XYI = XYII \xrightarrow{R} XR^{-1} \xrightarrow{\rho_{XY}I} XYI \tag{23}
\]

Proof. The conditions precisely define the 2-cells, since \(I\) is cancellable. \(\square\)

7.3 Lemma. In the situation of 7.1, there is a canonical family of invertible 2-cells (the Kelly cell)

\[
\begin{align*}
K : X\lambda_Y &\Rightarrow \rho_X Y,
\end{align*}
\]

natural in \(X\) and \(Y\).

Proof. This is analogous to the construction of the associator: \(K\) is defined as the unique 2-cell \(K : X\lambda_Y \Rightarrow \rho_X Y\) satisfying the equation

\[
\begin{align*}
XIIY \xrightarrow{X\alpha Y} XIY \xrightarrow{XL} XYY \xrightarrow{\rho_X Y} XY = XIIY \xrightarrow{\lambda_{XY}} XIY \xrightarrow{\rho_X Y} XY \tag{24}
\end{align*}
\]
André Joyal and Joachim Kock

This makes sense since $X\alpha Y$ is an equi-arrow, so we can cancel it away. Clearly $K$ is invertible since $L$ and $R$ are.

We constructed $K^\lambda$ and $K^\rho$ directly from $L$ and $R$. Meanwhile we also constructed $K$, and we know from classical theory (6.5) that this cell determines the two others. The following proposition shows that all these constructions match up, and in particular that the constructed Kelly cell is coherent on both sides:

7.4 Proposition. In the situation of 7.1, the families of 2-cells $K$, $K^\lambda$ and $K^\rho$ (constructed in 7.2 and 7.3) satisfy the GPS unit axioms (17) and (18):

$$
\begin{align*}
X\lambda YZ & \Rightarrow X\lambda YZ \\
X\rho YZ & \Rightarrow \rho_{XY}Z
\end{align*}
$$

Proof. We treat the left constraint (the right constraint being completely analogous). We need to establish

$$
XIYZ \Rightarrow XYZ
$$

and it is enough to establish this equation pre-whiskered with $X\alpha YZ$. In the diagram resulting from the left-hand side:

$$
\begin{align*}
 XIYZ \Rightarrow X\lambda YZ \Rightarrow X\lambda YZ & \Rightarrow X\lambda YZ \\
XK_{Y,Z}^X & \Rightarrow XK_{Y,Z}^X \\
X\lambda YZ & \Rightarrow X\lambda YZ
\end{align*}
$$

we can replace $(X\alpha YZ)\#(K_{X,Y,Z})$ by the expression that defined $K_{X,Y,Z}$
(cf. (24)), yielding altogether

Here we can move the cell $XK_{Y,Z}^{X}$ across the square, where it becomes $XIK_{Y,Z}^{X}$ and combines with $XLYZ$ to give altogether $XLYZ$ (cf. (22)). The resulting diagram

is nothing but

(by Equation (24) again) which is what we wanted to establish. \qed

Hereby we have concluded the construction of a GPS unit from $(I, \alpha)$. We will also need a result for morphisms:

7.5 Proposition. Let $(u, \mathcal{U}) : (I, \alpha) \to (J, \beta)$ be a morphism of units in the sense of 2.9, and consider the two canonical 2-cells $U^{\text{left}}$ and $U^{\text{right}}$ constructed
in Lemma 5.1. Then Equation (20) holds:

\[
\begin{align*}
&X \mapsto Y \\
&\xymatrix{ X \ar@{->}[r]^{X_{X_{Y}}} \ar@{->}[d]_{X_{\lambda_{Y}}} & X_{Y} \ar@{->}[r]^{X_{U_{Y}}} \ar@{->}[d]_{X_{\rho_{Y}}} & X_{Y} \ar@{->}[r]^{X_{X_{Y}}} & X_{Y} } \\
\xymatrix{ X \ar@{->}[r]^{X_{X_{Y}}} \ar@{->}[d]_{X_{\lambda_{Y}}} & X_{Y} \ar@{->}[r]^{X_{U_{Y}}} & X_{Y} & \xymatrix{ X \ar@{->}[r]^{X_{X_{Y}}} \ar@{->}[d]_{X_{\lambda_{Y}}} & X_{Y} \ar@{->}[r]^{X_{U_{Y}}} \ar@{->}[d]_{X_{\rho_{Y}}} & X_{Y} \ar@{->}[r]^{X_{X_{Y}}} & X_{Y} } \\
\xymatrix{ X \ar@{->}[r]^{X_{X_{Y}}} \ar@{->}[d]_{X_{\lambda_{Y}}} & X_{Y} \ar@{->}[r]^{X_{U_{Y}}} & X_{Y} & \xymatrix{ X \ar@{->}[r]^{X_{X_{Y}}} \ar@{->}[d]_{X_{\lambda_{Y}}} & X_{Y} \ar@{->}[r]^{X_{U_{Y}}} \ar@{->}[d]_{X_{\rho_{Y}}} & X_{Y} \ar@{->}[r]^{X_{X_{Y}}} & X_{Y} } }
\end{align*}
\]

(Hence \((u, U^\text{left}, U^\text{right})\) is a morphism of GPS units.)

**Proof.** It is enough to prove the equation obtained by pasting the 2-cell \(X_{U_{Y}}\) on top of each side of the equation. This enables us to use the characterising equation for \(K\) and \(H\). After this rewriting, we are in position to apply Equations (13) and (14), and after cancelling \(R\) and \(L\) cells, the resulting equation amounts to a cube, where it is easy to see that each side is just \(U^\text{left}_{X} U^\text{right}_{Y}\). \(\square\)

7.6. From GPS units to cancellable-idempotent units. Given a GPS unit \((I, \lambda, \rho, K)\), just put

\[
\alpha := \lambda_{I},
\]

then \((I, \alpha)\) is a unit object in the sense of 2.9. Indeed, we already observed that \(I\) is cancellable (6.4), and from the outset \(\lambda_{I}\) is an equi-arrow. That’s all! To construct it we didn’t even need the Kelly cell, or any of the auxiliary cells or their axioms.

7.7. Left and right actions from the Kelly cell. Start with natural left and right constraints \(\lambda\) and \(\rho\) and a Kelly cell \(K : X_{\lambda_{Y}} \Rightarrow \rho_{X_{Y}}\) (not required to be coherent on either side). Construct \(K^\lambda\) as in 6.5, put \(\alpha := \lambda_{I}\), and define left and right actions as follows. We define \(L_X\) as

\[
I_{\lambda_{X}} \xRightarrow{X_{\lambda_{X}}} \lambda_{I_{X}} \Rightarrow \lambda_{I} X = \alpha X,
\]

while we define \(R_X\) simply as

\[
X_{\alpha} = X_{\lambda_{I}} \xRightarrow{K_{X_{I}}} \rho_{X_{I}}.
\]

7.8 Proposition. For fixed \((I, \lambda, \rho, K)\), the following are equivalent:

(i) The left and right 2-cells \(L\) and \(R\) just constructed in 7.7 are compatible with the Kelly cell in the sense of Equation (24).

(ii) The Kelly cell \(K\) is coherent on the left (i.e. satisfies Axiom (17)).

(ii') The Kelly cell \(K\) is coherent on the right (i.e. satisfies Axiom (18)).

**Proof.** Proposition 7.4 already says that (i) implies both (ii) and (ii'). To prove (ii)\(\Rightarrow\)(i), we start with an auxiliary observation: by massaging the naturality
we find the equation

we find the equation
tailor-made to a substitution we shall perform in a moment.

Now for the main computation, assuming first that $K$ is coherent on the left, i.e. that Axiom (17) holds. Start with the left-hand side of Equation (24), and insert the definitions we made for $L$ and $R$ to arrive at
in which we can now substitute (25) to get

\[
\begin{array}{c}
X_{II}Y \\
\downarrow X_{II}Y \\
\downarrow K_{X,Y} \\
\downarrow \rho_{X,Y} \\
X \rightarrow XY
\end{array}
\]

Here finally the three 2-cells incident to the \(X_{II}Y\) vertex cancel each other out, thanks to Axiom (17), and in the end, remembering \(\alpha = \lambda_{I}\), we get

\[
\begin{array}{c}
X_{II}Y \\
\downarrow X_{II}Y \\
\downarrow K_{X,Y} \\
\downarrow \rho_{X,Y} \\
X \rightarrow XY
\end{array}
\]

as required to establish that \(K\) satisfies Equation (24). Hence we have proved (ii)⇒(i), and therefore altogether (ii)⇒(ii'). The converse, (ii')⇒(ii) follows now by left-right symmetry of the statements. (But note that the proof via (i) is not symmetric, since it relies on the definition \(\alpha = \lambda_{I}\). To spell out a proof of (ii')⇒(ii), use rather \(\alpha = \rho_{I}\), observing that the intermediate result (i) would refer to different \(L\) and \(R\).) \(\square\)

We have now given a construction in each direction, but both constructions involved choices. With careful choices, applying one construction after the other in either way gets us back where we started. It is clear that this should constitute an equivalence of 2-categories. However, the involved choices make it awkward to make the correspondence functorial directly. (In technical terms, the constructions are ana-2-functors.) We circumvent this by introducing an intermediate 2-category dominating both. With this auxiliary 2-category, the results we already proved readily imply the equivalence.
7.9. A correspondence of 2-categories of units. Let \( \mathcal{W} \) be following 2-category. Its objects are septuples

\[
(I, \alpha, \lambda, \rho, L, R, K),
\]

with equi-arrows

\[
\alpha : II \to I, \quad \lambda_X : IX \to X, \quad \rho_X : XI \to X,
\]

(and accompanying naturality 2-cell data), and natural invertible 2-cells

\[
L : I\lambda_X \Rightarrow \alpha X, \quad R : X\alpha \Rightarrow \rho_X I, \quad K : X\lambda_Y \Rightarrow \rho_X Y.
\]

These data are required to satisfy Equation (24) (compatibility of \( K \) with \( L \) and \( R \)).

The arrows in \( \mathcal{W} \) from \((I, \alpha, \lambda, \rho, L, R, K)\) to \((J, \beta, \ell, r, L', R', H)\) are quadruples

\[
(u, U_{\text{left}}, U_{\text{right}}, U),
\]

where \( u : I \to J \) is an arrow in \( \mathcal{E} \), \( U_{\text{left}} \) and \( U_{\text{right}} \) are as in 6.10, and \( U \) is a morphism of pseudo-idempotents from \((I, \alpha)\) to \((J, \beta)\). These data are required to satisfy Equation (20) (compatibility with Kelly cells) as well as Equations (13) and (14) in Lemma 5.1 (compatibility with the left and right 2-cells).

Finally a 2-cell from \((u, U_{\text{left}}, U_{\text{right}}, U)\) to \((v, V_{\text{left}}, V_{\text{right}}, V)\) is a 2-cell

\[
T : u \Rightarrow v
\]

required to be a 2-morphism of pseudo-idempotents (compatibility with \( U \) and \( V \) as in 2.5), and to satisfy Equation (15) (compatibility with \( U_{\text{left}} \) and \( V_{\text{left}} \)) as well as Equation (16) (compatibility with \( U_{\text{right}} \) and \( V_{\text{right}} \)).

Let \( \mathcal{E} \) denote the 2-category of cancellable-idempotent units introduced in 2.9, and let \( \mathcal{G} \) denote the 2-category of GPS units of 6.10. We have evident forgetful (strict) 2-functors

\[
\Phi : \mathcal{W} \to \mathcal{E}, \quad \Psi : \mathcal{W} \to \mathcal{G}.
\]

**Theorem E (Equivalence).** The 2-functors \( \Phi \) and \( \Psi \) are 2-equivalences. More precisely they are surjective on objects and strongly fully faithful (i.e. isomorphisms on hom categories).

**Proof.** The 2-functor \( \Phi \) is surjective on objects by Lemma 3.1 and Proposition 7.4. Given an arrow \((u, U)\) in \( \mathcal{E} \) and overlying objects in \( \mathcal{W} \), Lemma 5.1 says there are unique \( U_{\text{left}} \) and \( U_{\text{right}} \), and Proposition 7.5 ensures the required compatibility with Kelly cells (Equation (20)). Hence \( \Phi \) induces a bijection on
objects in the hom categories. Lemma 5.2 says we also have a bijection on the level of 2-cells, hence Φ is an isomorphism on hom categories. On the other hand, Ψ is surjective on objects by 7.7 and Proposition 7.8. Given an arrow \((u, U^\text{left}, U^\text{right})\) in \(\mathcal{W}\), Lemma 7.10 below says that for fixed overlying objects in \(\mathcal{W}\) there is a unique associated \(U\), hence \(Ψ\) induces a bijection on objects in the hom categories. Finally, Lemma 5.2 gives also a bijection of 2-cells, hence \(Ψ\) is strongly fully faithful. ✷

7.10 Lemma. Given a morphism of GPS units

\[
(I, \lambda, \rho, K) \xrightarrow{(u, U^\text{left}, U^\text{right})} (J, \ell, r, H)
\]

fix an equi-arrow \(α : II \simeq I\) with natural families \(L_X : I\lambda_X \Rightarrow αX\) and \(R_X : αX \Rightarrow ρ_X I\) satisfying Equation (24) (compatibility with \(K\)), and fix an equi-arrow \(β : JJ \simeq J\) with natural families \(L'_X : I\ell_X \Rightarrow βX\) and \(R'_X : βX \Rightarrow r_X I\) also satisfying Equation (24) (compatibility with \(H\)). Then there is a unique 2-cell

\[
\begin{array}{c}
II \\
\alpha \downarrow \\
I
\end{array}
\xrightarrow{ uu }
\begin{array}{c}
JJ \\
β \downarrow \\
J
\end{array}
\]

satisfying Equations (13) and (14) (compatibility with \(U^\text{left}\) and the left 2-cells, as well as compatibility with \(U^\text{right}\) and the right 2-cells).

Proof. Working first with left 2-cells, define a family \(W_X\) by the equation

\[
\begin{array}{c}
IIX \\
I\lambda_X \downarrow \\
IX
\end{array}
\xrightarrow{ uuX }
\begin{array}{c}
JJX \\
βX \downarrow \\
JX
\end{array} = \begin{array}{c}
IX \\
αX \downarrow \\
IX \\
\ell_X \downarrow \\
IX \\
\lambda_X \downarrow \\
IX
\end{array}
\xrightarrow{ uuX }
\begin{array}{c}
JJX \\
βX \downarrow \\
JX
\end{array}
\]

It follows readily from Equation (21) that the family has the property

\[W_{XY} = W_X Y\]

for all \(X, Y\), and it is a standard argument that since a unit object exists, for example \((I, \lambda_I)\), this implies that

\[W_X = U X\]

for a unique 2-cell

\[
\begin{array}{c}
II \\
\alpha \downarrow \\
I
\end{array}
\xrightarrow{ uu }
\begin{array}{c}
JJ \\
β \downarrow \\
J.
\end{array}
\]
and by construction this 2-cell has the required compatibility with \( U^{left} \) and the left constraints. To see that this \( U \) is also compatible with \( U^{right} \) and the right constraints we reason backwards: \((u, U)\) is now a morphisms of units from \((I, \alpha)\) to \((J, \beta)\) to which we apply the right-hand version of Lemma 5.1 to construct a new \( U^{new} \), characterised by the compatibility condition. By Proposition 7.5 this new \( U^{new} \) is compatible with \( U^{left} \) and the Kelly cells \( K \) and \( H \) (Equation (20)), and hence it must in fact be the original \( U^{right} \) (remembering from 6.10 that \( U^{left} \) and \( U^{right} \) determine each other via (20)). So the 2-cell \( U \) does satisfy both the required compatibilities.

\[ \square \]

REFERENCES


A Criterion for Flatness of Sections of Adjoint Bundle of a Holomorphic Principal Bundle over a Riemann Surface

Indranil Biswas

Received: August 8, 2013
Communicated by Edward Frenkel

Abstract. Let $E_G$ be a holomorphic principal $G$–bundle over a compact connected Riemann surface, where $G$ is a connected reductive affine algebraic group defined over $\mathbb{C}$, such that $E_G$ admits a holomorphic connection. Take any $\beta \in H^0(X, \text{ad}(E_G))$, where $\text{ad}(E_G)$ is the adjoint vector bundle for $E_G$, such that the conjugacy class $\beta(x) \in g/G$, $x \in X$, is independent of $x$. We give a sufficient condition for the existence of a holomorphic connection on $E_G$ such that $\beta$ is flat with respect to the induced connection on $\text{ad}(E_G)$.

2010 Mathematics Subject Classification: 14H60, 14F05, 53C07
Keywords and Phrases: Holomorphic connection, adjoint bundle, flatness, canonical connection

1. Introduction

A holomorphic vector bundle $E$ over a compact connected Riemann surface $X$ admits a holomorphic connection if and only if every indecomposable component of $E$ is of degree zero [We], [At]. This criterion generalizes to the holomorphic principal $G$–bundles over $X$, where $G$ is a connected reductive affine algebraic group defined over $\mathbb{C}$ [AB].

Let $E_G$ be a holomorphic principal $G$–bundle over $X$, where $X$ and $G$ are as above. Let $\mathfrak{g}$ denote the Lie algebra of $G$. Let $\beta$ be a holomorphic section of the adjoint vector bundle $\text{ad}(E_G) = E_G \times^G \mathfrak{g}$. Our aim here is to find a criterion for the existence of a holomorphic connection on $E_G$ such that $\beta$ is flat with respect to the induced connection on $\text{ad}(E_G)$. A sufficient condition is obtained in Theorem 3.4.

For $G = \text{GL}(r, \mathbb{C})$, Theorem 3.4 says the following:
Let $E$ be a holomorphic vector bundle of rank $r$ on $X$, and $\beta \in H^0(X, \text{End}(E))$. Let 

$$E = \bigoplus_{i=1}^t E_i$$

be the generalized eigen-bundle decomposition for $\beta$. So $\beta|_{E_i} = \lambda_i \cdot \text{Id}_{E_i} + N_i$, where $\lambda_i \in \mathbb{C}$, and either $N_i = 0$ or $N_i$ is nilpotent. If $N_i \neq 0$, then assume that the section $N_i^{r_i-1}$ is nowhere vanishing, where $r_i$ is the rank of the vector bundle $E_i$. Also, assume that $E$ admits a holomorphic connection. Then Theorem 3.4 says that $E$ admits a holomorphic connection $D$ such that $\beta$ is flat with respect to the connection on $\text{End}(E)$ induced by $D$.

One may ask whether the above condition that $N_i^{r_i-1}$ is nowhere vanishing whenever $N_i \neq 0$ can be replaced by the weaker condition that the conjugacy class of $\beta(x)$, $x \in X$, is independent of $x$. As example constructed by the referee shows that this cannot be done (see Example 3.6).

2. Flat sections of the adjoint bundle

Let $X$ be a compact connected Riemann surface. Let $G$ be a connected reductive affine algebraic group defined over $\mathbb{C}$. The Lie algebra of $G$ will be denoted by $\mathfrak{g}$. The set of all conjugacy classes in $\mathfrak{g}$ will be denoted by $\mathfrak{g}/G$.

Let

$$(2.1) \quad f : E_G \longrightarrow X$$

be a holomorphic principal $G$–bundle. Define the adjoint vector bundle 

$$\text{ad}(E_G) := E_G \times^G \mathfrak{g}.$$ 

In other words, $\text{ad}(E_G)$ is the quotient of $E_G \times \mathfrak{g}$ where any $(z, v) \in E_G \times \mathfrak{g}$ is identified with $(zg, \text{Ad}(g)(v))$, $g \in G$; here $\text{Ad}(g)$ is the automorphism of $\mathfrak{g}$ corresponding to the automorphism of $G$ defined by $g' \longmapsto g^{-1}g'g$. Therefore, we have a set-theoretic map

$$(2.2) \quad \phi : \text{ad}(E_G) \longrightarrow \mathfrak{g}/G$$

that sends any $(z, v) \in E_G \times \mathfrak{g}$ to the conjugacy class of $v$.

Let

$$\text{At}(E_G) := (f_\ast TE_G)^G \subset f_\ast TE_G$$

be the Atiyah bundle for $E_G$, where $f$ is the projection in (2.1), and $TE_G$ is the holomorphic tangent bundle of $E_G$ (the action of $G$ on $E_G$ produces an action of $G$ on the direct image $f_\ast TE_G$). The Atiyah bundle fits in a short exact sequence of vector bundles

$$(2.3) \quad 0 \longrightarrow \text{ad}(E_G) \longrightarrow \text{At}(E_G) \longrightarrow TX \longrightarrow 0;$$

the above projection $\text{At}(E_G) \longrightarrow TX$, where $TX$ is the holomorphic tangent bundle of $X$, is defined by the differential $df : TE_G \longrightarrow f^\ast TX$ of $f$. A holomorphic connection on $E_G$ is a holomorphic splitting of the sequence in (2.3).
A holomorphic connection \( D \) on \( E_G \) induces a holomorphic connection on each holomorphic fiber bundle associated to \( E_G \). In particular, the vector bundle \( \text{ad}(E_G) \) gets a holomorphic connection from \( D \). A section \( \beta \) of \( \text{ad}(E_G) \) is said to be flat with respect to \( D \) if \( \beta \) is flat with respect to the connection on \( \text{ad}(E_G) \) induced by \( D \).

**Lemma 2.1.** Take a holomorphic connection \( D \) on \( E_G \), and let \( \beta \in H^0(X, \text{ad}(E_G)) \) be flat with respect to \( D \). Then the element \( \phi \circ \beta(x) \in \mathfrak{g}/G \), where \( x \in X \), is independent of \( x \).

**Proof.** Any holomorphic connection on \( X \) is flat because \( \Omega^2_X = 0 \). Using the flat connection \( D \), we may holomorphically trivialize \( E_G \) on any connected simply connected open subset of \( X \). With respect to such a trivialization, the section \( \beta \) is a constant one because it is flat with respect to \( D \). This immediately implies that \( \phi \circ \beta(x) \in \mathfrak{g}/G \) is independent of \( x \in X \). \( \square \)

### 3. Holomorphic Connections on Principal \( G \)-Bundles

A nilpotent element \( v \) of the Lie algebra of a complex semisimple group \( H \) is called regular nilpotent if the dimension of the centralizer of \( v \) in \( H \) coincides with the rank of \( H \) [Hu, p. 53].

As before, \( G \) is a connected reductive affine algebraic group defined over \( \mathbb{C} \). Take \( E_G \) as in (2.1).

**Proposition 3.1.** Take any \( \beta \in H^0(X, \text{ad}(E_G)) \). Assume that

- \( E_G \) admits a holomorphic connection,
- the element \( \phi \circ \beta(x) \in \mathfrak{g}/G \), \( x \in X \), is independent of \( x \), where \( \phi \) is defined in (2.2), and
- for every adjoint type simple quotient \( H \) of \( G \), the section of the adjoint bundle \( \text{ad}(E_H) \) given by \( \beta \), where \( E_H := E_G \times^G H \) is the principal \( H \)-bundle over \( X \) associated to \( E_G \) for the projection \( G \twoheadrightarrow H \), has the property that it is either zero or it is regular nilpotent at some point of \( X \).

Then the principal \( G \)-bundle \( E_G \) admits a holomorphic connection for which \( \beta \) is flat.

**Proof.** Let \( Z := G/[G,G] \) be the abelian quotient. It is a product of copies of \( \mathbb{C}^* \). There are quotients \( H_1, \ldots, H_\ell \) of \( G \) such that

1. each \( H_i \) is simple of adjoint type (the center is trivial), and
2. the natural homomorphism

\[
\varphi : G \to Z \times \prod_{i=1}^\ell H_i
\]

is surjective, and the kernel of \( \varphi \) is a finite group.

Let

\[
E_Z := E_G \times^G Z \quad \text{and} \quad E_{H_i} := E_G \times^G H_i, \quad i \in [1, \ell],
\]
be the holomorphic principal \( Z \)-bundle and principal \( H_i \)-bundle associated to \( E_G \) for the quotient \( Z \) and \( H_i \) respectively. Let \( \text{ad}(E_Z) \) and \( \text{ad}(E_{H_i}) \) be the adjoint vector bundles for \( E_Z \) and \( E_{H_i} \) respectively. Since the homomorphism \( \varphi \) in (3.1) induces an isomorphism of Lie algebras, we have

\[
\text{ad}(E_{G}) = \text{ad}(E_{Z}) \oplus \bigoplus_{i=1}^{\ell} \text{ad}(E_{H_i}).
\]

(3.2)

Let \( \beta_Z \) (respectively, \( \beta_i \)) be the holomorphic section of \( \text{ad}(E_Z) \) (respectively, \( \text{ad}(E_{H_i}) \)) given by \( \beta \) using the decomposition in (3.2). Since the conjugacy class of \( \beta(x) \) is independent of \( x \in X \) (the second condition in the proposition), we conclude that the conjugacy class of \( \beta_i(x) \) is also independent of \( x \in X \). A holomorphic connection on \( E_G \) induces a holomorphic connection on \( E_Z \). Since \( E_Z \) admits a holomorphic connection, and \( Z \) is a product of copies of \( \mathbb{C}^* \), there is a unique holomorphic connection \( D_Z \) on \( E_Z \) whose monodromy lies inside the maximal compact subgroup of \( Z \). The connection on \( \text{ad}(E_Z) \) induced by this connection \( D_Z \) has the property that any holomorphic section of \( \text{ad}(E_Z) \) is flat with respect to it. In particular, the section \( \beta_Z \) is flat with respect to this induced connection on \( \text{ad}(E_Z) \).

Now take any \( i \in [1, \ell] \). A holomorphic connection on \( E_G \) induces a holomorphic connection on \( E_{H_i} \). If the section \( \beta_i \) is zero at some point, then \( \beta_i \) is identically zero because the conjugacy class of \( \beta_i(x) \) is independent of \( x \). Hence, in that case \( \beta_i \) is flat with respect to any connection on \( \text{ad}(E_{H_i}) \). Therefore, assume that \( \beta_i \) is not zero at any point of \( X \).

By the assumption in the proposition, \( \beta_i \) is regularly nilpotent over some point of \( X \). Since the conjugacy class of \( \beta_i(x), x \in X \), is independent of \( x \), we conclude that \( \beta_i \) is regular nilpotent over every point of \( X \). We will now show that the holomorphic principal \( H_i \)-bundle \( E_{H_i} \) is semistable.

For each point \( x \in X \), from the fact that \( \beta_i(x) \) is regular nilpotent we conclude that there is a unique Borel subalgebra \( \mathfrak{b}_x \) of \( \text{ad}(E_{H_i})_x \) such that \( \beta_i(x) \in \mathfrak{b}_x \). [Hu] p. 62, Theorem]. Let

\[
\widetilde{\mathfrak{b}} \subset \text{ad}(E_{H_i})
\]

be the Borel subalgebra bundle such that for every point \( x \) the fiber \( (\widetilde{\mathfrak{b}})_x \) is \( \mathfrak{b}_x \). Fix a Borel subgroup \( B \subset H_i \). Using \( \widetilde{\mathfrak{b}} \), we will construct a holomorphic reduction of structure group of \( E_{H_i} \) to the subgroup \( B \).

Let \( \mathfrak{b} \) be the Lie algebra of \( B \). The Lie algebra of \( H_i \) will be denoted by \( \mathfrak{h}_i \). We recall that \( \text{ad}(E_{H_i}) \) is the quotient of \( E_{H_i} \times \mathfrak{h}_i \) where two points \( (z_1, v_1) \) and \( (z_2, v_2) \) of \( E_{H_i} \times \mathfrak{h}_i \) are identified if there is an element \( h \in H_i \) such that \( z_2 = z_1 h \) and \( v_2 = \text{Ad}(h)(v_1) \), where \( \text{Ad}(h) \) is the automorphism of \( \mathfrak{h}_i \) corresponding to the automorphism \( y \mapsto h^{-1} y h \) of \( H_i \). For any point \( x \in X \), let \( \widetilde{E}_{B,x} \subset (E_{H_i})_x \) be the complex submanifold consisting of all \( z \in (E_{H_i})_x \) such that for all \( v \in \mathfrak{b} \), the image of \( (z, v) \) in \( \text{ad}(E_{H_i})_x \) lies in \( \widetilde{\mathfrak{b}}_x \). Since any two Borel subalgebras of \( \mathfrak{h}_i \) are conjugate, it follows that \( \widetilde{E}_{B,x} \) is nonempty. The normalizer of \( \mathfrak{b} \) in \( H_i \) coincides with \( B \). From this it follows that \( \widetilde{E}_{B,x} \) is preserved by the action of \( B \) on \( (E_{H_i})_x \), with the action of \( B \) on \( E_{B,x} \) being
transitive. Let

\[ E_B \subset E_{H_i} \]

be the complex submanifold such that \( E_B \cap (E_{H_i})_x = E_{B,x} \) for every \( x \in X \). From the above properties of \( E_{B,x} \) it follows immediately that \( E_B \) is a holomorphic reduction of structure group of the principal \( H_i \)-bundle \( E_{H_i} \) to the subgroup \( B \).

Consider the adjoint action of \( B \) on \( b_1 := b/[b, b] \). Let

\[ E_B(b_1) := E_B \times^B b_1 \to X \]

be the holomorphic vector bundle associated to \( E_B \) for the \( B \)-module \( b_1 \). Since \( \beta_i \) is everywhere regular nilpotent, it follows that the vector bundle \( E_B(b_1) \) is trivial. Consequently, for any character \( \chi \) of \( B \) which is a nonnegative integral combination of simple roots, the line bundle \( E_B(\chi) \to X \) associated to \( E_B \) for the character \( \chi \) is trivial \([AAB, p. 708, Theorem 5]\). Therefore, for any character \( \chi \) of \( B \), the line bundle \( E_B(\chi) \) associated to \( E_B \) for \( \chi \) is trivial.

Let \( d \) be the complex dimension of \( h_i \). Consider the adjoint action on \( h_i \). Note that \( \text{ad}(E_{H_i}) \) is identified with the vector bundle associated to the principal \( B \)-bundle \( E_B \) for this \( B \)-module \( h_i \). Since \( B \) is solvable, there is a filtration of \( B \)-modules

\[ 0 = V_0 \subset V_1 \subset \cdots \subset V_{d-1} \subset V_d = h_i \]

such that \( \dim V_j = j \) for all \( j \in [1, d] \). The corresponding filtration of vector bundles associated to \( E_B \) is a filtration of \( \text{ad}(E_{H_i}) \) such that the successive quotients are the line bundles \( E_B(V_j/V_{j-1}), i \in [1, d], \) associated to \( E_B \) for the \( B \)-modules \( V_j/V_{j-1} \). We noted above that the line bundles associated to \( E_B \) for the characters of \( B \) are trivial.

Therefore, we get a filtration of \( \text{ad}(E_{H_i}) \) such that each successive quotient is the trivial line bundle. This immediately implies that the vector bundle \( \text{ad}(E_{H_i}) \) is semistable. Hence the holomorphic principal \( H_i \)-bundle \( E_{H_i} \) is semistable \([AAB, p. 698, Lemma 3]\).

Since \( H_i \) is simple, and \( E_{H_i} \) is semistable, there is a natural holomorphic connection on \( E_{H_i} \) \([BG, p. 20, Theorem 1.1]\) (set the Higgs field in \([BG, Theorem 1.1]\) to be zero). Let \( D^{H_i} \) denote this connection. The vector bundle \( \text{ad}(E_{H_i}) \) being semistable of degree zero has a natural holomorphic connection \([Si, p. 36, Lemma 3.5]\). See also \([BG, p. 20, Theorem 1.1]\). (In both \([Si, Lemma 3.5]\) and \([BG, Theorem 1.1]\) set the Higgs field to be zero.) Let \( D^{\text{ad}} \) denote this holomorphic connection on \( \text{ad}(E_{H_i}) \). This connection \( D^{\text{ad}} \) coincides with the one induced by \( D^{H_i} \) (see the construction of the connection in \([BG]\)). Any holomorphic section of \( \text{ad}(E_{H_i}) \) is flat with respect to \( D^{\text{ad}} \). To see this, let

\[ \phi : O_X \to \text{ad}(E_{H_i}) \]

be the homomorphism given by a nonzero holomorphic section of \( \text{ad}(E_{H_i}) \).

Since image(\( \phi \)) is a semistable subbundle of \( \text{ad}(E_{H_i}) \) of degree zero, the connection \( D^{\text{ad}} \) preserves image(\( \phi \)), and, moreover, the restriction of \( D^{\text{ad}} \) to image(\( \phi \)) coincides with the canonical connection of image(\( \phi \)) \([Si, p. 36, Lemma 3.5]\).
The canonical connection on the trivial holomorphic line bundle image(φ) is the trivial connection (the monodromy is trivial).
In particular, the connection on ad(E_{H_1}) induced by D_{H_1} has the property that the section β is flat with respect to it.
Since the homomorphism of Lie algebras corresponding to φ (in (3.1)) is an isomorphism, if we have holomorphic connections on E_Z and E_{H_1}, [1, ℓ], then we get a holomorphic connection on E_G; simply pullback the connection form using the map

$$E_G \rightarrow E_Z \times X E_{H_1} \times X \cdots \times X E_{H_ℓ}.$$  

The connection on E_G given by the connections on E_Z and E_{H_1}, [1, ℓ], constructed above satisfies the condition that β is flat with respect to it. This completes the proof of the proposition. □

**Lemma 3.2.** Take any semisimple section β_s ∈ H^0(X, ad(E_G)) such that the element φ ◦ β_s(x) ∈ g/G, x ∈ X, is independent of x, where φ is defined in (2.2). Then β_s produces a holomorphic reduction of structure group of E_G to a Levi subgroup of a parabolic subgroup of G. The conjugacy class of the Levi subgroup is determined by φ ◦ β_s(x) ∈ g/G.

**Proof.** Fix an element

$$v_0 \in g$$

such that the image of v_0 in g/G coincides with φ ◦ β_s(x). Let L ⊂ G be the centralizer of v_0. It is known that L is a Levi subgroup of some parabolic subgroup of G [DM, p. 26, Proposition 1.22] (note that L is the centralizer of the torus in G generated by v_0). In particular, L is connected and reductive.
For any point x ∈ X, let F_x ⊂ (E_G)_x be the complex submanifold consisting of all points z such that the image of (z, v_0) in ad(E_G)_x coincides with β_s(x). (Recall that ad(E_G) is a quotient of E_G × g.) Let

$$F_L \subset E_G$$

be the complex submanifold such that F_L \cap (E_G)_x = F_x for all x ∈ X. It is straightforward to check that F_L is a holomorphic reduction of structure group of the principal G-bundle E_G to the subgroup L. □

**Remark 3.3.** If β_s ∈ H^0(X, ad(E_G)) is such that β_s(x) is semisimple for every x ∈ X, then the conjugacy class of β_s(x) is in fact independent of x. But we do not need this here.

From the Jordan decomposition of a complex reductive Lie algebra we know that for any holomorphic section θ of ad(E_G), there is a naturally associated semisimple (respectively, nilpotent) section θ_s (respectively, θ_n) such that θ = θ_s + θ_n.
Take any β ∈ H^0(X, ad(E_G)). Let

$$β = β_s + β_n$$

be the Jordan decomposition. Assume that the element φ ◦ β(x) ∈ g/G, x ∈ X, is independent of x, where φ is defined in (2.2). This implies that
A Criterion for Flatness of Sections of Adjoint Bundle

117

φ ◦ βs(x) ∈ g/G, x ∈ X, is also independent of x. Let (LL, FL) be the principal bundle constructed in Lemma 3.2 from βs. Let H be an adjoint type simple quotient of L. Let

\[ E_H := F_L \times^L H \rightarrow X \]

be the holomorphic principal H–bundle associated to FL for the projection \( \mathbb{L} \rightarrow H \).

Since \([β_s, β_n] = 0\), from the construction of FL it follows that

\[ β_n \in H^0(X, \text{ad}(FL)) \subset H^0(X, \text{ad}(EG)). \]

Therefore, using the natural projection \( \text{ad}(FL) \rightarrow \text{ad}(E_H) \), given by the projection of the Lie algebra \( \text{Lie}(L) \rightarrow \text{Lie}(H) \), the above section \( β_n \) produces a holomorphic section of \( \text{ad}(E_H) \). Let

\[ (3.3) \quad \tilde{β}_n \in H^0(X, \text{ad}(E_H)) \]

be the section constructed from \( β_n \).

THEOREM 3.4. Take any \( β \in H^0(X, \text{ad}(EG)) \). Let \( β = β_s + β_n \) be the Jordan decomposition. Assume that

- \( E_G \) admits a holomorphic connection,
- the element \( φ \circ β(x) \in g/G, x \in X \), is independent of \( x \), where \( φ \) is defined in (2.2), and
- for every adjoint type simple quotient \( H \) of \( L \), the section \( \tilde{β}_n \) in (3.3) of \( \text{ad}(E_H) \) has the property that it is either zero or it is regular nilpotent at some point of \( X \).

Then the principal \( G \)–bundle \( E_G \) admits a holomorphic connection for which \( β \) is flat.

Proof. Note that

\[ β_s \in H^0(X, \text{ad}(FL)) \subset H^0(X, \text{ad}(EG)). \]

In fact, for each point \( x \in X \), the element \( β_s(x) \in \text{ad}(FL)_x \) is in the center of \( \text{ad}(FL)_x \). Consider the abelian quotient

\[ Z_L = \mathbb{L}/[\mathbb{L}, \mathbb{L}]. \]

Let \( F_{Z_L} \) be the holomorphic principal \( Z_L \)–bundle over \( X \) obtained by extending the structure group of the principal \( \mathbb{L} \)–bundle \( F_L \) using the quotient map \( \mathbb{L} \rightarrow Z_L \). The adjoint vector bundle \( \text{ad}(F_{Z_L}) \) is a direct summand of \( \text{ad}(F_L) \). In fact, for each \( x \in X \), the subspace \( \text{ad}(F_{Z_L})_x \subset \text{ad}(F_L)_x \) is the center of the Lie algebra \( \text{ad}(F_L)_x \).

A holomorphic connection on \( F_{Z_L} \) induces a holomorphic connection on \( E_G \). We can now apply Proposition 3.1 to \( F_L \) to complete the proof of the theorem. But for that we need to show that \( F_{Z_L} \) admits a holomorphic connection.

Let \( \mathfrak{l} \) be the Lie algebra of \( \mathbb{L} \). Consider the inclusion of \( \mathbb{L} \)–modules \( \mathfrak{l} \rightarrow \mathfrak{g} \) given by the inclusion of \( \mathbb{L} \) in \( G \). Since \( \mathbb{L} \) is reductive, there is a sub \( \mathbb{L} \)–module \( S \subset \mathfrak{g} \) such that the natural homomorphism

\[ \mathfrak{l} \oplus S \rightarrow \mathfrak{g} \]
is an isomorphism (so $S$ is a complement of $0$). Let

$$p : g \rightarrow l$$

be the projection given by the above decomposition of $g$.

Let $D$ be a holomorphic connection on $E_G$. So $D$ is a holomorphic $1$–form on the total space of $E_G$ with values in the Lie algebra $g$. Let $D'$ be the restriction of this $1$–form to the complex submanifold $F_L \subset E_G$. Consider the $1$–valued $1$–form $p \circ D'$ on $E_L$, where $p$ is the projection in (3.4). This $1$–valued $1$–form on $F_L$ defines a holomorphic connection of the principal $L$–bundle $F_L$. Now Proposition 3.1 completes the proof of the theorem.

We recall that a holomorphic vector bundle $W$ on $X$ has a holomorphic connection if and only if each indecomposable component of $W$ is of degree zero [We], [At, p. 203, Theorem 10]. This criterion generalizes to holomorphic principal $G$–bundles on $X$ (see [AB] for details).

We now set $G = GL(r, \mathbb{C})$ in Theorem 3.4. Let $E$ be a holomorphic vector bundle of rank $r$ on $X$. Take any $\beta \in H^0(X, \text{End}(E))$.

Let

$$E = \bigoplus_{i=1}^{\ell} E_i$$

be the generalized eigen-bundle decomposition of $E$ for $\beta$. Therefore,

$$\beta|_{E_i} = \lambda_i \cdot \text{Id}_{E_i} + N_i,$$

where $\lambda \in \mathbb{C}$, and either $N_i = 0$ or $N_i$ is nilpotent.

Then Theorem 3.4 has the following corollary:

**Corollary 3.5.** For every $N_i \neq 0$, assume that the section $N_i^{r_i - 1}$ of $\text{End}(E_i)$ is nowhere vanishing, where $r_i$ is the rank of the vector bundle $E_i$ in (3.5). If the holomorphic vector bundle $E$ admits a holomorphic connection, then it admits a holomorphic connection $D$ such that the section $\beta$ is flat with respect to the connection on $\text{End}(E)$ induced by $D$.

Consider the condition on $\beta$ in Corollary 3.5, which says that $N_i^{r_i - 1}$ is nowhere vanishing whenever $N_i \neq 0$. This condition implies that the image of $\beta(x)$ in $M(r, \mathbb{C})/GL(r, \mathbb{C})$ is independent of $x \in X$ (here $GL(r, \mathbb{C})$ acts on its Lie algebra $M(r, \mathbb{C})$ via conjugation). Therefore, one may ask whether the above mentioned condition in Corollary 3.5 can be replaced by the weaker condition that the conjugacy class of $\beta(x)$ is independent of $x \in X$. Note that if this can be done, then the sufficient condition in Corollary 3.5 for the existence of a connection on $E$ such that $\beta$ is flat with respect to it actually becomes a necessary and sufficient condition. The following construction of the referee shows that the condition in Corollary 3.5 cannot be replaced by the weaker condition that the conjugacy class of $\beta(x)$ is independent of $x \in X$. 


Example 3.6 (Referee). Let $X$ be of sufficiently high genus. Let $L$ and $M$ be holomorphic line bundles on $X$ of degree 1 and degree $-2$ respectively. Then there exists an indecomposable holomorphic vector bundle $E$ of rank three on $X$ satisfying the following condition: it admits a filtration of holomorphic subbundles

$$L = E_1 \subset E_2 \subset E$$

such that $E_2/L = M$ and $E/E_2 = L$. We omit the detailed arguments given by the referee showing that such a vector bundle $E$ exists. Let $\beta$ denote the composition

$$E \rightarrow E/E_2 = L = E_1 \hookrightarrow E.$$ 

Clearly, the conjugacy class of $\beta(x)$ is independent of $x \in X$. The vector bundle $E$ admits a holomorphic connection because it is indecomposable of degree zero. If $D$ is a holomorphic connection on $E$ such that $\beta$ is flat with respect to the connection on $\text{End}(E)$ induced by $D$, then the subsheaf $\text{image}(\beta) \subset E$ is flat with respect to $D$. But $\text{image}(\beta) = L$ does not admit a holomorphic connection because it is of nonzero degree.

Acknowledgements

The referee pointed out an error in an earlier version of the paper, and also made detailed comments that included Example 3.6. This helped the author to find the right formulation of Theorem 3.4. The author is immensely grateful to the referee for this.

References


ABSTRACT. We show that points in the intersection of the tropicalizations of subvarieties of a torus lift to algebraic intersection points with expected multiplicities, provided that the tropicalizations intersect in the expected dimension. We also prove a similar result for intersections inside an ambient subvariety of the torus, when the tropicalizations meet inside a facet of multiplicity 1. The proofs require not only the geometry of compactified tropicalizations of subvarieties of toric varieties, but also new results about the geometry of finite type schemes over non-noetherian valuation rings of rank 1. In particular, we prove subadditivity of codimension and a principle of continuity for intersections in smooth schemes over such rings, generalizing well-known theorems over regular local rings. An appendix on the topology of finite type morphisms may also be of independent interest.

2010 Mathematics Subject Classification: 14T05, 14C17, 14M25, 14A15

Keywords and Phrases: tropical geometry, intersection theory, schemes over valuation rings

CONTENTS

1. Introduction 122
2. Preliminaries 125
  2.1. Initial degenerations 126
  2.2. Tropicalization 127
  2.3. Tropical multiplicities 128
  2.4. Minkowski weights and the fan displacement rule 129
  2.5. Toric intersection theory 130
  2.6. Stable tropical intersections 131
  2.7. Compatibility of toric and stable tropical intersections 132
3. The trivial valuation case 136
  3.1. Topology of extended tropicalizations 136
  3.2. Tropical subadditivity 138
1. Introduction

Tropical geometry studies the valuations of solutions to polynomial equations, and may be thought of as a generalization of the theory of Newton polygons to multiple polynomials in multiple variables. It is natural to consider what conditions guarantee that if two closed subvarieties $X$ and $X'$ in a torus $T$ have points with the same valuation, then $X \cap X'$ contains a point of the given valuation. In the context of tropical geometry, this is the question of when tropicalization commutes with intersection. The tropicalization of the intersection of two closed subvarieties of $T$ over a nonarchimedean field is contained in the intersection of their tropicalizations, but this containment is sometimes strict.

For instance, if $X$ is smooth in characteristic zero, and $X'$ is the translate of $X$ by a general torsion point $t$ in $T$, then $X$ and $X'$ are either disjoint or meet transversely, but their tropicalizations are equal. The dimension of $X \cap X'$ is then strictly less than the dimension of $\text{Trop}(X) \cap \text{Trop}(X')$, so there are tropical intersection points that do not lift to algebraic intersection points. In other cases, such as Example 6.2 below, the tropicalizations meet in a positive dimensional set that does not contain the tropicalization of any positive dimensional variety, so at most finitely many of the tropical intersection points can lift to algebraic intersection points. Our main result, in its most basic form, says that the dimension of the tropical intersection is the only obstruction to such lifting.

Following standard terminology from intersection theory in algebraic geometry, we say that $\text{Trop}(X)$ and $\text{Trop}(X')$ meet properly at a point $w$ if $\text{Trop}(X) \cap \text{Trop}(X')$ has codimension $\text{codim} X + \text{codim} X'$ in a neighborhood of $w$. We
say that $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly if they meet properly at every point of their intersection, which may be empty.\footnote{Codimension is subadditive for intersections of tropicalizations, which means that $\text{Trop}(X) \cap \text{Trop}(X')$ has codimension at most $j + j'$ at every point, so $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly if the intersection is either empty or has the smallest possible dimension. See Proposition \ref{proposition:codimension_subadditive}.}

**Theorem 1.1.** Suppose $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly at $w$. Then $w$ is contained in the tropicalization of $X \cap X'$.

In other words, tropicalization commutes with intersections when the intersections have the expected dimension. This generalizes a well-known result of Bogart, Jensen, Speyer, Sturmfels, and Thomas, who showed that tropicalization commutes with intersection when the tropicalizations meet transversely \cite[Lemma 3.2]{BJSST07}. Our proof, and the proofs of all of the other main results below, involves a reduction to the case where the base field is complete with respect to its nonarchimedean norm and the point $w$ is rational over the value group. The reduction is based on the results of Appendix A and should become a standard step in the rigorous application of tropical methods to algebraic geometry. We emphasize that the theorems hold in full generality, over an arbitrary algebraically closed nonarchimedean field, and at any real point in the intersection of the tropicalizations.

Although Theorem 1.1 improves significantly on previously known results, it is still too restrictive for the most interesting potential applications. Frequently, $X$ and $X'$ are closed subvarieties of an ambient variety $Y$ inside the torus. In this case, one cannot hope that $\text{Trop}(X)$ and $\text{Trop}(X')$ will meet properly in the above sense. Instead, we say that the tropicalizations $\text{Trop}(X)$ and $\text{Trop}(X')$ meet properly at a point $w$ in $\text{Trop}(Y)$ if the intersection $\text{Trop}(X) \cap \text{Trop}(X') \subset \text{Trop}(Y)$ has pure codimension $\operatorname{codim} Y X + \operatorname{codim} Y X'$ in a neighborhood of $w$. We extend Theorem 1.1 to proper intersections at suitable points of $\text{Trop}(Y)$, as follows.

Recall that $\text{Trop}(Y)$ is the underlying set of a polyhedral complex of pure dimension $\dim Y$, with a positive integer multiplicity assigned to each facet, by which we mean a maximal face. We say that a point in $\text{Trop}(Y)$ is simple if it is in the interior of a facet of multiplicity 1.

**Theorem 1.2.** Suppose $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly at a simple point $w$ in $\text{Trop}(Y)$. Then $w$ is contained in the tropicalization of $X \cap X'$.

Every point in the tropicalization of the torus is simple, so Theorem 1.2 is the special case of Theorem 1.1 where $Y$ is the full torus $T$. The hypothesis that $w$ be simple is necessary; see Section 6. We strengthen Theorem 1.2 further by showing that where $\text{Trop}(X)$ and $\text{Trop}(X')$ meet properly, the facets of the tropical intersection appear with the expected multiplicities, suitably interpreted. See Theorem 5.1.1 for a precise statement.
The proof of Theorem 1.2 is in two steps. Roughly speaking, we lift first from tropical points to points in the initial degeneration, and then from the initial degeneration to the original variety. Recall that the tropicalization of a closed subvariety $X$ of $T$ is the set of weight vectors $w$ such that the initial degeneration $X_w$ is nonempty in the torus torsor $T_w$ over the residue field.

**Theorem 1.3.** Suppose $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly at a simple point $w$ in $\text{Trop}(Y)$. Then $X_w$ and $X'_w$ have nonempty proper intersection in the smooth variety $Y_w$.

If $w$ is a simple point of $\text{Trop}(Y)$ then standard arguments show that $X_w$ meets $X'_w$ properly at every point in the intersection; the main content of the theorem is that $X_w \cap X'_w$ is nonempty. The proof, given in Section 3, uses extended tropicalizations and the intersection theory of toric varieties. In Section 4, we develop geometric techniques over valuation rings of rank 1 that permit lifting of such proper intersections at smooth points from the special fiber to the generic fiber.

**Theorem 1.4.** Suppose $X_w$ meets $X'_w$ properly at a smooth point $x$ of $Y_w$. Then $x$ is contained in $(X \cap X')_w$.

In particular, if $X_w$ meets $X'_w$ properly at a smooth point of $Y_w$ then $w$ is contained in $\text{Trop}(X \cap X')$. If $w$ is defined over the value group of the nonarchimedean field then surjectivity of tropicalization says that $x$ can be lifted to a point of $X \cap X'$. The dimension theory over valuation rings of rank 1 developed in Section 4 also gives a new proof of surjectivity of tropicalization, as well as density of tropical fibers.

Theorem 1.4 is considerably stronger than Theorems 1.1 and 1.2. It often happens that initial degenerations meet properly even when tropicalizations do not. The proof of this theorem follows standard arguments from dimension theory, but the dimension theory that is needed is not standard because the valuation rings we are working with are not noetherian. We develop the necessary dimension theory systematically in Section 4 using noetherian approximation to prove a version of the Krull principal ideal theorem and then deducing subadditivity of codimension for intersections in smooth schemes of finite type over valuation rings of rank 1. Furthermore, we prove a principle of continuity, showing that intersection numbers are well-behaved in families over valuation rings of rank 1. In Section 5, we apply this principle of continuity to prove a stronger version of Theorem 1.2 with multiplicities and then extend all of these results to proper intersections of three or more subvarieties of $T$.

One special case of these lifting results with multiplicities, for complete intersections, has been applied by Rabinoff in an arithmetic setting to construct canonical subgroups for abelian varieties over $p$-adic fields [Rab12a].

**Remark 1.5.** In addition to their intrinsic appeal, our results are motivated by the possibility of applications to proving correspondence theorems such as the one proved by Mikhalkin for plane curves [Mik05]. Mikhalkin considers curves of fixed degree and genus subject to constraints of passing through specified points.
points. He shows, roughly speaking, that a tropical plane curve that moves in a family of the expected dimension passing through specified points lifts to a predictable number of algebraic curves passing through prescribed algebraic points. Here, we consider points subject to the constraints of lying inside closed subvarieties $X$ and $X'$ and prove the analogous correspondence—if a tropical point moves in a family of the expected dimension inside $\text{Trop}(X) \cap \text{Trop}(X')$ then it lifts to a predictable number of points in $X \cap X'$. One hopes that Mikhalkin’s correspondence theorem will eventually be reproved and generalized using Theorem 1.2 on suitable tropicalizations of moduli spaces of curves. Similarly, one hopes that Schubert problems can be answered tropically using Theorem 1.2 on suitable tropicalizations of Grassmannians and flag varieties. The potential for such applications underlines the importance of having the flexibility to work with intersections inside ambient subvarieties of the torus.

Acknowledgments. We thank B. Conrad, D. Eisenbud, W. Heinzer, D. Rydh, and B. Ulrich for helpful conversations, and are most grateful to J. Rau and M. Rojas for insightful comments on an earlier draft of this work. We also thank the referee for a careful reading.

The first author was partially supported by a fellowship from the National Science Foundation. The second author was partially supported by the Clay Mathematics Institute and NSF DMS grant 1068689. Part of this research was carried out during the special semesters on Algebraic Geometry and Tropical Geometry at MSRI, in Spring and Fall 2009.

2. Preliminaries

Let $K$ be an algebraically closed field, and let $\nu : K^* \to \mathbb{R}$ be a valuation with value group $G$. Let $R$ be the valuation ring, with maximal ideal $\mathfrak{m}$, and residue field $k = R/\mathfrak{m}$. Since $K$ is algebraically closed, the residue field $k$ is algebraically closed and the valuation group $G$ is divisible. In particular, $G$ is dense in $\mathbb{R}$ unless the valuation $\nu$ is trivial, in which case $G$ is zero. Typical examples of such nonarchimedean fields in equal characteristic are given by the generalized power series fields $K = k((t^G))$, whose elements are formal power series with coefficients in the algebraically closed field $k$ and exponents in $G$, where the exponents occurring in any given series are required to be well-ordered [Poo93].

Let $T$ be an algebraic torus of dimension $n$ over $R$, with character lattice $M \cong \mathbb{Z}^n$, and let $N = \text{Hom}(M, \mathbb{Z})$ be the dual lattice. We write $N_G$ for $N \otimes G$, and $N_R$ for $N \otimes \mathbb{R}$, the real vector space of linear functions on the character lattice. We treat $M$ and $N$ additively and write $x^u$ for the character associated to a lattice point $u$ in $M$, which is a monomial in the Laurent polynomial ring $K[M]$. We write $T$ for the associated torus over $K$.

In this section, we briefly review the basic properties of initial degenerations and tropicalizations, as well as the relationship between tropical intersections and intersections in suitable toric compactifications of $T$. 
2.1. Initial degenerations. Each vector $w$ in $N_R$ determines a weight function on monomials, where the $w$-weight of $ax^u$ is $\nu(a) + \langle u, w \rangle$. The tilted group ring $R[M]^w \subset K[M]$ consists of the Laurent polynomials $\sum a_u x^u$ in which every monomial has nonnegative $w$-weight. We then define $T^w = \operatorname{Spec} R[M]^w$. If $\nu$ is nontrivial, then $T^w$ is an integral model of $T$, which means that it is a scheme over $R$ whose generic fiber $T^w \times_{\operatorname{Spec}R} \operatorname{Spec} K$ is naturally identified with $T$. In general, the scheme $T^w$ carries a natural $T$-action, and is a $T$-torsor if $w$ is in $N_G$.

Remark 2.1.1. If $\nu$ is nontrivial and $w$ is not in $N_G$, or if $\nu$ is trivial and $w$ does not span a rational ray, then $T^w$ is not of finite type over $\operatorname{Spec} R$. Some additional care is required in handling these schemes, but no major difficulties arise for the purposes of this paper. The special fiber $T^w_k$ is still finite type over $k$, and is a torsor over a quotient torus of $T_k$. The basic properties of the schemes $T^w$ and their closed subschemes may be understood by passing to a valued extension field with value group $G'$ such that $w$ is in $N_{G'}$ and analyzing how these schemes and their special fibers transform under such extensions. This analysis is carried out in Appendix A. See, in particular, Theorems A.3 and A.4 and Remark A.5.

For arbitrary $w$ in $N_R$, the scheme $T^w$ is reduced and irreducible, flat over $\operatorname{Spec} R$, and contains $T$ as a dense open subscheme. We define $T^w_w$ to be the closed subscheme cut out by monomials of strictly positive $w$-weight. If the valuation is nontrivial, then $T^w_w$ is the special fiber of $T^w$.

Let $X$ be a closed subscheme of $T$, of pure dimension $d$. Let $X^w$ denote the closure of $X$ in $T^w$.

**Definition 2.1.2.** The initial degeneration $X^w_w$ is the closed subscheme of $T^w_w$ obtained by intersecting with $X^w$.

The terminology reflects the fact that $X^w_w$ is cut out by residues of initial terms (lowest-weight monomials) of Laurent polynomials in the ideal of $X$. If the valuation is nontrivial, then $X^w_w$ is an integral model of $X$, and $X^w_w$ is the special fiber of this model.

One consequence of Gröbner theory is that the space of weight vectors $N_R$ can be decomposed into finitely many polyhedral cells so that the initial degenerations are essentially invariant at points of $N_G$ on the relative interior of each cell, as discussed in more detail in the following subsection. Roughly speaking, the cells are cut out by inequalities whose linear terms have integer coefficients, and whose constant terms are in the value group $G$. More precisely, the cells are integral $G$-affine polyhedra, defined as follows.

**Definition 2.1.3.** An integral $G$-affine polyhedron in $N_R$ is the solution set of a finite number of inequalities

$$\langle u, v \rangle \leq b,$$

with $u$ in the lattice $M$ and $b$ in the value group $G$. 
Lifting Tropical Intersections

An integral $G$-affine polyhedral complex $\Sigma$ is a polyhedral complex consisting entirely of integral $G$-affine polyhedra. In other words, it is a finite collection of integral $G$-affine polyhedra such that every face of a polyhedron in $\Sigma$ is itself in $\Sigma$, and the intersection of any two polyhedra in $\Sigma$ is a face of each. Note that if $G$ is zero, then an integral $G$-affine polyhedron is a rational polyhedral cone, and an integral $G$-affine polyhedral complex is a fan.

2.2. Tropicalization. Following Sturmfels, we define the tropicalization of $X$ to be

$$\text{Trop}(X) = \{w \in \mathbb{N}_G | X_w \text{ is nonempty}\}.$$ 

The foundational theorems of tropical geometry, due to the work of many authors, are the following.

1. The tropicalization $\text{Trop}(X)$ is the underlying set of an integral $G$-affine polyhedral complex of pure dimension $d$.

2. The integral $G$-affine polyhedral structure on $\text{Trop}(X)$ can be chosen so that the initial degenerations $X_w$ and $X_{w'}$ are $T_k$-affinely equivalent for any $w$ and $w'$ in $\mathbb{N}_G$ in the relative interior of the same face.

3. The image of $X(K)$ under the natural tropicalization map $\text{trop} : T(K) \to \mathbb{N}_R$ is exactly $\text{Trop}(X) \cap \mathbb{N}_G$. Here, two subschemes of the $T_k$-torsors $T_w$ and $T_{w'}$ are said to be $T_k$-affinely equivalent if they are identified under some $T_k$-equivariant choice of isomorphism $T_w \cong T_{w'}$. In fact, a stronger statement holds: for any points $w$ and $w'$ in $\mathbb{N}_G$, there exists a $T_k$-equivariant isomorphism $T_w \cong T_{w'}$ which sends $X_w$ to $X_{w'}$ for all $X$ such that $w$ and $w'$ are in the relative interior of the same face of $\text{Trop}(X)$. If the valuation is nontrivial, then it follows from (3) that $\text{Trop}(X)$ is the closure of the image of $X(K)$ under the tropicalization map.

For any extension of valued fields $L|K$, the tropicalization of the base change $\text{Trop}(X_L)$ is exactly equal to $\text{Trop}(X) \setminus \mathbb{N}_G$ in particular, for any algebraically closed extension $L|K$ with nontrivial valuation, $\text{Trop}(X)$ is the closure of the image of $X(L)$. Furthermore, if we extend to some $L$ that is complete with respect to its valuation then the tropicalization map on $X(L)$ extends naturally to a continuous map on the nonarchimedean analytification of $X_L$, in the sense of Berkovich [Ber90], whose image is exactly $\text{Trop}(X)$ [Pay09a]. It follows, by reducing to the case of a complete field, that $\text{Trop}(X)$ is connected if $X$ is connected, since the analytification of a connected scheme over a complete nonarchimedean field is connected.

Remark 2.2.1. The natural tropicalization map from $T(K)$ to $\mathbb{N}_G$ takes a point $t$ to the linear function $u \mapsto \nu \circ \text{ev}_t : x^u$, and can be understood as a coordinatewise valuation map, as follows. The choice of a basis for $M$ induces isomorphisms

\cite{BG84}, \cite{SS04}, and Propositions 2.1.4, Theorem 2.2.1, and Proposition 2.4.5 of [Spe05] for proofs of (1) and (2). The first proposed proof of (3), due to Speyer and Sturmfels, contained an essential gap that is filled by Theorem 4.2.5 below. Other proofs of (3) have appeared in [Dra08 Theorem 4.2], [Gub12 Proposition 4.14] and [Pay09b Pay12]. See also Remark 4.2.6.

Documenta Mathematica 18 (2013) 121–175
$T \cong (K^*)^n$ and $N_R \cong \mathbb{R}^n$. In such coordinates, the tropicalization map sends
$(t_1, \ldots, t_n)$ to $(\nu(t_1), \ldots, \nu(t_n))$.

There is no canonical choice of polyhedral structure on $\text{Trop}(X)$ satisfying (2) in
general, but any refinement of such a complex again satisfies (2). Throughout
the paper, we assume that such a polyhedral complex with underlying set $\text{Trop}(X)$ has been chosen. We refer to its faces and facets as faces and facets of $\text{Trop}(X)$, and refine the complex as necessary.

The tropicalization of a closed subscheme of a torus torsor over $K$ is well-
defined up to translation by $N_G$. In particular, if the value group is zero,
then the tropicalization is well-defined, and is the underlying set of a rational
polyhedral fan. An important special case is the tropicalization of an initial
degeneration. The valuation $\nu$ induces the trivial valuation on the residue
field $k$, and there is a natural identification of $\text{Trop}(X_w)$ with the star of $w$ in $\text{Trop}(X)$ [Spe05, Proposition 2.2.3]. This star is, roughly speaking, the fan one sees looking out from $w$ in $\text{Trop}(X)$; it is constructed by translating $\text{Trop}(X)$ so that $w$ is at the origin and taking the cones spanned by faces of $\text{Trop}(X)$ that contain $w$.

2.3. **Tropical multiplicities.** If $w$ and $w'$ are points of $N_G$ in the relative
interior of the same cone $\sigma$, then $X_w$ and $X_w'$ are $T_k$-affinely equiva-

lent. In particular, they are isomorphic as schemes, so the sum of the multiplicities of their irreducible components are equal. By Theorem [A.4], the sum of the multiplicities of the irreducible components of the initial degeneration at $w$ is invariant under extensions of valued fields, for any $w$ in $N_R$. Since any point in $N_G$ becomes rational over the value group after a suitable extension, it follows that this sum is independent of the choice of $w$ in the relative interior of $\sigma$. We will be concerned with this sum only in the case where $\sigma$ is a facet. For applications of tropical multiplicities at points that are not in the relative
interior of a facet, see [BPR11].

**Definition 2.3.1.** The tropical multiplicity $m(\sigma)$ of a facet $\sigma$ in $\text{Trop}(X)$ is the sum of the multiplicities of the irreducible components of $X_w$, for $w$ in the relative interior of $\sigma$.

For $w \in N_G$, the multiplicities on the facets of $\text{Trop}(X_w)$ agree with those on the facets of $\text{Trop}(X)$ that contain $w$. This is because the initial degeneration of $X_w$ at a point in the relative interior of a given facet is $T_k$-affinely equivalent to the initial degeneration of $X$ at a point in the relative interior of the corresponding facet of $\text{Trop}(X)$. See [Spe05, Proposition 2.2.3] and [Gub12, Proposition 10.9].

Points in the relative interiors of facets of multiplicity 1 will be particularly
important for our purposes.

**Definition 2.3.2.** A simple point in $\text{Trop}(X)$ is a point in the relative interior of a facet of multiplicity 1.

If $w \in N_G$ is a simple point, then the initial degeneration $X_w$ is isomorphic
to a $d$-dimensional torus [Spe05 Proposition 2.2.4], and initial degenerations
at arbitrary simple points are isomorphic to tori of dimension at most \(d\). The latter follows from the case where \(w\) is in \(N_G\) by choosing a suitable extension of valued fields and applying Theorem A.3. In particular, the initial degeneration at an arbitrary simple point is smooth.

The importance of simple points, roughly speaking, is that intersections of tropicalizations of closed subvarieties of \(X\) at simple points behave just like intersections in \(N_\mathbb{R}\) of tropicalizations of closed subvarieties of \(T\). See, for instance, the proof of Theorem 1.2 in Section 5.

### 2.4. Minkowski weights and the fan displacement rule.

We briefly review the language of Minkowski weights and their ring structure, in which the product is given by the fan displacement rule. These tools are the basis of both toric intersection theory, as developed in [FS97], and the theory of stable tropical intersections [RGST05, Mik06], discussed in the following sections. We refer the reader to these original papers for further details. Although the theory of Minkowski weights extends to arbitrary complete fans, we restrict to unimodular fans, corresponding to smooth toric varieties, which suffice for our purposes. Fix a complete unimodular fan \(\Sigma\) in \(N_\mathbb{R}\).

**Definition 2.4.1.** A **Minkowski weight** of codimension \(j\) on \(\Sigma\) is a function \(c\) that assigns an integer \(c(\sigma)\) to each codimension \(j\) cone \(\sigma\) in \(\Sigma\), and satisfies the following balancing condition at every codimension \(j + 1\) cone.

**Balancing Condition.** Let \(\tau\) be a cone of codimension \(j + 1\) in \(\Sigma\). Say \(\sigma_1, \ldots, \sigma_r\) are the codimension \(j\) cones of \(\Sigma\) that contain \(\tau\), and let \(v_i\) be the primitive generator of the image of \(\sigma_i\) in \(N_\mathbb{R}/\text{span}(\tau)\). Then \(c\) is balanced at \(\tau\) if

\[
c(\sigma_1)v_1 + \cdots + c(\sigma_r)v_r = 0
\]

in \(N_\mathbb{R}/\text{span}(\tau)\).

Tropicalizations are one interesting source of Minkowski weights. If \(X\) is a closed subscheme of pure codimension \(j\) in \(T\) and \(\text{Trop}(X)\) is a union of cones in \(\Sigma\), then the function \(c\) given by

\[
c(\sigma) = \begin{cases} m(\sigma) & \text{if } \sigma \subset \text{Trop}(X), \\ 0 & \text{otherwise}. \end{cases}
\]

satisfies the balancing condition, and hence is a Minkowski weight of codimension \(j\). This correspondence between tropicalizations and Minkowski weights is compatible with intersection theory in the toric variety \(Y(\Sigma)\). See Section 2.7 below.

We write \(\text{Mink}^j(\Sigma)\) for the group of Minkowski weights of codimension \(j\) on \(\Sigma\). The direct sum

\[
\text{Mink}^j(\Sigma) = \bigoplus_j \text{Mink}^j(\Sigma)
\]

naturally forms a graded ring, whose product is given by the fan displacement rule, as follows. Let \(v\) be a vector in \(N_\mathbb{R}\) that is sufficiently general so that for any two cones \(\sigma\) and \(\sigma'\) in \(\Sigma\), the displaced cone \(\sigma' + v\) meets \(\sigma\) properly. The
set of such vectors is open and dense in $N_R^\mathbb{R}$; it contains the complement of a finite union of linear spaces. For each cone $\sigma$ in $\Sigma$, let $N_{\sigma}$ denote the sublattice of $N$ generated by $\sigma \cap N$. Note that if $\sigma$ intersects the displaced cone $\sigma' + v$ then $\sigma$ and $\sigma'$ together span $N_R^\mathbb{R}$, so the index $[N : N_{\sigma} + N_{\sigma'}]$ is finite.

**Fan Displacement Rule.** The product $c \cdot c'$ is the Minkowski weight of codimension $j + j'$ given by

$$(c \cdot c')(\tau) = \sum_{\sigma, \sigma'} [N : N_{\sigma} + N_{\sigma'}] \cdot c(\sigma) \cdot c'(\sigma'),$$

where the sum is over cones $\sigma$ and $\sigma'$ containing $\tau$, of codimensions $j$ and $j'$ respectively, such that the intersection $\sigma \cap (\sigma' + v)$ is nonempty.

The balancing conditions on $c$ and $c'$ ensure that the product is independent of the choice of displacement vector $v$, and also satisfies the balancing condition.

2.5. Toric intersection theory. We briefly recall the basics of intersection theory in smooth complete toric varieties. As in the previous section, we fix a complete unimodular fan $\Sigma$ in $N_R^\mathbb{R}$. Let $Y(\Sigma)$ be the associated smooth complete toric variety. A codimension $j$ cycle in $Y(\Sigma)$ is a formal sum of codimension $j$ closed subvarieties, with integer coefficients, and we write $A^j(Y(\Sigma))$ for the Chow group of codimension $j$ cycles modulo rational equivalence. Let $X$ be a closed subscheme of pure codimension $j$ in $Y(\Sigma)$. We write $[X]$ for the class it represents in $A^j(Y(\Sigma))$: by definition, this is the sum over the irreducible components $Z$ of $X$ of the length of $X$ along $Z$ times the class $[Z]$.

Intersection theory gives a natural ring structure on the direct sum

$$A^\ast(Y(\Sigma)) = \bigoplus_j A^j(Y(\Sigma)).$$

If $X$ and $X'$ have complementary dimension, which means that $\dim X + \dim X' = \dim Y(\Sigma)$, then the product $[X] \cdot [X']$ is a zero-dimensional cycle class. The degree of this class, which is the sum of the coefficients of any representative cycle, is denoted $\deg([X] \cdot [X'])$. In particular, if $\sigma$ is a cone of codimension $j$ in $\Sigma$, with $V(\sigma)$ the associated closed $T$-invariant subvariety, then $X$ and $V(\sigma)$ have complementary dimension. We write $c_X$ for the induced function on codimension $j$ cones, given by

$$c_X(\sigma) = \deg([X] \cdot [V(\sigma)]).$$

The fact that the intersection product respects rational equivalence ensures that $c_X$ satisfies the balancing condition and is therefore a Minkowski weight of codimension $j$ on $\Sigma$. The main result of [FS97] then says that there is a natural isomorphism of rings

$$A^\ast(Y(\Sigma)) \xrightarrow{\sim} \text{Mink}^\ast(\Sigma)$$

taking the Chow class $[X]$ to the Minkowski weight $c_X$. The theory of refined intersections says that the product $[X] \cdot [X']$ of the classes of two pure-dimensional closed subschemes of a smooth variety $Y$ is not only
a well-defined cycle class in that smooth variety, but also the Gysin push-forward of a well-defined cycle class in the intersection $X \cap X'$. If $X$ and $X'$ meet properly, then this class has codimension zero, and hence can be written uniquely as a formal sum of the components of $X \cap X'$. The coefficient of a component $Z$ of the intersection is called the intersection multiplicity of $X$ and $X'$ along $Z$, and is denoted $i(Z; X, X')$. It is a positive integer, less than or equal to the length of the scheme theoretic intersection of $X$ and $X'$ along the generic point of $Z$, by [Ful98, Proposition 7.1], and can be computed by Serre’s alternating sum formula

$$i(Z; X, X') = \sum_j (-1)^j \text{length}_{\mathcal{O}_Y} \text{Tor}_j^{\mathcal{O}_Y, \mathcal{O}_Z}(\mathcal{O}_X, \mathcal{O}_{X'}).$$

See [Ful98, Chapter 8] for further details on the theory of refined intersections in smooth varieties.

**Definition 2.5.1.** If $X$ and $X'$ meet properly in a smooth variety $Y$, then the refined intersection cycle is

$$X \cdot X' = \sum_Z i(Z; X, X'; Y) Z,$$

where the sum is over all irreducible components $Z$ of $X \cap X'$.

We will consider such refined intersections inside the torus $T$, as well as in its smooth toric compactifications. Finally, it is useful to work not only with tropicalizations of closed subschemes of $T$, but also with tropicalizations of these refined intersection cycles.

**Definition 2.5.2.** Let $a_1Z_1 + \cdots + a_rZ_r$ be a pure-dimensional cycle in $T$, with positive integer coefficients $a_i$. The tropicalization of this cycle is the union

$$\text{Trop}(a_1Z_1 + \cdots + a_rZ_r) = \text{Trop}(Z_1) \cup \cdots \cup \text{Trop}(Z_r)$$

with multiplicities

$$m_{Z_i}(\tau) = a_1m_{Z_i}(\tau) + \cdots + a_rm_{Z_i}(\tau),$$

where $m_{Z_i}(\tau)$ is defined to be zero if $\text{Trop}(Z_i)$ does not contain $\tau$.

2.6. **Stable tropical intersections.** Many papers have introduced and studied analogues of intersection theory in tropical geometry, including [AK10, BldM11, Kat12, Mik06, RGST05, Tab08]. Here we are interested in only one basic common feature of these theories, the stable tropical intersection, and its compatibility with intersection theory in toric varieties. The key point is that stable tropical intersections are defined combinatorially, depending only on the polyhedral geometry of the tropicalizations, by a local displacement rule. Compatibility with the multiplication rule for Minkowski weights and algebraic intersections of generic translates is discussed in the following section.

Given closed subschemes $X$ and $X'$ of $T$, of pure codimensions $j$ and $j'$, respectively, the stable intersection $\text{Trop}(X) \cdot \text{Trop}(X')$ is a polyhedral complex
of pure codimension $j + j'$, with support contained in the set-theoretic intersection $\text{Trop}(X) \cap \text{Trop}(X')$, and with appropriate multiplicities on its facets, satisfying the balancing condition. Many different closed subschemes of $T$ may have the same tropicalization, but the stable intersection depends only on the tropicalizations.

Suppose the valuation is trivial, so $\text{Trop}(X)$ and $\text{Trop}(X')$ are fans, and let $\tau$ be a face of $\text{Trop}(X) \cap \text{Trop}(X')$ of codimension $j + j'$ in $\mathbb{N}$. The tropical intersection multiplicity of $\text{Trop}(X)$ and $\text{Trop}(X')$ along $\tau$, is given by

$$i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) = \sum_{\sigma, \sigma'} [N : N_{\sigma} + N_{\sigma'}] \cdot m(\sigma) \cdot m(\sigma'),$$

where $\sigma$ and $\sigma'$ are facets of $\text{Trop}(X)$ and $\text{Trop}(X')$, respectively, such that the intersection $\sigma \cap (\sigma' + v)$ is nonempty, and $v \in \mathbb{N}$ is a fixed generic displacement vector, as in Section 2.4.

In the general case, where the valuation may be nontrivial, the tropical intersection multiplicities are defined similarly, by a local displacement rule, as follows. Let $\tau$ be a face of codimension $j + j'$ in $\text{Trop}(X) \cap \text{Trop}(X')$, and for each face $\sigma$ containing $\tau$, let $N_{\sigma}$ be the sublattice of $\mathbb{N}$ parallel to the affine span of $\sigma$.

**Definition 2.6.1.** The tropical intersection multiplicity $i(\tau, \text{Trop}(X) \cdot \text{Trop}(X'))$ is given by the local displacement rule

$$i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) = \sum_{\sigma, \sigma'} [N : N_{\sigma} + N_{\sigma'}] \cdot m(\sigma) \cdot m(\sigma'),$$

where $v \in \mathbb{N}$ is a generic displacement vector, as above, and the sum is over facets $\sigma$ and $\sigma'$ of $\text{Trop}(X)$ and $\text{Trop}(X')$, respectively, such that the intersection $\sigma \cap (\sigma' + \epsilon v)$ is nonempty for $\epsilon$ sufficiently small and positive.

**Definition 2.6.2.** The stable tropical intersection, denoted $\text{Trop}(X) \cdot \text{Trop}(X')$, is the union of those faces $\tau$ such that $i(\tau, \text{Trop}(X) \cdot \text{Trop}(X'))$ is positive, weighted by their tropical intersection multiplicities.

2.7. **Compatibility of toric and stable tropical intersections.** As discussed in Section 2.4, tropical multiplicities satisfy the balancing condition, and hence give Minkowski weights. The stable tropical intersections are compatible with multiplication of Minkowski weights, and algebraic intersections of generic translates in the torus, as follows.

Let $X$ be a closed subscheme of pure codimension $j$ in $T$. Suppose the valuation is trivial, so $\text{Trop}(X)$ is a fan. After subdividing, we may assume this fan is unimodular, and extends to a complete unimodular fan $\Sigma$ in $\mathbb{N}$. Let $Y(\Sigma)$ be the associated smooth complete toric variety, and let $\overline{X}$ be the closure of $X$ in $Y(\Sigma)$. Because $\Sigma$ contains $\text{Trop}(X)$ as a subfan, the closure $\overline{X}$ meets every torus orbit $O_\tau$ properly, and the intersection is nonempty if and only if $\tau$ is contained in $\text{Trop}(X)$. This follows from the basic properties of extended tropicalizations, given in [Pay09a, Proposition 3.7].
Remark 2.7.1. Indeed the closure $\overline{X}$ is proper and meets every torus orbit properly if and only if $\Sigma$ contains a subfan whose support is $\text{Trop}(X)$. See [Gub12, Section 14].

Furthermore, for any facet $\sigma$ of $\text{Trop}(X)$, the tropical multiplicity $m(\sigma)$ is equal to the intersection number of the closure $\overline{X}$ with the torus invariant subvariety corresponding to $\sigma$, that is

$$m(\sigma) = \deg([X] \cdot V(\sigma)).$$

In other words,

$$c_X(\sigma) = \begin{cases} m(\sigma) & \text{if } \sigma \subset \text{Trop}(X), \\ 0 & \text{otherwise.} \end{cases}$$

See [ST08, Lemma 3.2] for the case of a tropical compactification and [KP11, Lemma 2.3] for the general case.

Lemma 2.7.2. Suppose the valuation is trivial, and let $X$ be a pure-dimensional closed subscheme of $T$. Then the tropicalization $\text{Trop}(X)$ is equal to the tropicalization of the fundamental cycle $\text{Trop}([X])$.

Recall that the multiplicities of facets in the tropicalization of a cycle are defined by linearity; see Definition 2.5.2.

Proof. The tropicalizations agree set-theoretically, because both are the closures of the images of $X(L)$ where $L/K$ is an algebraically closed valued extension field with nontrivial valuation. The multiplicities agree because the tropical multiplicity of a facet $\sigma$ in $\text{Trop}(X)$ is $m(\sigma) = \deg([X] \cdot V(\sigma))$, and this degree is a linear function of the cycle $[X]$. □

Remark 2.7.3. The tropicalization of a pure-dimensional closed subscheme of $T$ also agrees with the tropicalization of its underlying cycle in the case of a nontrivial valuation, but this requires considerably more work. This is deduced from our general results on intersection multiplicities over valuation rings of rank 1 in Corollary 4.4.9. See also [Gub12, Section 13] for an analytic proof using nonarchimedean GAGA and Gubler’s theory of cycles and Cartier divisors on affinoid spaces.

Now consider the general case, where the valuation may be nontrivial, and let $w \in N_G$ be a point of $\text{Trop}(X)$. After subdividing, we may assume that the star of $w$ in $\text{Trop}(X)$ is unimodular, and extends to a complete unimodular fan $\Sigma$. We write $X_w$ for the closure of $X_w$ in $Y(\Sigma)$.

Definition 2.7.4. If $\sigma$ is a facet of $\text{Trop}(X)$ that contains $w$ then let

$$\sigma_w = \mathbb{R}_{\geq 0}(\sigma - w)$$

be the corresponding cone in $\Sigma$.

Recall that the star of $w$ in $\text{Trop}(X)$ is the tropicalization of the initial degeneration $X_w$ and, as discussed in Section 2.4.3 the initial degeneration of $X_w$ at a point in the relative interior of $\sigma_w$ is $T_k$-affinely equivalent to the initial degeneration of $X$ at a point of $N_G$ in the relative interior of $\sigma$. In particular, the
tropical multiplicity \( m(\sigma_w) \) in \( \text{Trop}(X_w) \) is equal to the tropical multiplicity \( m(\sigma) \).

**Lemma 2.7.5.** Let \( \gamma \) be a face of codimension \( j \) in \( \Sigma \). Then

\[
\langle \alpha_{X_w} \cdot \alpha_{X_w} \rangle(\gamma) = \begin{cases} 
    m(\sigma) & \text{if } \gamma = \sigma_w, \text{ for some } \sigma \subset \text{Trop}(X), \\
    0 & \text{otherwise}.
\end{cases}
\]

**Proof.** This follows from the case of the trivial valuation and the fact that the tropical multiplicity \( m(\sigma) \) in \( \text{Trop}(X) \) is equal to the tropical multiplicity \( m(\sigma_w) \) in \( \text{Trop}(X_w) \).

This correspondence between tropical multiplicities and Minkowski weights is compatible with intersections; stable tropical intersections corresponding to products of Minkowski weights, as follows. Suppose \( X' \) is a closed subscheme of pure codimension \( j' \) in \( T \), and \( \tau \) is a face of codimension \( j + j' \) in \( \text{Trop}(X) \cap \text{Trop}(X') \) that contains \( w \). We may assume that the star of \( w \) in \( \text{Trop}(X') \) is a subfan of the unimodular fan \( \Sigma \), and we write \( X_w \) for the closure of \( X_w' \) in \( Y(\Sigma) \).

**Proposition 2.7.6.** Let \( \gamma \) be a cone of codimension \( j + j' \) in \( \Sigma \). Then

\[
\langle \alpha_{X_w} \cdot \alpha_{X_w} \rangle(\gamma) = \begin{cases} 
    i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) & \text{if } \gamma = \tau_w, \\
    0 & \text{otherwise},
\end{cases}
\]

**Proof.** Suppose \( \gamma = \tau_w \). We claim that the displacement rules giving \( \langle \alpha_{X_w} \cdot \alpha_{X_w} \rangle(\gamma) \) and \( i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) \) agree, term by term. If \( \sigma \) and \( \sigma' \) are facets of \( \text{Trop}(X) \) and \( \text{Trop}(X') \), respectively, that contain \( \tau \), then \( \sigma \) meets \( \sigma' + e\nu \) for small positive \( \nu \) if and only if \( \sigma_w \) meets \( \sigma_w' + v \). Then we have an equality of sums of displacements in the displacement rules,

\[
[N : N_{\sigma} + N_{\sigma'}] \cdot m(\sigma) \cdot m(\sigma') = [N : N_{\sigma_w} + N_{\sigma_w'}] \cdot \alpha_{X_w}(\sigma_w) \cdot \alpha_{X_w}(\sigma_w'),
\]

because \( N_{\sigma} + N_{\sigma'} \), \( m(\sigma) \), and \( m(\sigma') \) are equal to \( N_{\sigma_w} + N_{\sigma_w'} \), \( \alpha_{X_w}(\sigma_w) \), and \( \alpha_{X_w}(\sigma_w') \), respectively. The Minkowski weights \( \alpha_{X_w} \) and \( \alpha_{X_w} \) vanish on cones that do not come from facets of \( \text{Trop}(X) \) and \( \text{Trop}(X') \), respectively. Therefore, all nonzero summands in the displacement rules are accounted for in the equality above, and the proposition follows.

It follows that the star of \( w \) in the stable tropical intersection \( \text{Trop}(X) \cdot \text{Trop}(X') \) is exactly the union of the faces of \( \Sigma \) on which \( \langle \alpha_{X_w} \cdot \alpha_{X_w} \rangle \) is positive, with multiplicities given by this product of Minkowski weights.

We now return to the case of the trivial valuation. As above, we choose \( \Sigma \) to be a complete unimodular fan in \( N_{\mathbb{R}} \) that contains \( \text{Trop}(X) \) and \( \text{Trop}(X') \) as subfans.

**Proposition 2.7.7.** Suppose the valuation is trivial. Assume \( X \) meets \( X' \) properly in \( T \) and, moreover, \( X \cap V(\sigma) \) meets \( X' \cap V(\sigma) \) properly in \( V(\sigma) \) for each \( \sigma \in \Sigma \).
Then, for any face $\tau$ of $\text{Trop}(X) \cap \text{Trop}(X')$ of codimension $j + j'$ in $\mathbb{R}$, the tropical intersection multiplicity along $\tau$ is equal to the weighted sum of multiplicities of tropicalizations

$$i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) = \sum Z i(Z, X \cdot X'; T) m_Z(\tau),$$

where the sum is over components $Z$ of $X \cap X'$ such that $\text{Trop}(Z)$ contains $\tau$, with tropical multiplicity $m_Z(\tau)$.

**Proof.** First of all, by the preceding proposition,

$$i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) = (\text{c}_{X} \cdot \text{c}_{X'})(\tau).$$

Next, the isomorphism from $\text{Mink}^*(\Sigma)$ to $A^*(Y(\Sigma))$ identifies $(\text{c}_{X} \cdot \text{c}_{X'})(\tau)$ with the intersection number $\deg([X] \cdot [X'] \cdot V(\tau))$. By hypothesis, $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly in $Y(\Sigma)$, and the generic point of each component of the intersection lies in $T$, so we have the refined intersection

$$\text{Trop}(X) \cdot \text{Trop}(X') = \sum Z i(Z, X \cdot X'; T)[Z],$$

where the sum is taken over components $Z$ of $X \cap X'$. Therefore,

$$\deg([X] \cdot [X'] \cdot [V(\tau)]) = \sum Z i(Z, X \cdot X'; T) \deg([Z] \cdot [V(\tau)]).$$

Now, since $\text{Trop}(X)$ and $\text{Trop}(X')$ are subfans of $\Sigma$, the closures $\text{Trop}(X)$ and $\text{Trop}(X')$ meet $V(\sigma)$ properly in $Y(\Sigma)$, for all $\sigma \in \Sigma$. Furthermore, by hypothesis, $\text{Trop}(X) \cap V(\sigma)$ and $\text{Trop}(X') \cap V(\sigma)$ meet properly in $V(\sigma)$ for all $\sigma$. It follows that $Z$ meets all orbits of $Y(\Sigma)$ properly. Therefore $\text{Trop}(Z)$ is a union of cones of codimension $j + j'$ in $\Sigma$ (see Remark 2.7.1) and the intersection number $\deg([Z] \cdot [V(\tau)])$ is equal to the tropical multiplicity $m_Z(\tau)$.

The following proposition may be seen as a geometric explanation for the existence of well-defined stable tropical intersections, in the case of the trivial valuation; they are exactly the tropicalizations of intersections of generic translates in the torus.

**Proposition 2.7.8.** Suppose the valuation is trivial, and let $t$ be a general point in $T$. Then $X$ meets $tX'$ properly and the tropicalization of the intersection scheme $X \cap tX'$ is exactly the stable tropical intersection $\text{Trop}(X) \cdot \text{Trop}(X')$.

**Proof.** First, an easy Bertini argument shows that $X$ meets $tX'$ properly when $t$ is general. Furthermore, since $Y(\Sigma)$ is smooth, subadditivity of codimension says that every component of $\text{Trop}(X) \cap t\text{Trop}(X')$ has codimension at most $j + j'$. Say $Z$ is a component of this intersection, and let $O_\sigma$ be the torus orbit in $Y(\Sigma)$ that contains the generic point of $Z$. Now, $\text{Trop}(X) \cap t\text{Trop}(X')$ meets $O_\sigma$ properly, since their tropicalizations are subfans of $\Sigma$, and $\text{Trop}(X) \cap O_\sigma$ meets $t\text{Trop}(X') \cap O_\sigma$ properly in $O_\sigma$, since $t$ is general. Therefore $Z$ has codimension $j + j' + \text{codim} O_\sigma$, and hence $O_\sigma$ must be the dense torus $T$. 

*Documenta Mathematica 18 (2013) 121–175*
Next, the Cohen-Macaulay locus in an excellent scheme is open [Gro65, Scholium 7.8.3(iv)], and contains every generic point, so the Cohen-Macaulay loci of \(X\) and \(X'\) are open and dense. Therefore, since \(t\) is general, the generic point of every component of \(X \cap tX'\) must lie in the Cohen-Macaulay loci of both \(X\) and \(X'\). Then, by [Ful98, Example 8.2.7], the intersection multiplicity \(i(Z, X \cdot tX'; T)\) is equal to the length of the intersection of \(X\) and \(tX'\) along the generic point of \(Z\). Proposition 2.7.7 then says that

\[
i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) = \sum_Z \text{length}(\mathcal{O}_{Z, X \cap tX'}) m_Z(\tau).
\]

Now the right hand side is the multiplicity of \(\tau\) in the tropicalization of the cycle \([X \cap tX']\), and the proposition follows, by Lemma 2.7.2. \(\square\)

**Remark 2.7.9.** We apply this proposition in the proof of Lemma 3.3.3. All that is needed for that application is the standard fact that the intersection multiplicity at a point of proper intersection is strictly positive. In particular, the equality of the intersection multiplicity \(i(Z, X \cdot tX'; T)\) with the length of the scheme-theoretic intersection of \(X\) and \(X'\) along \(Z\) is not needed for the main lifting theorems.

In Section 5, we extend Proposition 2.7.8 to the general case, where the valuation may be nontrivial, as an application of tropical lifting theorems. See Theorem 5.3.3.

### 3. The trivial valuation case

The first step in our approach to tropical lifting theorems is to understand how the intersection of \(\text{Trop}(X)\) and \(\text{Trop}(X')\) near a point \(w\) relates to the intersection of the initial degenerations \(X_w\) and \(X'_w\). If \(w\) is in \(N_G\) then the stars of \(w\) in \(\text{Trop}(X)\) and \(\text{Trop}(X')\) are the tropicalizations of \(X_w\) and \(X'_w\), respectively, with respect to the trivial valuation, so this amounts to understanding how tropicalization relates to intersections when the tropicalizations meet properly and the valuation is trivial. Our proof that tropicalization commutes with intersection in this case uses both the theory of extended tropicalizations as projections of nonarchimedean analytifications and a new result about the support of stable intersections of tropicalizations (Lemma 3.3.3). We prove two necessary lemmas about the topology of extended tropicalizations in the general case, where the valuation is not necessarily trivial, since the arguments are identical to those in the case where the valuation is trivial.

#### 3.1. Topology of extended tropicalizations

The extended tropicalizations of \(Y(\Sigma)\) and its closed subschemes were introduced by Kajiwara [Kaj08], and their basic properties were studied further in Section 3 of [Pay09a], to which we refer the reader for further details. We recall the definition and prove that the extended tropicalization of a closed subscheme \(X\) in which the torus \(T\) is dense is the closure of the ordinary tropicalization of \(X \cap T\). We also characterize the closure in the extended tropicalization \(\text{Trop}(Y(\Sigma))\) of a cone \(\sigma\) in \(\Sigma\).
Let $\Sigma$ be a fan in $\mathbb{N}^\mathbb{R}$, and let $Y(\Sigma)$ be the associated toric variety. Recall that each cone $\sigma$ in $\Sigma$ corresponds to an affine open subvariety $U_\sigma$ of $Y(\Sigma)$ whose coordinate ring is the semigroup ring $K[\sigma^\vee \cap M]$ associated to the semigroup of lattice points in the dual cone $\sigma^\vee$ in $M_\mathbb{R}$. Let $\mathbb{R}$ be the real line extended in the positive direction

$$\mathbb{R} = \mathbb{R} \cup \{+\infty\},$$

which is a semigroup under addition, with identity zero. Then

$$\text{Trop}(U_\sigma) = \text{Hom}(\sigma^\vee \cap M, \mathbb{R})$$

is the space of all semigroup homomorphisms taking zero to zero, with its natural topology as a subspace of $\mathbb{R}^{\sigma^\vee \cap M}$. For each face $\tau$ of $\Sigma$, let $N(\tau)$ be the real vector space $N(\tau) = \mathbb{N}^\mathbb{R}/\text{span}(\tau)$. Then $\text{Trop}(U_\sigma)$ is naturally a disjoint union of the real vector spaces $N(\tau)$, for $\tau \preceq \sigma$, where $N(\tau)$ is identified with the subset of semigroup homomorphisms that are finite exactly on the intersection of $\tau^\perp$ with $\sigma^\vee \cap M$.

The tropicalization $\text{Trop}(Y(\Sigma))$ is constructed by gluing the spaces $\text{Trop}(U_\sigma)$ for $\sigma \in \Sigma$ along the natural open inclusions $\text{Trop}(U_\tau) \subseteq \text{Trop}(U_\sigma)$ for $\tau \preceq \sigma$. It is a disjoint union

$$\text{Trop}(Y(\Sigma)) = \bigcup_{\sigma \in \Sigma} N(\sigma),$$

just as $Y(\Sigma)$ is the disjoint union of the torus orbits $O_\sigma$. Points in $N(\sigma)$ may be seen as weight vectors on monomials in the coordinate ring of $O_\sigma$, and the tropicalization of a closed subscheme $X$ in $Y(\Sigma)$ is defined to be

$$\text{Trop}(X) = \bigcup_{\sigma \in \Sigma} \{w \in N(\sigma) \mid (X \cap O_\sigma)_w \text{ is not empty}\}.$$

In other words, $\text{Trop}(X)$ is the disjoint union of the tropicalizations of its intersections with the $T$-orbits in $Y(\Sigma)$. This space is compact when $\Sigma$ is complete.

**Lemma 3.1.1.** If every component of $X$ meets the dense torus $T$, then $\text{Trop}(X)$ is the closure in $\text{Trop}(Y(\Sigma))$ of the ordinary tropicalization $\text{Trop}(X \cap T)$.

**Proof.** Since tropicalizations are invariant under extensions of valued fields [Pay09a, Proposition 6.1], and the completion of the algebraically closed field $K$ is algebraically closed [BGR84, Proposition 3.4.1.3], we may assume the field $K$ is complete with respect to its valuation.

The tropicalization map from $X(K)$ to $\text{Trop}(Y(\Sigma))$ extends to a proper continuous map on the nonarchimedean analytification $X^\text{an}$ whose image is $\text{Trop}(X)$ [Pay09a, Section 2]. Since the open subset $(X \cap T)^\text{an}$ is dense in $X^\text{an}$ [Ber90, Corollary 3.4.5] and maps onto $\text{Trop}(X \cap T)$, the extended tropicalization $\text{Trop}(X)$ is contained in the closure of $\text{Trop}(X \cap T)$. The lemma follows, since $\text{Trop}(X)$ is closed. $\square$
Lemma 3.1.2. Let $\sigma$ and $\tau$ be faces of $\Sigma$, and let $\overline{\sigma}$ be the closure of $\sigma$ in $\text{Trop}(Y(\Sigma))$. Then

$$\overline{\sigma} \cap N(\tau) = \sigma,$$

if $\sigma$ contains $\tau$, and $\overline{\sigma}$ is disjoint from $N(\tau)$ otherwise.

Proof. Let $v$ be a point in $N(\tau)$, and let $\pi : \text{Trop}(U_\tau) \to N(\tau)$ be the continuous map that restricts to the canonical linear projections from $N(\gamma)$ onto $N(\tau)$ for $\gamma \preceq \tau$. If $v$ is not in $\sigma$, then the preimage under $\pi$ of $N(\tau) \setminus \sigma$ is an open neighborhood of $v$ that is disjoint from $\sigma$, so $v$ is not in $\overline{\sigma}$. Similarly, if $\sigma$ does not contain $\tau$ then $\pi^{-1}(N(\tau)) \setminus \sigma$ is an open neighborhood of $v$ that is disjoint from $\sigma$.

It remains to show that $\overline{\sigma} \cap N(\tau)$ contains $\sigma$ in the case where $\sigma$ contains $\tau$. Suppose $v$ is in $\sigma$, and let $w$ be a point in $\sigma$ that projects to $v$. Let $w'$ be a point in the relative interior of $\tau$. Then $w + \mathbb{R}_{>0}w'$ is a path in $\sigma$ whose limit is $v$, so $v$ is in $\overline{\sigma}$. □

3.2. Tropical subadditivity. Although not strictly necessary for the main results of the paper, it is helpful to know that codimension is subadditive for intersections of tropicalizations in $N_{\mathbb{R}}$, so $\text{Trop}(X)$ and $\text{Trop}(X')$ never intersect in less than the expected dimension. The proof is by a diagonal projection argument, similar to the proof of Lemma 3.1.2 below.

Proposition 3.2.1. Let $X$ and $X'$ be closed subschemes of pure codimension $j$ and $j'$, respectively, in $T$. If $\text{Trop}(X) \cap \text{Trop}(X')$ is nonempty then it has codimension at most $j + j'$ at every point.

Proof. Let $w$ be a point in $\text{Trop}(X) \cap \text{Trop}(X')$. Then, in a neighborhood of $w$, the intersection can be identified with a neighborhood of zero in $\text{Trop}(X_w) \cap \text{Trop}(X'_w)$. Therefore, replacing $X$ and $X'$ by their initial degenerations at $w$, we may assume the valuation is trivial, and it will suffice to show that the global codimension of $\text{Trop}(X) \cap \text{Trop}(X)$ is at most $j + j'$.

Suppose the valuation is trivial. Let $T'$ be the quotient of $T \times T$ by the diagonal subtorus, with $\pi : X \times X' \to T'$ the induced projection, and $p : \text{Trop}(X) \times \text{Trop}(X') \to N'_{\mathbb{R}}$ the tropicalization of $\pi$. Since $\text{Trop}(X)$ meets $\text{Trop}(X')$, the point zero is in the image of $p$, which is exactly the tropicalization of the closure of the image of $\pi$.

Let $y$ be a point in the image of $\pi$. The fiber $\pi^{-1}(y)$ has dimension at least $\dim X + \dim X' - \dim T'$ and, since the valuation is trivial, its tropicalization is contained in $p^{-1}(0)$. Since $p^{-1}(0)$ is naturally identified with $\text{Trop}(X) \cap \text{Trop}(X')$, it follows that the tropical intersection has dimension at least $\dim X' + \dim X - \dim T$, as required. □
3.3. Lower bounds on multiplicities. For the remainder of this section, we assume the valuation $\nu$ is trivial. Let $X$ and $X'$ be closed subschemes of pure codimension $j$ and $j'$ in $T$, respectively. After subdividing the tropicalizations, we choose a complete unimodular fan $\Sigma$ in $\mathbb{N}_\mathbb{R}$ such that each face of $\text{Trop}(X)$ and each face of $\text{Trop}(X')$ is a face of $\Sigma$. We write $\overline{X}$ and $\overline{X'}$ for the closures of $X$ and $X'$ in the smooth complete toric variety $Y(\Sigma)$.

Our goal is to prove the following version of Theorem 1.1 in the special case of the trivial valuation, with lower bounds on the multiplicities of the facets in $\text{Trop}(X \cap X')$. These lower bounds are extended to the general case in Section 5.

**Theorem 3.3.1.** Suppose $\nu$ is trivial and $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly. Then

$$\text{Trop}(X \cap X') = \text{Trop}(X) \cap \text{Trop}(X'),$$

and the multiplicity of any facet $\tau$ is bounded below by the tropical intersection multiplicity

$$m_{X \cap X'}(\tau) \geq i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')).$$

Furthermore, the tropical intersection multiplicity is equal to the weighted sum of algebraic intersection multiplicities

$$i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) = \sum_Z i(Z, X \cdot X'; Y(\Sigma)) m_Z(\tau),$$

where the sum is over components $Z$ of $X \cap X'$ whose tropicalizations contain $\tau$.

Given the compatibility of toric and stable tropical intersections in Proposition 2.7.7, the main step in the proof of Theorem 3.3.1 is showing that every component of $\overline{X} \cap \overline{X'}$ is the closure of a component of $X \cap X'$. This step can fail spectacularly when $\text{Trop}(X)$ and $\text{Trop}(X')$ do not meet properly. In such cases, $\overline{X} \cap \overline{X'}$ may contain components of larger than expected dimension, even if $X$ and $X'$ meet properly in $T$, and it is difficult to predict what $\text{Trop}(X \cap X')$ will look like.

**Proposition 3.3.2.** Suppose $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly, and let $\tau$ be a face of $\Sigma$. Then

1. The tropicalizations of $\overline{X} \cap O_\tau$ and $\overline{X'} \cap O_\tau$ meet properly in $N(\tau)$.
2. The subschemes $\overline{X} \cap V(\tau)$ and $\overline{X'} \cap V(\tau)$ meet properly in $V(\tau)$.
3. The closures $\overline{X}$ and $\overline{X'}$ meet properly in $Y(\Sigma)$, and every component of $\overline{X} \cap \overline{X'}$ is the closure of a component of $X \cap X'$.

**Proof.** First, if the tropicalizations of $\overline{X} \cap O_\tau$ and $\overline{X'} \cap O_\tau$ meet properly in $N(\tau)$, then the intersections themselves must meet properly in $O_\tau$, by part (1) of the foundational theorem in Section 2.2. If this holds for all cones $\sigma$ containing $\tau$, then $\overline{X} \cap V(\tau)$ meets $\overline{X'} \cap V(\tau)$ properly in $V(\tau)$. Therefore (1) implies (2).
Now, since $\text{Trop}(X)$ and $\text{Trop}(X')$ are subfans of $\Sigma$, the closures $\overline{X}$ and $\overline{X}'$ meet all orbits of $Y(\Sigma)$ properly (see Section 2.7). Hence, if (2) holds then every component of $\overline{X} \cap \overline{X}' \cap V(\tau)$ has codimension $j + j'$ in $V(\tau)$. In particular, if $O_{\tau}$ is not the dense torus $T$, then $V(\tau)$ does not contain a component of $\overline{X} \cap \overline{X}'$. If this holds for all $\tau$, then every component of $\overline{X} \cap \overline{X}'$ meets the dense torus $T$. This shows that (2) implies (3). We now prove (1).

The tropicalization of $\overline{X} \cap O_{\tau}$ is the intersection of the extended tropicalization $\text{Trop}(\overline{X})$ with $N(\tau)$ and, by Lemma 3.1.1, this extended tropicalization is the closure of $\text{Trop}(X)$. Therefore, by Lemma 3.1.2, $\text{Trop}(\overline{X} \cap O_{\tau})$ is the union of the projected faces $\sigma_{\tau}$ such that $\sigma$ is in $\text{Trop}(X)$ and contains $\tau$. Similarly, $\text{Trop}(\overline{X} \cap O_{\tau})'$ is the union of the projected faces $\sigma'_{\tau}$ such that $\sigma'$ is in $\text{Trop}(X')$ and contains $\tau$. Since the intersection of $\sigma_{\tau}$ and $\sigma'_{\tau}$ is exactly $(\sigma \cap \sigma')_{\tau}$, the codimension of $\sigma_{\tau} \cap \sigma'_{\tau}$ in $N(\tau)$ is equal to the codimension of $\sigma \cap \sigma'$ in $N_{\mathbb{R}}$. In particular, if $\text{Trop}(X)$ and $\text{Trop}(X')$ meet properly at $\tau$, then $\text{Trop}(\overline{X} \cap O_{\tau})$ and $\text{Trop}(\overline{X}' \cap O_{\tau})$ meet properly in $N(\tau)$, as required. □

The following lemma says that if $\text{Trop}(X)$ meets $\text{Trop}(X')$ properly then the support of their stable tropical intersection is equal to their set-theoretic intersection.

**Lemma 3.3.3.** Suppose $\text{Trop}(X)$ and $\text{Trop}(X')$ meet properly and let $\tau$ be a facet of their intersection. Then the tropical multiplicity $i(\tau, \text{Trop}(X) \cdot \text{Trop}(X'))$ is strictly positive.

**Proof.** First, we reduce to the case where $X$ and $X'$ have complementary dimension. Recall that the tropical multiplicity $m(\sigma)$ of a facet $\sigma$ in $\text{Trop}(X)$ is equal to the algebraic intersection number $\deg(\overline{X} \cdot V(\sigma))$ in $Y(\Sigma)$, as discussed in Section 2.7. and the intersection number $\deg(\overline{X} \cdot \overline{X}' \cdot V(\tau))$ in $Y(\Sigma)$ is equal to the intersection number $\deg ((\overline{X} \cdot V(\tau)) \cdot (\overline{X}' \cdot V(\tau)))$ in $V(\tau)$ [Ful98, Example 8.1.10], because $V(\tau)$ is smooth. Because $\text{Trop}(X)$ and $\text{Trop}(X')$ are subfans of $\Sigma$, the closures $\overline{X}$ and $\overline{X}'$ in $Y(\Sigma)$ meet all torus orbits properly, and it follows that the tropical intersection multiplicity $i(\tau, \text{Trop}(X) \cdot \text{Trop}(X'))$ is equal to the tropical intersection multiplicity at zero of $\text{Trop}(\overline{X} \cdot O_{\tau})$ and $\text{Trop}(\overline{X}' \cdot O_{\tau})$ inside $N(\tau)$. By Proposition 3.3.2 the intersection cycles $\overline{X} \cdot O_{\tau}$ and $\overline{X}' \cdot O_{\tau}$ have complementary dimension in $O_{\tau}$ and their tropicalizations meet properly in $N(\tau)$. The same is then true for each of the reduced components of $\overline{X} \cdot O_{\tau}$ and $\overline{X}' \cdot O_{\tau}$, and it suffices to prove the proposition in this case.

Assume $X$ and $X'$ have complementary dimension. Since their tropicalizations meet properly, by hypothesis, their intersection must be the single point zero. Consider the quotient $T'$ of $T \times T$ by its diagonal subtorus, and the induced projection

$$\pi : X \times X' \to T'.$$
Lifting Tropical Intersections

By Proposition 2.7.8, the tropical intersection number \( \text{Trop}(X) \cdot \text{Trop}(X') \) is equal to the length of \( X \cap tX' \) for a general point \( t \in T \), which is equal to the length of the general fiber of \( \pi \). In particular, to prove the proposition, it is enough to show that \( \pi \) is dominant. Now, since the valuation is trivial, the tropicalization of any fiber of \( \pi \) is contained in the preimage of zero under the induced projection

\[
\text{Trop}(X) \times \text{Trop}(X') \to N^r_x,
\]

which is the single point zero, by hypothesis. In particular, every nonempty fiber of \( \pi \) is zero-dimensional, and hence \( \pi \) is dominant, as required. \( \square \)

We can now easily prove the desired result.

Proof of Theorem 3.3.1. Let \( \tau \) be a facet of \( \text{Trop}(X) \cap \text{Trop}(X') \). By Propositions 2.7.7 and 3.3.2, the tropical intersection multiplicity \( i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) \) is a weighted sum over components of \( X \cap X' \)

\[
i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) = \sum_Z i(Z, X \cdot X'; T) m_Z(\tau).
\]

By Lemma 3.3.3 this multiplicity is strictly positive, so there is at least one component \( Z \) of \( X \cap X' \) such that \( \text{Trop}(Z) \) contains \( \tau \). This proves that \( \tau \) is contained in \( \text{Trop}(X \cap X') \), and hence \( \text{Trop}(X \cap X') \) is equal to \( \text{Trop}(X) \cap \text{Trop}(X') \), set-theoretically.

Finally, to see the inequality on multiplicities, recall that the multiplicity of \( \tau \) in \( \text{Trop}(X \cap X') \) is

\[
m_{X \cap X'}(\tau) = \sum_Z \text{length}_{O_{T, z}}(O_{X \cap X', z}) m_Z(\tau),
\]

by Lemma 2.7.2. The inequality \( m_{X \cap X'}(\tau) \geq i(\tau, \text{Trop}(X) \cdot \text{Trop}(X')) \) follows by comparing the two summation formulas term by term, because \( \text{length}_{O_{T, z}}(O_{X \cap X', z}) \) is at least as large as the intersection multiplicity \( i(Z, X \cdot X'; T) \) \cite{Ful98} Proposition 8.2. \( \square \)

4. Geometry over valuation rings of rank 1

In \cite{Nag66}, Nagata began extending certain results from dimension theory to non-noetherian rings. Here we continue in Nagata’s spirit with a view toward tropical geometry, focusing on valuation rings of rank 1 and showing that these rings have many of the pleasant properties of regular local rings.

Recall that a valuation ring of rank 1 is an integral domain \( R \) whose field of fractions \( K \) admits a nontrivial valuation \( \nu : K^* \to \mathbb{R} \) such that the nonzero elements of \( R \) are exactly the elements of \( K^* \) of nonnegative valuation. It is a local ring with exactly two prime ideals, the zero ideal and the maximal ideal. If \( K \) is algebraically closed or, more generally, if the valuation is not discrete, then \( R \) is not noetherian, because the maximal ideal is not finitely generated.

Throughout this section, \( R \) is a valuation ring of rank 1, with fraction field \( K \) and residue field \( k \). In this section only, we allow the possibility that \( K \)
is not algebraically closed, because the results of this section hold equally for nonclosed fields and the greater generality creates no additional difficulties. We fix 

\[ S = \text{Spec} \, R, \]

and use calligraphic notation for schemes over \( S \). If \( \mathcal{X} \) is a scheme over \( S \), we write \( \mathcal{X}_K = \mathcal{X} \times_S \text{Spec} \, K \) for the generic fiber and \( \mathcal{X}_k = \mathcal{X} \times_S \text{Spec} \, k \) for the special fiber.

A module is flat over \( R \) if and only if it is torsion-free \cite[Exercise 10.2]{Mat89}, so an integral scheme \( \mathcal{X} \) is flat over \( S \) if and only if it is dominant. Furthermore, a scheme that is flat and locally of finite type over an integral scheme is necessarily locally of finite presentation \cite[Theorem 3.4.6 and Corollary 3.4.7]{RG71}. In particular, if \( \mathcal{X} \) is an integral scheme that is surjective and locally of finite type over \( S \) then the special fiber \( \mathcal{X}_k \) is of pure dimension equal to the dimension of the generic fiber \( \mathcal{X}_K \). This follows from \cite[Theorem 12.1.1(i)]{Gro66}, applied to the generic point of each component of the special fiber. We will use these technical facts repeatedly throughout the section, without further mention.

Dimension theory over valuation rings of rank 1 involves some subtleties. For instance, if \( R \) is not noetherian then the formal power series ring \( R[[x]] \) has infinite Krull dimension \cite{Arn73}. Because dimensions are well-behaved under specialization over \( R \), such pathologies are always visible in at least one of the fibers. For \( \mathcal{X} = \text{Spec} \, R[[x]] \), the special fiber is \( \text{Spec} \, k[[x]] \), which is one dimensional, but \( R[[x]] \otimes K \) is the subring of \( K[[x]] \) consisting of formal power series whose coefficients have valuations bounded below. This ring, and hence the generic fiber of \( \text{Spec} \, R[[x]] \), is infinite dimensional. Nevertheless, for a scheme of finite type over \( S \), the generic fiber and special fiber are schemes of finite type over \( K \) and \( k \), respectively, and such pathologies do not occur.

Many of the results of this section have natural generalizations to more general base schemes. See Appendix B for details.

4.1. Subadditivity and lifting over valuation rings of rank 1. Serre famously proved that codimension is subadditive under intersections in regular schemes \cite[Theorem V.3]{Ser65}. In other words, for any irreducible closed subschemes \( X \) and \( X' \) of a regular scheme \( Y \), and any irreducible component \( Z \) of \( X \cap X' \), we have

\[ \text{codim}_Y Z \leq \text{codim}_Y X + \text{codim}_Y X'. \]

The same is then necessarily true for schemes smooth over \( Y \). We extend Serre’s theorem to smooth schemes over rank 1 valuation rings, as follows. Recall that we have fixed \( S = \text{Spec} \, R \), where \( R \) is a valuation ring of rank 1.

**Theorem 4.1.1.** Let \( \mathcal{Y} \) be smooth over \( S \). Then codimension is subadditive under intersection in \( \mathcal{Y} \).

In this respect, valuation rings of rank 1 behave like regular local rings.

**Remark** 4.1.2. Subadditivity of codimension is often used to deform points in families using dimension counting arguments. Such techniques are essential,
for instance, in the theory of limit linear series developed by Eisenbud and Harris [EH86]. Here, we use subadditivity of codimension to lift intersection points from the special fiber to the generic fiber in schemes of finite type over valuation rings of rank 1.

As an application of Theorem 4.1.1 we prove the following.

**Theorem 4.1.3.** Let $Y$ be smooth over $S$ with closed subschemes $X$ and $X'$ of pure codimension $j$ and $j'$, respectively. Suppose the special fibers $X_k$ and $X'_k$ intersect in codimension $j + j'$ at a point $x$ in $Y_k$. Then

1. There is a point in $X_K \cap X'_K$ specializing to $x$.
2. If $x$ is closed then there is a closed point in $X_K \cap X'_K$ specializing to $x$.

More generally, if $k(x)$ denotes the residue field of $x$, there is a point $x' \in X_K \cap X'_K$ specializing to $x$ and satisfying

$$\text{trdeg } k(x')/K = \text{trdeg } k(x)/k.$$

In Section 4.4 we also prove a principle of continuity for intersection multiplicities in families over $S$.

**Remark 4.1.4.** The theorem reduces easily to the case where $X$ and $X'$ are reduced and irreducible and then, using subadditivity, it is not difficult to show that both must meet the generic fiber $Y_K$ and are therefore flat over $S$. Our methods also give a more general lifting theorem for intersections of flat subschemes over an arbitrary base scheme (Theorem 4.3), even when subadditivity fails. Here we proceed through subadditivity because it eliminates the need for a flatness hypothesis and makes the lifting arguments particularly transparent, especially for those familiar with similar arguments over regular local rings.

4.2. A principal ideal theorem. Recall that Krull’s principal ideal theorem, translated into geometric terms, says that every component of a Cartier divisor on a noetherian scheme has codimension 1. Here we prove the following generalization to valuation rings of rank 1.

**Theorem 4.2.1.** Let $X$ be of finite type over $S$, and let $Z$ be a locally principal closed subscheme of $X$. Then every irreducible component of $Z$ has codimension at most 1 in each component of $X$ that contains it.

The same is not true for valuation rings that are not of rank 1. If $A$ is such a ring then Spec $A$ has principal subschemes of codimension greater than 1. A key step in the proof of the theorem is the following technical proposition, which we prove by noetherian approximation.

---

The assumption that schemes are of finite type over $S$ is crucial for the main results of this section. In the tropical setting, if $\nu$ is nontrivial and $w$ is not in $N_G$, then closed points of $X_w$ never lift to closed points of $X$. However, this does not contradict Theorem 4.2.5 because in this case $X^w$ is not of finite type. See Appendix A.
Proposition 4.2.2. Let \( \mathcal{X} \) be irreducible, locally of finite type, and flat over \( S \). Suppose that \( D \) is a locally principal closed subscheme in \( \mathcal{X} \) that does not meet the generic fiber \( \mathcal{X}_K \). Then every irreducible component of \( D_k \) is an irreducible component of \( \mathcal{X}_k \).

Proof. Since \( \mathcal{X} \) is flat and locally of finite type over \( S \), it is locally of finite presentation over \( S \). The question is local, so we may assume that \( \mathcal{X} \) is affine, and in particular of finite presentation over \( S \). Therefore, there exists a finitely generated subring \( R' \subset R \), with models \( X' \) and \( D' \) over \( S' = \text{Spec} R' \) of \( \mathcal{X} \) and \( D \), respectively. Then \( S' \) is irreducible by construction, and we may assume \( X' \) is irreducible as well.

Let \( s' \) be the image in \( S' \) of the closed point in \( S \). Since \( \mathcal{X}_K \) and \( \mathcal{X}_k \) are both pure of the same dimension \( d \), the fibers of \( X' \) over the generic point and \( s' \) are also pure of dimension \( d \). Similarly, \( D_k \) is pure of dimension \( d \) if and only if \( D'_s \) is, so to prove the proposition it is enough to show that \( D'_s \) is pure of dimension \( d \).

Let \( Z \) be a component of \( D' \) that meets the fiber over \( s' \). We will show that \( Z_s \) has pure dimension \( d \). Let \( s'' \) be the image in \( S' \) of the generic point of \( Z \). Then \( s'' \) is a specialization of the generic point of \( S' \) and specializes to \( s' \), so upper semicontinuity of fiber dimension [Gro66, Theorem 13.1.3] implies that \( X'_{s''} \) has pure dimension \( d \). We claim that \( Z_{s''} \) is a component of \( X'_{s''} \). To see this, note that \( s'' \) is not the generic point of \( S' \), since \( D \) does not meet the generic fiber of \( \mathcal{X} \), and hence every component of \( X'_{s''} \) has codimension at most \( 1 \) in \( X' \). Since \( X' \) is noetherian, Krull’s principal ideal theorem says that every component of \( D' \) has codimension at most \( 1 \) in \( X' \), and hence \( Z_{s''} \) is a component of \( X'_{s''} \), as claimed. In particular, \( Z_{s''} \) has pure dimension \( d \). Since \( s'' \) specializes to \( s' \) and \( X'_{s''} \) has pure dimension \( d \), it follows by upper semicontinuity that \( Z_{s'} \) has pure dimension \( d \), as required.

The following two lemmas are special cases of a more general altitude formula over valuation rings of finite rank [Nag66, Theorem 2]. These codimension formulas are used in the proof of the principal ideal theorem, and again in the proof of subadditivity of codimension over valuation rings of rank 1. We include proofs for the reader’s convenience.

Lemma 4.2.3. Suppose \( \mathcal{X} \) is irreducible, finite type, and flat over \( S \) with \( Z \subset \mathcal{X}_k \) an irreducible closed subset of the special fiber. Then

\[
\text{codim}_X Z = \text{codim}_{\mathcal{X}_k} Z + 1.
\]

Proof. A maximal chain of irreducible closed subsets between \( Z \) and \( \mathcal{X}_k \) can be extended by adding \( \mathcal{X} \) at the end, so \( \text{codim}_X Z \) is at least \( \text{codim}_{\mathcal{X}_k} Z + 1 \).

We now show that \( \text{codim}_X Z \) is at most \( \text{codim}_{\mathcal{X}_k} Z + 1 \).

If \( W \subset W' \) is a strict inclusion of irreducible closed subsets of \( \mathcal{X} \) that meet the special fiber then the dimension of \( W_k \) is less than or equal to the dimension of \( W'_k \). This inequality is strict if \( W \) and \( W' \) are both contained in the special fiber. Similarly, if \( W \) and \( W' \) both meet the generic fiber then \( \dim W_K \) is less than \( \dim W'_K \), and the dimensions of the special fibers are equal to the...
dimensions of the respective generic fibers, by flatness. Therefore, if \( \dim W_k \) is equal to \( \dim W'_k \) then \( W \) is contained in the special fiber and \( W' \) is not. This can happen at most once in any chain of inclusions, so any chain between \( Z \) and \( X \) has length at most \( \dim X_k - \dim Z + 1 \), as required. \( \square \)

**Lemma 4.2.4.** Suppose \( X \) is irreducible, finite type, and flat over \( S \), and let \( Z \subseteq Z' \) be irreducible closed subsets. Then

\[
\operatorname{codim}_X Z = \operatorname{codim}_X Z' + \operatorname{codim}_Z Z.
\]

**Proof.** If \( Z \) meets the generic fiber of \( X \) then this is a classical formula for codimension of varieties over \( K \). Suppose \( Z \) is contained in the special fiber. If \( Z' \) meets the generic fiber then \( \operatorname{codim}_X Z' = \dim X_K - \dim Z'_K \), and \( \operatorname{codim}_Z Z = \dim Z'_k - \dim Z + 1 \), by Lemma 4.2.3. Now \( \dim X_K \) and \( \dim Z'_K \) are equal to \( \dim X_k \) and \( \dim Z'_k \), respectively, so adding these two equations gives

\[
\operatorname{codim}_X Z' + \operatorname{codim}_Z Z = \dim X_k - \dim Z + 1,
\]

and Lemma 4.2.3 says that the right hand side is equal to \( \operatorname{codim}_X Z \). The proof in the case where \( Z' \) is contained in the special fiber is similar. \( \square \)

We now prove the principal ideal theorem over valuation rings of rank 1.

**Proof of Theorem 4.2.1.** We may assume that \( X \) is integral, and since the statement is local about the generic points of \( Z \), we may assume that \( Z \) is irreducible. If \( X \) is supported in the special fiber, the theorem reduces to the classical principal ideal theorem over \( k \). Otherwise \( X \) is dominant, and hence flat, over \( S \). If \( Z \) meets the generic fiber \( X_K \), then the theorem follows from the classical principal ideal theorem for \( X_K \). On the other hand, if \( Z \) is contained in the special fiber, then it must be a union of components of the special fiber, by Proposition 4.2.2, and hence has codimension 1 in \( X \), by Lemma 4.2.3. \( \square \)

We apply the principal ideal theorem to prove the following result on lifting closed points in the special fiber to closed points in the generic fiber. In fact, we prove a more general result, for points of arbitrary transcendence degree. Our argument is in the spirit of Katz’s proof of [Kat09, Lemma 4.15].

**Theorem 4.2.5.** Let \( X' \) be an irreducible scheme locally of finite type over \( S \), and let \( x \) be a closed point of the special fiber \( X_k \). If \( x \) is in the closure of \( X_K \) then the set of closed points \( x' \) in \( X_K \) specializing to \( x \) is Zariski dense in \( X_K \). More generally, if \( x \) is not necessarily closed in \( X_k \), and \( k(x) \) denotes the residue field of \( x \), we can choose \( x' \) to satisfy the identity

\[
\operatorname{trdeg} k(x')/K = \operatorname{trdeg} k(x)/k,
\]

and again the choices of \( x' \) are Zariski dense in \( X_K \).

**Proof.** Note that the statement for closed points is a special case of the statement involving transcendence degrees. We may assume \( X \) is integral and in particular flat over \( S \), and also affine.
By hypothesis, $x$ is a specialization of the generic point of $X$. Observe that given $x' \in X_K$ specializing to $x$, the closure of $x'$ is flat over $S$, so we obtain the inequality $\text{trdeg} k(x')/K \geq \text{trdeg} k(x)/k$. Now, if $\dim X_K = \text{trdeg} k(x)/k$, then we may take $x'$ to be the generic point of $X_K$. We thus proceed by induction on $d := \dim X_K - \text{trdeg} k(x)/k$.

Suppose now $d > 0$, and note that this implies that $x$ is not the generic point of any component of $X_k$. Let $W$ be a closed subset properly contained in $X_K$; by affineness, we can choose a regular function $f$ on $X$ that vanishes at $x$, but not on any component of $W \cup X_k$. Then let $D \subset X$ be the principal subscheme cut out by $f$. Passing to a smaller neighborhood of $x$, if necessary, we may assume every component of $D$ contains $x$. By Proposition 4.2.2, some component $Z$ of $D$ meets the generic fiber, and hence $\dim Z_K - \text{trdeg} k(x)/k = d - 1$. By induction, $Z_K$ contains a point $x'$ specializing to $x$, not contained in $W$, and with $\text{trdeg} k(x')/K = \text{trdeg} k(x)/k$. We conclude the desired statement. \hfill \Box

\textbf{Remark 4.2.6.} For the initial degeneration of a closed subscheme of a torus $T$ over $K$ associated to a weight vector $w \in N_G$, in the special case where $x$ is closed, Theorem 4.2.5 says that every point in $X_w(k)$ lifts to a point in $X(K)$, and the set of such lifts is Zariski dense. In particular, the theorem gives an algebraic proof of surjectivity of tropicalization, along the lines suggested by Speyer and Sturmfels in [SS04], as well as a proof of the density of tropical fibers stated as Theorem 4.1 and Corollary 4.2 in [Pay09b]. A gap in the proof of the former paper is explained in a footnote on the first page of the latter. The proof of Theorem 4.1 in the latter also contains a serious error, discovered by W. Buczynska and F. Sottile, which is explained and corrected in [Pay12].

\textbf{Remark 4.2.7.} When applied to non-closed points, Theorem 4.2.5 says that every closed subvariety of $X_w$ is a component of the special fiber of the closure in $X^w$ of some closed subvariety of $X$. So, at least in this weak sense, every curve in $X_w$ lifts to a curve in $X$, every surface in $X_w$ lifts to a surface in $X$, and so on.

\section{4.3. Proofs of Subadditivity and Lifting.} We now use the principal ideal theorem to prove subadditivity of codimension under intersection and deduce a lifting theorem for proper intersections, in smooth schemes over a valuation ring of rank 1. Our proof follows the classical argument for smooth varieties, which is simpler than Serre’s proof for regular schemes.

\textbf{Remark 4.3.1.} Codimension is not subadditive under intersection in smooth schemes over valuation rings of rank greater than 1, as shown by the following example. Nevertheless, this failure of subadditivity can be understood and controlled, and proper intersections in special fibers still lift to more general fibers over arbitrary valuation rings and, more generally, for flat families of subschemes in a smooth scheme over an arbitrary base. See Appendix B.

\textbf{Example 4.3.2.} Let $A$ be a valuation ring of finite rank $r$ greater than 1, and let $a$ be an element of $A$ that generates an ideal of height $r$. Then $A[x]$ has Krull dimension $r + 1$, by [Nag66, Theorem 2]. The ideals of $A[x]$ generated
by $x$ and $x - a$ each have height 1, but the ideal $(x, x - a) = (x, a)$ has height $r + 1$, which is greater than 2.

Our proof of Theorem 4.1.1 involves a reduction to the diagonal and requires the following lemma on dimensions of fiber products over $S$.

**Lemma 4.3.3.** Let $\mathcal{Y}$ be irreducible and finite type over $S$, and let $\mathcal{X}$ and $\mathcal{X}'$ be irreducible closed subschemes. Then for every irreducible component $Z$ of $\mathcal{X} \times_S \mathcal{X}'$,

$$\text{codim}_{\mathcal{Y} \times_S \mathcal{Y}} Z \leq \text{codim}\, \mathcal{X} + \text{codim}\, \mathcal{X}'$$

Furthermore, the inequality is strict if and only if $\mathcal{Y}$ is dominant over $S$ and $\mathcal{X}$ and $\mathcal{X}'$ are both contained in the special fiber.

**Proof.** If $\mathcal{X}$ and $\mathcal{X}'$ both meet the generic fiber then they are flat over $S$. In this case, $Z$ must also meet the generic fiber, and the proposition holds with equality, by classical dimension theory over $K$. Similarly, if $\mathcal{Y}$ is not dominant over $S$, then the proposition holds with equality by dimension theory over $k$. Suppose $\mathcal{Y}$ is dominant and $\mathcal{X}$ is contained in the special fiber. Then $Z$ is also contained in the special fiber. If, furthermore, $\mathcal{X}'$ is contained in the special fiber, then $Z$ has dimension $\dim \mathcal{X} + \dim \mathcal{X}'$. Therefore, by Lemma 4.2.3,

$$\text{codim}_{\mathcal{Y} \times_S \mathcal{Y}} Z = 2 \dim \mathcal{Y}_K - \dim \mathcal{X} - \dim \mathcal{X}' + 1.$$

The right hand side is equal to $\text{codim}\, \mathcal{X} + \text{codim}\, \mathcal{X}' - 1$, so the inequality is strict by exactly one in this case. A similar argument shows that the proposition holds with equality if $\mathcal{X}'$ meets the generic fiber. \qed

We will prove subadditivity of codimension using the previous proposition and a reduction to the diagonal.

**Lemma 4.3.4.** If $\mathcal{Y}$ is smooth over $S$ then the diagonal $\Delta$ in $\mathcal{Y} \times_S \mathcal{Y}$ is a local complete intersection subscheme.

Here, by local complete intersection we mean that the number of local generators for $I_\Delta$ is equal to the codimension.

**Proof.** First note that the diagonal morphism is locally of finite presentation, by [Gr64, Corollary 1.4.3.1]. The lemma follows, since a locally finite presentation immersion of a smooth scheme in another smooth scheme over an arbitrary base is a local complete intersection. See Proposition 7 of Section 2.2 in [BLR90]. \qed

We now proceed with the proof of subadditivity of codimension and lifting of proper intersections in the special fiber, over valuation rings of rank 1.

**Proof of Theorem 4.1.1.** The intersection $\mathcal{X} \cap \mathcal{X}'$ can be realized as the intersection of $\mathcal{X} \times_S \mathcal{X}'$ with the diagonal in $\mathcal{Y} \times_S \mathcal{Y}'$. Now, the codimension of any component $Z$ of $\mathcal{X} \times_S \mathcal{X}'$ is at most

$$\text{codim}_{\mathcal{Y} \times_S \mathcal{Y}} Z \leq \text{codim}\, \mathcal{X} + \text{codim}\, \mathcal{X}'$$
by Lemma 4.3.3. Then the principal ideal theorem (Theorem 4.2.1) and Lemma 4.2.4 together imply that codimension can only decrease when intersecting with a local complete intersection subscheme of \( Y \times_S Y \). Lemma 4.3.4 says that the diagonal is a local complete intersection, and the theorem follows.

\[ \square \]

**Proof of Theorem 4.1.3.** Let \( Z \) be a component of \( X \cap X' \) that contains \( x \). By Theorem 4.1.1,

\[ \text{codim}_Y Z \leq \text{codim}_Y X + \text{codim}_Y X' + 1, \]

by Lemma 1.2.3 so \( Z \) must meet the generic fiber. Therefore, the generic point of \( Z_K \) is a point of \( X_K \cap X'_K \), specializing to \( x \). If \( x \) is closed in its fiber, we can then find a closed point of \( Z_K \) specializing to \( x \), by Theorem 4.2.5 and similarly for the assertion on transcendence degrees.

\[ \square \]

4.4. Intersection multiplicities over valuation rings of rank 1. The principle of continuity says that intersection numbers are constant when cycles vary in flat families. See [Ful98, Section 10.2] for a precise statement and proof when the base is a smooth variety. For applications to tropical lifting with multiplicities, we need to apply a principle of continuity over the spectrum \( S \) of a possibly non-Noetherian valuation ring of rank 1. Lacking a suitable reference, we include a proof.

**Definition 4.4.1.** Suppose \( Y \) is smooth over a field \( k \), and let \( X \) and \( X' \) be closed subschemes of \( Y \) whose intersection \( X \cap X' \) is finite. Then the intersection number \( i(X \cdot X'; Y) \) of \( X \) and \( X' \) in \( Y \) is

\[ \dim Y \sum_{i=0}^\dim Y (-1)^i \dim_k \Tor^1_{\mathcal{O}_Y}(\mathcal{O}_X, \mathcal{O}_{X'}). \]

In other words, the intersection number is the sum of the local intersection multiplicities at the finitely many points of \( X \cap X' \), weighted by degrees of extension fields,

\[ i(X \cdot X'; Y) = \sum_{P \in X \cap X'} [k(P) : k] i_P(P, X \cdot X'; Y). \]

While we are primarily interested in the case where \( X \) and \( X' \) have complementary dimension, so the finiteness of \( X \cap X' \) means that \( X \) and \( X' \) meet properly in \( Y \), this hypothesis is not technically necessary, since the local intersection multiplicities vanish when the intersection is not proper [Ser65, Theorem V.C.1].

**Theorem 4.4.2.** Let \( Y \) be smooth and quasiprojective over \( S \), and let \( X \) and \( X' \) be closed subschemes of \( Y \) that are flat over \( S \). If \( X \cap X' \) is finite over \( S \) then

\[ i(X_K \cdot X'_K; Y_K) = i(X_k \cdot X'_k; Y_k). \]
In other words, intersection numbers are invariant under specialization over $S$. We now give a direct proof of this equality, by first establishing coherence properties for the structures sheaves of these schemes and then using free resolutions to compute the Tor groups. An alternative approach, using noetherian approximation and specialization properties of intersection theory over a noetherian base, is sketched in Remark 4.4.4.

In general, the structure sheaf of a non-noetherian scheme is not necessarily coherent. The following proposition asserts that such pathologies do not occur on the schemes we are considering.

**Proposition 4.4.3.** Let $X$ be locally of finite presentation over $S$. Then $\mathcal{O}_X$ is coherent.

**Proof.** We first claim that $\mathcal{O}_X$ is coherent in the special case where $X$ is the affine space $\text{Spec} R[t_1, \ldots, t_n]$. Since the question is local, it is enough to show that for any affine subset $U$ of $X$, and any homomorphism $f: \mathcal{O}_X|_U \to \mathcal{O}_X|_U$, the kernel of $f$ over $U$ is finitely generated. Now, the image of $f$ is finitely generated, and is the quasicoherent ideal sheaf $\tilde{I}$ associated to some ideal $I$ in $\mathcal{O}_X(U)$. The ideal $I$ is torsion-free and hence flat over $S$, so it is finitely presented by [RG71, Theorem 3.4.6]. Therefore, by [Mat89, Theorem 2.6], the kernel of the given presentation $\mathcal{O}_X(U)^m \to I$ is finitely generated, which proves the claim.

For the general case, since coherence is local, we may again assume that $X$ is affine, so $X = \text{Spec} A[t_1, \ldots, t_n]/I$ for some finitely generated ideal $I$. The same argument as in the special case above shows that $\tilde{I}$ is coherent on $\text{Spec} R[t_1, \ldots, t_n]$, and hence $\mathcal{O}_X$ is coherent [Gro60, Chapter 0, 5.3.10].

**Proof of Theorem 4.4.2.** Since $Y$ is quasiprojective and $X \cap X'$ is finite over $S$, there is an affine open subset of $Y$ containing the intersection, and hence we may assume $Y$ is affine. Now $\mathcal{O}_Y$ is coherent, by Proposition 4.4.3 and $\mathcal{O}_X$ and $\mathcal{O}_{X'}$, being flat and finitely generated, are finitely presented and hence coherent as $\mathcal{O}_Y$-modules. Let $P_\bullet$ be a resolution of $\mathcal{O}_X$ by free $\mathcal{O}_Y$-modules of finite rank. Then, since $\mathcal{O}_X$ is flat over $S$, the restrictions $P_\bullet \otimes K$ and $P_\bullet \otimes k$ to the generic and special fibers remain exact, giving free resolutions of $\mathcal{O}_{X_K}$ and $\mathcal{O}_{X'_k}$, respectively. Therefore, not only does the homology of $\mathcal{O}_X$ compute $\text{Tor}_Y(X, X')$, but also the homology of the base changes $Q_\bullet \otimes K$ and $Q_\bullet \otimes k$ computes $\text{Tor}_{Y_K}(X'_K, X'_K)$ and $\text{Tor}_{Y_K}(X_K, X'_K)$, respectively.

Now, $\text{Tor}_Y(X, X')$ is coherent, and supported on $X \cap X'$ [Gro63, 6.5.1]. It follows from [Gro63, Proposition 1.4.7] that the push forward of a quasicoherent, finitely presented module under a finite morphism of finite presentation is still quasicoherent and finitely presented. Applying this to $\pi: X \cap X' \to S$ it follows that the push forward $\pi_! \mathcal{O}_{X'}$ is coherent, and since $\pi$ is affine,
this push forward can be computed as the homology of the complex
\[ L \]  
As above, since \( O_X \) is flat over \( S \), the homology of \( L \otimes K \) and \( L \otimes k \) compute \( \text{Tor}^Y_k(X_K, X'_K) \) and \( \text{Tor}^Y_k(X_k, X'_k) \), respectively. Note also that since \( O_{X'} \) is flat over \( S \), the terms of \( Q \) and hence \( L \) are flat \( O_S \)-modules.
Since the homology of \( L \) is coherent, there is a quasi-isomorphic bounded below complex \( M \) of free \( O_S \)-modules of finite rank, and since \( L \) is flat over \( S \), the restrictions \( M \otimes K \) and \( M \otimes k \) are quasi-isomorphic to \( L \otimes K \) and \( L \otimes k \), respectively [Gro61b, Chapter 0, Proposition 11.9.1; see also Remark 11.9.3]. We claim that the homology of \( M \) vanishes in high degree: indeed, the homology of \( M \otimes k \) computes \( \text{Tor} \) on the smooth \( Y_k \), so vanishes in high degree, and the claim then follows from flatness of \( M \) and Nakayama’s lemma. Thus, we can truncate the complex, replacing some \( M_i \) by the image of \( M_{i+1} \), which is coherent and torsion-free and hence free of finite rank. In particular, we may assume that \( M \) is also bounded above. It is then clear that
\[ i(X_s \cdot X'_s, Y_s) = \sum_i (-1)^i \text{rk} M_i, \]
for \( s \) equal to either the generic or closed point of \( S \), and the theorem follows.

Remark 4.4.4. Theorem 4.4.2 can also be proved over an arbitrary base scheme as follows. It is easy to see that intersection numbers are invariant under extension of the base field, so the result behaves well under base change. Passing to the closure of the more general point, and then taking an affine neighborhood around the special point, we reduce to the case where the base is affine and irreducible. This ensures that \( X \) and \( X' \) are finitely presented over the base \( S \), because they are flat and finite type, and then we can proceed by noetherian approximation. Say the affine base is \( \text{Spec} A \). Then there is a finitely generated subalgebra \( A_0 \) of \( A \) over which \( X, X', \) and \( Y \) and all of the relevant morphisms are defined, and all can be chosen so that the relevant geometric properties are preserved, including that the models of \( X \) and \( X' \) are flat [Gro66 Sections 11 and 12], and the model of \( Y \) is quasiprojective [Gro66 Theorem 8.10.5] and smooth [Gro67 Section 17.7]. In particular, we may assume the base is noetherian. By [Gro61a Proposition 7.1.9] we can thus reduce to the case of a DVR, and the theorem follows from the well-known fact that intersection numbers are preserved by specialization over a DVR [Ful98 Section 20.3].

In the geometric case that we are interested in, where the fraction field \( K \) is algebraically closed and hence the residue field \( k \) is algebraically closed as well, we apply Theorem 4.4.2 on intersection numbers to get the following result on individual intersection multiplicities along components of the intersection of the special fibers. Note that this result is of a local nature, and holds for arbitrary expected dimension for the intersection.
Theorem 4.4.5. Assume \( K \) is algebraically closed. Let \( \mathcal{Y} \) be smooth over \( S \), and let \( \mathcal{X} \) and \( \mathcal{X}' \) be closed subschemes that are flat over \( S \). Suppose \( \mathcal{X}_k \) and \( \mathcal{X}'_k \) meet properly along a component \( Z \) of their intersection. Then
\[
i(Z, \mathcal{X}_k \cdot \mathcal{X}'_k; \mathcal{Y}_k) = \sum_{\tilde{Z}} m(Z, \tilde{Z}) \cdot i(\tilde{Z}, \mathcal{X}_K \cdot \mathcal{X}'_K; \mathcal{Y}_K),
\]
where the sum is over irreducible components \( \tilde{Z} \) of \( \mathcal{X}_K \cap \mathcal{X}'_K \) whose closures contain \( Z \), and \( m(Z, \tilde{Z}) \) is the multiplicity of \( Z \) in the special fiber of the closure of \( \tilde{Z} \) inside \( \mathcal{Y} \).

Proof. We first prove the special case where \( \mathcal{X} \) and \( \mathcal{X}' \) have complementary dimension. Suppose \( Z \) is a point. We claim that there is an affine open neighborhood \( U \) of \( Z \) such that \( \mathcal{X} \cap \mathcal{X}' \cap U \) is finite over \( S \) and has special fiber exactly \( Z \). By upper semicontinuity of fiber dimension \([\text{Gro}66, \text{Theorem 13.1.3}]\), the union \( \mathcal{W}_1 \) of the positive dimensional components of \( \mathcal{X}_K \cap \mathcal{X}'_K \) and \( \mathcal{X}_k \cap \mathcal{X}'_k \) is closed in \( \mathcal{Y} \). The intersection of \( \mathcal{X}_K \) and \( \mathcal{X}'_K \) in the complement of \( \mathcal{W}_1 \) is a finite set of \( K \)-points. Let \( \mathcal{W}_2 \) be the closure of those points in this set that do not specialize to \( Z \). We claim that any affine neighborhood of \( Z \) contained in \( \mathcal{Y} \setminus (\mathcal{W}_1 \cup \mathcal{W}_2) \) is the desired neighborhood. Now \( \mathcal{X} \cap \mathcal{X}' \cap \mathcal{U} \) is separated, quasifinite over \( S \) and has special fiber \( Z \), by construction, and it is locally of finite presentation, since \( \mathcal{X} \) and \( \mathcal{X}' \) are locally of finite type and flat, and hence locally of finite presentation. By \([\text{Gro}66, \text{Theorem 8.11.1}]\), to show that this intersection is finite it only remains to check that it is proper over \( S \). The generic fiber consists of finitely many \( K \)-points, each of which extends to an \( R \)-point, and one can check that this implies the valuative criterion for universal closedness \([\text{Gro}61a, \text{Theorem 7.3.8}]\), which proves the claim. Now, note that \( m(\tilde{Z}, Z) = 1 \) for every \( \tilde{Z} \), since each \( \tilde{Z} \) is a \( K \)-point whose closure is a section. The required equality of intersection numbers then follows immediately from Theorem 4.4.2.

We now prove the general case, where the intersection may have positive dimension, by reducing to the case of a zero-dimensional intersection. Since the statement is local on \( \mathcal{Y} \), we may assume \( \mathcal{Y} \) is affine, that \( Z \) is the only component of \( \mathcal{X}_k \cap \mathcal{X}'_k \), and that every component of \( \mathcal{X}_K \cap \mathcal{X}'_K \) contains \( Z \) in its closure. Let \( \tilde{Z}_1, \ldots, \tilde{Z}_m \) be the irreducible components of \( \mathcal{X}_K \cap \mathcal{X}'_K \). Localizing further, we may also assume that \( Z \) and the \( \tilde{Z}_i \) are all smooth (although we cannot assume that the closures of the \( \tilde{Z}_i \) are smooth, as \( Z \) may appear with multiplicity greater than 1).

Now, let \( \mathcal{L}_k \) be a general linear subspace of complementary dimension in the special fiber, so \( \mathcal{L}_k \) meets \( Z \) transversely at finitely many points. Choose a general lift \( \mathcal{L} \) of \( \mathcal{L}_k \) to \( S \); by the Zariski density of Theorem 1.4.2.3 applied to the Grassmannian over \( S \), the generic fiber \( \mathcal{L}_K \) meets each component of \( \mathcal{X}_K \cap \mathcal{X}'_K \) transversely, at isolated points. After further localizing, we may assume \( \mathcal{L}_k \) meets \( Z \) at a single point \( z \), and every point \( \tilde{z} \) of \( \mathcal{X}_K \cap \mathcal{X}'_K \cap \mathcal{L}_K \) specializes to \( z \). Since \( \mathcal{L} \) is a complete intersection, applying \([\text{Ful}98, \text{Example 8.1.10}]\) in the
special fiber, we have
\[ i(Z, \mathcal{X}_k \cdot \mathcal{X}_k'; Y_k) = i(z, (\mathcal{X}_k \cap L_k) \cdot (\mathcal{X}_k' \cap L_k); L_k). \]

Now, note that an inductive application of [Gro66, Proposition 11.3.7] implies that \( \mathcal{X} \cap L \) and \( \mathcal{X}' \cap L \) remain flat over \( S \) so, by the zero-dimensional case treated above, the right hand side is equal to
\[ \sum_{\tilde{z}} i(\tilde{z}, (\mathcal{X}_K \cap L_K) \cdot (\mathcal{X}_K' \cap L_K); L_K). \]

If \( \tilde{Z}_i \) is the component of \( \mathcal{X}_K \cap \mathcal{X}_K' \) containing \( \tilde{z} \), then a similar argument in the generic fiber gives
\[ i(\tilde{z}, (\mathcal{X}_K \cap L_K) \cdot (\mathcal{X}_K' \cap L_K); L_K) = i(\tilde{Z}_i, \mathcal{X}_K \cdot \mathcal{X}_K'; Y_K). \]

The theorem follows since, by the zero-dimensional case above, the multiplicity \( m(\tilde{Z}_i, Z) \) is the number of points \( \tilde{z} \) in \( \tilde{Z}_i \cap L \) specializing to \( z \).

\[ \square \]

**Corollary 4.4.6.** Let \( X \) be a pure-dimensional closed subscheme of \( T \). Then \( \text{Trop}(X) \) is equal to the tropicalization of the fundamental cycle \( \text{Trop}([X]) \).

**Proof.** We may assume the valuation is nontrivial. It is clear that the tropicalizations agree set-theoretically, since both are the closure of the image of \( X(K) \).

To see that the multiplicities agree, let \( w \) be a point in the relative interior of a facet of \( \text{Trop}(X) \), and apply Theorem 4.4.5 in the special case where \( Y \) and \( X' \) are both equal to \( T^w \).

\[ \square \]

5. **Lifting tropical intersections with multiplicities**

We now use the results of Sections 3 and 4 to prove the main tropical lifting theorems, both as stated in the introduction and in refined forms with multiplicities. We also prove generalizations to intersections of three or more subschemes and discuss weaker lifting results when the tropicalizations do not meet properly.

5.1. **Intersections of two subschemes.** Let \( Y \) be a closed subvariety of \( T \), and let \( X \) and \( X' \) be closed subschemes of pure codimension \( j \) and \( j' \), respectively, in \( Y \).

**Proof of Theorem 5.1.** Suppose \( \text{Trop}(X) \) meets \( \text{Trop}(X') \) properly at a simple point \( w \) of \( \text{Trop}(Y) \). Let \( K|K \) be an extension of valued fields such that \( w \in N_G \), where \( G \) is the value group of \( G \). Then the tropicalizations after base change \( \text{Trop}(\tilde{X}) \) and \( \text{Trop}(\tilde{X}') \) meet properly at \( w \), and Theorem 4.4.4 implies furthermore that \( X_w \) meets \( X'_w \) properly at a smooth point of \( Y_w \) if and only if \( \tilde{X}_w \) meets \( \tilde{X}'_w \) properly at a smooth point of \( \tilde{Y}_w \). Therefore, after an extension of valued fields, we may assume \( w \in N_G \).

The initial degeneration \( Y_w \) is a torus torsor containing \( X_w \) and \( X'_w \), and the tropicalizations \( \text{Trop}(X_w) \) and \( \text{Trop}(X'_w) \) with respect to the trivial valuation...
Lifting Tropical Intersections

are the stars of \( w \) in \( \text{Trop}(X) \) and \( \text{Trop}(X') \), respectively, and hence meet properly in the vector space \( \text{Trop}(Y_w) \). Theorem 3.3.1 then says that

\[
\text{Trop}(X_w \cap X'_w) = \text{Trop}(X_w) \cap \text{Trop}(X'_w).
\]

In particular, \( X_w \cap X'_w \) is nonempty of codimension \( j + j' \), and the theorem follows, since \( Y_w \) is smooth. \( \Box \)

**Proof of Theorem 1.4.** Suppose \( X_w \) meets \( X'_w \) properly at a smooth point \( x \) of \( Y_w \). After extending the valued field we may assume the valuation is nontrivial and \( w \in N_G \). In this case, the integral model \( Y_w \) is of finite type over \( \text{Spec} R \), by Proposition \ref{prop:finite-type}. We can then pass to an open neighborhood of \( x \) in \( Y_w \) in which the special fiber is smooth and apply Theorem 4.1.3 to deduce that there is a point of \( X \cap X' \) specializing to \( x \). Therefore \( x \) is contained in \( (X \cap X')_w \). \( \Box \)

As noted in the introduction, Theorem 1.2 follows immediately from Theorems 1.3 and 4.1.3, and Theorem 1.1 is the special case where \( Y \) is the torus \( T \). So this concludes the proof of the theorems stated in the introduction.

We now state and prove a refined version of Theorem 1.2 with multiplicities.

\[ m_{X \cap X'}(\tau) \geq i(\tau, \text{Trop}(X) \cdot \text{Trop}(X'); \text{Trop}(Y)), \]

and both are strictly positive. Furthermore, the tropical intersection multiplicity is equal to the weighted sum of algebraic intersection multiplicities

\[ i(\tau, \text{Trop}(X) \cdot \text{Trop}(X'); \text{Trop}(Y)) = \sum_Z i(Z, X \cdot X'; Y) m_Z(\tau), \]

where the sum is over components \( Z \) of \( X \cap X' \) whose tropicalizations contain \( \tau \).

Although the global intersection product \( X \cdot X' \) need not be defined, the intersection multiplicities \( i(Z, X \cdot X'; Y) \) (as defined in terms of Tor in Section 2.5) appearing in the statement of Theorems 5.1.1 and Theorem 5.2.3 below, are nonetheless well-defined. Indeed, for every component \( Z \) of \( X \cap X' \) whose tropicalization contains \( \tau \), we have that \( Y \) must be smooth on a non-empty open subset that intersects \( Z \) by the simple point hypothesis, and because \( \text{Trop}(X) \)
Brian Osserman and Sam Payne

is assumed to meet Trop(X′) properly along τ, it also follows that X meets X′ properly along Z.

Proof. First, by Theorem 1.2, the face τ is contained in Trop(X ∩ X′). We prove the identity between the tropical intersection multiplicity and the weighted sum of algebraic intersection multiplicities and then deduce the inequality for m_{X∩X′}(τ), as follows. Let w be a point in τ that is a simple point of Trop(Y).

After extending the valued field, we may assume w ∈ NG. Then the tropical intersection multiplicity along τ is equal to the local tropical intersection multiplicity at w

\[ i(\tau, \text{Trop}(X) \cdot \text{Trop}(X′) ; \text{Trop}(Y)) = i(\tau_w, \text{Trop}(X_w) \cdot \text{Trop}(X′_w) ; \text{Trop}(Y_w)), \]

where \( \tau_w = \mathbb{R}_{>0}(\tau - w) \), as in Section 2.7. Then, since Y_w is a torus torsor, Theorem 3.3.1 says that this local tropical intersection multiplicity is given by

\[ i(\tau_w, \text{Trop}(X_w) \cdot \text{Trop}(X′_w) ; \text{Trop}(Y_w)) = \sum_{Z_0} i(Z_0, X_w \cdot X′_w ; Y_w) m_{Z_0}(\tau_w), \]

where the sum is over components Z_0 of X_w ∩ X′_w whose tropicalizations contain \( \tau_w \). By Theorem 4.4.5 for each such Z_0,

\[ i(Z_0, X_w \cdot X′_w ; Y_w) = \sum_Z m(Z, Z_0) i(Z, X \cdot X′ ; Y), \]

where the sum is over components Z of X ∩ X′ and m(Z, Z_0) is the multiplicity of Z_0 in the special fiber of the closure of Z in Y_w. Combining these identities with the results of Section 2.7 gives the required equality, since

\[ m(Z, Z_0) = \sum_{Z_0} m(Z_0, Z) m_{Z_0}(\tau_w), \]

where the sum ranges over components Z_0 of the special fiber of the closure of Z.

The inequality follows immediately from this equality, since

\[ m_{X∩X′}(\tau) = \sum_Z \text{length}(O_{Z,X∩X′}) m_Z(\tau), \]

by Corollary 4.4.6 and \[ \text{length}(O_{Z,X∩X′}) \geq i(Z, X \cdot X′ ; Y) \] for each Z [Pit98 Proposition 7.1].

If Trop(X) meets Trop(X′) properly in N_R, then part (2) of Theorem 5.1.1 has a particularly simple statement equating the tropicalization of the refined intersection cycle with the stable tropical intersection.

**Corollary 5.1.2.** Suppose Trop(X) meets Trop(X′) properly in N_R. Then

\[ \text{Trop}(X \cdot X′) = \text{Trop}(X) \cdot \text{Trop}(X′). \]

We also note that the inequality in Theorem 5.1.1 can be replaced by an equality if X and X′ are smooth, or mildly singular.

**Corollary 5.1.3.** Under the hypotheses of Theorem 5.1.1 if X and X′ are Cohen-Macaulay then

\[ m_{X∩X′}(\tau) = i(\tau, \text{Trop}(X) \cdot \text{Trop}(X′) ; \text{Trop}(Y)). \]
Proof. If $X$ and $X'$ are Cohen-Macaulay, then the length of the scheme-theoretic intersection of $X$ and $X'$ along each component $Z$ of expected dimension is equal to the intersection multiplicity $i(Z, X \cdot X'; Y)$ [Ful98, Example 8.2.7]. The result then follows from the equality in Theorem 5.1.1, since

$$m_{X \cap X'}(\tau) = \sum_{\mathcal{Z}} \text{length}(\mathcal{O}_{\mathcal{Z}, X \cap X'}) m_{\mathcal{Z}}(\tau),$$

by Lemma 2.7.2, where the sum is over components $\mathcal{Z}$ of $X \cap X'$ such that $\text{Trop}(\mathcal{Z})$ contains $\tau$. $\square$

5.2. Intersections of three or more subschemes. In applications, and particularly in the context of enumerative geometry, one frequently wants to intersect more than two subschemes. The algebraic intersection product of several closed subschemes $X_1 \cdots X_r$ may be treated either by induction from results on intersection of pairs or by a standard reduction to the diagonal argument. For the reduction to the diagonal, one uses the facts that $X_1 \cap \cdots \cap X_r$ is canonically identified with the intersection $X_1 \times \cdots \times X_r \cap \Delta$ in $Y^r$, where $\Delta$ is the diagonal subscheme, and that $X_1 \times \cdots \times X_r \cdot \Delta$ is the push forward of $X_1 \cdots X_r$ under the diagonal embedding.

Definition 5.2.1. Let $X_1, \ldots, X_r$ be closed subschemes in $T$, of pure codimension $j_1, \ldots, j_r$, respectively. Then the stable tropical intersection $\text{Trop}(X_1) \cdots \text{Trop}(X_r)$ is the iterated pairwise stable tropical intersection

$$\text{Trop}(X_1) \cdots \text{Trop}(X_r) = (\cdots ((\text{Trop}(X_1) \cdot \text{Trop}(X_2)) \cdot \text{Trop}(X_3)) \cdots \text{Trop}(X_r)).$$

The following proposition shows that stable tropical intersections for three or more closed subschemes can be computed as a pairwise intersection with the tropicalization of the diagonal. In particular, the stable tropical intersection is independent of the order of the factors.

Proposition 5.2.2. The stable tropical intersection $\text{Trop}(X_1 \times \cdots \times X_r) \cdot \text{Trop}(\Delta)$ is the image of $\text{Trop}(X_1) \cdots \text{Trop}(X_r)$ under the diagonal embedding of $N^r_\mathbb{R}$ in $N^r_\mathbb{R}$.

Proof. Let $\Sigma$ be a complete unimodular fan that contains $\text{Trop}(X_1), \ldots, \text{Trop}(X_r)$ as subfans. By Proposition 2.7.6 the stable tropical intersections $\text{Trop}(X_1 \times \cdots \times X_r) \cdot \text{Trop}(\Delta)$ and $\text{Trop}(X_1) \cdots \text{Trop}(X_r)$ correspond to the products of Minkowski weights $\omega_{X_1} \cdots \omega_{X_r}$ and $\omega_{X_1} \cdots \omega_{X_r}$, on $\Sigma'$ and $\Sigma$, respectively. These products of Minkowski weights are equal to $\omega_{X_1} \cdots \omega_{X_r}$ and $\omega_{X_1} \cdots \omega_{X_r}$ by the identification of rings of Minkowski weights with Chow rings, discussed in Section 2.5. The proposition follows, since $X_1 \times \cdots \times X_r \cdot \Delta$ is the push forward of $X_1 \cdots X_r$ under the diagonal embedding. $\square$

As in the case of pairwise intersections, when $X_1, \ldots, X_r$ are subschemes of pure codimension $j_1, \ldots, j_r$ in an irreducible variety $Y$, we define the stable tropical intersection multiplicity along a face $\tau$ of codimension $j_1 + \cdots + j_r$ in $\text{Trop}(Y)$, provided that $\tau$ contains a simple point $w$ of $\text{Trop}(Y)$, as the multiplicity of $\tau_w$.
in the stable tropical intersection of the stars of $w$ in $\text{Trop}(X_1), \ldots, \text{Trop}(X_r)$ inside the star of $w$ in $\text{Trop}(Y)$.

We now generalize Theorem 5.1.1 to intersections of several closed subschemes by reducing to the case of a pairwise intersection with the diagonal in $Y^r$.

**Theorem 5.2.3.** Let $X_1, \ldots, X_r$ be closed subschemes of pure codimension $j_1, \ldots, j_r$ in $Y$, respectively. Suppose $\tau$ is a facet of $\text{Trop}(X_1) \cap \cdots \cap \text{Trop}(X_r)$ of codimension $j_1 + \cdots + j_r$ in $\text{Trop}(Y)$ that contains a simple point of $\text{Trop}(Y)$. Then the tropicalization $\text{Trop}(X_1 \cap \cdots \cap X_r)$ contains $\tau$ with multiplicity bounded below by the tropical intersection multiplicity
\[ m_{X_1 \cap \cdots \cap X_r}(\tau) \geq i(\tau, \text{Trop}(X_1) \cdots \text{Trop}(X_r); \text{Trop}(Y)), \]
and both are strictly positive. Furthermore, the tropical intersection multiplicity is equal to the weighted sum of algebraic intersection multiplicities
\[ i(\tau,\text{Trop}(X_1) \cdots \text{Trop}(X_r);\text{Trop}(Y)) = \sum_Z i(Z,X_1 \cdots X_r;Y)m_Z(\tau), \]
where the sum is over components of $X_1 \cap \cdots \cap X_r$ such that $\text{Trop}(Z)$ contains $\tau$.

**Proof.** Let $\tau$ be a facet of $\text{Trop}(X_1) \cap \cdots \cap \text{Trop}(X_r)$ of codimension $j_1 + \cdots + j_r$ in $\text{Trop}(Y)$ that contains a simple point $w$ of $\text{Trop}(Y)$. Then the image of $\tau$ under the diagonal embedding $\delta$ of $\text{Trop}(Y)$ in $\text{Trop}(Y^r)$ is a facet of $\text{Trop}(X_1 \times \cdots \times X_r) \cap \text{Trop}(\Delta)$ of codimension $(r-1)\dim(Y) + j_1 + \cdots + j_r$, in $\text{Trop}(Y^r)$, where $\Delta$ is the diagonal subscheme in $Y^r$, containing a simple point of $\text{Trop}(Y^r)$. By Theorem 5.1.1, $\delta(\tau)$ is a face of $\text{Trop}(X_1 \times \cdots \times X_r \cap \Delta)$. Furthermore, the multiplicity of $\delta(\tau)$ in the stable tropical intersection $\text{Trop}(X_1 \times \cdots \times X_r \cap \text{Trop}(\Delta))$ inside $\text{Trop}(Y^r)$ is strictly positive, equal to
\[ \sum_Z i(Z,X_1 \times \cdots \times X_r;\Delta; Y^r)m_Z(\delta(\tau)), \]
and less than or equal to $m_{X_1 \times \cdots \times X_r \cap \Delta}(\delta(\tau))$. Now $X_1 \times \cdots \times X_r \cap \Delta$ is canonically identified with $X_1 \cap \cdots \cap X_r$, and both the tropical multiplicities and the local intersection multiplicities agree. In other words, we have
\[ i(Z,X_1 \times \cdots \times X_r;\Delta; Y^r) = i(Z,X_1 \cdots X_r;Y) \]
and
\[ m_Z(\delta(\tau)) = m_Z(\tau). \]

The theorem then follows, by Proposition 5.2.2. \qed

The following application of Theorem 5.2.3, giving a formula for counting points in zero-dimensional complete intersections, has been applied by Rabinoff to construct canonical subgroups of abelian varieties over $p$-adic fields, via a suitable generalization for power series [Rab12a, Rab12b].

Let $f_i = \sum a_i(u)x^u$ be an equation defining $X_i$, with $P_i = \text{conv}\{u|a_i \text{ is nonzero}\}$ its Newton polytope. Projecting the lower faces of the lifted Newton polytope $\text{conv}\{\{u,\nu(a_i(u))\}\}$ in $M_\mathbb{R} \times \mathbb{R}$ gives the Newton
subdivision of $P_i$. This Newton subdivision is dual to $\text{Trop}(X_i)$, in the sense that there is a natural polyhedral structure on $\text{Trop}(X_i)$ whose faces are in order-reversing bijection with the positive dimensional faces of the Newton subdivision. A face $\tau$ of $\text{Trop}(X_i)$ corresponds to the convex hull of the lattice points $u$ such that $a_i(u)x^n$ is a monomial of minimal $w$-weight, for $w$ in the relative interior of $\tau$.

Our formula is phrased in terms of mixed volumes of faces of the Newton subdivision. Recall that, for lattice polytopes $Q_1, \ldots, Q_n$ in $M_\mathbb{R} \cong \mathbb{R}^n$, the euclidean volume of the Minkowski sum $b_1Q_1 + \cdots + b_nQ_n$ is a polynomial of degree $n$ in $b_1, \ldots, b_n$, and the mixed volume $V(Q_1, \ldots, Q_n)$ is the coefficient of $b_1 \cdots b_n$ divided by $n!$. If $\Sigma$ is the smallest common refinement of the inner normal fans of the $Q_i$, then each $Q_i$ corresponds to a nef line bundle $L_i$ on the toric variety $Y(\Sigma)$, and the mixed volume is equal to the intersection number, divided by $n!$,

$$V(Q_1, \ldots, Q_n) = (c_1(L_1) \cdots c_1(L_n))/n!.$$ 

See [Ful93, p. 116] for further details on mixed volumes and their relation to toric intersection theory and [KK12, Section 4] for generalizations to arbitrary projective varieties via Newton-Okounkov bodies. Mixed volume formulas for tropical stable complete intersections are standard in the case of the trivial valuation. The earliest reference we know of relating mixed volumes to tropical complete intersections for a nontrivial (discrete) valuation is due to Smirnov [Smi96].

**Corollary 5.2.4.** Let $X_1, \ldots, X_n$ be hypersurfaces in $T$, and suppose $w$ is an isolated point in $\text{Trop}(X_1) \cap \cdots \cap \text{Trop}(X_n)$. Let $Q_i$ be the face of the Newton subdivision corresponding to the minimal face of $\text{Trop}(X_i)$ that contains $w$. Then the number of points in $X_1 \cap \cdots \cap X_n$ with tropicalization $w$, counted with multiplicities, is exactly $n!V(Q_1, \ldots, Q_n)$.

**Proof.** Fix a complete unimodular fan $\Sigma$ in $N_\mathbb{R}$ such that, for $i = 1, \ldots, n$, every face of $\text{Star}_w \text{Trop}(X_i)$ is a union of faces of $\Sigma$. Since $\text{Star}_w \text{Trop}(X_i)$ is the codimension 1 skeleton of the possibly degenerate inner normal fan of $Q_i$, it follows that $Q_i$ corresponds to a nef line bundle $L_i$ on $Y(\Sigma)$. Furthermore, the Minkowski weight of codimension 1 on $\Sigma$ given by the tropical multiplicities on $\text{Star}_w \text{Trop}(X_i)$ corresponds to $c_1(L_i)$. It follows, by the compatibility of toric and stable multiplicity the tropical intersection multiplicity

$$i(w, \text{Trop}(X_1) \cdots \text{Trop}(X_n)) = (c_1(L_1) \cdots c_1(L_n)),$$

and the latter is $n!V(Q_1, \ldots, Q_n)$. By Theorem 5.2.3, this tropical intersection multiplicity is equal to the number of points in $X_1 \cap \cdots \cap X_n$ with tropicalization $w$, counted with multiplicities, as required. \[\square\]

---

There is a minor misstatement in the definition of mixed volumes, in the text in between displayed formulas (1) and (2) of [Ful93, p. 116]. The mixed volume is the coefficient of $v_1 \cdots v_n$ divided by $n!$, not multiplied by $n!$. The displayed formulas (1), (2), and (3) are correct. Formulas (1) and (2) uniquely determine the mixed volumes of rational polytopes, as does (3), which also characterizes mixed volumes of arbitrary convex bodies.
Remark 5.2.5. In Corollary 5.2.4 the multiplicities on the points $x$ in $X_1 \cap \cdots \cap X_n$ with tropicalization $w$ are equal to the lengths of the scheme-theoretic intersection of $X_1, \ldots, X_n$ along $x$, as in Corollary 5.1.3 since complete intersections are Cohen-Macaulay.

5.3. Non-proper intersections. Our main tropical lifting results require the tropicalizations to meet properly. Nevertheless, our results on lifting from the special fiber to the generic fiber of $Y^w$, such as Theorem [14] still yield nontrivial statements for nonproper tropical intersections. In many cases, such as Example 6.2 the initial degenerations meet properly even when the tropicalizations do not.

Proposition 5.3.1. Suppose that for each $w \in \text{Trop}(X) \cap \text{Trop}(X')$, we have that $w$ is smooth and $X_w$ meets $X'_w$ properly in $Y_w$. Then $X$ meets $X'$ properly in $Y$. Furthermore, the set of $w$ such that $X_w \cap X'_w$ is nonempty is either empty or the underlying set of a polyhedral complex of pure codimension $j + j'$ in $\text{Trop}(Y)$.

Proof. If $w$ is in $\text{Trop}(X \cap X')$, then $X_w \cap X'_w$. Conversely, if $X_w$ meets $X'_w$ properly at some smooth point $x$ of $Y_w$ then $x$ is contained in $(X \cap X')_w$, by Theorem 1.3 and hence $w$ is in $\text{Trop}(X \cap X')$. Therefore, the hypotheses of the proposition ensure that the set of $w$ such that $X_w$ meets $X'_w$ is exactly $\text{Trop}(X \cap X')$, and $(X \cap X')_w$ is equal to $X_w \cap X'_w$ for all $w$. If $X_w \cap X'_w$ is nonempty then it has pure codimension $j + j'$ in $Y_w$. It follows that $X \cap X'$ has pure dimension $j + j'$ in $Y$, and hence $\text{Trop}(X \cap X')$ is a polyhedral complex of pure codimension $j + j'$ in $\text{Trop}(Y)$, as required. □

In the special case where $Y$ is the ambient torus $T$, one can say even more. If all initial degenerations meet properly, then the set of weight vectors where the intersection is nonempty is exactly the stable tropical intersection, as was suggested to us by J. Rau.

Proposition 5.3.2. Suppose $X_w$ meets $X'_w$ properly in $T_w$ for all $w$. Then the set of $w \in N_\mathbb{R}$ such that $X_w \cap X'_w$ is nonempty is exactly the underlying set of the stable tropical intersection $\text{Trop}(X) \cdot \text{Trop}(X')$.

Proof. Let $\Sigma$ be a complete unimodular fan that contains the stars of $w$ in $\text{Trop}(X)$ and in $\text{Trop}(X')$ as subfans. First we claim that, for any $w$ in $N_\mathbb{R}$, the closures $\overline{X}_w$ and $\overline{X}'_w$ meet properly in $Y(\Sigma)$ and, furthermore, $\overline{X}_w \cap O_\sigma$ and $\overline{X}'_w \cap O_\sigma$ meet properly in $O_\sigma$ for every $\sigma$ in $\Sigma$. Indeed, after choosing an extension of valued fields such that $w$ is rational over the value group and subdividing $\Sigma$, we may assume that it contains tropical fans, in the sense of [Lew07], for $X_w$ and $X'_w$ as subfans. Then, for $v \in N_G$ in the relative interior of $\sigma$ and $\epsilon \in G$ sufficiently small and positive, the initial degenerations $X_{w + \epsilon v}$ and $X'_{w + \epsilon v}$ agree with the initial degenerations of $X_w$ and $X'_w$, respectively, for the weight vector $v$ [Gub12 Corollary 10.12]. The images of these initial degenerations under projection to $O_\sigma$ are identified with $\overline{X}_w \cap O_\sigma$ and $\overline{X}'_w \cap O_\sigma$, up to simultaneous translation in $O_\sigma$ [Gub12 Remark 12.7]. Since $X_{w + \epsilon v}$ and
$X_{w+tv}$ meet properly in $T_{w+tv}$, it follows that $\overline{X}_w \cap O_\sigma$ and $\overline{X}_w \cap O_\sigma$ meet properly in $O_\sigma$, as claimed.

Since $\overline{X}_w \cap O_\sigma$ and $\overline{X}_w \cap O_\sigma$ meet properly in $O_\sigma$ for every $\sigma$ in $\Sigma$, Proposition 2.7.7 says that a cone $\tau_w$ of codimension $j + j'$ in $\Sigma$ appears in the tropicalization of $X_w \cdot X'_w$ with multiplicity equal to the stable tropical intersection multiplicity $i(\tau_w, \text{Trop}(X_w) \cdot \text{Trop}(X'_w))$ which agrees with the stable tropical intersection multiplicity $i(\tau, \text{Trop}(X) \cdot \text{Trop}(X'))$, by Proposition 2.7.6.

In particular, $X_w \cap X'_w$ is nonempty if and only if $w$ is contained in the stable tropical intersection $\text{Trop}(X) \cdot \text{Trop}(X')$. □

For closed subschemes of the torus $T$, when a tropical intersection is not proper we can translate the subschemes by a suitable element of $T(K)$ so that the initial degenerations meet properly, without changing the tropicalizations. The following theorem extends Proposition 2.7.8 to the general case, where the valuation may not be trivial, and gives a geometric meaning to the stable tropical intersection, as the tropicalization of an intersection with a translate by a general point $t$ such that $\text{Trop}(t)$ is zero.

If $t$ is in $T(K)$ then the tropicalization of the translate $tX$ is the translate of $\text{Trop}(X)$ by the vector $\text{Trop}(t)$ in $N_G$. In particular, the tropicalization is invariant under translation by points $t$ such that $\text{Trop}(t)$ is zero. The set of such points is Zariski dense in $T(K)$ [Pay09, Corollary 4.2] (see also [Pay12, Remark 2]), but not Zariski open if the valuation is nontrivial. Nevertheless, we say that a property holds for a general point $t$ such that $\text{Trop}(t)$ is zero if there is a Zariski open subset $U_0$ of the initial degeneration $T_0$ such that the property holds for all $t$ with $t_0 \in U_0$. This set is again Zariski dense in $T(K)$.

**Theorem 5.3.3.** Let $X$ and $X'$ be pure-dimensional closed subschemes of $T$. Then, for a general point $t$ such that $\text{Trop}(t)$ is zero,

$$\text{Trop}(X \cap tX') = \text{Trop}(X) \cdot \text{Trop}(X').$$

**Proof.** If the valuation is trivial then the theorem is given by Proposition 2.7.8. Assume the valuation is nontrivial. Let $\tau$ be a face of $\text{Trop}(X) \cap \text{Trop}(X')$, and let $w \in N_G$ be a point in $\tau$. The initial degeneration $(tX')_w$ is the translate of $X'_w$ by $t_0$. Since $t_0$ is general in $T_0$, Proposition 2.7.8 says that

$$\text{Trop}(X_w \cap (tX')_w) = \text{Trop}(X_w) \cdot \text{Trop}(X'_w).$$

In particular, if $w$ is not in the stable tropical intersection then the initial degenerations $X_w$ and $(tX')_w$ are disjoint, so $w$ is not in $\text{Trop}(X \cap tX')$. On the other hand, if $w$ is in the stable intersection then $X_w$ meets $(tX')_w$ properly in $T_w$, by Proposition 2.7.8 and all points of $X_w \cap (tX')_w$ lift to the generic fiber $X \cap tX'$, by Theorem 4.1.3. Note that for $w$ and $w'$ in the relative interior of $\tau$, because the isomorphisms $X_w \cong X_{w'}$ and $tX'_w \cong tX'_w$ can be simultaneously induced by a single isomorphism $T_w \cong T_{w'}$, if $X_w$ meets $tX'_w$, then $X_{w'}$ likewise meets $tX'_w$ properly. Since there are only finitely many faces of $\text{Trop}(X) \cap \text{Trop}(X')$, the point $t$ can be chosen sufficiently general so that this holds in every face.
This shows that the underlying sets of $\text{Trop}(X \cap tX')$ and $\text{Trop}(X) \cdot \text{Trop}(X')$ are equal. Now, let $\tau$ be a facet of $\text{Trop}(X) \cdot \text{Trop}(X')$ and let $w \in N_G$ be a point in the relative interior of $\tau$. The tropical intersection multiplicity along $\tau$ is equal to the local tropical intersection multiplicity at $w$

\[ i(\tau, \text{Trop}(X) \cdot \text{Trop}(X'); \text{Trop}(T)) = i(\tau, \text{Trop}(X_w) \cdot \text{Trop}(X'_w); \text{Trop}(T_w)), \]

and, by Proposition 2.7.8 the right hand side is equal to $m_{X_w \cap tX'_w}(\tau_w)$. Applying Theorem 4.4.5 to the intersection of $X^w$ and $(tX')^w$ in $T^w$, as in the proof of Theorem 5.1.1 then shows that $m_{X_w \cap tX'_w}(\tau_w)$ is equal to $m_{X \cdot tX'}(\tau)$. As in the proof of Proposition 2.7.8, the genericity of $t$ guarantees that the cycle $X \cdot tX'$ is equal to the fundamental cycle of $X \cap tX'$. Therefore, $m_{X \cdot tX'}(\tau)$ is equal to $m_{X \cap tX'}(\tau)$, and the theorem follows. 

6. Examples

Here we give a number of examples illustrating our tropical lifting theorems and the necessity of their hypotheses. The first example involves tropicalizations that meet properly, but not necessarily in the interiors of maximal faces.

\textbf{Example 6.1. }Inside $Y = (K^*)^2$, let $X$ and $X'$ be given by $y = x + 1$ and $y = ax^2$, respectively, with $a \in K^*$. We consider three cases, according to whether or not $\nu(a)$ is zero, and its sign if it is nonzero.

If $\nu(a)$ is positive then $\text{Trop}(X)$ meets $\text{Trop}(X')$ at two points, each with multiplicity 1. In this case, Theorem 5.1.1 says that $X$ meets $X'$ at two points, each with multiplicity 1, and one of these intersection points lies over each of the points of $\text{Trop}(X) \cap \text{Trop}(X')$. It is also straightforward to check this directly. If $\nu(a)$ is negative, then $\text{Trop}(X)$ meets $\text{Trop}(X')$ at a single point, but with tropical intersection multiplicity 2. In this case, Theorem 5.1.1 says that $X$ meets $X'$ at either a single point with multiplicity 2, or at two points of
multiplicity 1, and one can check that the intersection is always two points of multiplicity 1. In both of these cases, the nonemptiness of \( X \cap X' \) also follows from the transverse tropical lifting result of [BJS+07], since the tropicalizations meet properly in the interiors of maximal faces.

Suppose \( \nu(a) \) is zero. Then \( \text{Trop}(X) \) and \( \text{Trop}(X') \) meet at a single point with tropical multiplicity 2, but the intersection is in a nonmaximal face of \( \text{Trop}(X) \), so transverse lifting results do not apply. Nevertheless, Theorem 5.1.1 still says that \( X \) and \( X' \) meet at either a single point with multiplicity 2, or at two points of multiplicity 1. Either possibility can occur; the algebraic intersection is a single point of multiplicity 2 when the characteristic is not 2 and \( a = -1/4 \), and two points of multiplicity 1 otherwise.

In the following example, the tropicalizations meet nonproperly along a positive dimensional set that does not contain the the tropicalization of any curve.

**Example 6.2.** Inside \( Y = (K^*)^2 \), let \( X \) and \( X' \) be given by \( y = x + 1 \) and \( y = ax + b \), respectively, with \( a \) and \( b \) in \( K^* \). Assume \( a \) and \( b \) are not both 1, so the closures of \( X \) and \( X' \) are distinct lines in \( K^2 \). In particular, \( X \) and \( X' \) intersect in at most one point.

Suppose \( \nu(a) \) is zero and \( \nu(b) \) is positive. Then \( \text{Trop}(X) \) and \( \text{Trop}(X') \) intersect nonproperly, along the ray \( R_{\leq 0} \cdot (1, 1) \).

If \( a = 1 \), then the closures of \( X \) and \( X' \) are parallel lines in \( K^2 \), so none of the tropical intersection points lift.

Suppose \( a \) is not 1. Then the unique algebraic intersection point is

\[
X \cap X' = ((1 - b)/(a - 1), (a - b)/(a - 1)),
\]

and the unique point of \( \text{Trop}(X) \cap \text{Trop}(X') \) that lifts is \((\nu(a - 1), -\nu(a - 1))\).

For a suitable choice of \( a \) congruent to 1 modulo \( m \), any nonzero \( G \)-rational point in the tropical intersection can lift. For such \( a \), the initial degenerations \( X_w \) and \( X'_w \) coincide for all nonzero \( w \) in \( \text{Trop}(X) \cap \text{Trop}(X') \), and are disjoint otherwise. In particular, none of the initial degenerations meet properly. On
the other hand, if \( a \) is not congruent to 1 modulo \( \mathfrak{m} \), then the initial degenerations \( X_w \) and \( X'_w \) meet transversely at a single point for \( w = (0,0) \), and are disjoint otherwise. In this case \((0,0)\) is the unique tropical intersection point that lifts, as it must be by Theorem 1.4. Note that, even though \( \text{Trop}(X \cap X') \) and the stable tropical intersection are both zero-dimensional, neither is necessarily contained in the other.

In the remaining examples, we assume the characteristic of \( K \) is not 2 and consider tropicalizations of skew lines inside a smooth quadric surface. These examples demonstrate the necessity of requiring the point of proper intersection to be a simple point of \( \text{Trop}(Y) \) or a smooth point of \( Y_w \), in Theorems 1.2, 1.3, and 1.4.

**Example 6.3.** Let \( Y \) be the surface in \((K^*)^3\) given by
\[
z^2 + xy + x + y = 0.
\]
Then \( Y \) contains the curves \( X \) and \( X' \) given by \( x + 1 = z - 1 = 0 \) and \( x + 1 = z + 1 = 0 \), respectively. Now, \( \text{Trop}(X) \) and \( \text{Trop}(X') \) still meet properly at \( w = (0,0,0) \), as shown.

However, this tropical intersection point does not lift because the closures of \( X \) and \( X' \) are skew lines in \( K^3 \). This is consistent with Theorems 1.2 and 1.3 because \( w \) is not in the relative interior of a facet, and the initial degenerations \( X_w \) and \( X'_w \) are disjoint. The closures of \( X_w \) and \( X'_w \) in \( k^3 \) are skew lines in the closure of \( Y_w \), which is a smooth quadric surface.

**Example 6.4.** Let \( a \in K^* \) be an element of positive valuation, and let \( Y' \) be the surface given by
\[
z^2 - 1 + a(xy + x + y + 1) = 0,
\]
and let $X$ and $X'$ be as in Example 6.3. Then $X$ and $X'$ are contained in $Y'$, and their tropicalizations still meet properly in $\text{Trop}(Y')$ at the origin $w$, which is in the relative interior of a facet $\sigma$ of $\text{Trop}(Y')$, as shown.

The tropical intersection point $w$ does not lift to $X \cap X'$, but this is still consistent with Theorems 1.2 and 1.3 because the multiplicity of the facet $\sigma$ is 2, and hence $w$ is not a simple point of $\text{Trop}(Y')$. The initial degeneration $Y''_w$ has two disjoint components, each of which contains the initial degeneration of one of the curves.

In particular, this tropical intersection point does not lift even to the intersection of the initial degenerations.
Example 6.5. Let \( a \in K^* \) be an element of positive valuation, with \( Y'' \) be the surface in \( (K^*)^3 \) given by
\[
(x + 1)(y + 1) + (x + z)(y + z + a) = 0.
\]
Let \( X \) again be as in Example 6.3 and let \( X'' \) be the curve given by \( y + 1 = z - 1 + a = 0 \). Once again, \( \text{Trop}(X) \) and \( \text{Trop}(X'') \) meet properly at \( w = (0, 0, 0) \). Furthermore, the initial degenerations \( X_w \) and \( X''_w \) meet properly at a single point in \( Y''_w \).

This intersection point lifts to the initial degenerations but not to the general fiber, because the closures of \( X \) and \( X'' \) are skew lines in \( K^3 \). This is still consistent with Theorem 1.4 because \( Y''_w \) is a cone, and \( X_w \) meets \( X''_w \) at the singular point.

**Appendix A. Initial degenerations, value groups, and base change**

Here we study how initial degenerations behave with respect to arbitrary extensions of valued fields. These basic results are used throughout the paper to reduce our main lifting theorems to the case where \( w \) is in \( N_G \), by extending the ground field.

Let \( M_{G,w} \) be the maximal sublattice of \( M \) on which \( w \) is \( G \)-rational. In other words,
\[
M_{G,w} = \{ u \in M \mid \langle u, w \rangle \text{ is in } G \}.
\]
Since \( K \) is algebraically closed, the value group \( G \) is divisible, and hence \( M_{G,w} \) is saturated in \( M \). If the weight of a monomial \( ax^u \) is zero, then \( \langle u, w \rangle = -\nu(a) \), and hence \( u \) is in \( M_{G,w} \). In particular, \( M_{G,w} \) contains all exponents of monomials that restrict to nonzero functions on \( T_w \), and \( T_w \) is a torsor over the torus associated to \( M_{G,w} \).
Proposition A.1. Suppose the valuation is nontrivial. Then the integral model $T^w$ is of finite type over $\text{Spec} \mathcal{R}$ if and only if $w$ is in $N_G$. Furthermore, if $w$ is in $N_G$, then $T^w$ is of finite presentation over $\text{Spec} \mathcal{R}$.

Proof. Suppose $w$ is in $N_G$. Let $u_1, \ldots, u_r$ be a basis for $M$, and choose $a_1, \ldots, a_r$ in $K^*$ such that $\nu(a_i) = -(u_i, w)$. Then $R[M]^w$ is generated over $R$ by
\[
\{a_1x^{u_1}, (a_1x^{u_1})^{-1}, \ldots, a_rx^{u_r}, (a_rx^{u_r})^{-1}\},
\]
and hence is of finite type. Furthermore, the relations are generated by
\[
\{1 - a_ix^{u_i}, (a_ix^{u_i})^{-1}\},
\]
so $R[M]^w$ is of finite presentation.

For the converse, suppose $w$ is not in $N_G$. Then, $M_{G,w}$ is a proper sublattice of $M$, and hence the special fiber $T^w$ has dimension strictly less than the dimension of $T$. Since $T^w$ is irreducible and fiber dimension is semicontinuous in irreducible families of finite type, it follows that $T^w$ is not of finite type. \qed

Remark A.2. Suppose the valuation is trivial, and let $S_w$ be the set of lattice points $u \in M$ such that $(u, w) \geq 0$. Then $S_w$ is a subsemigroup of $M$, and $R[M]^w$ is naturally identified with the semigroup ring $K[S_w]$. The semigroup $S_w$ is finitely generated if and only if the ray spanned by $w$ has rational slope, and it follows that $T^w$ is of finite type over $K$ if and only if $w$ is a scalar multiple of a lattice point.

When the schemes $T^w$ and $X^w$ are not of finite type, there are many technical difficulties in handling them directly. These technical difficulties can be overcome by extending scalars, since these schemes become finite type after a suitable base change, as follows.

Suppose the value group $G$ is a proper subgroup of $\mathbb{R}$, and $b$ is a real number that is not in $G$. Then there is a unique valuation $\tilde{\nu}$ on the function field $\tilde{K}(t)$ such that $\tilde{\nu}(t) = b$. This valuation extends to the algebraic closure $\overline{\tilde{K}}$ of $\tilde{K}(t)$. Iterating this procedure finitely many times, we can ensure that an arbitrary $w$ is rational over the value group of a suitable extension of $K$. In particular, for a suitable choice of extension $K'|K$, the scheme $\tilde{T}_w$ is of finite type over the valuation ring $\tilde{R}$.

Let $\tilde{K}|K$ be an arbitrary extension of valued fields, and let $\tilde{G}$ be the value group of $\tilde{K}$, with $\tilde{R}$ the valuation ring in $\tilde{K}$. For any closed subscheme $X$ of $T$ over $K$, the tropicalization of the base change $\tilde{X}$ is equal to the tropicalization of $X$ [Pay09a, Proposition 6.1], so the initial degeneration $\tilde{X}_w$ is nonempty if and only if $X_w$ is nonempty. Here we give a more precise geometric relationship between these initial degenerations. First we treat the associated schemes over the valuation rings. Let
\[
\varphi : \tilde{T}_w \to T_w
\]
be the natural map induced by the inclusion of tilted group rings, which is equivariant over the projection of tori induced by the inclusions $M_{G,w} \hookrightarrow M_{\tilde{G},w}$ and $R \hookrightarrow \tilde{R}$.
The scheme $\tilde{X}^w$ is the preimage of $X^w$ under $\varphi$.

**Proof.** It is clear that $\varphi$ maps $\tilde{X}^w$ into $X^w$. To show that $\tilde{X}^w$ is the full preimage of $X^w$, we must prove that any Laurent polynomial in the tilted group ring $\tilde{R}[M]^w$ that vanishes on $\tilde{X}$ is in the ideal generated by $I_X \cap R[M]^w$.

Let $f = \sum u \alpha(u)x^u$ be a nonzero Laurent polynomial over $\tilde{K}$ in $I_{\tilde{X}}$. Then $f$ can be written as

$$f = \alpha_1 f_1 + \cdots + \alpha_n f_n,$$

with $f_1, \ldots, f_n$ in $I_X$ and linearly independent over $K$, and $\alpha_i$ in $\tilde{K}$ all nonzero. Say $\alpha_i(u)$ is the coefficient of $x^u$ in $f_i$, so the coefficient of $x^u$ in $f$ is

$$\alpha(u) = \alpha_1 \alpha_1(u) + \cdots + \alpha_n \alpha_n(u).$$

Now, suppose $f$ is in $\tilde{R}[M]^w$. If each $\alpha_i f_i$ is in $\tilde{R}[M]^w$, then it is easy to see that $f$ is in the ideal generated by $I_X \cap R[M]^w$, by applying the case $n = 1$, below. The difficulty is that there may be some cancellation of leading terms in the above expression for $f$, by which we mean that the valuation of some coefficient $\alpha(u)$ may be strictly larger than $\min_i \{\tilde{v}(\alpha_1 \alpha_1(u))\}$. Roughly speaking, this means that the vector $(\alpha_1, \ldots, \alpha_n)$ is nearly orthogonal to $(\alpha_1(u), \ldots, \alpha_n(u))$.

When $n$ is greater than 1, we proceed by carefully eliminating one term in the summation, writing $f$ as a $\tilde{K}$-linear combination of $f_1, \ldots, f_{n-1}$ plus a single element of $\tilde{K}$ times a $K$-linear combination of $f_1, \ldots, f_n$. A suitable choice in the elimination procedure ensures that each of these two terms is in $\tilde{R}[M]^w$, and we deduce that they are in the ideal generated by $I_X \cap R[M]^w$, by induction on $n$. Roughly speaking, we choose $u_1, \ldots, u_{n-1}$ so that $(\alpha_1(u_i), \ldots, \alpha_n(u_i))$ are as close to orthogonal as possible to $(\alpha_1, \ldots, \alpha_n)$, for $1 \leq i \leq n - 1$. Then we replace $(\alpha_1, \ldots, \alpha_n)$ by the unique vector in $K^n$ that is orthogonal to $(\alpha_1(u_i), \ldots, \alpha_n(u_i))$ for $1 \leq i \leq n - 1$ and with $n$th coordinate 1. The corresponding $K$-linear combination of $f_1, \ldots, f_n$ is then multiplied by $\alpha_n$ and subtracted from $f$. The details are as follows.

Suppose $n = 1$. Let $m = \alpha_1(u)x^u$ be a monomial of lowest weight in $f_1$, and let $g = f_1/m$. Then $f$ can be expressed as

$$f = \alpha_1 mg,$$

with $g$ in $I_X \cap R[M]^w$, and $\alpha_1 m$ in $\tilde{R}[M]^w$, as required.

We proceed by induction on $n$. Given $u_1, \ldots, u_{n-1}$ in $M$, consider the matrix whose $(i, j)$th entry is $a_i(u_j)$, and let $\delta_i$ be the $i$th maximal minor

$$\delta_i = \begin{vmatrix} a_1(u_1) & \cdots & a_1(u_{n-1}) \\ \vdots & \ddots & \vdots \\ a_i(u_1) & \cdots & a_i(u_{n-1}) \\ \vdots & \ddots & \vdots \\ a_n(u_1) & \cdots & a_n(u_{n-1}) \end{vmatrix}$$
Since $f_1, \ldots, f_{n-1}$ are linearly independent over $K$, we may choose $u_1, \ldots, u_{n-1}$ so that $\delta_n$ is nonzero and  
\[ \tilde{\nu}(\alpha(u_1)) + \cdots + \tilde{\nu}(\alpha(u_{n-1})) - \nu(\delta_n) \]
is as large as possible. This choice is essential in the proof of the following claim. 
We claim that  
\[ h = \frac{\alpha_n}{\delta_n} \sum_{i=1}^{n} (-1)^{n-i} \delta_i f_i \]
is in $\tilde{R}[M]^w$. The claim implies that $h$ is in the ideal generated by $I_X \cap R[M]^w$, by the case $n = 1$, above, and also that the difference $f - h$ is in $\tilde{R}[M]^w$. The coefficient of $f_n$ in the above expression is equal to $\alpha_n$, by construction, so the difference $f - h$ can be written as a $\tilde{K}$-linear combination of $f_1, \ldots, f_{n-1}$. It follows by induction that $f - h$ is also in the ideal generated by $I_X \cap R[M]^w$, and this proves the theorem. 
It therefore remains to show that $h = \sum \beta(u)x^w$ is in $\tilde{R}[M]^w$, which means that the $w$-weight of each monomial $\beta(u)x^w$ is nonnegative. Fix one such monomial, write $u_n = u$ for its exponent, and let $A$ be the square $n \times n$ matrix whose $(i, j)$th entry is $a_i(u_j)$,  
\[ A = \begin{pmatrix} a_1(u_1) & \cdots & a_1(u_n) \\ \vdots & \ddots & \vdots \\ a_n(u_1) & \cdots & a_n(u_n) \end{pmatrix}. \]
Let $A_{ij}$ be the $(i, j)$th minor of $A$, the determinant of the submatrix obtained by deleting the $i$th row and $j$th column. So $\delta_i = A_{in}$. Expanding $\det A$ in the last column shows that  
\[ \beta(u_n) = \frac{\alpha_n}{A_{nn}} \det A. \]
Therefore, the valuation of $\beta(u_n)$ is  
\[ \tilde{\nu}(\beta(u_n)) = \nu(\det A) - \nu(A_{nn}) + \tilde{\nu}(\alpha_n). \]
Since $f$ is in $\tilde{R}[M]^w$ by hypothesis, it will be enough to show that $\tilde{\nu}(\beta(u_n))$ is at least as large as $\tilde{\nu}(\alpha(u_n))$. To compare $\beta(u_n)$ with $\alpha(u_n)$, we consider the matrix  
\[ A' = \begin{pmatrix} \alpha_1 a_1(u_1) & \cdots & \alpha_1 a_2(u_n) \\ \vdots & \ddots & \vdots \\ \alpha_{n-1} a_{n-1}(u_1) & \cdots & \alpha_{n-1} a_{n-1}(u_n) \\ \alpha(u_1) & \cdots & \alpha(u_n) \end{pmatrix}. \]
Recall that the coefficient $\alpha(u_j)$ in the bottom row is $\alpha_1 a_1(u_j) + \cdots + \alpha_n a_n(u_j)$, so the determinant of $A'$ is $\alpha_1 \cdots \alpha_n \det A$. Expanding $\det A'$ in the last row gives also  
\[ \det A' = \alpha_1 \cdots \alpha_{n-1} \left( \sum_{i=1}^{n} (-1)^{n-i} \alpha(u_i) A_{ni} \right), \]
and comparing these two expressions for $\det A'$ yields

$$\det A = \frac{1}{\alpha_n} \left( \sum_{i=1}^{n} (-1)^{n-i} \alpha(u_i) A_{ni} \right).$$

Therefore,

$$\nu(\det A) \geq \min_i \{ \nu(\alpha(u_i)) + \nu(A_{ni}) \} - \nu(\alpha_n).$$

Now $u_1, \ldots, u_{n-1}$ were chosen so that this minimum occurs at $i = n$. Substituting the resulting inequality for $\nu(\det A)$ into the expression for $\tilde{\nu}(\beta(u_n))$ above shows that $\tilde{\nu}(\beta(u_n))$ is greater than or equal to $\tilde{\nu}(\alpha(u_n))$. This proves the claim, and the theorem follows.

We now pass to the initial degenerations. Let

$$\phi: \tilde{T}_w \to T_w$$

be the natural projection of torus torsors induced by the inclusions of tilted group rings, modulo monomials of strictly positive $w$-weight.

**Theorem A.4.** The initial degeneration $\tilde{X}_w$ is the preimage of $X_w$ under $\phi$.

**Proof.** It is clear that $\phi$ maps $\tilde{X}_w$ into $X_w$. Any element $f_w$ of the ideal of $\tilde{X}_w$ is the residue of a Laurent polynomial $f$ in $I_{\tilde{X}} \cap \tilde{R}[M]^w$, which is then in the ideal generated by $I_X \cap \tilde{R}[M]^w$, by Theorem A.3. Taking residues shows that $f_w$ in the ideal generated by the pullback of $I_{X_w}$, and it follows that $\tilde{X}_w$ is the full preimage of $X_w$. \qed

**Remark A.5.** Since $\phi$ is smooth and has connected fibers, it follows that many geometric properties of initial degenerations are preserved under extensions of valued fields. For instance, the sum of the multiplicities of the irreducible components of $X_w$ is equal to that of $\tilde{X}_w$, which is helpful for defining tropical multiplicities. Most importantly for our purposes, a point $\tilde{x}$ is smooth in $\tilde{Y}_w$ if and only if $\phi(\tilde{x})$ is smooth in $Y_w$, $\tilde{X}_w$ meets $\tilde{X}'_w$ properly at $\tilde{x}$ if and only if $X_w$ meets $X'_w$ properly at $\phi(\tilde{x})$, and $\tilde{x}$ is in $(\tilde{X} \cap \tilde{X}')_w$ if and only if $\phi(\tilde{x})$ is in $(X \cap X')_w$.

**Appendix B. Topology of finite type morphisms**

Because the results may be of independent interest, we explain how Theorem 4.2.5 on the existence of closed points in fibers, and Theorem 4.1.3 on lifting points of intersection, can both be extended to an arbitrary base scheme. The proof of Proposition 4.2.2 does not use that the base scheme is the spectrum of a valuation ring of rank 1, and in fact yields the following result.

**Proposition B.1.** Let $X \to S$ be a flat morphism of finite type of irreducible schemes, and suppose that $D$ is a locally principal closed subscheme of $X$ that does not meet the generic fiber. Then, for every $s \in S$, every irreducible component of the fiber $D_s$ is an irreducible component of $X_s$. 
Remark B.2. Some hypothesis such as flatness is necessary for such results on locally principal subschemes over general base schemes, as shown by the following example. Suppose $X \to S$ is the blowup of the affine plane at the origin. Then the strict transform $D$ of a line through the origin is locally principal and does not meet the generic fiber, but its intersection with the exceptional fiber is a single point.

We can now prove the first stated result.

**Theorem B.3.** Let $X \to S$ be a morphism locally of finite type, and let $s$ be a specialization of $s'$ in $S$. Suppose $x'$ is a point in $X_{s'}$ specializing to a closed point $x$ in $X_s$. Then there is a closed point $x''$ in $X_{s'}$ which specializes to $x$, and such that $x'$ specializes to $x''$. Moreover, the set of such $x''$ is Zariski dense in the closure of $x'$ inside $X_{s'}$. More generally, if $x$ is not necessarily closed in $X_s$, and $k(x)$ denotes the residue field of $x$, we can choose $x''$ to satisfy the inequality

$$ (B.1) \quad \text{trdeg} k(x'')/k(s') \leq \text{trdeg} k(x)/k(s), $$

and again the choices of $x''$ are Zariski dense in the closure of $x'$ inside $X_{s'}$. Moreover, if $S$ is the spectrum of a valuation ring, we have equality in (B.1).

Note that the example of Remark B.2 shows that we cannot do better than the inequality (B.1) for a general base scheme.

**Proof.** First, in light of the generalized Proposition B.1, the argument of Theorem 4.2.5 goes through to prove the desired result in the case that $S$ is an arbitrary valuation ring, with $s'$ the generic point of $S$. The only subtlety is that in the general case, the components of $D$ not containing $x$ do not necessarily form a closed subset. Nonetheless, we can pass to an open neighborhood of $x$ such that every component of $D$ meeting the generic fiber (if there are any) must contain $x$, and this suffices for the argument. To remove the restriction that $s'$ be the generic point of $S$, we simply note that the closure of $s'$ is again the spectrum of a valuation ring.

We thus wish to reduce to the valuation ring case. Replacing $X$ by the closure of $x'$, we may assume $X$ is integral and $x'$ is its generic point. By [Gro61a Proposition 7.1.4(ii)], there is a valuation ring $A$ in $k(x')$ with a dominant morphism $\text{Spec } A \to X$

mapping the closed point to $x$. Let $R$ be the valuation ring in the residue field $k(s')$ given by intersecting with $A$ in $k(x')$, and let $X' = X \times_S \text{Spec } R$. Then the following diagram is commutative.

$$
\begin{array}{ccc}
X' & \longrightarrow & X \\
\downarrow & & \downarrow \\
\text{Spec } R & \longrightarrow & S
\end{array}
$$

Documenta Mathematica 18 (2013) 121–175
By construction the map from Spec $A$ to $X$ factors through $X'$. So, by the case we have already handled, there is a point $x''$ in the generic fiber of $X'$ over Spec $R$ specializing to the image $\tilde{x}$ in $X'$ of the closed point of Spec $A$ and satisfying the desired equality of residue field extensions; moreover, such points are Zariski dense in the generic fiber. Since $k(s')$ is identified with the fraction field of $R$, by construction, the generic fiber of $X'$ maps isomorphically to the generic fiber of $X$. Finally, observing that we have the inequality

$$\text{trdeg} \kappa(\tilde{x})/(R/m_R) \leq \text{trdeg} \kappa(x)/k(s),$$

we conclude the desired statement. □

Naive statements about global codimension and subadditivity do not extend from valuation rings of rank 1 to valuation rings of higher rank, as shown by Example 4.3.2. Nevertheless, hypotheses on codimension of intersection can still yield lifting results even when subadditivity of codimension fails, as demonstrated by the following theorem.

**Theorem B.4**. Let $Y \to S$ be a smooth morphism, and let $X$ and $X'$ be closed subschemes of $Y$, flat over $S$, such that the codimension in $Y'$ of every component of $X$ and every component of $X'$ is less than or equal to $j$ and $j'$, respectively. Suppose that, for some $s \in S$, the fibers $X_s$ and $X'_s$ meet in codimension $j + j'$ at a point $x$ in $Y_s$. Then for any $s' \in S$ specializing to $s$:

1. There is a point $x'$ in $X_s' \cap X'_s$ specializing to $x$.
2. If $x$ is closed in $Y_s$, then $x'$ may be chosen to be closed in $Y'_s$. More generally, we may choose $x'$ so that we have

$$\text{trdeg} \kappa(x')/k(s') \leq \text{trdeg} \kappa(x)/k(s).$$

**Proof.** In light of Theorem B.3, the second assertion follows immediately from the first. Observe that the flatness hypotheses mean that the hypotheses of the theorem are preserved under arbitrary base change. In particular, we reduce to the case that $S$ is the spectrum of a valuation ring, with $s'$ the generic point. We then prove the desired statement with a reduction to the diagonal and inductive application of Proposition B.1, using that $S$ is the spectrum of a valuation ring to preserve the flatness hypothesis. □

**Remark B.5.** In Example 4.3.2 the special fibers of the two subschemes coincide, and the generic fibers are disjoint. This does not contradict Theorem B.4 because the special fibers do not meet properly, and the failure of subadditivity is for simple numerical reasons. The intersection that does not lift has dimension one larger than expected, but is supported in a fiber of codimension $r$, which is greater than 1.

**Appendix C. An application to tropical elimination theory**

Let $X$ be an irreducible closed subscheme of $T$ and let $\varphi : T \to T'$ be a homomorphism of tori that induces a generically finite morphism from $X$ to the closure of its image, which we denote $X'$. Then, set theoretically, $\text{Trop}(X')$
is the image of $\text{Trop}(X)$ under the induced linear map $\phi : N_R \to N'_{R'}$. The fundamental problem of tropical elimination theory, solved by Sturmfels and Tevelev for the special case where the valuation is trivial, is to determine the multiplicities on the facets of $\text{Trop}(X')$. Here we use tropical lifting theorems to generalize [ST08, Theorem 1.1] to the case of a nontrivial valuation. See also [BPR11, Section 8] for an analytic proof of this result and applications to curves.

After subdividing, we may assume that $\phi$ maps each face of $\text{Trop}(X)$ onto a face of $\text{Trop}(X')$.

**Theorem C.1.** The multiplicity of a facet $\sigma'$ in $\text{Trop}(X')$ is

$$m(\sigma') = \frac{1}{\delta} \sum_{\phi(\sigma) = \sigma'} m(\sigma) \cdot [N'_{\sigma} : \phi(N_\sigma)],$$

where $\delta$ is the degree of $\phi$.

Theorem C.1 and Corollary [4.4.6] together imply that tropicalization of cycles commutes with push forward. See [Gub12, Theorem 13.17].

**Proof.** We prove the theorem by intersecting $X$ and $X'$ with suitable translates of subtori and then counting points using tropical intersection theory and lifting theorems. First, we choose the translated subtori to ensure that these intersections occur in a locus where $\alpha$ is well-behaved.

Since $\phi$ is generically finite, there is a dense open subset $U' \subset X'$ such that the induced map $\varphi^{-1}(U') \rightarrow U'$ is finite of degree $\delta$ [Har77, Exercise II.3.7]. Shrinking $U'$ further, if necessary, we may assume that $U'$ is smooth and $\varphi^{-1}(U')$ is flat over $U'$, so the preimage of a zero-dimensional subscheme of length $m$ in $U$ is a zero-dimensional subscheme of length $\delta \cdot m$.

Let $\Lambda'$ be a sublattice of $N'$ complementary to $N'_{\sigma'}$, so $N'$ splits as a direct sum

$$N' = N'_{\sigma'} \oplus \Lambda'.$$

We write $T'_{\Lambda'}$ for the subtorus of $T'$ whose lattice of one-parameter subgroups is $\Lambda'$. Let $\Lambda \subset N$ be the preimage of $\Lambda'$, with $T_{\Lambda}$ the associated subtorus of $T$. Let $\tilde{T}_{\Lambda}$ be the preimage of $T'_{\Lambda'}$, which is the product of $T_{\Lambda}$ with a zero-dimensional scheme of length $\ell = [N' : \phi(N)] / [\Lambda' : \phi(\Lambda)]$.

If the characteristic of $K$ is zero, or prime to $\ell$, then $\tilde{T}_{\Lambda}$ is a union of translates of $T_{\Lambda}$ by $\ell$ distinct torsion points.

We claim that, for any nonempty open subset $U \subset X$ there is an open dense set of $t \in T$ such that $t \tilde{T}_{\Lambda} \cap X$ is contained in $U$. To see this, consider the incidence subscheme $W$ in $T \times X$ parametrizing pairs $(t, x)$ such that $x$ is in $t \tilde{T}_{\Lambda}$. Then the first projection is dominant and generically finite, while the second projection is flat and maps $W$ surjectively onto $X$. Therefore, the preimage of $X \setminus U$ has positive codimension in $W$ and hence projects into a set of positive codimension in $T$. Therefore the complement of the closure of $p_1(p_2^{-1}(X \setminus U))$...
of these intersections that live in $\text{Trop}(X)$.

We fix
d and choose $v \in N_\mathbb{G}$ such that $v' = \phi(v)$ is in the relative interior of $\sigma'$. Since $\text{Trop}^{-1}(v)$ is Zariski dense in $T$, we can choose $t \in \text{Trop}^{-1}(v)$ in the open dense subset of $T$ described above, such that $t \tilde{T}_\Lambda \cap X$ is contained in $U$. Note that, since $v'$ is in the relative interior of the maximal face $\sigma'$ of $\text{Trop}(X')$, it has finitely many preimages $v_1, \ldots, v_r$ in $\text{Trop}(X)$, one in each maximal face $\sigma_i$ mapping onto $\sigma'$. Let $\sigma_i$ be the maximal face of $\text{Trop}(X)$ containing $v_i$.

Let $t' = \varphi(t)$. We now consider $t'T'_{\Lambda'} \cap X'$ and $t \tilde{T}_\Lambda \cap X$, and especially the parts of these intersections that live in $\text{Trop}^{-1}(v')$ and $\text{Trop}^{-1}(v_i)$, for $1 \leq i \leq r$, respectively. By construction, $\text{Trop}(t'T'_{\Lambda'})$ is the affine linear space $A_{\mathbb{G}} + v'$ with multiplicity 1, and meets $\text{Trop}(X')$ transversally at $v'$. The translation of $t'T'_{\Lambda'}$, by a sufficiently small vector in $N_{\mathbb{G}}$, also meets $\sigma'$ transversally at a single point, so the fan displacement rule gives the local tropical intersection multiplicity as

$$i(v', \text{Trop}(X') \cdot \text{Trop}(t'T'_{\Lambda'})) = m(\sigma') \cdot [N' : N_{\sigma'} + \Lambda'].$$

By the choice of $t$, the intersection of $t'T'_{\Lambda'}$ with $X'$ is contained in the smooth locus of $X'$. In particular, both are Cohen-Macaulay along their intersection in $\text{Trop}^{-1}(v')$. Therefore, Corollary 5.1.3 says that the intersection of $X'$ with $t'T'_{\Lambda'}$, in $\text{Trop}^{-1}(v')$ is a zero-dimensional scheme $Z'$ of length $i(v', \text{Trop}(X') \cdot \text{Trop}(t'T'_{\Lambda'}))$. Similarly, $\text{Trop}(t \tilde{T}_\Lambda)$ is the affine linear space $A_{\mathbb{G}} + v$ with multiplicity $\ell$, and meets $\text{Trop}(X)$ transversally at $v_i$ with local tropical intersection multiplicity

$$i(v, \text{Trop}(X) \cdot \text{Trop}(t \tilde{T}_\Lambda)) = \ell \cdot m(\sigma_i) \cdot [N : N_{\sigma_i} + \Lambda].$$

Both $t \tilde{T}_\Lambda$ and $X$ are smooth and hence Cohen-Macaulay along their intersection, so the intersection of $t \tilde{T}_\Lambda$ with $X$ in $\text{Trop}^{-1}(v_i)$ is a zero-dimensional scheme $Z_i$ of length $i(v, \text{Trop}(X) \cdot \text{Trop}(t \tilde{T}_\Lambda))$.

By the choice of $t$, the map $\varphi$ is finite of degree $\delta$ in a neighborhood of $Z'$. Furthermore, the preimage $\varphi^{-1}(Z')$ is exactly $Z_1 \cup \cdots \cup Z_r$. Therefore,

$$\text{length}(Z') = \frac{1}{\delta} \cdot \text{length}(Z_1) + \cdots + \text{length}(Z_r)).$$

Substituting the above tropical intersection multiplicities for these lengths gives the identity

$$m(\sigma') \cdot [N' : N_{\sigma'} + \Lambda'] = \frac{\ell}{\delta} \cdot \sum_{i=1}^{r} m(\sigma_i) \cdot [N : N_{\sigma_i} + \Lambda].$$

Now $\sigma_1, \ldots, \sigma_r$ are exactly the faces of $\text{Trop}(X)$ that map onto $\sigma$, and $\ell = [N' : \phi(N)]/[N : \phi(\Lambda)]$. By rearranging terms, one then sees that, to prove the theorem, it suffices to show

$$[N' : \phi(N)] \cdot [N : N_{\sigma} + \Lambda] = [N'_{\sigma} : \phi(N_{\sigma})] \cdot [\Lambda' : \phi(\Lambda)].$$
for $1 \leq i \leq r$. Both sides are equal to $[N' : \phi(N_\sigma + \Lambda)]$, and the theorem follows. □

**References**


Brian Osserman and Sam Payne


[Har77], R. Hartshorne, Algebraic geometry, Springer-Verlag, New York, 1977, Graduate Texts in Mathematics, No. 52.


Cycle Classes for $p$-adic Étale Tate Twists and the Image of $p$-adic Regulators

Kanetomo Sato$^1$

Received: December 23, 2011
Revised: February 8, 2013
Communicated by Stephen Lichtenbaum

Abstract. In this paper, we construct Chern class maps and cycle class maps with values in $p$-adic étale Tate twists $[Sa2]$. We also relate the $p$-adic étale Tate twists with the finite part of Bloch-Kato. As an application, we prove that the integral part of $p$-adic regulator maps has values in the finite part of Galois cohomology under certain assumptions.

2000 Mathematics Subject Classification: Primary 19F27, 14F30; Secondary 14C25

1 Introduction

Let $p$ be a rational prime number. Let $A$ be a Dedekind ring whose fraction field has characteristic zero and which has a residue field of characteristic $p$. Let $X$ be a regular scheme of pure-dimension which is flat of finite type over $S := \text{Spec}(A)$ and which is a smooth or semistable family around its fibers over $S$ of characteristic $p$. Extending the idea of Schneider [Sc], the author defined in $[Sa2]$ the objects $\Xi_n(r)_X (r, n \geq 0)$ of the derived category of étale $\mathbb{Z}/p^n$-sheaves on $X$ playing the role of the $r$-th Tate twist with $\mathbb{Z}/p^n$-coefficients, which are endowed with a natural product structure with respect to $r$ and both contravariantly and covariantly functorial (i.e., there exist natural pull-back and trace morphisms) for arbitrary separated morphisms of finite type of such schemes. Those pull-back and trace morphisms satisfy the projection formula.

The first aim of this paper is to construct the following Chern class map and cycle class map for $m, r \geq 0$:

$$
\begin{array}{ccc}
K_m(X) & \xrightarrow{\text{(missing)}} & C_{r,m} \quad \text{(Chern class)} \\
\text{CH}^r(X, m) & \xrightarrow{\sigma^r_{X,m}} & H^{2r-m}_{\text{ét}}(X, \Xi_n(r)_X),
\end{array}
$$

$^1$supported by JSPS Core-to-Core Program 18005 and Grant-in-Aid for Scientific Research B-23340003
where $K_m(X)$ denotes the algebraic $K$-group $[Q]$ and $\text{CH}^r(X, m)$ denotes the higher Chow group $[B2]$. The Chern classes with values in higher Chow groups have not been defined in this arithmetic situation for the lack of a product structure on them, which we do not deal with in this paper. We will prepare the following results to construct the above maps:

(a) The Dold-Thom isomorphism, i.e., the projective bundle formula for $p$-adic étale Tate twists. See Theorem 4.1 below for details. We note here that the $p$-adic étale Tate twists do not satisfy homotopy invariance.

(b) For a regular closed immersion $i : X_\ast \hookrightarrow X'_\ast$ of codimension $c$ of simplicial schemes for which the $p$-adic étale Tate twists are defined, we construct a Gysin morphism $\text{gys}_i : \mathcal{I}(r)_{X_\ast} \longrightarrow \text{Ri}^!\mathcal{I}(r+c)_{X'_\ast} \otimes [2c]$ in $D((X_\ast)_{\text{et}}, \mathbb{Z}/p^n)$ satisfying transitivity and projection formula. See Proposition 5.4 below for details.

(c) We introduce a version of $p$-adic étale Tate twists with log poles along horizontal normal crossing divisors (see §3 below), and prove a $p$-adic analogue of the usual homotopy invariance (see Corollary 4.3 below) and a semi-purity property (see Theorem 6.5 below) for this new coefficient.

The existence of $C_{r,m}$ will be verified by (a) and the general framework due to Gillet [Gi1]. The additivity of $C_{r,m}$ for $m \geq 1$ will follow from (b) (cf. §5). On the other hand, we will need the results in (c) to construct $\text{cl}^m_{X}$ (cf. §7). This ‘higher’ cycle class map will be a fundamental object to study in ‘higher’ higher classfield theory [Sai]. We will mention a local behavior of $\text{cl}^m_X$ in Remark 7.2 below.

The second aim of this paper is to relate the $p$-adic étale Tate twists with the finite part of Galois cohomology [BK2], using the Fontaine-Jannsen conjecture proved by Hyodo, Kato and Tsuji ([HK], [K4], [Ts1], cf. [Ni2]). We assume here that $A$ is a $p$-adic integer ring and that $X$ is projective over $A$ with strict semistable reduction. Let $K$ be the fraction field of $A$ and put $X_K := X \otimes_A K = X[p^{-1}]$. We define

$$H^i(X, \mathcal{I}_{Q_p}(r)_X) := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \geq 1} H^i_{\text{et}}(X, \mathcal{I}_n(r)_X),$$

$$H^i(X_K, \mathcal{I}_{Q_p}(r)) := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \geq 1} H^i_{\text{et}}(X_K, \mu_{p^n} \otimes r),$$

where $\mu_{p^n}$ denotes the étale sheaf of $p^n$-th roots of unity on $X_K$, i.e., the usual Tate twist on $X_K$. We have a natural restriction map $H^i(X, \mathcal{I}_{Q_p}(r)_X) \rightarrow H^i(X_K, \mathcal{I}_{Q_p}(r))$ and a canonical descending filtration $F^\ast$ on $H^i(X_K, \mathcal{I}_{Q_p}(r))$ resulting from the Hochschild-Serre spectral sequence for the covering $X_K \leftarrow X_K \otimes K \rightarrow X_K$ (cf. (9.0.1)). We define a (not necessarily exhaustive) filtration $F^\ast$ on $H^i(X, \mathcal{I}_{Q_p}(r)_X)$ as the inverse image of $F^\ast$ on $H^i(X_K, \mathcal{I}_{Q_p}(r))$, which induces obvious inclusions for $m \geq 0$

$$\text{gr}^m F^i H^i(X, \mathcal{I}_{Q_p}(r)_X) \subseteq \text{gr}^m F^i H^i(X_K, \mathcal{I}_{Q_p}(r)) \approx H^m(K, H^{i-m}(X_K, \mathcal{I}_{Q_p}(r))).$$
Here $H^i(K, -)$ denotes the continuous Galois cohomology of the absolute Galois group $G_K = \text{Gal}(\overline{K}/K)$ defined by Tate [Ta]. We will prove that

$$\text{gr}_p^1 H^i(X, \mathcal{T}_{\mathbb{Q}_p}(r)_{\chi}) \subset H^i_f(K, H^{i-1}(X_{\mathbb{P}^1}, \mathbb{Q}_p(r))) \quad (f = \text{finite part})$$

assuming that $p$ is sufficiently large and that the monodromy-weight conjecture [Mo] holds for the log crystalline cohomology of the reduction of $X$ in degree $i - 1$ (see Theorem 9.1 below). This result is an extension of the $p$-adic point conjecture ([Sc], [LS], [Ne1]) to the semistable reduction case and gives an ‘unramified’ version of results of Langer [La] and Nekovář [Ne2] relating the log syntomic cohomology of $X$ with the geometric part $H^i_f(K, H^{i-1}(X_{\mathbb{P}^1}, \mathbb{Q}_p(r)))$.

There is an application of the above results as follows. Let $K$ be a number field and let $V$ be a proper smooth geometrically integral variety over $K$. Put $\overline{V} := V \otimes_K \overline{K}$. Let $i$ and $r$ be non-negative integers with $2r \geq i + 1$, and let $p$ be a prime number. The étale Chern characters (cf. [So]) induce $p$-adic regulator maps

$$\text{reg}^{2r-i-1,r}_p : K_{2r-i-1}(V)_\text{o} \longrightarrow H^i_f(K, H^i(\overline{V}, \mathbb{Q}_p(r))) \quad (2r > i + 1),$$

$$\text{reg}^{0,r}_p : K_0(V)_{\text{hom}} \longrightarrow H^1_f(K, H^{2r-1}(\overline{V}, \mathbb{Q}_p(r))) \quad (2r = i + 1).$$

Here $K_0(V)_{\text{hom}}$ denotes the homologically trivial part of $K_0(V)$, and $K_{m}(V)_o$ denotes the integral part of $K_{m}(V)$ in the sense of Scholl (see §10 below). Motivated by the study of special values of $L$-functions, Bloch and Kato [BK2] conjecture that the image of $\text{reg}^{2r-i-1,r}_p$ is contained in the finite part $H^i_f(K, H^i(\overline{V}, \mathbb{Q}_p(r)))$ and spans it over $\mathbb{Q}_p$. In the direction of this conjecture, we will prove the following result, which extends a result of Nekovář [Ne2] Theorem 3.1 on $\text{reg}^{0,r}_p$ to the case $2r > i + 1$ and extends a result of Niziol [Ni1] on the potentially good reduction case to the general case:

**Theorem 1.1** (§10) **Assume $r \leq p - 2$ and the monodromy-weight conjecture for the log crystalline cohomology of degree $i$ of projective strict semistable varieties over $\mathbb{F}_p$. Then we have**

$$\text{Im}(\text{reg}^{2r-i-1,r}_p) \subset H^i_f(K, H^i(\overline{V}, \mathbb{Q}_p(r))).$$

Here a projective strict semistable variety over $\mathbb{F}_p$ means the reduction of a regular scheme which is projective flat over a $p$-adic integer ring with strict semistable reduction.

We use the alteration theorem of de Jong [dJ] to prove Theorem 1.1, and the projective strict semistable varieties concerned in the assumption mean those obtained from alterations of scalar extensions of $V$ to the completion of $K$ at places dividing $p$.

This paper is organized as follows. In §2, we introduce cohomological and logarithmic Hodge-Witt sheaves with horizontal log poles on normal crossing varieties over a field of characteristic $p > 0$. In §3, we define $p$-adic étale Tate twists with horizontal log poles, and construct a localization sequence using this object (Theorem 3.12). In §4 and §5, we prove the Dold-Thom isomorphisms and define the Chern
class maps for $p$-adic étale Tate twists. The sections 6 and 7 will be devoted to the construction of cycle class maps for $p$-adic étale Tate twists. In §8 we will introduce Hodge-Witt cohomology and homology of normal crossing varieties and prove that the monodromy-weight conjecture implies a certain invariant cycle theorem. In §9, we establish the comparison between $p$-adic étale Tate twists and the finite part of Bloch-Kato. We will prove Theorem 1.1 in §10. In the appendix, we will formulate a continuous crystalline cohomology and a continuous syntomic cohomology to prove several technical compatibility results which will have been used in §9.

The author expresses his gratitude to Professors Shuji Saito and Takeshi Tsuji, and Masanori Asakura for valuable comments and discussions on the subjects of this paper.

Notation

For an abelian group $M$ and a positive integer $n$, $nM$ and $M/n$ denote the kernel and the cokernel of the map $M \to M$, respectively. For a field $k$, $\overline{k}$ denotes a fixed separable closure, and $G_k$ denotes the absolute Galois group $\text{Gal}(\overline{k}/k)$. For a topological $G_k$-module $M$, $H^*(k, M)$ denote the continuous Galois cohomology groups $H^*_{\text{cont}}(G_k, M)$ in the sense of Tate [Ta]. If $M$ is discrete, then $H^*(k, M)$ agree with the étale cohomology groups of $\text{Spec}(k)$ with coefficients in the étale sheaf associated with $M$.

Unless indicated otherwise, all cohomology groups of schemes are taken over the étale topology. For a scheme $X$, an abelian sheaf $\mathcal{F}$ on $X_{\text{ét}}$ (or more generally an object in the derived category of abelian sheaves on $X_{\text{ét}}$) and a point $x \in X$, we often write $H^*_e(X, \mathcal{F})$ for $H^*_e(\text{Spec}(\mathcal{O}_{X,x}), \mathcal{F})$. For a prime number $p$ which is invertible on $X$, we write $\mu_p$ for the étale sheaf of the $p$-th roots of unity on $X$. For a prime number $p$ which is invertible on $X$ and integers $m, r \geq 0$, we define

$$H^m(X, \mathbb{Z}_p(r)) := \lim_{\substack{\to \atop n \geq 1}} H^i(X, \mu_p^\otimes r^n),$$
$$H^m(X, \mathbb{Q}_p(r)) := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} H^m(X, \mathbb{Z}_p(r)).$$

For an equi-dimensional scheme $X$ and a non-negative integer $q$, we write $X^q$ for the set of all points on $X$ of codimension $q$.

For a scheme (or a simplicial scheme) $X$ and an integer $n \geq 2$, $\text{Shv}(X_{\text{ét}}, \mathbb{Z}/n)$ denotes the category of étale $\mathbb{Z}/n$-sheaves on $X_{\text{ét}}$, and $D(X_{\text{ét}}, \mathbb{Z}/n)$ denotes the derived category of complexes of objects of $\text{Shv}(X_{\text{ét}}, \mathbb{Z}/n)$.

2 Logarithmic Hodge-Witt sheaves

We first fix the following terminology.

**Definition 2.1**

1. A normal crossing variety over a field $k$ is a pure-dimensional scheme which is separated of finite type over $k$ and everywhere étale locally
isomorphic to
\[ \text{Spec}(k[t_0, \ldots, t_N]/(t_0 \cdots t_a)) \quad \text{for some} \quad 0 \leq a \leq N := \dim(Y). \]

(2) We say that a normal crossing variety \( Y \) is simple if all irreducible components of \( Y \) are smooth over \( k \).

(3) An admissible divisor on a normal crossing variety \( Y \) is a reduced effective Cartier divisor \( D \) such that the immersion \( D \hookrightarrow Y \) is everywhere étale locally isomorphic to
\[ \text{Spec}(k[t_0, \ldots, t_N]/(t_0 \cdots t_a, t_a+1 \cdots t_{a+b})) \hookrightarrow \text{Spec}(k[t_0, \ldots, t_N]/(t_0 \cdots t_a)) \]
for some \( a, b \geq 0 \) with \( a + b \leq N = \dim(Y) \).

Let \( p \) be a prime number, and let \( k \) be a field of characteristic \( p \). Let \( Y \) be a normal crossing variety over \( k \), and let \( D \) be an admissible divisor on \( Y \). Put \( V := Y - D \) and let \( f \) and \( g \) be as follows:
\[ D \twoheadrightarrow f \rightarrow \rightarrow V \quad \quad \quad \quad V \twoheadleftarrow g \leftarrow \leftarrow Y = Y - D. \]

For \( r \in \mathbb{Z} \), we define étale sheaves \( \nu^r_{(Y,D),n} \) and \( \lambda^r_{(Y,D),n} \) on \( Y \) as follows:
\[
\lambda^r_{(Y,D),n} := \ker \left( \partial : \bigoplus_{x \in V} i_x \ast W^r_x \Omega^q_{x, \log} \rightarrow \bigoplus_{x \in V} i_x \ast W^r_{x-1} \Omega^q_{x, \log} \right).
\]
Here for \( x \in V \), \( i_x \) denotes the composite map \( x \hookrightarrow V \hookrightarrow Y \), and \( W^q_x \Omega^q_{x, \log} \) denotes the étale subsheaf of the logarithmic part of the Hodge-Witt sheaf \( W^q_x \Omega^q_x \) (cf. [B1], [I1]), which means the zero sheaf for \( q < 0 \). The map \( \partial \) is the sum of boundary maps due to Kato [K1]. By definition, we have
\[
\lambda^r_{(Y,D),n} = \nu^r_{(Y,D),n} = 0 \quad \text{for} \quad r < 0.
\]
When \( D = \emptyset \), we put
\[
\nu^r_{Y,n} := \nu^r_{(Y,\emptyset),n} \quad \text{and} \quad \lambda^r_{Y,n} := \lambda^r_{(Y,\emptyset),n},
\]
which are considered in [Sa1]. Although we assumed the perfectness of \( k \) in [Sa1], the results in loc. cit. \( \S \S \) 2.1–2.4 and \( \S \S \) 3.1–3.2 are extended to the case that \( k \) is not necessarily perfect by results of Shiho [Sh] Theorems 3.2, 4.1 and the compatibility mentioned in [AS] Remark 7.3 (1). We have \( \nu^r_{(Y,D),n} = g_* \nu^r_{Y,n} \) by the left exactness of \( g_* \), and \( \nu^{r-1}_{D,n} \cong R^1 f_* \nu^r_{Y,n} \) by the purity of \( \nu^r_{Y,n} \) ([Sa1] Theorem 2.4.2). Hence there is a short exact sequence
\[
0 \twoheadrightarrow \nu^r_{Y,n} \twoheadrightarrow \nu^r_{(Y,D),n} \twoheadrightarrow f_* \nu^{r-1}_{D,n} \twoheadrightarrow 0. \quad (2.1.1)
\]
By this fact, we have

\[ \lambda_r^r(Y,D) = \nu^r(Y,D) = g_* W_n \Omega^r_{Y,\log} \]

if \( Y \) is smooth (loc. cit. (2.4.9)), which we denote by \( W_n \Omega^r_{Y,\log} \). The following proposition will be useful later, where \( Y \) is not necessarily smooth:

**Proposition 2.2** Assume that \( Y \) is simple, and let \( Y_1, Y_2, \ldots, Y_q \) be the distinct irreducible components of \( Y \). Then there is an exact sequence on \( Y_\mathfrak{et} \)

\[ 0 \longrightarrow \lambda^r_{(Y,D),n} \longrightarrow \bigoplus_{|I|=1} W_n \Omega^r_{(Y_1,D_1),\log} \longrightarrow \bigoplus_{|I|=2} W_n \Omega^r_{(Y_1,D_1),\log} \longrightarrow \cdots \longrightarrow \bigoplus_{|I|=q} W_n \Omega^r_{(Y_1,D_1),\log} \longrightarrow 0, \]

where the notation \(|I| = t\) means that \( I \) runs through all subsets of \( \{1, 2, \ldots, q\} \) consisting of \( t \) elements, and for such \( I = \{i_1, i_2, \ldots, i_t\} \) (\( i_j \)'s are pair-wise distinct), we put

\[ Y_I := Y_{i_1} \cap Y_{i_2} \cap \cdots \cap Y_{i_t} \quad \text{and} \quad D_I := D \times_Y Y_I. \]

The arrow \( r^0 \) denotes the natural restriction map. For \( (I, I') \) with \( I = \{i_1, i_2, \ldots, i_t\} \) (\( i_1 < i_2 < \cdots < i_t \)) and \( |I'| = t + 1 \), the \( (I, I') \)-factor of \( r^t \) is defined as

\[ \left\{ \begin{array}{ll}
0 & \text{(if } I \not\subset I') \\
(-1)^{t-n} \cdot (\beta_{I'I})^t & \text{(if } I' = I \cup \{i_{t+1}\} \text{ and } i_1 < \cdots < i_t < i_{t+1} < \cdots < i_{t+n}).
\end{array} \right. \]

where \( \beta_{I'I} \) denotes the closed immersion \( Y_{I'} \hookrightarrow Y_I \).

We need the following lemma to prove this proposition:

**Lemma 2.3** Assume that \( (Y, D, V) \) fits into cartesian squares of schemes

\[ \begin{array}{ccc}
D & \xrightarrow{f} & Y \\
\downarrow \circlearrowleft & & \downarrow \circlearrowleft \\
\mathcal{D} & \xrightarrow{h} & \mathcal{Y}
\end{array} \]

\[ \begin{array}{ccc}
\mathcal{D} & \xrightarrow{\psi} & \mathcal{V} \\
\downarrow \circlearrowleft & & \downarrow \circlearrowleft \\
\mathcal{D} & \xrightarrow{h} & \mathcal{Y}
\end{array} \]

such that \( \mathcal{Y} \) is regular, such that the vertical arrows are closed immersions and such that \( Y, \mathcal{D} \) and \( Y \cup \mathcal{D} \) are simple normal crossing divisors on \( \mathcal{Y} \), where we put \( \mathcal{Y}' := \mathcal{Y} \setminus \mathcal{D} \). Then the pull-back map \( \psi^* h_* \mathcal{O}_{\mathcal{Y}}^r \rightarrow g_* \mathcal{O}_{\mathcal{V}}^r \) on \( Y_\mathfrak{et} \) is surjective.

**Proof of Lemma 2.3.** We use the same notation as in Proposition 2.2. For a Cartier divisor \( E \) on a scheme \( Z \), let \( c_2^r(E) \in H^1_{\mathcal{E}l}(Z, \mathcal{O}_Z^r) \) be the localized first Chern class of the invertible sheaf \( \mathcal{O}_Z(E) \). Since \( \mathcal{Y} \) is regular, \( h_* \mathcal{O}_{\mathcal{Y}}^r \) is generated by \( \mathcal{O}_{\mathcal{Y}}^r \) and local uniformizers of the irreducible components \( \{ \mathcal{D}_j \}_{j \in J} \) of \( \mathcal{D} \). Put

\[ D_j := Y \times_{\mathcal{Y}} \mathcal{D}_j \quad (j \in J), \]

\[ \text{Documenta Mathematica 18 (2013) 177–247} \]
which is an admissible divisor on $Y$. Since $(g_*\mathcal{O}_Y^x)/\mathcal{O}_Y^x \simeq R^1f'_!\mathcal{O}_Y^x$, it is enough to show that the Gysin map
\[ \varphi : \bigoplus_{j \in J} \mathbb{Z}_{D_j} \rightarrow R^1f'_!\mathcal{O}_Y^x \]
sending $1 \in \mathbb{Z}_{D_j}$ to $c_{Y_1}(D_j)$ is bijective on $D_{\text{et}}$. By [Sa1] Lemma 3.2.2, there is an exact sequence on $Y_{\text{et}}$
\[ 0 \rightarrow \mathcal{O}_Y^x \stackrel{\varphi^0}{\rightarrow} \bigoplus_{|I|=1} \mathcal{O}_{Y_I}^{\times} \stackrel{\varphi^1}{\rightarrow} \bigoplus_{|I|=2} \mathcal{O}_{Y_I}^{\times} \cdots \stackrel{\varphi^{q-1}}{\rightarrow} \bigoplus_{|I|=q} \mathcal{O}_{Y_I}^{\times} \rightarrow 0, \quad (2.3.1) \]
where $\varphi^i$'s are defined in the same way as $\varphi^i$'s in Proposition 2.2, and $q$ denotes the number of the distinct irreducible components of $Y$. Since $f'_!\mathcal{O}_Y^x = 0$ for any non-empty subset $I \subset \{1, 2, \ldots, q\}$, the exactness of (2.3.1) implies that of the lower row of the following commutative diagram with exact rows:

\[
\begin{array}{cccccccc}
0 & \rightarrow & \bigoplus_{j \in J} \mathbb{Z}_{D_j} & \stackrel{\varphi^0}{\rightarrow} & \bigoplus_{|I|=1} \mathbb{Z}_{D_I \cap Y_I} & \stackrel{\varphi^1}{\rightarrow} & \bigoplus_{|I|=2} \mathbb{Z}_{D_I \cap Y_I} & \rightarrow & R^1f'_!\mathcal{O}_Y^x \\
& & \downarrow & & \downarrow & & \downarrow & & \\
0 & \rightarrow & \bigoplus_{j \in J} \mathbb{Z}_{D_j} & \stackrel{\varphi^0}{\rightarrow} & \bigoplus_{|I|=1} \mathbb{Z}_{D_I \cap Y_I} & \stackrel{\varphi^1}{\rightarrow} & \bigoplus_{|I|=2} \mathbb{Z}_{D_I \cap Y_I} & \rightarrow & R^1f'_!\mathcal{O}_Y^x \\
\end{array}
\]

where $D_j \cap Y_I$ is regular for each $j \in J$ and $I \subset \{1, 2, \ldots, q\}$ by the assumption that $Y' \cup \mathcal{D}$ has simple normal crossings on $X$. The middle and the right vertical arrows are defined in the same way as for $\varphi$, and bijective by the standard purity for $\mathcal{O}^x$ ([Gr1] III §6). Hence $\varphi$ is bijective as well.

**Proof of Proposition 2.2.** Since the problem is étale local on $Y$, we may assume that $(Y, D, V)$ fits into a diagram as in Lemma 2.3. Then there is an exact sequence
\[ W_n \Omega^r_{(Y, D), \log} \stackrel{\varphi^0}{\rightarrow} \bigoplus_{|I|=1} W_n \Omega^r_{(Y_I, D_I), \log} \stackrel{\varphi^1}{\rightarrow} \cdots \stackrel{\varphi^{q-1}}{\rightarrow} \bigoplus_{|I|=q} W_n \Omega^r_{(Y_I, D_I), \log} \rightarrow 0 \]
on $\mathfrak{m}$ by Lemma 2.3 and induction on the number of components of $Y$ which is similar as for [Sa1] Lemma 3.2.2. The assertion follows from this exact sequence. \[ \square \]

3 \hspace{1em} \textit{p-adic étale Tate twists with log poles}

In §§3–7, we are mainly concerned with the following setting.

**Setting 3.1** Let $A$ be a Dedekind domain whose fraction field has characteristic 0 and which has a maximal ideal of positive characteristic. Put
\[ S := \text{Spec}(A). \]
Let $p$ be a prime number which is not invertible in $A$. Let $X$ be a regular scheme which is flat of finite type over $A$ and whose fibers over the closed points of $S$ of characteristic $p$ are empty or reduced normal crossing divisors on $X$. We write $Y \subset X$ for the union of those fibers, which may be empty.

Let $D \subset X$ be a normal crossing divisor such that $D \cup Y$ has normal crossings on $X$ ($D$ may be empty). Put $U := X - (Y \cup D)$ and $V := Y - (Y \cap D)$, and consider a diagram of immersions

$$
\begin{array}{ccc}
V & \subset & X - D \\
\subset & \quad & \quad \\
\cap & \quad & \quad \\
Y & \subset & X
\end{array}
$$

Let $n$ and $r$ be positive integers. We first state the Bloch-Kato-Hyodo theorem on the structure of the sheaf $M_n^r := i^*R^qj_*\mu_p^\otimes r$, which will be useful in this paper. We define the étale sheaf $\mathcal{M}_r^M$ on $Y$ as

$$
\mathcal{M}_r^M := (i^*j_*\mathcal{O}_U^\times)^\otimes n / J,
$$

where $J$ denotes the subsheaf of $(i^*j_*\mathcal{O}_U^\times)^\otimes n$ generated by local sections of the form $a_1 \otimes a_2 \otimes \cdots \otimes a_r$ with $a_s + a_t = 0$ or $1$ for some $1 \leq s < t \leq n$. There is a homomorphism of étale sheaves ([BK1] 1.2)

$$
\mathcal{M}_r^M \longrightarrow M_n^r,
$$

(3.1.1)

which is a geometric version of Tate’s Galois symbol map. For local sections $a_1, a_2, \ldots, a_r \in i^*j_*\mathcal{O}_U^\times$, we denote the class of $a_1 \otimes a_2 \otimes \cdots \otimes a_r$ in $\mathcal{M}_r^M$ by $\{a_1, a_2, \ldots, a_r\}$, and denote the image of $\{a_1, a_2, \ldots, a_r\}$ in $\mathcal{M}_r^M$ under (3.1.1) again by $\{a_1, a_2, \ldots, a_r\}$. We define filtrations $\mathcal{H}^*$ and $\mathcal{Y}^*$ on $M_n^r$ as follows.

**Definition 3.2** Put $p := \ker(\mathcal{O}_X \to i_*\mathcal{O}_Y)$ and $1 + p^q := \ker(\mathcal{O}_X^\times \to (\mathcal{O}_X/p^q)^\times)$ for $q \geq 1$.

1. We define $\mathcal{H}^0\mathcal{M}_n^r$ as the full sheaf $\mathcal{M}_n^r$. For $q \geq 1$, we define $\mathcal{H}^q\mathcal{M}_n^r \subset \mathcal{M}_n^r$ as the image of $i^*(1 + p^q) \otimes (i^*j_*\mathcal{O}_U^\times)^{\otimes r - 1}$.

2. For $q \geq 0$, we define $\mathcal{H}^qM_n^r$ as the image of $\mathcal{M}_n^r$ under the map (3.1.1).

3. When $A$ is local and its residue field $k$ has characteristic $p$, we fix a prime element $\pi \in A$ and define $\mathcal{Y}^rM_n^r \subset M_n^r$ as the part generated by $\mathcal{H}^{r+1}M_n^r$ and the image of $\mathcal{H}^rM_{n-1}^r \otimes (\pi)$ under (3.1.1).

Let $L_{X,s}$ be the log structure on $X$ associated with the normal crossing divisor $Y \cup D$ ([K3]), and let $L_{Y,s}$ be its inverse image log structure onto $Y_s$ (loc. cit. (1.4)). The following theorem is a variant of theorems of Bloch-Kato-Hyodo ([BK1] Theorem 1.4, [Hy1] Theorem 1.6), and the case $D = \emptyset$ corresponds to their theorems.

**Theorem 3.3** (1) The symbol map (3.1.1) is surjective, i.e., $\mathcal{H}^0M_n^r = M_n^r$.  

**Documenta Mathematica 18 (2013) 177–247**
(2) Assume that $A$ is local and that its residue field $k$ has characteristic $p$. Then there are isomorphisms
\[
M_n^r / \mathcal{Y}_n^r \overset{\sim}{\longrightarrow} W_n \omega_{Y_n, \log}^r,
\]
\[
\mathcal{Y}_n^r / \mathcal{Y}_n^{r+1} \overset{\sim}{\longrightarrow} W_n \omega_{Y_n, \log}^{r-1},
\]
where $W_n \omega_{Y_n, \log}^m$ denotes the image of the logarithmic differential map
\[
d \log : (L_{Y_n}^p)^{m} \longrightarrow \bigoplus_{y \in Y_n} i_y W_n \Omega_{y / \log}^m,
\]
and for a point $y \in Y$, $i_y$ denotes the natural map $y \mapsto Y$.

(3) Under the same assumption as in (2), let $e$ be the absolute ramification index of $A$, and let $L_k$ be the log structure on $\text{Spec}(k)$ associated with the pre-log structure $\mathbb{N} \rightarrow k$ sending $1 \mapsto 0$. Put $e' := pe/(p - 1)$. Then for $1 \leq q < e'$, there are isomorphisms
\[
\mathcal{Y}_1^r / \mathcal{Y}_1^{r+1} \overset{\sim}{\longrightarrow} \begin{cases} \omega_{Y_1}^{r-1} / \mathcal{Y}_{Y_1}^{r-1} & (p | q), \\ \omega_{Y_1}^{-1} / \mathcal{Y}_{Y_1}^{r-2} & (p \not| q), \end{cases}
\]
\[
\mathcal{Y}_1^r / \mathcal{Y}_1^{r+1} \overset{\sim}{\longrightarrow} \omega_{Y_1}^{r-2} / \mathcal{Y}_{Y_1}^{r-1}.
\]
Here $\omega_{Y_1}^m$ denotes the differential module of $(Y, L_{Y_1})$ over $(\text{Spec}(k), L_k)$ ([K3] (1.7)), and $\mathcal{Y}_{Y_1}^m$ (resp. $\mathcal{Y}_{Y_1}^{r-2}$) denotes the image of $d : \omega_{Y_1}^{m-1} \rightarrow \omega_{Y_1}^m$ (resp. the kernel of $d : \omega_{Y_1}^m \rightarrow \omega_{Y_1}^{m+1}$).

(4) Under the same assumption and notation as in (3), we have $\mathcal{Y}_1^0 \mathcal{Y}_1^r = \mathcal{Y}_1^r = 0$ for $q \geq e'$.

Proof. Note that the irreducible components of $D$ are semistable families around the fibers of characteristic $p$ by the assumption that $Y \cup D$ has normal crossings on $X$. The assertions (1) and (2) are reduced to the case that $X$ is smooth over $S$ and that $D = \emptyset$ (i.e., the Bloch-Kato theorem) by Tsuji’s trick in [Ts2] Proof of Theorem 5.1 and a variant of Hyodo’s lemma [Hy1] Lemma 3.5, whose details will be explained in a forthcoming paper [KSS]. The assertion (4) follows from [BK1] Lemma 5.1.

We prove (3). Let $\pi \in A$ be the fixed prime element. Let $\pi \in L_{Y_n}$ be the image of $\pi$, and let $[\pi] \subset L_{Y_n}$ be the subsheaf of monoids generated by $\pi$. The quotient $L_{Y_n}/[\pi]$ is a subsheaf of monoids of $\mathcal{O}_Y$ (with respect to the multiplication of functions) generated by $\mathcal{O}_Y^r$ and local equations defining $D$ and irreducible components of $Y$. There is a surjective homomorphism
\[
\delta_m : \mathcal{O}_Y \otimes (L_{Y_n}/[\pi])^p \otimes m \longrightarrow \omega_{Y_n}^m,
\]
defined by the local assignment
\[
z \otimes y_1 \otimes \cdots \otimes y_m \longmapsto z \cdot d \log(y_1) \wedge \cdots \wedge d \log(y_m),
\]
with $z \in \mathcal{O}_Y$ and each $y_i \in (L_{Y_n}/[\pi])^p$. The kernel of $\delta_m$ is generated by local sections of the following forms (cf. [Hy1] Lemma 2.2):
(1) \( z \otimes y_1 \otimes \cdots \otimes y_m \) such that \( y_s \) belongs to \( \mathcal{O}_{\text{Spec}(k)}^X \) for some \( 1 \leq s \leq m \).

(2) \( z \otimes y_1 \otimes \cdots \otimes y_m \) such that \( y_s = y_t \) for some \( 1 \leq s < t \leq m \).

(3) \( \sum_{i=1}^t (a_i \otimes a_i \otimes y_1 \otimes \cdots \otimes y_{m-1}) = \sum_{j=1}^\ell (b_j \otimes b_j \otimes y_1 \otimes \cdots \otimes y_{m-1}) \) with each \( a_i, b_j \in L_{Y, r} \) such that the sums \( \sum_{i=1}^t a_i \) and \( \sum_{j=1}^\ell b_j \) taken in \( \mathcal{O}_Y \) belong to \( L_{Y, r} \) and satisfy \( \sum_{i=1}^t a_i = \sum_{j=1}^\ell b_j \).

Hence the assertion follows from the arguments in loc. cit. p. 551.

We define the étale subsheaf \( FM_r^e \subset M_r^e \) as the part generated by \( \mathfrak{P}^1 M_r^e \) and the image of \((i^* h_* \mathcal{O}_{X_D}^e)^{\otimes r} \), where \( h \) denotes the open immersion \( X - D \hookrightarrow X \). By Theorem 3.3(2), Proposition 2.2 and the same arguments as in [Sa2] §3.4, we obtain the following theorem:

**Theorem 3.4** There are short exact sequences of sheaves on \( Y_{\text{ét}} \)

\[
0 \rightarrow FM_n^e \rightarrow M_n^e \xrightarrow{\sigma} \nu_{(Y, D \cap Y)_n}^{-1} \rightarrow 0,
0 \rightarrow \mathfrak{P}^1 M_n^e \rightarrow FM_n^e \xrightarrow{\tau} \nu_{(Y, D \cap Y)_n}^{-1} \rightarrow 0,
\]

where \( \sigma \) is induced by the boundary map of Galois cohomology groups due to Kato [K1], and \( \tau \) is given by the local assignment

\[
\{a_1, a_2, \ldots, a_r\} \mapsto d\log(\overline{a_1} \otimes \overline{a_2} \otimes \cdots \otimes \overline{a_r}).
\]

Here \( a_1, a_2, \ldots, a_r \) are local sections of \( i^* h_* \mathcal{O}_{X_D}^e \) and for \( \alpha \in i^* h_* \mathcal{O}_{X_D}^e, \) \( \pi \) denotes its residue class in \( g_* \mathcal{O}_Y^e \).

Now we define the \( p \)-adic étale Tate twists.

**Definition 3.5** For \( n \geq 1 \) and \( r \in \mathbb{Z} \), we define a cochain complex \( C_n(r)^{\bullet}_{(X, D)} \) of sheaves on \( X_D \) as follows.

(1) For \( r = 0 \), we define \( C_n(0)^{\bullet}_{(X, D)} := \mathbb{Z}/p^n \), the constant sheaf \( \mathbb{Z}/p^n \) placed in degree 0. For \( r < 0 \), we define \( C_n(r)^{\bullet}_{(X, D)} := j_* \mathcal{H}om_U (\mu_{p^n}^{\otimes r}, \mathbb{Z}/p^n) \), the sheaf \( j_* \mathcal{H}om_U (\mu_{p^n}^{\otimes r}, \mathbb{Z}/p^n) \) placed in degree 0.

(2) Assume \( r \geq 1 \), and let \( \mathcal{F}^e \) be the Godement resolution of \( \mu_{p^n}^{\otimes r} \) on \( U_{\text{ét}} \). We define \( C_n(r)^{\bullet}_{(X, D)} \) as

\[
\begin{align*}
\sigma^n_r : & \quad \ker(d : j_* \mathcal{F}^e \rightarrow j_* \mathcal{F}^e) \\
& \quad \xrightarrow{\sigma^n_r} j_* \mathcal{F}^e \\
& \quad \xrightarrow{\tau^n_r} \nu_{(Y, D \cap Y)_n}^{-1} \rightarrow 0,
\end{align*}
\]

where \( j_* \mathcal{F}^e \) is placed in degree 0 and \( i_* \nu_{(Y, D \cap Y)_n}^{-1} \) is placed in degree \( r + 1 \). The last arrow \( \sigma^n_r \) is defined as the composite map

\[
\sigma^n_r : \quad \ker(d : j_* \mathcal{F}^e_U \rightarrow j_* \mathcal{F}^e_U) \xrightarrow{R^* j_* \mu_{p^n}^{\otimes r} \xrightarrow{i_* \pi} i_* \nu_{(Y, D \cap Y)_n}^{-1}},
\]

and \( \sigma \) denotes the surjective map in Theorem 3.4.
We often write $\Sigma_n(r)(X,D)$ for $C_n(r)(X,D)$ regarded as an object of $D^b(X_{et},\mathbb{Z}/p^n)$. When $D = \emptyset$, we often write $C_n(r)_X$ and $\Sigma_n(r)_X$ for $C_n(r)(X,\emptyset)$ and $\Sigma_n(r)(X,\emptyset)$, respectively.

**Proposition 3.6** For $r \geq 0$, $\Sigma_n(r)(X,D)$ is concentrated in $[0,r]$, and there is a distinguished triangle in $D^b(X_{et},\mathbb{Z}/p^n)$

$$i_*\nu_{(Y,D\cap Y),n}^{r-1} \xrightarrow{\tau} \Sigma_n(r)(X,D) \xrightarrow{\tau_{\leq r}Rj_*\mu_{p^n}^{\otimes r}} i_*\nu_{(Y,D\cap Y),n}^{r-1}.$$

**Proof.** The first assertion follows from the surjectivity of $\sigma^r_n$ in the definition of $C_n(r)(X,D)$. The second assertion is straight-forward. \hfill \Box

**Remark 3.7** Proposition 3.6 implies that $\Sigma_n(r)(X,D)$ defined here agrees with that in [Sa2] §4 by a unique isomorphism compatible with the identity map of $\mu_{p^n}^{\otimes r}$ on $U$ (loc. cit. Lemma 4.2.2). Although we assumed that the residue fields of $\Lambda$ of characteristic $p$ are perfect in loc. cit., the results in loc. cit. §4–§7.2 hold true without this assumption by results of Shiho [Sh] Theorems 3.2, 4.1 and the compatibility result in [AS] Remark 7.3 (1).

**Proposition 3.8** The complex $C_n(r)(X,D)$ is contravariantly functorial in the pair $(X,D)$. Here a morphism of pairs $(X,D) \to (X',D')$ means a morphism of schemes $f : X \to X'$ satisfying $f(X - D) \subset X' - D'$.

**Proof.** The case $r \leq 0$ is clear. As for the case $r \geq 1$, it is enough to show that the map

$$i_*\sigma : R^rj_*\mu_{p^n}^{\otimes r} \longrightarrow i_*\nu_{(Y,D\cap Y),n}^{r-1}$$

is contravariant in $(X,D)$. Let $f : (X,D) \to (X',D')$ be a morphism of pairs, and consider the following diagram of immersions:

$$Y' \xleftarrow{i} X' \xrightarrow{j'} U' := X' - D',$$

where $Y'$ denotes the union of the fibers of $X' \to S$ of characteristic $p$. By the first exact sequence in Theorem 3.4, $\sigma'$ ($:= \sigma$ for $(X',D')$) is surjective and $\text{Ker}(i_*\sigma')$ maps into $\text{Ker}(i_*\sigma)$ under the base-change map

$$f^* : f^* R^rj_*\mu_{p^n}^{\otimes r} \longrightarrow R^rj_*\mu_{p^n}^{\otimes r}.$$

Hence this map induces a pull-back map

$$f^* : f^* i_*\nu_{(Y,D\cap Y),n}^{r-1} \longrightarrow i_*\nu_{(Y,D\cap Y),n}^{r-1} \quad (3.8.1)$$

These maps are obviously compatible with $\sigma'$s and satisfy transitivity. Thus we obtain the proposition. \hfill \Box
Corollary 3.9  (1) Let $V$ be the category whose objects are schemes $X$ as in Setting 3.1 and whose morphisms are morphisms of schemes. Then the complexes $C_{n}(r)^{•}_{X} = C_{n}(r)^{•}_{(X,∅)}$ with $X \in \text{Ob}(V)$ form a complex $C_{n}(r)^{•}$ of sheaves on the big étale site $V_{\text{ét}}$.

(2) The Godement resolution $G_{n}(r)^{•}_{(X,D)}$ on $X_{\text{ét}}$ of $C_{n}(r)^{•}_{(X,D)}$ is contravariantly functorial in $(X,D)$.

Remark 3.10 The object $\mathfrak{T}_{n}(r)_{(X,D)}$ is also contravariantly functorial in the pair $(X,D)$. More precisely, for a morphism of pairs $f : (X,D) \to (X',D')$, there is a natural morphism

$$f^{•} : f^{•}\mathfrak{T}_{n}(r)_{(X',D')} \to \mathfrak{T}_{n}(r)_{(X,D)} \quad \text{in} \quad D^{b}(X_{\text{ét}},\mathbb{Z}/p^{n})$$

by Proposition 3.8, which is, in fact, the unique morphism that extends the pull-back isomorphism of $\mu_{p^{n}}^{\otimes}$ for $U \to U'$ (cf. [Sa2] Proposition 4.2.8).

Remark 3.11 Let $V$ and $C_{n}(r)^{•}$ be as in Corollary 3.9, and let $\mathfrak{T}_{n}(r)$ be the complex $C_{n}(r)^{•}$ regarded as an object of the derived category $D^{b}(V_{\text{ét}},\mathbb{Z}/p^{n})$. Let $V' \subset V$ be the full subcategory consisting of schemes $X \in \text{Ob}(V)$ with $p^{-1} \in \Gamma(X,O_{X})$. The following facts will be useful later in §5:

(1) There exists a unique product structure

$$\mathfrak{T}_{n}(q) \otimes^{L} \mathfrak{T}_{n}(r) \longrightarrow \mathfrak{T}_{n}(q + r) \quad \text{in} \quad D(V_{\text{ét}},\mathbb{Z}/p^{n})$$

that extends the isomorphism $\mu_{p^{n}}^{\otimes q} \otimes \mu_{p^{n}}^{\otimes r} \cong \mu_{p^{n}}^{\otimes q + r}$ on the big étale site $(V')_{\text{ét}}$, which follows from the same arguments as in [Sa2] Proposition 4.2.6.

(2) There exists a unique isomorphism

$$\mathfrak{G}_{n} \otimes^{L} \mathbb{Z}/p^{n}[-1] \cong \mathfrak{T}_{n}(1) \quad \text{in} \quad D(V_{\text{ét}},\mathbb{Z}/p^{n})$$

that extends the canonical isomorphism $\mathfrak{G}_{n} \otimes^{L} \mathbb{Z}/p^{n}[-1] \longrightarrow \mu_{p^{n}}$ on $(V')_{\text{ét}}$, which follows from the same arguments as in loc. cit. Proposition 4.5.1.

Theorem 3.12 When $D$ is regular, there is a canonical morphism

$$\delta : \mathfrak{T}_{n}(r)_{(X,D)} \longrightarrow \alpha_{•}\mathfrak{T}_{n}(r-1)_{D}[-1] \quad \text{in} \quad D^{b}(X_{\text{ét}},\mathbb{Z}/p^{n})$$

fitting into a distinguished triangle

$$\alpha_{•}\mathfrak{T}_{n}(r-1)_{D}[-2] \xrightarrow{\alpha_{•}} \mathfrak{T}_{n}(r)_{X} \xrightarrow{\beta_{•}} \mathfrak{T}_{n}(r)_{(X,D)} \xrightarrow{\delta} \alpha_{•}\mathfrak{T}_{n}(r-1)_{D}[-1], \quad (3.12.1)$$

where $\alpha$ denotes the closed immersion $D \hookrightarrow X$, and $\beta$ denotes the natural morphism of pairs $(X,D) \to (X,∅)$. The arrow $\alpha_{•}$ denotes the Gysin morphism [Sa2] Theorem 6.1.3.
Proof. The case $r < 0$ immediately follows from the absolute purity [FG]. To prove the case $r \geq 0$, we first construct the morphism $\delta$. Consider a diagram of immersions

\[
\begin{array}{ccc}
Y & \xrightarrow{i} & X \\
\downarrow \alpha & & \downarrow \alpha \\
E & \xrightarrow{\epsilon} & D
\end{array}
\]

There is a distinguished triangle on $X[p^{-1}]_{et}$

\[
v_*\mu_{p^n}^{(r-1)}[-2] \xrightarrow{u_*} \mu_{p^n}^{(r)} \xrightarrow{u^*} Ru_*\mu_{p^n}^{(r)} \xrightarrow{\delta_1} v_*\mu_{p^n}^{(r-1)}[-1],
\]

where $\delta_1$ is defined as the composite

\[
\delta_1 : Ru_*\mu_{p^n}^{(r)} \xrightarrow{d^\loc} v_*Rv^!\mu_{p^n}^{(r)}[1] \xleftarrow{} v_*\mu_{p^n}^{(r-1)}[-1]
\]

and $d^\loc$ denotes the connecting morphism of a localization sequence. We used the absolute purity [FG] for the last isomorphism. Applying $Rw_*$, we get a distinguished triangle on $X_{et}$

\[
\alpha_*R\varphi_*\mu_{p^n}^{(r-1)}[-2] \xrightarrow{u_*} Rw_*\mu_{p^n}^{(r)} \xrightarrow{u^*} Rj_*\mu_{p^n}^{(r)} \xrightarrow{\delta_2} \alpha_*R\varphi_*\mu_{p^n}^{(r-1)}[-1].
\]  

(3.12.2)

Consider a diagram

\[
\begin{array}{ccc}
\Sigma_n(r)(X,D) & \xrightarrow{t} & \tau_{\leq r} Rj_*\mu_{p^n}^{(r)} \\
\downarrow \delta & & \downarrow \delta_3 = \tau_{\leq r}(\delta_2) \\
\alpha_\Sigma_n(r-1)(D[-1]) & \xrightarrow{\tau_{\leq r}(\alpha_*R\varphi_*\mu_{p^n}^{(r-1)}[-1])} & \gamma_*\nu_{E,n}^{-2}[-r],
\end{array}
\]

where the lower row is a part of a distinguished triangle (cf. Proposition 3.6 for $(D, \emptyset)$) and $\gamma$ denotes the composite morphism $\alpha \circ \iota' : E \hookrightarrow X$. Since $\Sigma_n(r)(X,D)$ is concentrated in $[0, r]$, we see that the composite $b' \circ \delta_3 \circ t$ is zero by Theorem 3.4 and a simple computation on symbols. On the other hand, we have

\[
\text{Hom}_{D^p(X_{et}, E/p^n)}(\Sigma_n(r)(X,D), \gamma_*\nu_{E,n}^{-2}[-r-1]) = 0,
\]

again by the fact that $\Sigma_n(r)(X,D)$ is concentrated in $[0, r]$. Hence there exists a unique morphism $\delta$ fitting into the above diagram (cf. [Sa2] Lemma 2.1.2(1)), which is the desired morphism. Finally the triangle (3.12.1) is distinguished by (3.12.2) and a commutative diagram with exact rows on $X_{et}$ ($\gamma = \alpha \circ \iota'$)

\[
\begin{array}{ccc}
R^*w_*\mu_{p^n}^{(r)} & \xrightarrow{\delta_3} & \alpha_*R^{r-1}\varphi_*\mu_{p^n}^{(r-1)} \xrightarrow{0} \\
\downarrow \sigma_X & & \downarrow \sigma_{(X,D)} \\
0 & \xrightarrow{\iota_*\nu_{Y,n}^{-1}} & \iota_*\nu_{(Y,E),n}^{-1} \xrightarrow{0} \\
\end{array}
\]

\[
\begin{array}{ccc}
\end{array}
\]

\textbf{Documenta Mathematica} 18 (2013) 177–247
where the surjectivity of $\delta_3$ in the upper row follows from Theorem 3.3 (1) for $(D, \emptyset)$, and the exactness of the lower row follows from (2.1.1).

\[ \square \]

\textbf{Remark 3.13} Let $\zeta_p$ be a primitive $p$-th root of unity, and let $A'$ be the normalization of $A[\zeta_p]$. Then all the definitions and results for the pair $(X, D)$ in this section are extended to the scalar extension $(X \otimes_A A', D \otimes_A A')$. Indeed, Theorem 3.3 (1) and (2) will be proved in [KSS] for this generalized situation. One can check Theorem 3.3 (3) and (4) for $(X \otimes_A A', D \otimes_A A')$ by the same arguments as for $(X, D)$. See [Sa2] §3.5 for an argument to extend Theorem 3.4.

4 Dold-Thom isomorphism

Let $S, p$ and $X$ be as in Setting 3.1, and let $n$ and $r$ be integers with $n \geq 1$. In this section we prove the Dold-Thom isomorphism for $p$-adic étale Tate twists. Let $E$ be a vector bundle of rank $a + 1$ on $X$, and let $f : \mathbb{P} := \mathbb{P}(E) \to X$ be the associated projective bundle, which is a projective smooth morphism of relative dimension $a$.

Let $\mathcal{O}(1)_E$ be the tautological invertible sheaf on $\mathbb{P}$, and let $\xi \in H^2(\mathbb{P}, \mathfrak{T}_n(1)_{\mathbb{P}})$ be the value of the first Chern class $c_1(\mathcal{O}(1)_E) \in H^1(\mathbb{P}, \mathcal{O}_{\mathbb{P}})$ under the connecting map associated with the Kummer distinguished triangle.

\[ \mathcal{O}_{\mathbb{P}} \to \mathcal{O}_{\mathbb{P}} \to \mathfrak{T}_n(1)_{\mathbb{P}}[1] \to \mathcal{O}_{\mathbb{P}}[1] \]

(cf. [Sa2] Proposition 4.5.1). The composite morphisms

\[ \mathfrak{T}_n(r - q)X[-2q] \xrightarrow{f^*} Rf_* \mathfrak{T}_n(r - q)_\mathbb{P}[-2i] \xrightarrow{-\cup \xi} Rf_* \mathfrak{T}_n(r)_\mathbb{P} \]

induce a canonical morphism

\[ \gamma_E : \bigoplus_{q=0}^a \mathfrak{T}_n(r - q)X[-2q] \to Rf_* \mathfrak{T}_n(r)_\mathbb{P} \quad \text{in} \quad D^b(X_\text{ét}, \mathbb{Z}/p^n). \]

\textbf{Theorem 4.1 (Dold-Thom isomorphism)} $\gamma_E$ is an isomorphism.

\textit{Proof.} $\gamma_E$ is an isomorphism outside of $Y$ by [M] VI Theorem 10.1. The case $r < 0$ follows from this fact and the proper base-change theorem. To prove the case $r \geq 0$, we consider a diagram of schemes

\[ \begin{array}{ccc}
  P & \xrightarrow{\gamma} & \mathbb{P} \\
  g \downarrow & \square & \downarrow f \\
  Y & \xrightarrow{i} & X \xrightarrow{j} X[p^{-1}].
\end{array} \]

We have to show that $i^*(\gamma_E)$ is an isomorphism:

\[ i^*(\gamma_E) : \bigoplus_{q=0}^a i^* \mathfrak{T}_n(r - q)X[-2q] \xrightarrow{\sim} Rg_* i^* \mathfrak{T}_n(r)_\mathbb{P}, \quad (4.1.1) \]
where we identified $i^* Rf_* \Sigma_n(r)_{\mathbb{F}}$ with $Rg_* \gamma^* \Sigma_n(r)_{\mathbb{F}}$ by the proper base-change theorem. By a standard norm argument (cf. [Sa2] §10.3) using Bockstein triangles (loc. cit. §4.3), we are reduced to the case that $n = 1$ and that $\mathcal{I}(X, \mathcal{O}_X)$ contains a primitive $p$-th root of unity (see Remark 3.13). We need the following lemma:

**Lemma 4.2** Let $\xi \in H^1(P, \lambda_{P,1}^1)$ be the image of $\xi$ under the pull-back map

$$\gamma^*: H^2(P, \Sigma_1(1)_{\mathbb{F}}) \longrightarrow H^1(P, \lambda_{P,1}^1)$$

(cf. [Sa2] Proposition 4.4.10). Then there are isomorphisms in $D^b(Y_{\text{et}}, \mathbb{Z}/p)$

\[
\bigoplus_{q=0}^{\alpha} \xi^q \cup - : \bigoplus_{q=0}^{\alpha} \lambda_{Y,1}^{r-q}[-q] \xrightarrow{\sim} Rg_* \lambda_{P,1}^1 \tag{1}
\]

\[
\bigoplus_{q=0}^{\alpha} \xi^q \cup - : \bigoplus_{q=0}^{\alpha} \nu_{Y,1}^{r-q}[-q] \xrightarrow{\sim} Rg_* \nu_{P,1}^r \tag{2}
\]

\[
\bigoplus_{q=0}^{\alpha} \xi^q \cup - : \bigoplus_{q=0}^{\alpha} \omega_{Y}^{r-q}[-q] \xrightarrow{\sim} Rg_* \omega_{P}^q \tag{3}
\]

\[
\bigoplus_{q=0}^{\alpha} \xi^q \cup - : \bigoplus_{q=0}^{\alpha} \mathcal{Z}_Y^{r-q}[-q] \xrightarrow{\sim} Rg_* \mathcal{Z}_P^r. \tag{4}
\]

Here $\mathcal{Z}_Y^m$ (resp. $\mathcal{Z}_P^m$) denotes the kernel of $d : \omega_Y^m \rightarrow \omega_Y^{m+1}$ (resp. $\omega_P^m \rightarrow \omega_P^{m+1}$).

**Proof of Lemma 4.2.** Note that $\xi$ agrees with the first Chern class of the tautological invertible sheaf on $P = \mathbb{P}(i^*E)$. Since the problems are étale local on $Y$, we may assume that $Y$ is simple. If $Y$ is smooth, then (1) and (2) are due to Gros [Gr1] I Théorème 2.1.11. The general case is reduced to the smooth case by [Sa1] Proposition 3.2.1, Corollary 2.2.7. As for (3), since we have

$$Rg_* \omega_P^{r-q} \simeq \bigoplus_{q=0}^{r} Rg_* (\Omega^q_{P/Y} \otimes \mathcal{O}_Y) \omega_Y^{r-q} \simeq \bigoplus_{q=0}^{r} (Rg_* \Omega^q_{P/Y}) \otimes_{\mathcal{O}_Y} \omega_Y^{r-q}$$

by projection formula, the assertion follows from the isomorphisms

$$\xi^q \cup - : \mathcal{O}_Y [-q] \xrightarrow{\sim} Rg_* \Omega^q_{P/Y} \quad (0 \leq q \leq \alpha).$$

(4) follows from the same arguments as for [Gr1] I (2.2.3). \qed

We return to the proof of (4.1.1) for $r \geq 0$. The case $r = 0$ follows from Lemma 4.2(1) with $r = 0$. To prove the case $r > 1$, we use the objects $\mathcal{K}(r-q)_{X} \in D^b(Y_{\text{et}}, \mathbb{Z}/p)$ and $\mathcal{K}(r)_{P} \in D^b(P_{\text{et}}, \mathbb{Z}/p)$ defined in [Sa2] Lemma 10.4.1, which fit into distinguished triangles

$$\mathcal{K}(r-q)_{X}[-1] \rightarrow \mu^\vee \otimes^L i^* \Sigma_1(r-q-1)_{X} \rightarrow i^* \Sigma_1(r-q)_{X} \rightarrow \mathcal{K}(r-q)_{X},$$

$$\mathcal{K}(r)_{P}[-1] \rightarrow g^* \mu^\vee \otimes^L \gamma^* \Sigma_1(r-1)_{P} \rightarrow \gamma^* \Sigma_1(r)_{P} \rightarrow \mathcal{K}(r)_{P}.$$
Here $\mu'$ denotes the constant sheaf $i^*j_*\mu_p(z/p)$ on $Y_\et$ and the central arrows are induced by the product structure of Tate twists. By induction on $r \geq 0$, our task is to show that the morphism

$$
\bigoplus_{q=0}^a \xi^q \cup - : \bigoplus_{q=0}^a K(r-q)_X[-2q] \to Rg_*K(r)_P
$$

(4.2.1)

is an isomorphism, where we have used the pull-back morphisms

$$
f^* : K(r-q)_X \to K(r-q)_P \quad (0 \leq q \leq a)
$$

induced by the pull-back morphisms for Tate twists (loc. cit. Proposition 4.2.8, Lemma 2.1.2 (2)). By loc. cit. Lemma 10.4.1 (2), we have

$$
H_m(K(r-q)_X) \simeq \begin{cases} 
\mu' \otimes \nu_r-q-2 & (m = r - q - 1) \\
FM_{r-q}^m & (m = r - q) \\
0 & (\text{otherwise}) 
\end{cases}
$$

and similar facts holds for $K(r)_P$ (see §3 for $FM_n^q$). Therefore (4.2.1) is an isomorphism by Lemma 4.2 and Theorems 3.3 (3), (4) and 3.4 with $D = \emptyset$ (see also the projection formula in loc. cit. 4.4.10). This completes the proof of Theorem 4.1. □

The following corollary 4.3 follows immediately from Theorems 3.12, 4.1 and the projection formula ([Sa2] Corollary 7.2.4), which is a $p$-adic version of homotopy invariance and plays an important role in our construction of cycle class maps (see §7 below).

**Corollary 4.3** Let $E$ be a vector bundle on $X$ of rank $a$, and put $P = P(E \oplus 1)$, the projective completion of $E$. Let $P'$ be the projective bundle $P(E)$ regarded as the infinite hyperplane, and let $f : P \to X$ be the natural projection. Then the composite morphism

$$
\xi_n(r)_X \xrightarrow{f_*} Rf_*\xi_n(r)_P \xrightarrow{Rf_*} D^b(X_\et, \mathbb{Z}/p^n)
$$

is an isomorphism in $D^b(X_\et, \mathbb{Z}/p^n)$.

5 **Chern Class**

The main aim of this section is to construct the Chern class map (5.6.1) below. Let $S$ and $p$ be as in Setting 3.1, and let $\mathcal{V}$ be the category whose objects are $S$-schemes satisfying the conditions in Setting 3.1 for $X$ and whose morphisms are morphisms of schemes. Let $X_\star$ be a simplicial object in $\mathcal{V}$, i.e., a contravariant functor

$$
X_\star : \Delta^{op} \to \mathcal{V},
$$

where $\Delta$ denotes the simplex category. For a morphism $\gamma : [a] \to [b]$ in $\Delta$, we write

$$
\gamma^X : X_b \to X_a \quad (X_a := X_\star([a]))
$$
for $X_\star(\gamma)$, which is a morphism in $\mathcal{V}$. For $0 \leq i \leq a$, let $d^i$ be the coface map in $\Delta$:

$$d^i : [a] \rightarrow [a + 1], \quad j \mapsto \begin{cases} j & (0 \leq j < i) \\ j + 1 & (i \leq j \leq a). \end{cases}$$

For integers $0 \leq i \leq a$, we often write $d^i : X_{a + 1} \rightarrow X_a$

for $(d^i)^X$. See [Fr] §1 for the definition of the small étale site $(X_\star)_{\text{ét}}$ on $X_\star$.

**Definition 5.1**

(1) For $r \in \mathbb{Z}$, we define a complex $C_n(r)_{X_\star}^\bullet$ of sheaves on $(X_\star)_{\text{ét}}$ by restricting the complex $C_n(r)^\bullet$ on $\mathcal{V}_{\text{ét}}$, cf. Corollary 3.9. We will often write $T_n(r)_{X_\star}$ for the complex $C_n(r)_{X_\star}^\bullet$ regarded as an object of $D^b((X_\star)_{\text{ét}})$.

(2) We define a canonical morphism

$$\varrho : \mathbb{G}_m[-1] \rightarrow \mathcal{T}_n(1)_{X_\star} \quad \text{in} \quad D^b((X_\star)_{\text{ét}})$$

by the composite morphism

$$\mathbb{G}_m[-1] \rightarrow \mathbb{G}_m \otimes \mathbb{Z}/p^n[-1] \xrightarrow{\sim} \mathcal{T}_n(1)_{X_\star},$$

where the left arrow denotes the canonical morphism induced by $\mathbb{Z} \rightarrow \mathbb{Z}/p^n$ and the right arrow is the restriction of the isomorphism in Remark 3.11 (2).

We next review the following basic notions:

**Definition 5.2**

(1) A vector bundle over $X_\star$ is a morphism $f : E_\star \rightarrow X_\star$ of simplicial schemes such that $f_a : E_a \rightarrow X_a$ is a vector bundle for any $a \geq 0$ and such that the commutative diagram

$$\begin{array}{c}
E_b \\
\gamma^X \\
E_a
\end{array} \xrightarrow{f} \begin{array}{c}
X_b \\
\gamma^X \\
X_a
\end{array}$$

induces an isomorphism $E_b \cong \gamma^X E_a := E_a \times_{X_a} X_b$ of vector bundles over $X_a$ for any morphism $\gamma : [a] \rightarrow [b]$ in $\Delta$ (cf. [Gi2] Example 1.1).

(2) A regular closed immersion of simplicial schemes is a morphism $i : X_\star \rightarrow Y_\star$ such that $i_a : X_a \rightarrow Y_a$ is a regular closed immersion for any $a \geq 0$ and such that the diagram

$$\begin{array}{c}
X_b \\
\gamma^X \\
X_a
\end{array} \xrightarrow{i} \begin{array}{c}
Y_b \\
\gamma^Y \\
Y_a
\end{array}$$

is cartesian for any morphism $\gamma : [a] \rightarrow [b]$ in $\Delta$. An effective Cartier divisor $X_\star$ on $Y_\star$ is a regular closed immersion $X_\star \rightarrow Y_\star$ of pure codimension 1.
We now define the first Chern classes of effective Cartier divisors and line bundles.

**Definition 5.3**

1. For an effective Cartier divisor \( D \) on \( X \), we define the first Chern class \( c_1(D) \in H^1_D((X)_{\text{Zar}}, \mathcal{O}^\times) \) as the value of the first Chern class \( c_1(D) \in H^1((X)_{\text{ét}}, \mathcal{O}) \) under the composite map
   \[
   H^1_D((X)_{\text{Zar}}, \mathcal{O}^\times) \xrightarrow{\epsilon^*} H^1_D((X), \mathcal{O}_m) \xrightarrow{\varphi} H^2_D((X), T_n(1)),
   \]
   where \( \epsilon : (X)_{\text{ét}} \to (X)_{\text{Zar}} \) denotes the continuous map of small sites. The arrow \( \varphi \) denotes that in Definition 5.1 (2).

2. For a line bundle \( L \) on \( X \), we define the first Chern class \( c_1(L) \in H^2((X), T_n(1)) \) as the value of the isomorphism class \([L] \in H^1((X)_{\text{ét}}, \mathcal{O}_m)\) under the composite map
   \[
   H^1((X)_{\text{ét}}, \mathcal{O}_m) \xrightarrow{\epsilon^*} H^1((X), G^\text{m}) \xrightarrow{\varphi} H^2((X), T_n(1)).
   \]

The following proposition plays a key role in the proof of the Whitney sum formula in Proposition 5.5 (3) below.

**Proposition 5.4**

Let \( f : X \hookrightarrow X' \) be a regular closed immersion of simplicial objects in \( \mathcal{V} \) of pure codimension \( c \geq 1 \). Let \( r \) be a non-negative integer. Then there exists a Gysin morphism
   \[
gys_f : \mathfrak{T}_n(r)_{X'} \to Rf^!\mathfrak{T}_n(r+c)_{X'}[2c]
   \]
   in \( \mathcal{D}^+(((X)_{\text{ét}}, \mathcal{O}/\mathbb{Z}/p^n) \)

satisfying the following three properties:

(a) **(Consistency with the first Chern class)** If \( r = 1 \), then the value of \( 1 \in \mathbb{Z}/p^n = H^0(X, \mathfrak{T}_n(0)) \) under the Gysin map
   \[
gys_f : H^0(X, \mathfrak{T}_n(0)) \to H^2_F((X)_{\text{ét}}, \mathfrak{T}_n(1))
   \]
   agrees with the first Chern class \( c_1(X) \) in Definition 5.3 (1).

(b) **(Transitivity)** For another regular closed immersion \( g : X' \hookrightarrow X'' \) of simplicial objects in \( \mathcal{V} \) of pure codimension \( c' \geq 1 \), the composite morphism
   \[
   \mathfrak{T}_n(r)_{X'} \xrightarrow{Rf^!(\text{gys}_f)} Rf^!\mathfrak{T}_n(r+c)_{X'}[2c] \xrightarrow{Rf^g} Rf^g\mathfrak{T}_n(r+c+c')_{X''}[2c+2c'] \xrightarrow{R(g \circ f)^!} \mathfrak{T}_n(r+c+c')_{X''}[2(c+c')]
   \]
   agrees with \( \text{gys}_{g \circ f} \).
(c) (Projection formula) The following diagram commutes in $D((X_\ast)_{\text{et}}, \mathbb{Z}/p^n)$:

$$
\begin{array}{ccc}
Rf^!\mathcal{I}_n(q)X' \otimes \mathcal{I}_n(r)X & \xrightarrow{id \otimes \text{gys}_f} & Rf^!\mathcal{I}_n(q)X' \otimes \mathcal{I}_n(n + r)X' [2r] \\
\pi & \downarrow & \text{product} \\
\mathcal{I}_n(q + r)X & \xrightarrow{\text{gys}_f} & Rf^!\mathcal{I}_n(q + r + c)X' [2c],
\end{array}
$$

where the left vertical arrow $\pi$ is the composite morphism

$$Rf^!\mathcal{I}_n(q)X' \otimes \mathcal{I}_n(r)X \xrightarrow{f^* \otimes \text{id}} \mathcal{I}_n(q)X \otimes \mathcal{I}_n(r)X \xrightarrow{\text{product}} \mathcal{I}_n(q + r)X,$$

and the product structures on $\{\mathcal{I}_n(r)X, r \in \mathbb{Z}\}$ and $\{\mathcal{I}_n(r)X, r \in \mathbb{Z}\}$ are obtained from that in Remark 3.11.

Proof. Put $U_\ast := X_\ast \otimes \mathbb{Z}[p^{-1}]$ and $V_\ast := X'_\ast \otimes \mathbb{Z}[p^{-1}]$. Let $\varphi : U_\ast \to V_\ast$ be the regular closed immersion induced by $f$. By the absolute purity [FG] and the spectral sequence

$$E_1^{a,b} = H_{U_\ast}^b(V_\ast, \mu_p^\otimes \epsilon) \Longrightarrow H_{U_\ast}^{a+b}(V_\ast, \mu_p^\otimes \epsilon),$$

we have

$$H_{U_\ast}^{2e}(V_\ast, \mu_p^\otimes \epsilon) \simeq \text{Ker}(d_0^* - d_1^* : H_{U_\ast}^{2e}(V_\ast, \mu_p^\otimes \epsilon) \to H_{U_\ast}^{2e}(V_\ast, \mu_p^\otimes \epsilon)).$$

By the definition of regular closed immersions in Definition 5.2 (2), the cycle class $\text{cl}_{V_\ast}(U_0) \in H_{U_\ast}^{2e}(V_\ast, \mu_p^\otimes \epsilon)$ lies in the group on the right hand side, loc. cit. Proposition 1.1.3. We thus define the cycle class

$$\text{cl}_{V_\ast}(U_\ast) \in H_{U_\ast}^{2e}(V_\ast, \mu_p^\otimes \epsilon)$$

as the element corresponding to $\text{cl}_{V_\ast}(U_0)$. Since $\varphi^* \mu_p^\otimes \epsilon_{V_\ast} \simeq \mu_p^\otimes \epsilon_{U_\ast}$ on $(U_\ast)_{\text{et}}$, the cup product with $\text{cl}_{V_\ast}(U_\ast)$ defines a Gysin morphism

$$\text{gys}_\varphi : \mu_p^\otimes \epsilon_{U_\ast} \simeq \varphi^* \mu_p^\otimes \epsilon_{V_\ast} \xrightarrow{d_{V_\ast}(U_\ast) \cup -} R\varphi^! \mu_p^\otimes \epsilon_{V_\ast} [2c] \quad \text{in} \quad D^+((U_\ast)_{\text{et}}, \mathbb{Z}/p^n),$$

which satisfies the three properties (a)–(c) listed above (see loc. cit. Proposition 1.2.1 for (b)). We show that there exists a unique morphism

$$\text{gys}_f : \mathcal{I}_n(r)X \to Rf^!\mathcal{I}_n(r + c)X' [2c] \quad \text{in} \quad D^+((X_\ast)_{\text{et}}, \mathbb{Z}/p^n)$$

that extends $\text{gys}_\varphi$. Put

$$Y_\ast := X_\ast \otimes \mathbb{Z}/p\mathbb{Z}, \quad \mathcal{L} := \mathcal{I}_n(r)X, \quad \text{and} \quad \mathcal{M} := Rf^!\mathcal{I}_n(r + c)X' [2c].$$

and let $\alpha$ and $\beta$ be as follows:

$$U_\ast \xrightarrow{\beta} X_\ast \xrightarrow{\alpha} Y_\ast.$$
Consider the following diagram in $D^+(\mathcal{X}_*; \mathbb{Z}/p^n)$ whose lower row is distinguished:

$$
\begin{array}{ccc}
\mathcal{L} & \xrightarrow{\tau_{\leq r}} & R\beta_*\mu_{p^n}^\otimes \\
\downarrow R\beta_*(\text{gys}_*{\text{et}}) & & \downarrow R\beta_*\beta^*\mathcal{M} \\
\alpha_*\mathcal{M} & \xrightarrow{\alpha_*} & \mathcal{M} \xrightarrow{\beta_*} R\beta_*\beta^*\mathcal{M} \\
\end{array}
$$

(5.4.1)

Here the upper horizontal arrow is the canonical morphism (cf. Propositions 3.6 and 3.8), and the lower row is the localization distinguished triangle for $\mathcal{M}$ (cf. [Sa2] (1.9.2)). We have

$$
\tau_{\leq r}\alpha_*\mathcal{M} = 0
$$

(5.4.2)

by the purity in loc. cit. Theorem 4.4.7, which implies that

$$
\text{Hom}_{D^+(\mathcal{X}_*; \mathbb{Z}/p^n)}(\tau_{\leq r}R\beta_*\mu_{p^n}^\otimes, \alpha_*\mathcal{M}[1])
$$

$$
= \text{Hom}_{\text{Shv}(\mathcal{X}_*; \mathbb{Z}/p^n)}(R\beta_*\mu_{p^n}^\otimes, \alpha_*R^{r+1}\alpha^!\mathcal{M})
$$

where $\text{Shv}(\mathcal{X}_*; \mathbb{Z}/p^n)$ denotes the category of étale sheaves of $\mathbb{Z}/p^n$-modules on $\mathcal{X}_*$. By this fact and the compatibility fact in loc. cit. Theorem 6.1.1, one can easily check that the composite $(-\delta) \circ R\beta_*(\text{gys}_*)$ is zero in $D^+(\mathcal{X}_*; \mathbb{Z}/p^n)$. Therefore we obtain a unique morphism $\text{gys}_f$ that extends $\text{gys}_*$ again by (5.4.2) and by loc. cit. Lemma 2.1.2(1). The property (a) of $\text{gys}_f$ is straight-forward, and the property (b) follows from the uniqueness of $\text{gys}_f$. The property (c) follows from the same argument as for loc. cit. Corollary 7.2.4.

Following the method of Grothendieck [G] and Gillet [Gi1], we define Chern classes

$$
c(E_*) = (c_i(E_*))_{i \geq 0} \in \bigoplus_{i \geq 0} H^{2i}(\mathcal{X}_*, \tau(i))
$$

of a vector bundle $E_*$ over $\mathcal{X}_*$ as follows. Let $E_*$ be of rank $a$, and let $f$ be the natural projection $\mathbb{P}(E_*) \to \mathcal{X}_*$. Let $\xi \in H^2(\mathbb{P}(E_*), \tau(1))$ be the value of the first Chern class of the tautological line bundle on $\mathbb{P}(E_*)$, cf. Definition 5.3(2). Noting the Dold-Thom isomorphism

$$
\bigoplus_{i=1}^a H^{2i}(\mathcal{X}_*, \tau(i)) \simeq H^{2a}(\mathbb{P}(E_*), \tau(a)), \quad (b_i)_{i=1}^a \mapsto \sum_{i=1}^a f^*(b_i) \cup \xi^{a-i}
$$

(5.4.3)

obtained from Theorem 4.1 and [Gi1] Lemma 2.4, we define

$$
c_0(E_*) := 1 \quad \text{and} \quad c_i(E_*) := 0 \quad \text{for} \quad i > a,
$$

and define $c_i = c_i(E_*)$ for $i = 1, 2, \ldots, a$ by the equation

$$
\xi^a + f^*(c_1) \cup \xi^{a-1} + \cdots + f^*(c_{a-1}) \cup \xi + f^*(c_a) = 0
$$

in $H^{2a}(\mathbb{P}(E_*), \tau(a))$. 

---

Documenta Mathematica 18 (2013) 177–247
Moreover, the Chern classes \( c(E_*) \) are characterized by these three properties.

Proof. The properties (1) and (2) immediately follow from this definition of Chern classes. The last assertion on the uniqueness follows from the splitting principle of vector bundles, whose details are straight-forward and left to the reader.

We prove the property (3) using the arguments of Grothendieck in [G] p. 144 Theorem 1 (iii), as follows. Let \( \pi^*: D'_* \to X_* \) and \( \pi'^*: D''_* \to X_* \) be the (simplicial) flag schemes of \( E'_* \) and \( E''_* \), respectively, and put

\[
D_* := D'_* \times_X D''_*,
\]

which is identified with the flag scheme of \( \pi''_* E'_* \) over \( D''_* \). Let \( f: D_* \to X_* \) be the natural projection. Since the pull-back map

\[
f^*: H^{2i}(X_*, \mathbb{Z}_n(i)) \to H^{2i}(D_*, \mathbb{Z}_n(i)),
\]

is injective by (5.4.3), we may replace \( (X_*, E_*, E'_*, E''_*) \) with \( (D_*, f^* E_*, f^* E'_*, f^* E''_*) \) to assume that \( E_* \) has a filtration by subbundles

\[
E_* = E^0_* \supset E^1_* \supset \cdots \supset E^a_* = 0 \quad (a := \text{rank}(E_*))
\]

such that the quotient \( E^i_*/E^{i+1}_* \) is a line bundle over \( X_* \) for \( 0 \leq i \leq a-1 \) and such that \( E^b_* = E'_* \) for \( b = \text{rank}(E''_*) \). Now let

\[
g: X'_* := \mathbb{P}(E'_*) \to X_*
\]

Moreover, the Chern classes \( c(E_*) \) are characterized by these three properties.
be the projective bundle associated with $E_*$. Let $L_*$ be the tautological line bundle over $X' = \mathbb{P}(E_*)$ and let $s : X'_* \to g^*E_* \otimes L_* =: F_*$ be the section induced by the canonical inclusion $(L_*)^i \hookrightarrow g^*E_*$:

$$s : X'_* \to \mathbb{P}(X'_*) \to g^*E_* \otimes L_* = F_*.$$

Put $F_i^i := g^*E_i^i \otimes L$ and $V_+^i := s^{-1}(F_i^i)$ for $0 \leq i \leq a$. Then $V_+^a$ is empty because $s$ does not vanish. Moreover $V_+^i$ is a simplicial object in $\mathcal{V}$ and the section

$$s_i : V_+^i \to F_+^i \to (F_+^i/F_+^{i+1})|_{V_+^i}$$

meets the zero section transversally for $0 \leq i \leq a - 1$, cf. [G] p. 147. On the other hand, to prove the Whitney sum formula, it is enough to show

$$\prod_{i=0}^{m+n-1} c_i(F_+^i/F_+^{i+1}) = 0.$$

We are thus reduced to the following simplicial analogue of loc. cit. p.141 Lemma 2:

**Lemma 5.6** Let $X_*$ be a simplicial object in $\mathcal{V}$. Let $E_*$ be a vector bundle of rank $a$ over $X_*$. Let $(E_*^i)_{0 \leq i \leq a}$ be a descending filtration consisting of subbundles on $E_*$ such that $E_*^i$ has rank $a - i$. For $1 \leq i \leq a$, put

$$\xi_i := c_1(E_*^{i-1}/E_*^i) \in H^2(X_*, \mathbb{T}_n(1)).$$

Let $s : X_* \to E_*$ be a section of $E_* \to X_*$. For $0 \leq i \leq a$, put

$$V_*^i := s^{-1}(E_*^i), \quad L_*^i := (E_*^i/E_*^{i+1})|_{V_*^i} \quad \text{(restriction of } E_*^i/E_*^{i+1 \text{ onto } V_*^i),}$$

and let $s_i : V_*^i \to L_*^i$ be the section induced by $s$. Assume the following condition:

- $V_*^i$ is a simplicial object in $\mathcal{V}$ for $0 \leq i \leq a$, and $s_i$ intersects the zero section transversally for $0 \leq i \leq a - 1$.

Then we have

$$\text{cl}_{X_*}(V_*^a) = \prod_{i=1}^a \xi_i \quad \text{in} \quad H^{2a}(X_*, \mathbb{T}_n(a)),$$

where $\text{cl}_{X_*}(V_*^a)$ denotes the value of 1 under the Gysin map

$$\mathbb{Z}/p^n = H^0(V_*^a, \mathbb{T}_n(0)) \to H^{2a}(X_*, \mathbb{T}_n(a)).$$

One can easily check this lemma by the properties of the Gysin morphisms in Proposition 5.4 and the arguments in loc. cit. p.141 Lemma 2. This completes the proof of Proposition 5.5.

Now let $X$ be a scheme which belongs to $\mathcal{V}$. Applying the construction of Chern classes to the case $X_* = B_s \text{GL}_{s,X}$ and $E_* = \text{universal rank } s$ bundle over $B_s \text{GL}_{s,X}$, we obtain Chern classes

$$c_j(E_*) \in H^{2j}(B_s \text{GL}_{s,X}, \mathbb{T}_n(j)) \quad (j \geq 0),$$

**Documenta Mathematica** 18 (2013) 177–247
which are called the universal rank $s$ Chern classes. On the other hand, let $I_n(j)^X$ be an injective resolution of the complex $C_n(j)^X$ on $X_{\text{ét}}$ defined in Definition 3.5, and consider the following complex of abelian sheaves on $X_{\text{Zar}}$:

$$
\cdots \to \epsilon_* I_n(j)^{2j-2} \to \epsilon_* I_n(j)^{2j-1} \to \epsilon_* I_n(j)^{2j} \to \ker(p^{2j}) \to \epsilon_* I_n(j)^{2j+1}.
$$

Here $\epsilon : X_{\text{ét}} \to X_{\text{Zar}}$ denotes the continuous map of small sites, and we regarded this sequence as a chain complex with the most right term placed in degree 0. We apply the Dold-Puppe construction [DP] to this complex to obtain a sheaf of simplicial abelian groups, which we denote by $K$. Let $\tau$ apply the Dold-Puppe construction [DP] to this complex to obtain a sheaf of simplicial abelian groups, which we denote by $K$. For a closed subset $Z$ of $X$ and non-negative integers $i, r \geq 0$, we define the Chern class map

$$
C_{r,i}^Z : K^Z(X) \to H_{Z}^{2r-i}(X, \mathbb{T}_n(r)) \tag{5.6.1}
$$

as the following composite map (cf. [Gi1] Definition 2.22):

$$
K^Z(X) \to H_{Z}^{1}(X_{\text{Zar}}, \mathbb{Z} \times \mathbb{Z}_\infty \mathbb{B}, \mathbb{GL}(\mathcal{O}_X)) \xrightarrow{pr} H_{Z}^{1}(X_{\text{Zar}}, \mathbb{Z}_\infty \mathbb{B}, \mathbb{GL}(\mathcal{O}_X))
$$

$$
\pi_1(\mathbb{Z}_\infty c_r) H_{Z}^{i-1}(X_{\text{Zar}}, \mathbb{Z}_\infty \mathbb{K}(\mathcal{O}_n(r), 2r)) \cong H_{Z}^{2r-i}(X_{\text{ét}}, \mathcal{T}_n(r)).
$$

Here $\mathbb{Z}_\infty$ denotes the Bousfield-Kan completion [BoK], and we have used the universal rank $s$ Chern class $c_r(E_s)$ for a sufficiently large $s$ to define the arrow $\pi_1(\mathbb{Z}_\infty c_r)$. See [Gi1] Proposition 2.15 for the first arrow and see loc. cit. p. 226 for the last isomorphism. The map $C_{r,i}^X$ agrees with $c_r$ for $X_s = X$ (constant simplicial scheme) defined before.

**Theorem 5.7**

1. $C_{r,i}^Z$ is contravariantly functorial in the pair $(X, Z)$, that is, for a morphism $f : X \to X'$ in $\mathcal{V}$ and a closed subset $Z' \subset X'$ with $f^{-1}(Z') \subset Z$, there is a commutative diagram

$$
\begin{array}{ccc}
K^Z_{r,i}(X') & \xrightarrow{C_{r,i}^{Z'}} & H_{Z}^{2r-i}(X', \mathcal{T}_n(r)) \\
\downarrow{f^*} & & \downarrow{f^*} \\
K^Z_{r,i}(X) & \xrightarrow{C_{r,i}} & H_{Z}^{2r-i}(X, \mathcal{T}_n(r)).
\end{array}
$$

2. $C_{r,i}^Z$ is additive for $i > 0$.

3. The induced Chern character

$$
\text{ch} : \bigoplus_{i \geq 0} K_i(X) \to \prod_{i,r \geq 0} H^i(X, \mathcal{T}_{Q_p}(r))
$$

with $H^i(X, \mathcal{T}_{Q_p}(r)) := \mathcal{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \geq 1} H^i(X, \mathcal{T}_n(r))$ is a ring homomorphism.

**Proof.** (1) follows from the functoriality results in Proposition 5.5 (2) and Proposition 3.8. The assertion (2) follows from Proposition 5.5 (3) and [Gi1] Lemma 2.26. See loc. cit. Definition 2.34 and Proposition 2.35 for the assertion (3).
6 Purity along log poles

This and the next section are devoted to the construction of cycle class maps from higher Chow groups to $p$-adic étale Tate twists.

Let $k$ be a field of characteristic $p > 0$, and let $Y$ be a normal crossing variety over $k$. Let $D$ be a non-empty admissible divisor on $Y$, and let $Z$ be a reduced closed subscheme of $Y$ which has codimension $\geq c$ and contained in $D$. Let $i : Z \hookrightarrow Y$ be the natural closed immersion. In this section, we first prove the following purity result:

**Theorem 6.1 (Purity)** We have $R^q i_! \nu^r_{(Y,D),n} = 0$ for $q \leq c$.

See Theorem 6.5 below for a consequence of this theorem. Note that the assertion for $q = c$ does not follow directly from Gros' purity [Gr1] II Théorème 3.5.8, even when $Y$ is smooth. The case that $Y$ is smooth has been considered and proved independently by Mieda ([Mi1] Proof of Theorem 2.4, [Mi2]). Although the proof of Theorem 6.1 given below is essentially a variant of Gros’ proof of [Gr1] II Théorème 3.5.8, we include detailed computations for the convenience of the reader.

**Proof of Theorem 6.1.** Since the problem is étale local, we may assume that $Y$ is simple and that there exists a pair $(Y, D)$ satisfying the following (1) and (2):

1. $Y$ is smooth over $k$ and contains $Y$ as a normal crossing divisor,
2. $D$ is a normal crossing divisor on $Y$ such that $Y \cup D$ has normal crossings on $Y$ and such that $Y \cap D = D$.

Then we have a short exact sequence on $\mathcal{Y}_{\text{et}}$

$$0 \longrightarrow W_n \Omega^{r+1}_{(Y,D),\log} \longrightarrow W_n \Omega^{r+1}_{(Y \cup D),\log} \longrightarrow \nu^r_{(Y,D),n} \longrightarrow 0$$

(a variant of (2.1.1)), which reduces the assertion to the case that $Y$ is smooth over $k$ and that $D$ has normal crossings on $Y$. Furthermore, we may assume that $Z$ is regular of pure codimension $c$ by a standard dévissage argument. In what follows, we prove

$$R^q i_! W_n \Omega^r_{(Y,D),\log} = 0 \quad \text{for} \quad q \neq c + 1,$$

(6.1.1)

assuming that $Y$ is smooth over $k$ (and that $Z$ is regular and contained in $D$). By the short exact sequences on $Y_{\text{et}}$,

$$0 \longrightarrow W_{n-1} \Omega^r_{(Y,D),\log} \longrightarrow W_n \Omega^r_{(Y,D),\log} \longrightarrow \Omega^r_{(Y,D),\log} \longrightarrow 0,$$

$$0 \longrightarrow \Omega^r_{(Y,D),\log} \longrightarrow \mathcal{Z}^r_Y(\log D) \longrightarrow \Omega^r_Y(\log D) \longrightarrow 0$$

(cf. (2.1.1) and [Sa1] Corollary 2.2.5 (2), Lemma 2.4.6), it is enough to show that

$$R^q i_! \mathcal{Z}^r_Y(\log D) = 0 = R^q i_! \Omega^r_Y(\log D) \quad \text{for} \quad q \neq c,$$

(6.1.2)

and

$$1 - C : R^c i_! \mathcal{Z}^r_Y(\log D) \longrightarrow R^c i_! \Omega^r_Y(\log D) \quad \text{is injective.}$$

(6.1.3)
We define an ascending filtration $\mathcal{F}^i$ where we regarded the sheaves on the right hand side as sheaves on $\mathcal{Y}_r$. The assertion (6.1.2) follows from the smoothness of $\mathcal{Y}_r$ and the fact that $\mathcal{Y}_r$ is affine, and there exists a regular sequence $(t_1, \ldots, t_c) \in \mathcal{Y}_r$ such that the ideal $(t_1, \ldots, t_c) \subset \mathcal{O}_{\mathcal{Y}_r}$ defines $Z$ and such that $t_1, \ldots, t_c (1 \leq a \leq c)$ are uniformizers of the irreducible components of $D$. Moreover, $\mathcal{Y}_r$ is free over $\mathcal{O}_{\mathcal{Y}_r}$ and has a basis $\{dt_a\}_{a=1}^c$ which contains $dt_1, \ldots, dt_c$.

We prove here the following lemma.

**Lemma 6.2** We recall here the following standard fact (cf. [Gr1] II (3.3.6)):

For $j = 1, \ldots, c$, let $\sigma_j$ be the natural open immersion $\sigma_j : \mathcal{Y}_r[t_1^{-1}, \ldots, t_c^{-1}] \to \mathcal{Y}_r$. We have isomorphisms of sheaves on $\mathcal{Y}_r$.

\[
\tau : \mathcal{Y}_r[t_1^{-1}, \ldots, t_c^{-1}] \to \mathcal{Y}_r.
\]

For $j = 1, \ldots, c$, let $\rho_j$ be the natural open immersion $\rho_j : \mathcal{Y}_r[t_1^{-1}, \ldots, t_j^{-1}, t_{j+1}^{-1}, \ldots, t_c^{-1}] \to \mathcal{Y}_r$.

We recall here the following standard fact (cf. [Gr1] II (3.3.6)):

**Lemma 6.2** We have isomorphisms of sheaves on $\mathcal{Y}_{\text{et}}$

\[
R^c i_! \Omega_{\mathcal{Y}_r}^\tau (\log D) \simeq \tau_* \tau^* \Omega_{\mathcal{Y}_r}^\tau (\log D) / \sum_{j=1}^c \sigma_j , \sigma_j^* \Omega_{\mathcal{Y}_r}^\tau (\log D) (6.2.1)
\]

\[
= \tau_* \tau^* \Omega_{\mathcal{Y}_r}^\tau / \left( \sum_{j=1}^c \sigma_j , \sigma_j^* \Omega_{\mathcal{Y}_r}^\tau (\log D) + \sum_{j=a+1}^c \sigma_j , \sigma_j^* \Omega_{\mathcal{Y}_r}^\tau \right),
\]

\[
R^c i_! \mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D) \simeq \tau_* \tau^* \mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D) / \sum_{j=1}^c \sigma_j , \sigma_j^* \mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D) (6.2.2)
\]

\[
= \tau_* \tau^* \mathcal{Z}_{\mathcal{Y}_r}^\tau / \left( \sum_{j=1}^c \sigma_j , \sigma_j^* \mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D) + \sum_{j=a+1}^c \sigma_j , \sigma_j^* \mathcal{Z}_{\mathcal{Y}_r}^\tau \right),
\]

where we regarded the sheaves on the right hand side as sheaves on $\mathcal{Y}_{\text{et}}$ naturally.

We define an ascending filtration $\mathcal{F}_m (m \geq 0)$ on $\tau_* \tau^* \Omega_{\mathcal{Y}_r}$ as

\[
\mathcal{F}_m (\tau_* \tau^* \Omega_{\mathcal{Y}_r}) := \left\{ \frac{1}{(t_1 t_2 \cdots t_c)^m} \omega \in \tau_* \tau^* \Omega_{\mathcal{Y}_r} \mid \omega \in \Omega_{\mathcal{Y}_r} \right\}.
\]

Let $\mathcal{F}_m (R^c i_! \Omega_{\mathcal{Y}_r}^\tau (\log D))$ be the induced filtration, and let $\mathcal{F}_m (R^c i_! \mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D))$ be its inverse image under the canonical map

\[
R^c i_! \mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D) \to R^c i_! \Omega_{\mathcal{Y}_r}^\tau (\log D) (6.2.3)
\]

induced by the natural inclusion $\mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D) \hookrightarrow \Omega_{\mathcal{Y}_r}^\tau (\log D))$. Note that

\[
R^c i_! \mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D) = \bigcup_{m=0}^{\infty} \mathcal{F}_m (R^c i_! \mathcal{Z}_{\mathcal{Y}_r}^\tau (\log D)).
\]

We prove here the following lemma.
Lemma 6.3  (1) The map (6.2.3) is injective.

(2) $\text{Fil}_m(R^{c_1} \mathcal{L}_Y^r (\log D))$ is generated by elements of the form
\[
\left[ \frac{1}{(t_1 t_2 \cdots t_c)^{pm}} \omega \right] \quad \text{with} \quad \omega \in \mathcal{L}_Y^r,
\]
where $[-]$ denotes the residue class in $\text{Fil}_m(R^{c_1} \mathcal{L}_Y^r (\log D))$ via (6.2.2).

(3) The kernel of the projection via (6.2.1)
\[
\text{Fil}_m(\tau_* \tau^* \Omega_Y^r) \longrightarrow \text{Fil}_m(R^{c_1} \Omega_Y^r (\log D))
\]
agrees with the subgroup
\[
\left\{ \sum_{j=1}^c \frac{1}{(t_1 \cdots t_{j-1} t_{j+1} \cdots t_c)^{pm}} \omega_j \mid \begin{array}{l}
\omega_j \in \Omega_Y^r (\log D_j) \quad (1 \leq j \leq a) \\
\omega_j \in \Omega_Y^r \quad (a < j \leq c)
\end{array} \right\},
\]
where $D_j \subset Y$ denotes the regular divisor defined by $t_j$ for $1 \leq j \leq a$.

Proof. We prove (1). By the short exact sequence
\[
0 \longrightarrow \mathcal{L}_Y^r (\log D) \longrightarrow \Omega_Y^r (\log D) \longrightarrow \mathcal{B}_Y^{r+1} (\log D) \longrightarrow 0,
\]
we have a long exact sequence on $\mathbb{Z}_{et}$
\[
\cdots \longrightarrow R^{c_1} \mathcal{B}_Y^{r+1} (\log D) \longrightarrow R^{c_1} \mathcal{L}_Y^r (\log D) \longrightarrow R^{c_1} \Omega_Y^r (\log D) \longrightarrow \cdots.
\]
The assertion follows from the smoothness of $Y$ and the fact that $\mathcal{B}_Y^{r+1} (\log D)$ is locally free over $(\mathcal{O}_Y)^p$. The assertion (2) is a consequence of (1), and (3) is straightforward.

We return to the proof of (6.1.3), and compute the map $1 - C$ in (6.1.3):
\[
1 - C : R^{c_1} \mathcal{L}_Y^r (\log D) \longrightarrow R^{c_1} \Omega_Y^r (\log D), \quad (6.3.1)
\]
using the filtration $\text{Fil}_\bullet$. Put
\[
\text{gr}_m(-) := \text{Fil}_m(-)/\text{Fil}_{m-1}(-).
\]
Since we have
\[
C\left( \frac{1}{(t_1 t_2 \cdots t_c)^{pm}} \omega \right) = \frac{1}{(t_1 t_2 \cdots t_c)^m} C(\omega) \quad \text{for} \quad \omega \in \mathcal{L}_Y^r,
\]
the Cartier operator preserves $\text{Fil}_\bullet$ (by Lemma 6.3 (2)) and induces a map
\[
\text{gr}_m(C) : \text{gr}_m(R^{c_1} \mathcal{L}_Y^r (\log D)) \longrightarrow \text{gr}_m(R^{c_1} \Omega_Y^r (\log D)).
\]
which is the zero map for \( m \geq 2 \). Hence we have

the kernel of (6.3.1) \( \subset \text{Fil}_1(\mathcal{R}^c i^! \mathcal{F}_Y^c(\log D)) \).

Let \( D_j \subset Y \) be as we defined in Lemma 6.3 (3) for \( 1 \leq j \leq a \). Fix an arbitrary \( \omega \in \mathcal{F}_Y^c \), and assume that

\[
\begin{align*}
x := \left[ \frac{1}{(t_1t_2 \cdots t_c)^p} \omega \right] \in \text{Fil}_1(\mathcal{R}^c i^! \mathcal{F}_Y^c(\log D))
\end{align*}
\]

belongs to the kernel of (6.3.1). To show (6.1.3), we have to prove that

\[
\omega = \sum_{j=1}^{c} t_j^p \eta_j \quad \text{for some} \quad \eta_j \in \Omega_Y^r(\log D_j) \quad (1 \leq j \leq a)
\]

\[
\eta_j \in \Omega_Y^r \quad (a < j \leq c)
\]

in \( \Omega_Y^r(\log D) \), which is equivalent to that \( x = 0 \) in \( \mathcal{R}^c i^! \mathcal{F}_Y^c(\log D) \) by Lemma 6.3 (1) and (6.2.1). Since \( x = C(x) \) in \( \text{Fil}_1(\mathcal{R}^c i^! \mathcal{F}_Y^c(\log D)) \) by assumption, we have

\[
\frac{1}{(t_1t_2 \cdots t_c)^p} \omega - \frac{1}{t_1t_2 \cdots t_c} C(\omega) = \sum_{j=1}^{c} \frac{1}{(t_1t_2 \cdots t_{j-1}t_{j+1} \cdots t_c)^p} \alpha_j
\]

in \( \tau \cdot \tau^* \Omega_Y^r \) for some \( \alpha_j \in \Omega_Y^r(\log D_j) \) (\( 1 \leq j \leq a \)) and \( \alpha_j \in \Omega_Y^r \) (\( a < j \leq c \)) by Lemma 6.3 (3). This implies

\[
\omega = (t_1t_2 \cdots t_c)^{p-1}C(\omega) + \sum_{j=1}^{c} t_j^p \alpha_j \quad \text{in} \quad \Omega_Y^r,
\]

and our task is to prove that

\[
(t_1t_2 \cdots t_c)^{p-1}C(\omega) = \sum_{j=1}^{c} t_j^p \zeta_j \quad \text{for some} \quad \zeta_j \in \Omega_Y^r(\log D_j) \quad (1 \leq j \leq a)
\]

\[
\zeta_j \in \Omega_Y^r \quad (a < j \leq c).
\]

(6.3.4)

Take a basis \( \{ dt_\lambda \}_{\lambda \in A} \) of \( \mathcal{F}_Y^c \) over \( \mathcal{O}_Y^c \) which contain \( dt_1, \ldots, dt_c \) (see the condition (\( \ast \))). Fix an ordering on \( A \) and let \( J \) be the set of all \( r \)-tuples \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r) \) of elements of \( A \) satisfying

\[
\lambda_1 < \lambda_2 < \cdots < \lambda_r.
\]

For \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r) \in J \), put

\[
dt_\lambda := dt_{\lambda_1} \wedge dt_{\lambda_2} \wedge \cdots \wedge dt_{\lambda_r} \in \Omega_Y^r.
\]

Using the basis \( \{ dt_\lambda \}_{\lambda \in J} \) of \( \Omega_Y^r \) over \( \mathcal{O}_Y^c \), we decompose \( C(\omega) \) as

\[
C(\omega) = \epsilon + \delta \quad (\epsilon, \delta \in \Omega_Y^r).
\]

(6.3.5)
Here $\epsilon$ is an $O_Y$-linear combination of $dt_\lambda$’s which contain $dt_j$ for some $1 \leq j \leq a$, and $\delta$ is an $O_Y$-linear combination of $dt_\lambda$’s which contain none of $dt_1, \ldots, dt_a$. We have
\[(t_1 t_2 \cdots t_c)^{p-1} \epsilon = \sum_{j=1}^{a} t_j^p \beta_j \text{ for some } \beta_j \in \Omega_Y^r(\log D_j) \quad (1 \leq j \leq a), \quad (6.3.6)\]
which implies (6.3.4) if $\delta = 0$. Otherwise, we proceed as follows. Put 
\[y := (t_2 t_3 \cdots t_c) t_1^{p-1} \delta \in \Omega_Y^r.
\]

Since $d\omega = 0$, the equalities (6.3.3) and (6.3.6) imply 
\[t_1^{p-2} dt_1 \wedge y = t_1^{p-1} dy + \sum_{j=1}^{c-1} t_j^p \beta'_j \text{ for some } \begin{cases} \beta'_{m} \in \Omega_Y^r(\log D_j) & (1 \leq j \leq a) \\ \beta'_j \in \Omega_Y^r(a < j \leq c) \end{cases} \quad (6.3.7)
\]

By this equality and the assumption on $\delta$, one can easily check that 
\[t_1^{p-2} dt_1 \wedge y = \sum_{j=2}^{c} t_j^p \beta''_j \text{ for some } \begin{cases} \beta''_{m} \in \Omega_Y^r(\log D_1) \\ \beta''_j \in \Omega_Y^r(2 \leq j \leq c) \end{cases} \quad (6.3.7)
\]
To proceed with the proof of (6.3.4), we need the following lemma:

**Lemma 6.4** Let $f$ and $g$ be the following $O_Y$-linear maps, respectively:
\[f : \Omega_Y^r \longrightarrow \Omega_Y^{r+1}, \quad z \mapsto dt_1 \wedge z,
\]
\[g : \Omega_Y^r(\log D_1) \longrightarrow \Omega_Y^{r+1}(\log D_1), \quad z \mapsto \frac{dt_1}{t_1} \wedge z.
\]
Then the following holds:

1. For $m \geq 1$, we have 
\[\operatorname{Im}(f) \cap \left( t_1^m \Omega_Y^{r+1}(\log D_1) + \sum_{j=2}^{c} t_j^m \Omega_Y^{r+1} \right) = t_1^m \operatorname{Im}(g) + \sum_{j=2}^{c} t_j^m \operatorname{Im}(f).
\]
2. For $m, n \geq 0$, we have 
\[t_1^m \Omega_Y^{r+1} \cap \left( \sum_{j=2}^{c} t_j^n \operatorname{Im}(f) \right) = \sum_{j=2}^{c} t_1^m t_j^n \operatorname{Im}(f).
\]
3. For $m \geq 0$, the sequence 
\[t_1^m \Omega_Y^{r-1} \xrightarrow{f} t_1^m \Omega_Y^r \xrightarrow{f} t_1^m \Omega_Y^{r+1}\]
is exact.
Proof. These assertions follow from linear algebra over $\mathcal{O}_Y$. The details are straightforward and left to the reader. □

By (6.3.7) and Lemma 6.4 (1) for $m = p$, we have

$$t^{p-2}_1 dt_1 \wedge y = t^{p-1}_1 dt_1 \wedge \gamma_1 + \sum_{j=2}^{c} t^{p}_j dt_1 \wedge \gamma_j$$  \hspace{1cm} (6.4.1)

for some $\gamma_1 \in \Omega^r_Y(\log D_1)$ and some $\gamma_j \in \Omega^r_Y$ $(2 \leq j \leq c)$. If $p = 2$, this equality implies

$$t^{p-2}_1 dt_1 \wedge y = t^{p-1}_1 dt_1 \wedge \left( t^{p}_1 \gamma_1 + \sum_{j=2}^{c} t^{p}_j \gamma_j \right).$$ \hspace{1cm} (6.4.2)

If $p \geq 3$, we obtain (6.4.2) from (6.4.1) by replacing $\gamma_2, \ldots, \gamma_c$ suitably in $\Omega^r_Y(2 \leq j \leq c)$, where we used Lemma 6.4 (2) for $(m,n) = (p,p-2)$. Noting that $t^{p}_1 \gamma_1$ and $t^{p}_j \gamma_j (2 \leq j \leq c)$ belong to $\Omega^r_Y$, we have

$$t^{p-2}_1 t^{p}_1 \gamma_1 + \sum_{j=2}^{c} t^{p}_j (t^{p-2}_1 \gamma_j)$$

by (6.4.2) and Lemma 6.4 (3) for $m = p - 2$. Thus we have

$$(t_1 t_2 \cdots t_c)^{p-1} \delta = t^{p-1}_1 y = t^{p}_1 \left( \frac{dt_1}{t_1} \wedge \vartheta + \gamma_1 \right) + \sum_{j=2}^{c} t^{p}_j (t^{p-2}_1 \gamma_j),$$

which implies (6.3.4) (see also (6.3.5), (6.3.6)). This completes the proof of (6.3.2), (6.1.3) and Theorem 6.1. □

**Theorem 6.5** Let $S, p$ and $X$ be as in Setting 3.1, and let $c$ be a positive integer. Let $D$ be a non-empty normal crossing divisor on $X$, and let $Z$ be a closed subset of $X$ which has codimension $\geq c$. Put $U := X - D$. Then we have

$$H^q_Z(X, \mathbb{T}_n(r)_{(X,D)}) \simeq \begin{cases} 0 & (q < r + c) \\ H^{q+c}_{Z \cap U}(U, \mathbb{T}_n(r)_{U}) & (q = r + c). \end{cases}$$ \hspace{1cm} (6.5.1)

In particular, when $Z$ has pure codimension $r$ on $X$, we have

$$H^q_Z(X, \mathbb{T}_n(r)_{(X,D)}) \simeq \begin{cases} 0 & (q < 2r) \\ \mathbb{Z}/p^n[Z^0 \cap U] & (q = 2r). \end{cases}$$ \hspace{1cm} (6.5.2)

where $\mathbb{Z}/p^n[Z^0 \cap U]$ means the free $\mathbb{Z}/p^n$-module generated by the set $Z^0 \cap U$.

*Proof.* Admitting (6.5.1), we obtain (6.5.2) from the purity of $\mathbb{T}_n(r)_{U}$ ([Sa2] Theorem 4.4.7, Corollary 4.4.9) and the absolute purity of $\mu^\infty_{(0)}$ on $U[p^{-1}]$ ([FG]). To show (6.5.1), we divide the problem into the following 4 cases:
(1) $Z \subset D \cap Y$  
(2) $Z \subset D$ and $Z \not\subset Y$

(3) $Z$ is arbitrary and $q < r + c$  
(4) $Z$ is arbitrary and $q = r + c$.

The case (1) follows from Theorems 3.3, 6.1 (see also Remark 3.13) and the same arguments as in [Sa2] Theorem 4.4.7. The case (2) follows from the case (1) and the same arguments as in loc. cit. Corollary 4.4.9. The case (3) also follows from similar arguments as for the previous cases. To prove the case (4), write $Z$ as the union of closed subsets

$$Z = Z_1 \cup Z_2,$$

where $Z_1$ has pure codimension $c$ on $X$, $Z_2$ has codimension $\geq c + 1$ on $X$ and we suppose that no irreducible components of $Z$ are contained in $Z_1$. Since $Z_1 \cap Z_2$ has codimension $\geq c + 2$ on $X$, the assertion in the cases (1)--(3) and a Mayer-Vietoris long exact sequence

$$\cdots \rightarrow H^{r+\epsilon}_{Z_1 \cap Z_2} (X, \mathfrak{T}_n(r)(X,D)) \rightarrow \bigoplus_{j=1,2} H^{r+\epsilon}_{Z_j} (X, \mathfrak{T}_n(r)(X,D))$$

$$\rightarrow H^{r+\epsilon}_{Z} (X, \mathfrak{T}_n(r)(X,D)) \rightarrow H^{r+\epsilon+1}_{Z_1 \cap Z_2} (X, \mathfrak{T}_n(r)(X,D)) \rightarrow \cdots$$

imply isomorphisms

$$H^{r+\epsilon}_{Z} (X, \mathfrak{T}_n(r)(X,D)) \cong \bigoplus_{j=1,2} H^{r+\epsilon}_{Z_j} (X, \mathfrak{T}_n(r)(X,D)) \cong H^{r+\epsilon}_{Z_1} (X, \mathfrak{T}_n(r)(X,D)).$$

Similarly, we have $H^{r+\epsilon}_{Z \cap U} (U, \mathfrak{T}_n(r)(U)) \cong H^{r+\epsilon}_{Z_1 \cap U} (U, \mathfrak{T}_n(r)(U))$ by the purity of $\mathfrak{T}_n(r)(U)$ ([Sa2] Corollary 4.4.9). Hence we may assume that $Z$ has pure codimension $c$ on $X$. Moreover we may assume that no irreducible components of $Z$ are contained in $D$ by the cases (1)+(2) and a similar devissage argument. Then noting that $Z \cap D$ has codimension $\geq c + 1$ on $X$, we obtain the case (4) from the cases (1)+(2) and a long exact sequence

$$\cdots \rightarrow H^{r+\epsilon}_{Z \cap D} (X, \mathfrak{T}_n(r)(X,D)) \rightarrow H^{r+\epsilon}_{Z} (X, \mathfrak{T}_n(r)(X,D)) \rightarrow H^{r+\epsilon}_{Z \cap U} (U, \mathfrak{T}_n(r)(U))$$

$$\rightarrow H^{r+\epsilon+1}_{Z \cap D} (X, \mathfrak{T}_n(r)(X,D)) \rightarrow \cdots$$

This completes the proof of Theorem 6.5. $\square$

Definition 6.6 Let $X$ and $D$ be as in Theorem 6.5, and let $C$ be a cycle on $U := X - D$ of codimension $r$. Let $W \subset U$ be the support of $C$ and let $\overline{W}$ be its closure in $X$. Then we define the cycle class

$$\text{cl}_X(C) \in H^{2r}_{\overline{W}}(X, \mathfrak{T}_n(r)(X,D))$$

as the inverse image of $\text{cl}_U(C)$ ([Sa2] 5.1.2) under the isomorphism in Theorem 6.5

$$H^{2r}_{\overline{W}}(X, \mathfrak{T}_n(r)(X,D)) \xrightarrow{\sim} H^{2r}_{\overline{U}}(U, \mathfrak{T}_n(r)(U)).$$

Documenta Mathematica 18 (2013) 177–247
7 Cycle class map

Let $S, p$ and $X$ be as in Setting 3.1. In this section, we construct a cycle class morphism
\[
\cl_X : \Z(r)_X^\et \otimes_{\Z/p^n} \to \Sigma_n(r)_X \iso D(X_\et, \Z/p^n),
\]
(7.0.1)
following the method of Bloch [B3] § 4 (cf. [GL] § 3, see also Remark 7.3 below). Here $\Z(r)_X^\et$ denotes the étale sheafification on $X$ of the presheaf of cochain complexes
\[
U \mapsto z^r(U, \ast)[-2r],
\]
and $z^r(U, \ast)$ denotes Bloch’s cycle complex
\[
\cdots \longrightarrow z^r(U, q) \xrightarrow{d_q} z^r(U, q-1) \xrightarrow{d_{q-1}} \cdots \xrightarrow{d_1} z^r(U, 0).
\]

We review the definition of this complex briefly, which will be useful later. Let $\Delta^q$ be the standard cosimplicial scheme over $\Spec(\Z)$:
\[
\Delta^q := \Spec(\Z[t_0, t_1, \ldots, t_q] / (t_0 + t_1 + \cdots + t_q = 1)).
\]
A face of $\Delta^q$ (of codimension $a \geq 1$) is a closed subscheme defined by the equation
\[
t_{i_1} = t_{i_2} = \cdots = t_{i_a} = 0 \quad \text{for some} \quad 0 \leq i_1 < i_2 < \cdots < i_a \leq q.
\]
Now $z^r(U, q)$ is defined as the free abelian group generated by the set of all integral closed subschemes on $U \times \Delta^q$ of codimension $r$ which meet all faces of $U \times \Delta^q$ properly. Here a face of $U \times \Delta^q$ means the product of $U$ and a face of $\Delta^q$. Noting that the faces of codimension 1 are effective Cartier divisors, we define the differential map $d_q$ as the alternating sum of pull-back maps along the faces of codimension 1, which defines the above complex $z^r(U, \ast)$.

We fix a projective completion $\overline{\Delta^q}$ of $\Delta^q$ as follows:
\[
\overline{\Delta^q} := \Proj(\Z[T_0, T_1, \ldots, T_q, T_{\infty}] / (T_0 + T_1 + \cdots + T_q = T_{\infty})).
\]
Let $D^q \subset \overline{\Delta^q}$ be the hyperplane at infinity:
\[
D^q : T_{\infty} = 0.
\]

The following proposition plays a key role in our construction of the morphism (7.0.1).

**Proposition 7.1** Let $q$ and $r$ be integers with $q, r \geq 0$, and let $U$ be étale of finite type over $X$. Let $\Sigma^{r,q}$ be the set of all closed subsets on $U \times \Delta^q$ of pure codimension $r$ which meet all faces of $U \times \Delta^q$ properly. Here a face of $U \times \Delta^q$ means the product of $U$ and a face of $\Delta^q$. Noting that the faces of codimension 1 are effective Cartier divisors, we define the differential map $d_q$ as the alternating sum of pull-back maps along the faces of codimension 1, which defines the above complex $z^r(U, \ast)$.

We fix a projective completion $\overline{\Delta^q}$ of $\Delta^q$ as follows:
\[
\overline{\Delta^q} := \Proj(\Z[T_0, T_1, \ldots, T_q, T_{\infty}] / (T_0 + T_1 + \cdots + T_q = T_{\infty})).
\]
Let $D^q \subset \overline{\Delta^q}$ be the hyperplane at infinity:
\[
D^q : T_{\infty} = 0.
\]

The following proposition plays a key role in our construction of the morphism (7.0.1).

**Proposition 7.1** Let $q$ and $r$ be integers with $q, r \geq 0$, and let $U$ be étale of finite type over $X$. Let $\Sigma^{r,q}$ be the set of all closed subsets on $U \times \Delta^q$ of pure codimension $r$ which meet all faces of $U \times \Delta^q$ properly. For $W \in \Sigma^{r,q}$, let $\overline{W}$ be the closure of $W$ in $U \times \overline{\Delta^q}$. Then:

1. There is an isomorphism
\[
\cl^{r,q} : z^r(U, q) \otimes \Z/p^n \iso \lim_{W \in \Sigma^{r,q}} H^{2r}_{\overline{\Delta^q}}(U \times \overline{\Delta^q}, \Sigma_n(r)(U \times \overline{\Delta^q} \setminus D^q))
\]
sending a cycle $C \in z^r(U, q)$ to the cycle class $\cl_{U \times \overline{\Delta^q}}(C)$ (see Definition 6.6).
(2) For \( W \in \Sigma^{r,q} \), the natural morphism

\[
\tau \leq 2r \ R\Gamma_{\mathcal{W}} \left( U \times \Delta^q, \mathcal{I}(r)_{U \times \Delta^q, U \times D^q} \right) \\
\longrightarrow H^{2r}_{U \times \Delta^q, \mathcal{I}(r)_{U \times \Delta^q, U \times D^q}} \left( U \times \Delta^q, \mathcal{I}(r)_{U \times \Delta^q, U \times D^q} \right) \left[ -2r \right]
\]

is an isomorphism in the derived category of \( \mathbb{Z}/p^n \)-modules.

(3) Let \( C \) be a cycle which belongs to \( z'(U, q) \), and let \( W \) be the support of \( C \) (note that \( W \) belongs to \( \Sigma^{r,q} \)). Let \( \overline{i} : U \times \Delta^q \hookrightarrow U \times \Delta^q \) be the closure of a face map \( i : U \times \Delta^{q-1} \hookrightarrow U \times \Delta^q \). Then the pull-back map

\[
\overline{i}^* : H^{2r}_{U \times \Delta^q, \mathcal{I}(r)_{U \times \Delta^q, U \times D^q}} \left( U \times \Delta^q, \mathcal{I}(r)_{U \times \Delta^q, U \times D^q} \right) \\
\longrightarrow H^{2r}_{U \times \Delta^{q-1}, \mathcal{I}(r)_{U \times \Delta^{q-1}, U \times D^{q-1}}} \left( U \times \Delta^{q-1}, \mathcal{I}(r)_{U \times \Delta^{q-1}, U \times D^{q-1}} \right)
\]

sends the cycle class \( \text{cl}_{U \times \Delta^q}(C) \) to \( \text{cl}_{U \times \Delta^{q-1}}(i^* C) \), where \( i^* C \) denotes the pull-back of the cycle \( C \) along \( i \).

Proof. (1) and (2) follow from Theorem 6.5 and the definition of \( z'(U, q) \). We show (3). By Theorem 6.5, it is enough to show that the pull-back map

\[
i^* : H^{2r}_{U \times \Delta^q, \mathcal{I}(r)_{U \times \Delta^q}} \left( U \times \Delta^q, \mathcal{I}(r)_{U \times \Delta^q} \right) \\
\longrightarrow H^{2r}_{U \times \Delta^{q-1}, \mathcal{I}(r)_{U \times \Delta^{q-1}}} \left( U \times \Delta^{q-1}, \mathcal{I}(r)_{U \times \Delta^{q-1}} \right)
\]

sends the cycle class \( \text{cl}_{U \times \Delta^q}(C) \) to \( \text{cl}_{U \times \Delta^{q-1}}(i^* C) \). Put

\[
\mathcal{U} := U \times \Delta^q, \quad \mathcal{D} := i(U \times \Delta^{q-1}) \subset \mathcal{U} \quad \text{and} \quad \mathcal{Y} := \mathcal{U} - \mathcal{D}
\]

and let \( t \in \Gamma(\mathcal{U}, \mathcal{O}_\mathcal{U}) \) be a defining equation of \( \mathcal{D} \). Noting that \( i^{-1}(W) = W \cap \mathcal{D} \), consider the following diagram:

\[
\begin{array}{ccc}
H^{2r}_{U \times \Delta^q, \mathcal{I}(r)_{U \times \Delta^q}}(\mathcal{Y}, \mathcal{I}(r)_{U \times \Delta^q}) & \overset{i^*}{\longrightarrow} & H^{2r}_{U \times \Delta^{q-1}, \mathcal{I}(r)_{U \times \Delta^{q-1}}}(\mathcal{Y}, \mathcal{I}(r)_{U \times \Delta^{q-1}}) \\
\downarrow \alpha & & \downarrow i_* \\
H^{2r+1}_{U \times \Delta^q}(\mathcal{Y}, \mathcal{I}(r+1)_{U \times \Delta^q}) & \overset{-\delta}{\longrightarrow} & H^{2r+2}_{U \times \Delta^q}(\mathcal{Y}, \mathcal{I}(r+1)_{U \times \Delta^q}) \\
\bigoplus_{y \in W^o} H^{2r+1}_{U \times \Delta^q}(\mathcal{Y}, \mathcal{I}(r+1)_{U \times \Delta^q}) & \overset{-\delta}{\longrightarrow} & \bigoplus_{x \in (W \cap \mathcal{D})^o} H^{2r+2}_{U \times \Delta^q}(\mathcal{Y}, \mathcal{I}(r+1)_{U \times \Delta^q}) \\
\bigoplus_{y \in W^o} \kappa(y)^{x/p^n} & \overset{\text{div}}{\longrightarrow} & \bigoplus_{x \in (W \cap \mathcal{D})^o} \mathbb{Z}/p^n.
\end{array}
\]

Here we defined \( \alpha \) by sending \( \omega \in H^{2r}_{U \times \Delta^q}(\mathcal{Y}, \mathcal{I}(r)_{U \times \Delta^q}) \) to \( \{ t \} \cup \omega \mid \mathcal{Y} \), where \( \{ t \} \mid \mathcal{Y} \) denotes the class of \( t \mid \mathcal{Y} \in \Gamma(\mathcal{Y}, \mathcal{O}_\mathcal{Y}) \) in \( H^1(\mathcal{Y}, \mathcal{I}(1)_{\mathcal{Y}}) \). The arrows \( i_* \), \( g \) and \( g' \) are
Gysin isomorphisms (cf. [Sa2] Theorem 4.4.7). The arrows $\delta$ are boundary maps of localization long exact sequences, and the central square commutes obviously. The arrow $\delta$ denotes the divisor map, and the bottom square commutes by the compatibility in loc. cit. Theorem 6.1.1 and [JSS] Theorem 3.1.1 (see also Remark 3.7 of this paper). The top square commutes by the following equalities ($\omega \in H_W^0(\mathcal{Y}, \mathcal{T}_n(r))$):

$$i_* \circ i^*(\omega) = c_{\mathcal{Y}}(\mathcal{D}) \cup \omega = -\delta(t \cdot \mathcal{Y}) \cup \omega = -\delta(t \cdot \mathcal{Y} \cup \omega) = -\delta \circ \alpha(\omega),$$

where the first (resp. second) equality follows from the projection formula in [Sa2] Corollary 7.2.4 (resp. the same compatibility as mentioned before). Hence the assertion follows from the transitivity of Gysin maps for $W \cap \mathcal{D} \hookrightarrow \mathcal{D} \hookrightarrow \mathcal{Y}$ (loc. cit. Corollary 6.3.3) and the fact that the divisor map $\text{div}$ sends $t|_y \in \kappa(y)^*/p^n$ to the cycle $i^*[y]$, where $[y]$ means the cycle on $\mathcal{Y}$ given by the closure of $y$. This completes the proof. □

We construct the morphism (7.0.1). For $U$ as in the proposition, there is a diagram of cochain complexes concerning $\bullet$ (see Corollary 3.9 (2) for $G_n(r)\bullet$):

$$z^*(U, q) \otimes \mathbb{Z}/p^n[-2r] \xrightarrow{d^r \downarrow} \lim_{W \in \Sigma^r q} H_W^2(U \times \Delta^r, \mathcal{T}_n(r)(U \times \Delta^r, U \times D^n))[-2r]$$

$$\xleftarrow{\alpha^r \downarrow} \lim_{W \in \Sigma^r q} \tau_{\leq 2r}^{-1} \Gamma(W \times \Delta^r, G_n(r)^\bullet(U \times \Delta^r, U \times D^n))$$

$$\xrightarrow{\beta^r \downarrow} \Gamma(U \times \Delta^r, G_n(r)^\bullet(U \times \Delta^r, U \times D^n)).$$

Here $\alpha^r \downarrow$ and $\beta^r \downarrow$ are natural maps of complexes, which are obviously contravariant for the face maps $U \times \Delta^{r-1} \hookrightarrow U \times \Delta^r$. The map $e^r \downarrow$ is bijective and contravariant for these face maps by Proposition 7.1 (1) and (3). Hence we get homomorphisms of double complexes concerning $(\bullet, \bullet)$:

$$z^*(U, \bullet) \otimes \mathbb{Z}/p^n[-2r] \xrightarrow{(d^r \downarrow)^{-1} \circ \alpha^r \downarrow} \lim_{W \in \Sigma^r \bullet} \tau_{\leq 2r}^{-1} \Gamma(W \times \Delta^r, G_n(r)^\bullet(U \times \Delta^r, U \times D^n))$$

$$\xrightarrow{\beta^r \downarrow} \Gamma(U \times \Delta^r, G_n(r)^\bullet(U \times \Delta^r, U \times D^n))$$

$$\xleftarrow{} \Gamma(U, G_n(r)_U),$$

where the differentials in the $\bullet$-direction are alternating sums of pull-back maps along the faces of codimension 1, and the last arrow is the inclusion to the factor of $\bullet = 0$. The first and the last arrows are quasi-isomorphisms on the associated total complexes by Proposition 7.1 (2) and Corollary 4.3, respectively. We thus obtain the desired morphism (7.0.1) in $D(X_{et}, \mathbb{Z}/p^n)$ by sheafifying the diagram of total complexes.

Remark 7.2 The Rost-Voevodsky theorem ([Vo1], [Vo2], [We], cf. [BK1] §3) and the Suslin-Voevodsky theorem [SV] (cf. [GL]) imply that the morphism (7.0.1) induces isomorphisms (cf. [Sa2] Conjecture 1.4.1):

$$\tau_{\leq r}(\mathbb{Z}(r)_X^t \otimes \mathbb{Z}/p^n) \xrightarrow{\sim} \mathcal{T}_n(r)_X \quad \text{in} \ D^b(X_{et}, \mathbb{Z}/p^n),$$

$$\tau_{\leq r}(\mathbb{Z}(r)_X^{2r} \otimes \mathbb{Z}/p^n) \xrightarrow{\sim} \tau_{\leq r} R\varepsilon_* \mathcal{T}_n(r)_X \quad \text{in} \ D^b(\mathcal{Z}_{zar}, \mathbb{Z}/p^n),$$

$$\tau_{\leq r}(\mathbb{Z}(r)_X^{3r} \otimes \mathbb{Z}/p^n) \xrightarrow{\sim} \tau_{\leq r} R\varepsilon_* \mathcal{T}_n(r)_X \quad \text{in} \ D^b(\mathcal{Z}_{zar}, \mathbb{Z}/p^n),$$

$$\tau_{\leq r}(\mathbb{Z}(r)_X^{4r} \otimes \mathbb{Z}/p^n) \xrightarrow{\sim} \tau_{\leq r} R\varepsilon_* \mathcal{T}_n(r)_X \quad \text{in} \ D^b(\mathcal{Z}_{zar}, \mathbb{Z}/p^n).$$
which one can easily check by similar arguments as in [SS] §A.2. Here $\mathcal{Z}(r)^{\text{zar}}_X$ denotes the complex of Zariski sheaves on $X$ defined as $U \mapsto z^r(U, \ast)[-2r]$, and $\varepsilon$ denotes the continuous map of sites $X_{\text{et}} \to X_{\text{zar}}$.

**Remark 7.3** If $X$ is smooth over $S$, then $\mathcal{Z}(r)^{\text{zar}}_X$ is concentrated in degrees $\leq r$ by a result of Geisser [Ge] Corollary 4.4. Hence his arguments in loc. cit. §6, Proof of Theorem 1.3 gives an alternative construction of (7.0.1) in this case.

8 de Rham-Witt cohomology and homology

In this section we define two kinds of de Rham-Witt complexes for normal crossing varieties playing the roles of cohomology and homology, and relate them with the modified de Rham-Witt complex of Hyodo [Hy2]. The main results of this section are Theorems 8.8 and 8.12 below.

Let $k$ be a field of characteristic $p > 0$, and let $Y$ be a normal crossing variety over $k$. Let $U \subset Y$ be a dense open subset which is smooth over $k$, and let $\sigma : U \hookrightarrow Y$ be the natural open immersion.

**Definition 8.1** We define the complex $(W_n A^\bullet_Y, d)$ of étale sheaves on $Y$ as the differential graded subalgebra of $\sigma^\ast W_n \Omega^\bullet_U$ generated by

$$W_n \theta_Y \quad \text{and} \quad \lambda^1_{n,\sigma} (\subset \sigma^\ast W_n \Omega^1_U),$$

which we call the cohomological de Rham-Witt complex of $Y$. It is easy to see that $W_n A^\bullet_Y$ does not depend on the choice of $U$. For $n = 1$, $W_1 A^\bullet_Y$ is the same as the complex $A^\bullet_Y$ defined in [Sa1] §3.3.

By the relations

$$\begin{align*}
V(aR(x)) &= V(a)x \quad (a \in W_n \theta_U, x \in W_{n+1} \Omega^{\text{log}}_U) \\
F(ax) &= F(a)R(x) \quad (a \in W_{n+1} \theta_U, x \in W_{n+1} \Omega^{\text{log}}_U) \\
R(ax) &= R(a)R(x) \quad (a \in W_{n+1} \theta_U, x \in W_{n+1} \Omega^{\text{log}}_U)
\end{align*}$$

(cf. [I1] I Théorème 2.17, Proposition 2.18), the operators $V, F, R$ on $W_n \Omega^\bullet_U$ induce operators

$$\begin{align*}
V : W_n A_Y^1 &\longrightarrow W_{n+1} A_Y^1, \\
F : W_{n+1} A_Y^1 &\longrightarrow W_n A_Y^1, \\
R : W_{n+1} A_Y^1 &\longrightarrow W_n A_Y^1,
\end{align*}$$

which satisfy relations

$$\begin{align*}
VF &= VF = p, \quad FdV = d, \quad dF = pFd, \quad Vd = pdV, \\
RV &= VR, \quad RF = FR, \quad Rd = dR.
\end{align*}$$

The local structure of $W_n A^1_Y$ can be written in terms of the usual Hodge-Witt sheaves of the strata of $Y$. 

**Documenta Mathematica** 18 (2013) 177–247
Proposition 8.2  
(1) Assume that \( Y \) is simple, and let \( Y_1, Y_2, \ldots, Y_q \) be the irreducible components of \( Y \). Then there is an exact sequence on \( Y_{\text{et}} \):

\[
0 \rightarrow W^n A_Y^r \xrightarrow{\sigma^r} \bigoplus_{|I|=1} W^n \Omega^r_{\overline{Y}} \xrightarrow{\theta^1} \bigoplus_{|I|=2} W^n \Omega^r_{Y_I} \xrightarrow{\theta^2} \cdots \xrightarrow{\theta^{q-1}} \bigoplus_{|I|=q} W^n \Omega^r_{Y_I} \rightarrow 0,
\]

where the notation is the same as in Proposition 2.2.

(2) Assume that \( Y \) is embedded into a smooth \( k \)-variety \( \mathcal{Y} \) as a normal crossing divisor. Let \( i : Y \hookrightarrow \mathcal{Y} \) be the closed immersion. Then there is a short exact sequence on \( Y_{\text{et}} \):

\[
0 \rightarrow W^n \Omega^r_{\mathcal{Y}}(\log Y) \rightarrow W^n \Omega^r_{\mathcal{Y}} \rightarrow i_* W^n A_Y^r \rightarrow 0.
\]

Proof. (1) follows from the contravariant functoriality of Hodge-Witt sheaves of smooth varieties and the same arguments as for [Sa1] Proposition 3.2.1. The assertion (2) follows from (1) and [Mo] Lemma 3.15.1.

We next introduce homological Hodge-Witt sheaves.

Definition 8.3 For \( r \geq 0 \), we define the étale sheaf \( W^n \Xi^r_Y \) on \( Y \) as the \( W^n \mathcal{O}_Y \)-submodule of \( \sigma_* W^n \Omega^r_U \) generated by \( \nu^r_{Y,n} \). Similarly as for \( W^n \Lambda^r_Y \), the sheaf \( W^n \Xi^r_Y \) does not depend on the choice of \( U \). For \( n = 1 \), \( W^1 \Xi^r_Y \) agrees with the sheaf \( \Xi^r_Y \) defined in [Sa1] §2.5.

Since \( \lambda^r_{Y,n} \subset \nu^r_{Y,n} \) by definition, we have \( W^n A^r_Y \subset W^n \Xi^r_Y \), which agree with \( W^n \Omega^r_Y \) when \( Y \) is smooth (cf. [Sa1] Remark 3.1.4). The following local description will be useful later.

Proposition 8.4  
(1) Under the same setting and notation as in Proposition 8.2 (1), there is a canonical ascending filtration \( F_\alpha (\alpha \geq 0) \) on \( W^n \Xi^r_Y \) satisfying

\[
F_0(W^n \Xi^r_Y) = 0 \quad \text{and} \quad \bigoplus_{|I|=\alpha} W^n \Omega^r_{Y_I} \simeq G^r_{\alpha} W^n \Xi^r_Y,
\]

where the isomorphism for \( \alpha \geq 2 \) depends on the fixed ordering of the irreducible components of \( Y \).

(2) Under the same setting and notation as in Proposition 8.2 (2), there is a short exact sequence on \( Y_{\text{et}} \):

\[
0 \rightarrow W^n \Omega^r+1_{\mathcal{Y}} \rightarrow W^n \Omega^r+1_{\mathcal{Y}}(\log Y) \xrightarrow{\theta_0} i_* W^n \Xi^r_Y \rightarrow 0,
\]

where \( \theta_0 \) is induced by the Poincaré residue mapping

\[
\theta_0 : W^n \Omega^r+1_{\mathcal{Y}}(\log Y) \rightarrow (i\sigma)_* W^n \Omega^r_U.
\]
Proof. We first recall the following fact due to Mokrane [Mo] Proposition 1.4.5. Under the setting of (2), let $F_a W_n^r \Omega_{Y,log}^m (\log Y) (a \geq 0)$ be the image of the product

$$W_n^r \Omega_{Y,log}^m (\log Y) \times W_n^r \Omega_{Y,log}^{m-a} \longrightarrow W_n^r \Omega_{Y,log}^m (\log Y).$$

Put $F_{-1} W_n^r \Omega_{Y,log}^m (\log Y) := 0$. Then there are isomorphisms

$$\text{gr}_a^F W_n^r \Omega_{Y,log}^m (\log Y) \simeq \begin{cases} W_n^r \Omega_{Y,log}^m & (a = 0) \\ \bigoplus_{|I|=a} W_n^r \Omega_{Y,log}^{m-a} & (a \geq 1) \end{cases} \quad (8.4.1)$$

which are induced by Poincaré residue mappings for $a \geq 1$ and depend on the fixed ordering of the irreducible components of $Y$. We prove (2). Because the complement $\Sigma := Y - U$ has codimension $\geq 2$ on $\mathfrak{M}$ and the sheaves $W_n^r \Omega_{Y,log}^{m+1}$ and $W_n^r \Omega_{Y,log}^{m+1} (\log Y)$ are finitely successive extensions of locally free $\mathcal{O}_U$-modules of finite rank ([III] I Corollaire 3.9), the sequence

$$0 \longrightarrow W_n^r \Omega_{Y,log}^{m+1} \longrightarrow W_n^r \Omega_{Y,log}^{m+1} (\log Y) \overset{d_0}{\longrightarrow} (i \sigma)_* W_n^r \Omega_{Y,log}^m$$

is exact by (8.4.1) for $\mathfrak{M} - \Sigma$. Hence it is enough to show that $\text{Im}(d_0) = i_* W_n^r \Xi_Y^r$. Let $F_a (\nu_{Y,r,n}^r) (a \geq 0)$ be the filtration on $\nu_{Y,r,n}^r$ in [Sa1] Proposition 2.2.1, which is defined under the assumption that $Y$ is simple and satisfies

$$F_0 (\nu_{Y,r,n}^r) = 0 \quad \text{and} \quad \text{gr}_a^F \nu_{Y,r,n}^r \simeq \bigoplus_{|I|=a} W_n^r \Omega_{Y,log}^{r-a+1} \quad \text{(non-canonically)}. \quad (8.4.2)$$

Hence comparing this local description of $\nu_{Y,r,n}^r$ with (8.4.1) for $m = r + 1$, we see that $\text{Im}(d_0) = i_* W_n^r \Xi_Y^r$. As for (1), we define the desired filtration $F_a (W_n^r \Xi_Y^r)$ as the $W_n^r \mathcal{O}_Y$-submodule generated by $F_a (\nu_{Y,r,n}^r)$. This filtration satisfies the required properties by (2) and again by (8.4.1).

Since the residue mapping $g_0$ in Proposition 8.4 (2) commutes with the operators $d, V, F, R$ (cf. (8.1.1)), these operators on $W_n^r \Omega_{Y,log}^m$ induce operators

$$\begin{align*}
  d &: W_n^r \Xi_Y^r \longrightarrow W_n^r \Xi_Y^{r+1}, \\
  V &: W_n^r \Xi_Y^r \longrightarrow W_{n+1}^r \Xi_Y^r, \\
  F &: W_{n+1}^r \Xi_Y^r \longrightarrow W_n^r \Xi_Y^r, \\
  R &: W_{n+1}^r \Xi_Y^r \longrightarrow W_n^r \Xi_Y^r
\end{align*}$$

by Proposition 8.4 (2) (without the assumptions in Proposition 8.4 (1) or (2)), which satisfy the relations listed in (8.1.2) and (8.1.3). We call the resulting complex $(W_n^r \Xi_Y^r, d)$ the homological de Rham-Witt complex of $Y$.

Remark 8.5 There are natural injective homomorphisms of complexes

$$\bigoplus_{r=0}^{\dim(Y)} \lambda_{Y,n}^r [-r] \hookrightarrow W_n^r A_Y^r, \quad \bigoplus_{r=0}^{\dim(Y)} \nu_{Y,n}^r [-r] \hookrightarrow W_n^r \Xi_Y^r,$$

where the differentials on the complexes on the left hand side are defined as zero. Indeed the differentials on $W_n^r \Omega_{Y,log}^m$ are zero on $W_n^r \Omega_{Y,log}^m$.

Documenta Mathematica 18 (2013) 177–247
The following Proposition 8.6 and Theorem 8.8 explain a fundamental relationship between $W_n^\bullet_A Y$ and $W_n^\bullet\Xi Y$.

**Proposition 8.6** The complex $W_n^\bullet\Xi Y$ is a $W_n^\bullet_A Y$-module, that is, there is a natural $W_n^\bullet_O Y$-bilinear pairing
\[ W_n^\bullet_A Y \times W_n^\bullet\Xi Y \to W_n^\bullet\Xi Y^{r+s} \quad (r, s \geq 0) \] (8.6.1)
satisfying the Leibniz rule
\[ d(x \cdot y) = (dx) \cdot y + (-1)^r x \cdot dy \quad (x \in W_n^\bullet A Y, \ y \in W_n^\bullet\Xi Y). \]
Moreover this pairing is compatible with $R$ and satisfies relations
\[ F(x \cdot y) = F(x) \cdot F(y) \]
\[ x \cdot V(y) = V(F(x) \cdot y) \]
\[ V(x) \cdot y = V(x \cdot F(y)) \]
\[ (x \in W_n^\bullet A Y, \ y \in W_n^\bullet\Xi Y). \]

**Proof.** The pairing (8.6.1) is induced by the product of $W_n^\bullet \Omega_U^\bullet$ and the biadditive pairing $\lambda^{r+s}_{Y,n} \to \nu^{r+s}_{Y,n}$ defined in [Sa1] Definition 3.1.1. The properties of the pairing (8.6.1) follow from the corresponding properties of the product of $W_n^\bullet \Omega_U^\bullet$ ([I1] I Théorème 2.17, Proposition 2.18).

**Remark 8.7** When $k$ is perfect and $Y$ is the special fiber of a regular semistable family $X$ over a discrete valuation ring with residue field $k$, we have
\[ W_n^\bullet A Y \subset W_n^\bullet \omega^\bullet Y \subset W_n^\bullet \Xi Y \]
by [Sa1] Proposition 4.2.1. Here $W_n^\bullet \omega^\bullet Y$ denotes the modified de Rham-Witt complex associated with $X$ [Hy2]. The complex $W_n^\bullet \Xi Y$ does not in general have a product structure unless $Y$ is smooth.

**Theorem 8.8** Assume that $k$ is perfect and that $Y$ is proper over $k$. Put $b := \dim(Y)$ and $W_n := W_n(k)$. Then:

1. There is a canonical trace map $\text{tr}_n : H^{2b}(Y, W_n^\bullet \Xi Y) \to W_n$, which is bijective if $Y$ is geometrically connected over $k$.

2. The pairing
\[ H^i(Y, W_n^\bullet A Y) \times H^{2b-i}(Y, W_n^\bullet \Xi Y) \to H^{2b}(Y, W_n^\bullet \Xi Y) \xrightarrow{\text{tr}_n} W_n \]
is a non-degenerate pairing of finitely generated $W_n$-modules for any $i \geq 0$.  

*Documenta Mathematica* 18 (2013) 177–247
Proof. (1) Let $f_n$ be the canonical morphism $W_n Y \to W_n$, where $W_n Y$ means the scheme consisting of the topological space $Y$ and the structure sheaf $W_n \mathcal{O}_Y$. Since $H^b(Y, W_n \Xi^\bullet_Y) \simeq H^b(Y, W_n \Xi^\bullet_Y)$, our task is to show that

$$W_n \Xi^\bullet_Y \xrightarrow{f_n} f_n^! W_n$$

in $D_{qc}(W_n Y, W_n \mathcal{O}_Y)$, (8.8.1)

where the subscript $qc$ means the triangulated subcategory consisting of complexes with quasi-coherent cohomology sheaves and $f_n^!$ means the twisted inverse image functor of Hartshorne [Ha]. If $Y$ is smooth, (8.8.1) is a theorem of Ekedahl [E]. In the general case, (8.8.1) is reduced to the smooth case by Proposition 8.4 (2) and the same arguments as for [Sa1] Proposition 2.5.9.

(2) The case $n = 1$ follows from [Sa1] Corollary 2.5.11, Proposition 3.3.5. The case $n \geq 2$ follows from a standard induction argument on $n$ and the following lemma: □

Lemma 8.9  (1) There are injective homomorphisms

$$p : W_{n-1} \Lambda_Y^\bullet \xhookrightarrow{} W_n \Lambda_Y^\bullet$$

and

$$\bar{p} : W_{n-1} \Xi_Y^\bullet \xhookrightarrow{} W_n \Xi_Y^\bullet$$

induced by the multiplication by $p$ on $W_n \Lambda_Y^\bullet$ and $W_n \Xi_Y^\bullet$, respectively.

(2) The following natural projections of complexes are quasi-isomorphisms:

$$W_n \Lambda_Y^\bullet / \bar{p}(W_{n-1} \Lambda_Y^\bullet) \longrightarrow \Lambda_Y^\bullet$$

and

$$W_n \Xi_Y^\bullet / \bar{p}(W_{n-1} \Xi_Y^\bullet) \longrightarrow \Xi_Y^\bullet.$$

Proof. When $Y$ is smooth, the assertions are due to Illusie [I1] Proposition 3.4 and Corollaire 3.15. In the general case, (1) follows from that for the dense open subset $U \subset Y$ we fixed before. To prove (2), we may assume that $Y$ is simple. Then the assertions follow from those for the strata of $Y$ and Propositions 8.2 (1) and 8.4 (1). □

In the rest of this section, we work under the following setting. Let $A$ be a discrete valuation ring with perfect residue field $k$, and let $X$ be a regular semistable family over $A$. Put $Y := X \otimes_A k$. Recall that

$$W_n \Lambda_Y^\bullet \subset W_n \omega_Y^\bullet \subset W_n \Xi_Y^\bullet$$

by Remark 8.7. We define a Frobenius endomorphism $\varphi$ on these complexes by $p^m F$ on degree $m \geq 0$. There is a short exact sequence of complexes with Frobenius action

$$0 \longrightarrow W_n \omega_Y^\bullet (-1)[-1] \xrightarrow{\varphi} W_n \omega_Y^\bullet \longrightarrow W_n \omega_Y^\bullet \longrightarrow 0$$

(8.9.1)

by [Hy2] (1.4.3), where the complex $W_n \omega_Y^\bullet (-1)$ means the complex $W_n \omega_Y^\bullet$ with Frobenius endomorphism $p \cdot \varphi$. The monodromy operator

$$N : W_n \omega_Y^\bullet \longrightarrow W_n \omega_Y^\bullet (-1)$$

is defined as the connecting morphism associated with this sequence.
Proposition 8.10 There is a short exact sequence of complexes with Frobenius action

\[
0 \rightarrow W_n A_Y^\bullet \rightarrow W_n \Omega_Y^r \rightarrow W_n \Xi_Y^r (-1)[-1] \rightarrow 0
\]  
(8.10.1)

fitting into a commutative diagram of complexes

\[
\begin{array}{cccc}
W_n \omega^r_Y (-1)[-1] \\
\downarrow \downarrow \downarrow \\
W_n A_Y^\bullet & \rightarrow & W_n \Omega_Y^r & \rightarrow & W_n \Xi_Y^r (-1)[-1] \\
\downarrow & & \downarrow & & \downarrow \\
W_n \omega^r_Y \\
\end{array}
\]

Consequently, we obtain a complex of \(W_n(k)\)-modules with Frobenius action for \(i \geq 0\)

\[
H^i(Y, W_n A_Y^\bullet) \rightarrow H^i(Y, W_n \omega_Y^r) \rightarrow H^i(Y, W_n \Xi_Y^r(-1)[-1]) \rightarrow H^i(Y, W_n \Omega_Y^r(-1). \)  
(8.10.2)

Proof. Let \(\sigma : U \rightarrow Y\) be as before and define a homomorphism \(\partial : W_n \tilde{\omega}_Y \rightarrow \sigma_* W_n \Omega_Y^r(-1)\) as the composite

\[
\partial : W_n \tilde{\omega}_Y \rightarrow \sigma_* \sigma^* W_n \tilde{\omega}_Y \rightarrow \sigma_* (W_n \Omega_Y^r \oplus W_n \Omega_Y^{r-1}) \cong \sigma_* W_n \Omega_Y^{r-1},
\]

where the second arrow is obtained from the fact that the sequence (8.9.1) splits on \(U\).

It is easy to see that \(\partial\) satisfies

\[
d\partial = \partial d \quad \text{and} \quad \partial F = pF\partial,
\]

which induces a map of complexes \(\partial^* : W_n \tilde{\omega}_Y \rightarrow \sigma_* W_n \Omega_Y^r(-1)[-1]\). To show that \(\partial^*\) induces the exact sequence (8.10.1), we may assume the assumption in Proposition 8.2 (2). Then we have an isomorphism of complexes

\[
W_n \omega^r_Y \cong W_n \Omega^r_Y (\log Y)/W_n \Omega^r_Y (-\log Y) \quad \text{([Hy2] p. 247 Lemma)}
\]

and the assertion follows from Propositions 8.2 (2) and 8.4 (2). The commutativity of the diagram in the proposition is straightforward and left to the reader. \(\square\)

We prove that the monodromy-weight conjecture for log crystalline cohomology implies an invariant cycle 'theorem' (cf. [I2] 2.4.5), which will be useful in §§9–10 below. Let \(A\) be a discrete valuation ring with finite residue field \(k\), and \(X\) be a regular scheme which is projective flat over \(A\) with strict semistable reduction. Put \(K_0 := \text{Frac}(W(k))\) and \(Y := X \otimes_A k\). For integer \(i \geq 0\), we define

\[
C^i := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{\longrightarrow n \geq 1} H^i(Y, W_n A_Y^\bullet),
\]

(8.10.3)

\[
D^i := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{\longrightarrow n \geq 1} H^i(Y, W_n \omega_Y^r),
\]

(8.10.4)

\[
E^i := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{\longrightarrow n \geq 1} H^i(Y, W_n \Xi_Y^r),
\]

(8.10.5)
which are finite-dimensional over $K_0$ and $F$-isocrystals over $\text{Spec}(k)$. The content of the monodromy-weight conjecture due to Mokrane is the following:

**Conjecture 8.11** ([Mo] Conjecture 3.27) The filtration on $D^i$ induced by the weight spectral sequence (loc. cit. 3.25, cf. [Na] (2.0.9; p)) agrees with the monodromy filtration induced by the monodromy operator $N : D^i \to D^i(-1)$ ([Mo] 3.26).

**Theorem 8.12** Let $i \geq 0$ be an integer, and assume Conjecture 8.11 for $D^i$. Then the following sequence of $F$-isocrystals over $\text{Spec}(k)$ induced by (8.10.2) is exact:

$$C^i \xrightarrow{N} D^i \xrightarrow{N} D^i(-1) \xrightarrow{\partial} E^i(-1).$$

**Proof.** Suppose that $\#(k) = p^a$. Then $\varphi^a$ is $K_0$-linear and we use the notion of weights concerning $\varphi^a$. Because the groups

$$\mathbb{H}^i(Y, W_n A_Y^\ast), \quad \mathbb{H}^i(Y, W_n \omega_Y^\ast), \quad \mathbb{H}^i(Y, W_n \Xi_Y^\ast), \quad \mathbb{H}^i(Y, W_n \varpi_Y^\ast)$$

are finite for any $n \geq 1$, there is a diagram of $F$-isocrystals with exact rows by (8.9.1) and Proposition 8.10

$$C^n \xrightarrow{\partial} D^i \xrightarrow{\partial} E^i(-1) \xrightarrow{\partial} C^{i+1} \xrightarrow{\partial} D^i(-1) \xrightarrow{\partial} E^i(-1),$$

where we put

$$\tilde{D}^i := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \geq 1} \mathbb{H}^i(Y, W_n \varpi_Y^\ast).$$

By the assumption on $D^i$, $\text{Ker}(N)$ has weights $\leq i$ and $\text{Coker}(N)$ has weights $\geq i + 2$. Hence for the reason of weights, it is enough to show the following lemma, where we do not need Conjecture 8.11:

**Lemma 8.13** For integers $i \geq 0$, $C^i$ has weights $\leq i$ and $E^i$ has weights $\geq i$.

**Proof of Lemma 8.13.** When $Y$ is smooth, the lemma is a consequence of the Katz-Messing theorem [KM]. In the general case, $C^i$ has weights $\leq i$ by the spectral sequence of $F$-isocrystals obtained from Proposition 8.2 (1)

$$E_1^{s,t} = \bigoplus_{|I| = s + 1} D_I^t \implies C^{s+t},$$

where we put

$$D_I^t := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \geq 1} \mathbb{H}^t(Y_I, W_n \Omega_{Y_I}^\ast).$$

Moreover $E^i$ has weights $\geq i$ by the duality result in Theorem 8.8. Thus we obtain Lemma 8.13 and Theorem 8.12. 

□
Remark 8.14 One can show Lemma 8.13 for $E^1$ alternatively by the Katz-Messing theorem and the spectral sequence obtained from Proposition 8.4(1)

$$E_1^{a,t} = \bigoplus_{|I|=1-s} D_{I,s}^{2a+1}(s) \Rightarrow E^{a,t}.$$ 

9 Comparison with the finite part of Bloch-Kato

In this section we relate the cohomology of $p$-adic étale Tate twists with the finite part of Bloch-Kato [BK2]. Let $p$ be a prime number, and let $K$ be a $p$-adic local field, i.e., a finite extension of $\mathbb{Q}_p$. Let $O_K$ be the integer ring of $K$, and let $k$ be the residue field of $O_K$. In this section we consider Tate twists only in the sense of $G_K$-modules (and not in the sense of $F$-crystals). For a topological $G_K$-module $M$ and an integer $r \in \mathbb{Z}$, we define

$$M(r) := \begin{cases} M \otimes_{\mathbb{Z}_p} \mathbb{Z}_p(1)^{\otimes r} & (r \geq 0), \\ \text{Hom}_{\mathbb{Z}_p}(\mathbb{Z}_p(1)^{\otimes (-r)}, M) & (r < 0), \end{cases}$$

where $\mathbb{Z}_p(1)$ denotes the topological $G_K$-module $\lim_{n \geq 1} \mu_{p^n}(\overline{K})$.

Let $X$ be a regular scheme which is projective flat over $O_K$ with strict semistable reduction, and put $X_K := X \otimes_{O_K} K$ and $Y := X \otimes_{O_K} k$. Let $j : X_K \hookrightarrow X$ be the natural open immersion, and let $i$ and $r$ be non-negative integers. Put $\mathcal{V}^i := H^i(X_K, \mathbb{Q}_p)$ (see also Notation) and

$$H^{i+1}(X, T_{\mathbb{Q}_p}(r)) := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \geq 1} H^{i+1}(X, T_n(r)),$$

which are finite-dimensional over $\mathbb{Q}_p$. Let $F^\bullet$ be the filtration on $H^{i+1}(X_K, \mathbb{Q}_p(r))$ resulting from the Hochschild-Serre spectral sequence

$$E_2^{a,b} = H^a(K, \mathcal{V}^b(r)) \Rightarrow H^{a+b}(X_K, \mathbb{Q}_p(r)). \quad (9.0.1)$$

Let $F^q H^{i+1}(X, T_{\mathbb{Q}_p}(r))$ be the inverse image of $F^q H^{i+1}(X_K, \mathbb{Q}_p(r))$ under the canonical map $j^* : H^{i+1}(X, T_{\mathbb{Q}_p}(r)) \to H^{i+1}(X_K, \mathbb{Q}_p(r))$. We have

$$\Phi^{i,r} := \text{gr}_{F^r} H^{i+1}(X, T_{\mathbb{Q}_p}(r)) \subset \text{gr}_{F^r} H^{i+1}(X_K, \mathbb{Q}_p(r)) = H^1(K, \mathcal{V}^i(r))$$

by definition. We relate $\Phi^{i,r}$ with the finite part of Bloch-Kato [BK2]:

$$H^1_f(K, \mathcal{V}^i(r)) := \text{Ker}(\epsilon : H^1(K, \mathcal{V}^i(r)) \to H^1(K, \mathcal{V}^i \otimes_{\mathbb{Q}_p} B_{\text{cryst}})).$$

where $B_{\text{cryst}}$ denotes the period ring of crystalline representations defined by Fontaine [Fo1], and $\epsilon$ is induced by the natural inclusion $\mathbb{Q}_p(r) \hookrightarrow B_{\text{cryst}}$. The main result of this section is as follows

Theorem 9.1 Assume $0 \leq r \leq p - 2$ and Conjecture 8.11 for $D^i := D^i(Y)$. Then we have

$$\Phi^{i,r} \subset H^1_f(K, \mathcal{V}^i(r)).$$

Further if $r \geq \dim(X_K) - p + 3$, then we have $\Phi^{i,r} = H^1_f(K, \mathcal{V}^i(r))$. 

Documenta Mathematica 18 (2013) 177–247
Remark 9.2 When $X$ is smooth over $\mathcal{O}_K$, Theorem 9.1 is a part of the $p$-adic point conjecture raised by Schneider [Sc] and proved by Langer-Saito [LS] and Nekovár [Ne1].

Put

$$H^{i+1}(X, \tau_{\leq r} R_j^* \mathbb{Q}_p(r)) := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{\leftarrow n \geq 1} H^{i+1}(X, \tau_{\leq r} R_j^* \mu_{p^n}^{(r)})$$

and let $F^q H^{i+1}(X, \tau_{\leq r} R_j^* \mu_{p^n}^{(r)})$ be the inverse image of $F^q H^{i+1}(X, \mathbb{Q}_p(r))$ under the map

$$j^* : H^{i+1}(X, \tau_{\leq r} R_j^* \mu_{p^n}^{(r)}) \to H^{i+1}(X, \mathbb{Q}_p(r)).$$

We first prepare the following key commutative diagram to prove Theorem 9.1, assuming $0 \leq r \leq p-2$:

$$\begin{array}{c}
\xymatrix{
H^1(K, V^i (r)) \ar[r]^-{i} \ar[d]^-{\delta} & F^1 H^{i+1}(X, \tau_{\leq r} R_j^* \mathbb{Q}_p(r)) \ar[d]^-{\alpha} \\
H^1(K, V^i \otimes_{\mathbb{Q}_p} B_{\text{crys}}) \ar[r]^-{\delta} & D^i / ND^i \ar[r]^-{\epsilon} & E^i,
}\end{array}$$

where $E^i := E^i(Y)$ is as in (8.10.5), and $D^i / ND^i$ denotes the cokernel of the monodromy operator $N : D^i \to D^i$. The top arrow is induced by $j^*$ and an edge homomorphism of the spectral sequence (9.0.1). The arrow $\epsilon$ is induced by the complex (8.10.2), and the arrow $\sigma'$ is induced by the composite map

$$H^{i+1}(X, \tau_{\leq r} R_j^* \mu_{p^n}^{(r)}) \xrightarrow{\sigma} H^{i+1}(Y, \nu_{Y,n}^{-1}) \to H^i(Y, W_n \Xi_Y^*) \quad (n \geq 1).$$

See Theorem 3.4 for $\sigma$ and see Remark 8.5 for the last map. The constructions of the arrows $\alpha$ and $\delta$ and the commutativity of the square (A) deeply rely on the $p$-adic Hodge theory. The commutativity of the triangle (B) is rather straight-forward once we define $\alpha$ (see Proposition 9.10 below). The proof of Theorem 9.1 will be given after Proposition 9.10.

Remark 9.3 In his paper [La], Langer considered the diagram (A) assuming that $\mathcal{O}_K$ is absolutely unramified (loc. cit. Proposition 2.9). We remove this assumption and the assumption (*) stated in loc. cit. p. 191, using a continuous version of crystalline cohomology (see Appendix A below). For this purpose, we include a detailed construction of $\alpha$ and a proof of the commutativity of (A).

Construction of $\delta$. We first construct the map $\delta$. There is an exact sequence of topological $G_K$-$\mathbb{Q}_p$-modules

$$0 \to \mathcal{V}^i \otimes_{\mathbb{Q}_p} B_{\text{crys}} \to \mathcal{V}^i \otimes_{\mathbb{Q}_p} B_{\text{cris}} \xrightarrow{1 \otimes N} \mathcal{V}^i \otimes_{\mathbb{Q}_p} B_{\text{cris}} \to 0,$$  

(9.3.1)
where \( N \) denotes the monodromy operator on \( B_{st} \). Taking continuous Galois cohomology groups, we obtain a long exact sequence

\[
0 \rightarrow \left( \mathcal{Y}^i \otimes \mathbb{Q}_p \right) \otimes_{B_{st}} B_{crys}^G \rightarrow \left( \mathcal{Y}^i \otimes \mathbb{Q}_p \right) \otimes_{B_{st}} B_{crys}^G \rightarrow \left( \mathcal{Y}^i \otimes \mathbb{Q}_p \right) \otimes_{B_{st}} B_{crys}^G \rightarrow H^1(G_K, \mathcal{Y}^i \otimes \mathbb{Q}_p) \rightarrow \cdots.
\]

By the \( C_{st} \)-conjecture proved by Hyodo, Kato and Tsuji ([HK], [K4], [Ts1], cf. [Ni2]), the last arrow of the top row is identified with the map \( N : D^i \rightarrow D^i \) (see (9.3.3) below for the construction of \( C_{st} \)-isomorphisms). We thus define the desired map \( \delta \) as the connecting map of this sequence.

**\( C_{st} \)-isomorphism.** Before constructing \( \alpha \), we recall the construction of \( C_{st} \)-isomorphisms, which will be useful later ([K4], [Ts1] §4.10). See [K3] for the general framework of log structures and log schemes. Put

\[
W := W(k) \quad \text{and} \quad K_0 := \text{Frac}(W) = W[p^{-1}].
\]

For a log scheme \((Z, M_Z)\) and an integer \( n \geq 1 \), we define

\[
(Z_n, M_{Z_n}) := (Z, M_Z) \times_{\text{Spec}(Z)} \text{Spec}(\mathbb{Z}/p^n),
\]

where we regarded \( \text{Spec}(Z) \) and \( \text{Spec}(\mathbb{Z}/p^n) \) as log schemes by endowing them with the trivial log structures and the fiber product is taken in the category of log schemes. Let \( M \) be the log structure on \( X \) associated with the normal crossing divisor \( Y \). Let \( \overline{O}_K \) be the integral closure of \( O_K \) in \( K \), and let \( \overline{M} \) be the log structure on \( \overline{X} := X \otimes_{\overline{O}_K} O_K \) define by base-change (see (A.0.2) below). Put \( \overline{Y} := Y \otimes_k \overline{K} \). For \( n \geq 1 \) and \( s \geq 0 \), let \( J_{n}^{-1}(s)(\overline{X}, \overline{M}) \in D^+(\overline{Y}_{\overline{k}}, \mathbb{Z}/p^n) \) be Tsuji’s version of syntomic complex of \((\overline{X}, \overline{M})\) (§A.5). Let \( j : X_{\overline{M}} \rightarrow \overline{X} \) be the natural open immersion. For \( m, s \geq 0 \), we define the syntomic cohomology of \((\overline{X}, \overline{M})\) as

\[
H^m_{\text{syn}}((\overline{X}, \overline{M}), J_{n}^{-1}(s)) := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \geq 1} H^m(\overline{Y}, J_{n}^{-1}(s)(\overline{X}, \overline{M})).
\]

The following isomorphism due to Tsuji ([Ts1] Theorem 3.3.4(2)) plays a crucial role:

\[
H^m_{\text{syn}}((\overline{X}, \overline{M}), J_{s}^{-1}(s)) \simeq H^m(\overline{X}, J_{s}^{-1}(s) \otimes_{\mathbb{Q}_p} \mathbb{Q}_p). \quad (9.3.2)
\]

The right hand side is isomorphic to \( Y^m(s) = H^m(X_{\overline{M}}, \mathbb{Q}_p(s)) \) if \( s \geq \dim(X_K) \) or \( s \geq m \). We now introduce the crystalline cohomology of \((\overline{X}, \overline{M})\) over \( W \):

\[
H^m_{\text{crys}}((\overline{X}, \overline{M})/W) := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \geq 1} H^m_{\text{crys}}((\overline{X}_n, \overline{M}_n)/W_n),
\]

where \( H^m_{\text{crys}}((\overline{X}_n, \overline{M}_n)/W_n) \) denotes the crystal cohomology of \((\overline{X}_n, \overline{M}_n)\) over \( W_n \) (see §A.1 below). We define a \( G_K \)-equivariant homomorphism

\[
g^m : Y^m \rightarrow D^m \otimes_{K_0} B_{st}
\]
as the composite map \((d := \dim(X_K))\)

\[
\mathcal{Y}^m \xrightarrow{(1)} H^m_{\text{syn}}(\bar{X}, \bar{M}), \mathcal{Y}_p(d) \otimes \mathbb{Q}_p(-d) \quad (9.3.3)
\]

\[
\xrightarrow{(2)} H^m_{\text{cris}}((\bar{X}, \bar{M})/W) \otimes \mathbb{Q}_p(-d)
\]

\[
\xrightarrow{(3)} (D^m \otimes_{K_0} B^+_{\text{st}})^{N=0} \otimes \mathbb{Q}_p(-d)
\]

\[
\xrightarrow{(4)} D^m \otimes_{K_0} B_{\text{st}},
\]

where (1) is obtained from the isomorphism (9.3.2) with \(s = d\), the arrow (2) is the \(p^d\)-times of the canonical map induced by (A.5.2) below, and (3) is the crystalline interpretation of \((D^m \otimes_{K_0} B^+_{\text{st}})^{N=0}\) due to Kato ([K4] §4, cf. Proposition A.3.1 (3) below). The map (4) is induced by the natural inclusion \(\mathbb{Q}_p(-d) \hookrightarrow B_{\text{st}}\). The \(B_{\text{st}}\)-linear extension of \(g^m\) is an isomorphism by [Ts2] Theorem 4.10.2:

\[
\mathcal{Y}^m \otimes_{\mathbb{Q}_p} B_{\text{st}} \xrightarrow{\sim} D^m \otimes_{K_0} B_{\text{st}}, \quad (9.3.4)
\]

which we call the \(C_{\text{st}}\)-isomorphism. Now we have a commutative diagram of \(G_K\)-modules for \(m, s \geq 0\)

\[
\begin{array}{ccc}
H^m_{\text{syn}}((\bar{X}, \bar{M}), \mathcal{Y}_p(s)) & \xrightarrow{(1)} & H^m_{\text{cris}}((\bar{X}, \bar{M})/W) \otimes \mathbb{Q}_p \xrightarrow{\sim} (D^m \otimes_{K_0} B^+_{\text{st}})^{N=0} \\
\xrightarrow{(9.3.2)} & & \xrightarrow{(9.3.4)} \\
\mathcal{Y}^m(s) & \xrightarrow{\mathbb{Q}_p(s) \otimes B_{\text{st}}} & \mathcal{Y}^m \otimes_{\mathbb{Q}_p} B_{\text{st}} & \xrightarrow{\sim} & D^m \otimes_{K_0} B_{\text{st}}
\end{array}
\]

(9.3.5)

where the arrow (1) is the \(p^j\)-times of the canonical map (A.5.2), and (2) is the crystalline interpretation of \((D^m \otimes_{K_0} B^+_{\text{st}})^{N=0}\) mentioned in (9.3.3). This diagram commutes by the construction of (9.3.4) and the compatibility facts in [Ts1] Corollaries 4.8.8 and 4.9.2.

**Remark 9.4.** We modified the \(C_{\text{st}}\)-isomorphisms in [Ts1] by constant multiplications to make them consistent with the canonical map from syntomic cohomology of Kato to (continuous) crystalline cohomology (see the map (1) of (9.5.1) and the equality (A.5.4) below). Under this modification, the element \(t_{\text{syn}}\) in [Ts1] p. 378 is replaced with \(p^{-1}t_{\text{syn}}\), and Corollary 4.8.8 of loc. cit. remains true.

**Construction of \(\alpha\).** We start the construction of \(\alpha\) assuming \(0 \leq r \leq p - 2\), which will be complete in Corollary 9.8 below. For \(n \geq 1\), let \(\mathcal{Y}_n(r)_{(X, M)} \in D^+(Y_{\text{et}}, \mathbb{Z}/p^n)\) be the syntomic complex of Kato (see §A.5 below). Since \(0 \leq r \leq p - 2\), there is an isomorphism

\[
(\tau_{\leq r} R_{\mathcal{I}_n})|_Y \simeq \mathcal{Y}_n(r)_{(X, M)} \quad \text{in} \quad D^+(Y_{\text{et}}, \mathbb{Z}/p^n) \quad (9.4.1)
\]

for each \(n \geq 1\) ([Ts2] Theorem 5.1), and \(H^{i+1}(X, \tau_{\leq r} R_{\mathcal{I}_n})\) is isomorphic to the syntomic cohomology

\[
H^{i+1}_{\text{syn}}((X, M), \mathcal{Y}_p(r)) := \mathbb{Q}_p \otimes \mathbb{Z}_p \lim_{n \geq 1} H^{i+1}_{\text{syn}}(Y, \mathcal{Y}_n(r)_{(X, M)}).
\]
by the proper base-change theorem. Therefore we consider $H^i_{\text{syn}}((X, M), \mathcal{I}_{Q_p}(r))$ instead of $H^i(X, \tau_{\leq r} R_j, \mathbb{Q}_p(r))$, which is the major reason we assume $r \leq p - 2$.

We first prove:

**Lemma 9.5** The kernel of the composite map

$$\xi : H^m_{\text{syn}}((X, M), \mathcal{I}_{Q_p}(r)) \to H^m_{\text{cris}}((X, M)/W)_{Q_p} \to H^m_{\text{cris}}((\overline{X}, \overline{M})/W)_{Q_p}$$

agrees with that of the composite map

$$\eta : H^m_{\text{syn}}((X, M), \mathcal{I}_{Q_p}(r)) \overset{(9.4.1)}{\approx} H^m(X, \tau_{\leq r} R_j, \mathbb{Q}_p(r)) \to \mathcal{V}^m(r) = H^m(X, K, \mathbb{Q}_p(r)).$$

**Proof.** The map $\eta$ factors through (9.3.2) with $s = r$, and the map $\xi$ factors through the map (1) in (9.3.5) with $s = r$ by (A.5.4) below. Hence the diagram

$$\begin{array}{ccc}
H^m_{\text{syn}}((X, M), \mathcal{I}_{Q_p}(r)) & \xrightarrow{\xi} & H^m_{\text{cris}}((\overline{X}, \overline{M})/W)_{Q_p} \\
\eta \downarrow & & \downarrow \\
\mathcal{V}^m(r) & \rightarrow & D^m \otimes_{Q_p} B_{\text{st}}
\end{array}$$

commutes by (9.3.5), which implies the assertion. \qed

Let $H^m_{\text{cont-cr}}((X, M)/W)$ be the continuous crystalline cohomology of $(X, M)$ over $W$ (see §A.2 below). Put

$$H^m_{\text{syn}}((X, M), \mathcal{I}_{Q_p}(r))^0 := \ker(\eta : H^m_{\text{syn}}((X, M), \mathcal{I}_{Q_p}(r)) \to \mathcal{V}^m(r)),$$

$$H^m_{\text{cont-cr}}((X, M)/W)_{Q_p} := Q_p \otimes_{Q_p} H^m_{\text{cont-cr}}((X, M)/W),$$

$$H^m_{\text{cont-cr}}((X, M)/W)_0^0 := \ker(H^m_{\text{cont-cr}}((X, M)/W)_{Q_p} \to H^m_{\text{cris}}((\overline{X}, \overline{M})/W)_{Q_p}).$$

There are canonical maps

$$H^i_{\text{syn}}((X, M), \mathcal{I}_{Q_p}(r))^0 \xrightarrow{(1)} H^i_{\text{cont-cr}}((X, M)/W)^0_{Q_p} \xrightarrow{(2)} H^i(K, H^i_{\text{cris}}((\overline{X}, \overline{M})/W)_{Q_p})$$

$$\xrightarrow{(3)} H^i(K, (D^f \otimes_{K_0} B_{\text{st}}^+)_{N=0}) \xrightarrow{(4)} H^i(K, \mathcal{V}^i \otimes_{Q_p} B_{\text{cris}}),$$

where (1) is the canonical map obtained from Lemma 9.5 and Proposition A.6.1 (2) below, the arrow (3) is the isomorphism (2) in (9.3.5) for $m = i$, and (4) is obtained from the $C_{\text{st}}$-isomorphism (9.3.4) and (9.3.1). The most crucial map (2) is obtained from an edge homomorphism of the spectral sequence in Theorem A.4.2 (see also Corollary A.4.5 below). To proceed with the construction of $\alpha$, we need to introduce some notation.
Definition 9.6  Put $S := \text{Spec}(O_K)$, and fix a prime $\pi$ of $O_K$. Let $M_S$ be the log structure on $S$ associated with the closed point \{ $\pi = 0$ \}. Put $V := \text{Spec}(W[t])$, and let $M_V$ be the log structure on $V$ associated with the divisor \{ $t = 0$ \}. Let $t : (S, M_S) \rightarrow (V, M_V)$ be the exact closed immersion over $W$ sending $t \mapsto \pi$. For each $n \geq 1$, let $(\delta_n, M_{\delta_n})$ be the PD-envelope of the exact closed immersion $\iota_n : (V_n, M_{V_n}) \hookrightarrow (S_n, M_{S_n})$ compatible with the canonical PD-structure on $pW_n$. We define

$$H^m_{\text{crys}}((X, M)/(\delta', M_{\delta'}))_{Q_p} := \lim_{n \rightarrow \infty} H^m_{\text{crys}}((X_n, M_n)/(\delta_n, M_{\delta_n})),$$

which is an $R_{\delta'} := \lim_{n \rightarrow \infty} \Gamma(\delta_n, O_{\delta_n})$-module endowed with a $\varphi_{\delta'}$-semilinear endomorphism $\varphi$ and a $K_0$-linear endomorphism $N$ satisfying $N\varphi = p\varphi N$. Here $\varphi_{\delta'}$ is a certain canonical Frobenius operator on $R_{\delta'}$. See [Ts1] p. 253 for $\varphi_{\delta'}$, loc. cit. §4.3 for $N$ and $\varphi$, and see loc. cit. Remark 4.3.9 for $N$.

The following lemma plays a key role in our construction of $\alpha$, where we do not need to assume $0 \leq r \leq p - 2$ (compare with the assumption (+) in [La] p. 191).

Lemma 9.7  (1) There exists a long exact sequence

$$\cdots \rightarrow H^i_{\text{cont-cr}}((X, M)/W)_{Q_p} \rightarrow H^i_{\text{crys}}((X, M)/(\delta', M_{\delta'}))_{Q_p} \rightarrow H^i_{\text{cont-cr}}((X, M)/W)_{Q_p} \rightarrow \cdots$$

$$N \rightarrow H^i_{\text{crys}}((X, M)/(\delta', M_{\delta'}))_{Q_p} \rightarrow H^{i+1}_{\text{cont-cr}}((X, M)/W)_{Q_p} \rightarrow \cdots$$

and the image of $\partial$ agrees with $H^{i+1}_{\text{cont-cr}}((X, M)/W)_{Q_p}$.

(2) There is a commutative diagram whose lower row is exact

$$\begin{array}{ccc}
D^i & \longrightarrow & H^1_{\text{cont-cr}}((X, M)/W)_{Q_p}^0 \\
N & \downarrow & \delta \\
D^i & \longrightarrow & H^1(K, \mathcal{Y} \otimes_{Q_p} B_{\text{crys}}) \\
\end{array}$$

Here the right vertical arrow is the composite of (2)–(4) in (9.5.1), and the left vertical arrow is the specialization maps with $t \mapsto 0$ (cf. [Ts1] (4.4.4)).

Proof. (1) follows from Proposition A.2.4 and Corollaries A.3.2 (1) and A.3.4 below. As for (2), the exactness of the lower row has been explained in the construction of $\delta$. See Corollary A.4.5 below for the commutativity of the square.

Corollary 9.8  The composite map (9.5.1) factors as follows:

$$H^{i+1}_{\text{syn}}((X, M), \mathcal{Y}_{Q_p}(r))^0 \longrightarrow D^i/N D^i \delta \longrightarrow H^1(K, \mathcal{Y} \otimes_{Q_p} B_{\text{crys}}).$$

We define $\alpha$ as the first arrow of this decomposition.
Commutativity of the diagram (9.2.1). We first prove that the square (A) in (9.2.1) is commutative. By the construction of $\alpha$ and $\delta$, it is enough to show

**Theorem 9.9** The following square is commutative:

$$
\begin{array}{c}
H^{i+1}_\text{syn}(X, M), \mathcal{S}_p(r))_0 \xrightarrow{\varphi} F^1 H^{i+1}(X, \tau_{\leq r}, R_j \mathbb{Q}_p(r)) \\
\downarrow (9.5.1) \\
H^1(K, \mathcal{V}^i \otimes \mathbb{Q}_p B_{\text{crys}}) \xrightarrow{B_{\text{crys}} \leftarrow \mathbb{Q}_p(r)} H^1(K, \mathcal{V}^i(r)),
\end{array}
$$

where the right vertical arrow is the top arrow in (9.2.1).

**Proof.** Put $d := \dim(X_K)$. We define a $G_K$-homomorphism

$$
\beta^{i,r} : \mathcal{V}^i(r) \longrightarrow \begin{cases} H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}(r - d) & \text{(if } r < d) \\ H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p} & \text{(if } r \geq d) \end{cases}
$$

as the following composite map for $r \geq d$:

$$
\mathcal{V}^i(r) \xrightarrow{9.3.2} H^{i+1}_\text{syn}(X, M), \mathcal{S}_p(r))_0 \xrightarrow{9.5.1(1)} H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p},
$$

and as the following map for $r < d$:

$$
\mathcal{V}^i(r) \simeq \mathcal{V}^i(d) \otimes_{\mathbb{Q}_p} \mathbb{Q}_p(r - d) \xrightarrow{\beta^{i,d} \otimes \text{id}} H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}(r - d).
$$

We next define a homomorphism

$$
\gamma^{i,r} : H^{i+1}_\text{syn}(X, M), \mathcal{S}_p(r))_0 \longrightarrow \begin{cases} H^1(K, H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}(r - d)) & \text{(if } r < d) \\ H^1(K, H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}) & \text{(if } r \geq d) \end{cases}
$$

as the following composite map, if $r < d$:

$$
H^{i+1}_\text{syn}(X, M), \mathcal{S}_p(r))_0 \xrightarrow{(i)} H^1(K, H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}) \simeq H^1(K, H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}(r - d) \otimes_{\mathbb{Q}_p} \mathbb{Q}_p(d - r)) \xrightarrow{(ii)} H^1(K, H^i_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}(r - d)),
$$

where (i) is the composite of (1) and (2) in (9.5.1), and (ii) is induced by the inclusion $\mathbb{Q}_p(d - r) \hookrightarrow H^0_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}$ (cf. (A.6.5) below) and the cup product of crystalline cohomology. If $r \geq d$, then we define $\gamma^{i,r}$ as the composite of (1) and (2) in (9.5.1).
We prove the commutativity of (9.9.1) assuming \( r < d \). In this case, the following triangle commutes by Theorem A.6.7 below:

\[
H_{\text{syn}}^{i+1}((X, M), \mathscr{F}_p(r))^0 \xrightarrow{\gamma^{i, r}} H^1(K, \mathscr{F}^i(r)) \xrightarrow{\beta^{i, r}} H^1(K, H^1_{\text{crys}}(\overline{(X, M)}/W)_{Q_p}(r - d)),
\]  

(9.9.2)

where the left vertical arrow is the composite of the top horizontal and the right vertical arrows of (9.9.1). Moreover the following square commutes by the construction of the \( C_{\text{st}} \)-isomorphism (9.3.4) (compare with (9.3.5)):

\[
\begin{array}{ccc}
\gamma^{i, r} & \rightarrow & \beta^{i, r} \\
\downarrow & & \downarrow \\
H^1(K, \mathscr{F}^i(r)) & \rightarrow & H^1(K, \mathscr{F}^i(r)) \otimes_{Q_p} B_{\text{crys}}(Q_p(r) - d) \\
\end{array}
\]  

(9.9.3)

The commutativity of (9.9.1) follows from these commutative diagrams and the fact that the composite map (9.5.1) agrees with the composite of \( \gamma^{i, r} \) and the map

\[
H^1(K, H^1_{\text{crys}}(\overline{(X, M)}/W)_{Q_p}(r - d)) \rightarrow H^1(K, \mathscr{F}^i(r) \otimes_{Q_p} B_{\text{crys}}(Q_p(r)) -d)
\]

induced by the right vertical arrow and the bottom isomorphism in (9.9.3).

When \( r \geq d \), the following triangle commutes by Corollary A.6.3 below:

\[
H_{\text{syn}}^{i+1}((X, M), \mathscr{F}_p(r))^0 \xrightarrow{\gamma^{i, r}} H^1(K, \mathscr{F}^i(r)) \xrightarrow{\beta^{i, r}} H^1(K, H^1_{\text{crys}}((X, M)/W)_{Q_p}),
\]

where the left vertical arrow is defined in the same way as for that of (9.9.2). The commutativity of (9.9.1) follows from this diagram and the diagram (9.3.5). This completes the proof of Theorem 9.9.

Proposal 9.10 The triangle (B) in (9.2.1) is commutative.

Proof. There is a commutative diagram by the definition of \( \partial \) in Proposition A.2.4

\[
\begin{array}{ccc}
H^i_{\text{crys}}((X, M)/ (\mathscr{E}, M_{\mathscr{E}}))_{Q_p} & \xrightarrow{\partial} & D^i \\
\downarrow & & \downarrow \delta \wedge \\
H^i_{\text{cont}}((X, M)/ W)_{Q_p} & \xrightarrow{\partial} & D^{i+1}
\end{array}
\]

\[\text{Documenta Mathematica 18 (2013) 177–247}\]
where $\tilde{D}^{i+1} = \tilde{D}^{i+1}(Y)$ is as in the proof of Theorem 8.12, the left (resp. right) vertical arrow is as in Lemma 9.7 (1) (resp. given by the product with $d\log(t)$, cf. (8.9.1)), and the horizontal arrows are the specialization maps by $t \mapsto 0$. Therefore in view of the constructions of $\alpha$ and $\sigma'$, it is enough to check that the following diagram of canonical morphisms commutes in $D^+(Y_{\text{ét}}, \mathbb{Z}/p^n)$ for each $n \geq 0$:

\[
\begin{array}{c}
\tau \downarrow \\
\mathcal{F}_n(r)(X,M) \rightarrow \mathbb{Q}_p^{\geq r} \rightarrow W_n \mathbb{Q}_p^{r} \\
\sigma \downarrow \\
\mathbb{Q}_{Y/n}^{r-1}[-r] \rightarrow W_n \mathbb{Q}_p^{\geq r-1}[-1] \rightarrow W_n \mathbb{Q}_p[-1],
\end{array}
\]

where $\mathbb{Q}_{Y/n}^{r-1}$ is as in §A.5 below, and $s$ denotes the specialization map by $t \mapsto 0$. The right square commutes obviously in the category of complexes of sheaves.

As for the left square, we have only to show the commutativity on the diagram of the $r$-th cohomology sheaves because $W_n \mathbb{Q}_p^{\geq r-1}[-1]$ is concentrated in degrees $\geq r$ and $\mathcal{F}_n(r)(X,M)$ is concentrated in degrees $\leq r$ by (9.4.1). Since we have

\[
\mathcal{H}^r(\mathcal{F}_n(r)(X,M)) \simeq (R^r j_* \mathbb{Q}_p^{\geq r})||(Y)
\]

and the sheaf on the right hand side is generated by symbols (cf. the case $D = \emptyset$ of Theorem 3.3 (1)), one can check this commutativity of the diagram of sheaves by a straightforward computation on symbols using the compatibility in [Ts1] Proposition 3.2.4 (2).

**Proof of Theorem 9.1.** We have the following commutative diagram by (9.2.1):

\[
\begin{array}{c}
F^1 H^{i+1}(X, T) \rightarrow F^1 H^{i+1}(X, T \otimes \mathbb{Q}_p^r) \rightarrow E^1 \\
\alpha \downarrow \uparrow \downarrow \alpha' \\
H^1(K, \mathcal{F}^i(r)) \rightarrow D^1/N D^1 \\
\iota \downarrow \downarrow \downarrow \delta \\
H^1(K, \mathcal{F}^i \otimes \mathbb{Q}_p \mathcal{B}_{\text{crys}}) \\
\end{array}
\]

where the left triangle commutes obviously. The kernel of $\iota$ is $H^1_f(K, \mathcal{F}^i(r))$ by definition. The top row is a complex by the definitions of $\sigma'$ and $\Sigma_n(r)$. The map $\epsilon$ is injective by Theorem 8.12 and Conjecture 8.11 for $D^1$. We obtain the first assertion

\[
\Phi_i \subset H^1_f(K, \mathcal{F}^i(r))
\]

by a simple diagram chase on this diagram. To obtain the second assertion, we repeat the same arguments for $\Phi^{2d-i,d-r+1}$, where we put $d := \dim(X_K)$. The map $\epsilon$ is injective by Theorem 8.12 and Conjecture 8.11 for $D^1$. We obtain the first assertion

\[
\Phi_i \subset H^1_f(K, \mathcal{F}^i(r))
\]

by a simple diagram chase on this diagram. To obtain the second assertion, we repeat the same arguments for $\Phi^{2d-i,d-r+1}$, where we put $d := \dim(X_K)$. The map $\epsilon$ is injective by Theorem 8.12 and Conjecture 8.11 for $D^1$. We obtain the first assertion

\[
\Phi_i \subset H^1_f(K, \mathcal{F}^i(r))
\]

by a simple diagram chase on this diagram. To obtain the second assertion, we repeat the same arguments for $\Phi^{2d-i,d-r+1}$, where we put $d := \dim(X_K)$. The map $\epsilon$ is injective by Theorem 8.12 and Conjecture 8.11 for $D^1$. We obtain the first assertion

\[
\Phi_i \subset H^1_f(K, \mathcal{F}^i(r))
\]

by a simple diagram chase on this diagram. To obtain the second assertion, we repeat the same arguments for $\Phi^{2d-i,d-r+1}$, where we put $d := \dim(X_K)$. The map $\epsilon$ is injective by Theorem 8.12 and Conjecture 8.11 for $D^1$. We obtain the first assertion

\[
\Phi_i \subset H^1_f(K, \mathcal{F}^i(r))
\]

by a simple diagram chase on this diagram. To obtain the second assertion, we repeat the same arguments for $\Phi^{2d-i,d-r+1}$, where we put $d := \dim(X_K)$. The map $\epsilon$ is injective by Theorem 8.12 and Conjecture 8.11 for $D^1$. We obtain the first assertion

\[
\Phi_i \subset H^1_f(K, \mathcal{F}^i(r))
\]

by a simple diagram chase on this diagram. To obtain the second assertion, we repeat the same arguments for $\Phi^{2d-i,d-r+1}$, where we put $d := \dim(X_K)$. The map $\epsilon$ is injective by Theorem 8.12 and Conjecture 8.11 for $D^1$. We obtain the first assertion

\[
\Phi_i \subset H^1_f(K, \mathcal{F}^i(r))
\]
is injective by Conjecture 8.11 for $D^i$, Theorems 8.12 and 8.8 and the Poincaré duality between $D^i$ and $D^{2d-i}$ [Hy2]. Hence we have

$$\Phi^{2d-i,d-r+1} \subset H^1_f(K, \mathcal{R}^{2d-i}(d-r+1)).$$  \hfill (9.10.2)

The results (9.10.1) and (9.10.2) imply $\Phi^{i,r} = H^1_f(K, \mathcal{R}^i(r))$, because $\Phi^{i,r}$ and $\Phi^{2d-i,d-r+1}$ (resp. $H^1_f(K, \mathcal{R}^i(r))$ and $H^1_f(K, \mathcal{R}^{2d-i}(d-r+1))$) are the exact annihilators of each other under the non-degenerate pairing of the Tate duality

$$H^1_f(K, \mathcal{R}^i(r)) \times H^1_f(K, \mathcal{R}^{2d-i}(d-r+1)) \rightarrow \mathbb{Q}_p$$

by [Sa1] Corollary 10.6.1 (resp. [BK2] Proposition 3.8). \hfill \Box

10 \hskip 50pt \textbf{IMAGE OF P-ADIC REGULATORS}

In this section, we prove Theorem 1.1. We first review the definitions of the finite part of Galois cohomology [BK2] and the integral part of algebraic $K$-groups [Sch] over number fields.

Let $K$ be a number field and let $\mathcal{R}$ be a topological $G_K$-$\mathbb{Q}_p$-module which is finite-dimensional over $\mathbb{Q}_p$. Let $\mathfrak{o}$ be the integer ring of $K$. For a place $v$ of $K$, let $K_v$ (resp. $\mathfrak{o}_v$) be the completion of $K$ (resp. $\mathfrak{o}$) at $v$. The finite part $H^1_f(K, \mathcal{R}) \subset H^1(K, \mathcal{R})$ is defined as

$$H^1_f(K, \mathcal{R}) := \ker \left( H^1(K, \mathcal{R}) \rightarrow \prod_{v \text{ finite}} H^1(K_v, \mathcal{R}) \right),$$

where $v$ runs through all finite places of $K$, and $H^1_f(K_v, \mathcal{R})$ is defined as follows:

$$H^1_f(K_v, \mathcal{R}) := \begin{cases} \ker(H^1(K_v, \mathcal{R}) \rightarrow H^1(I_v, \mathcal{R})) & \text{if } \mathfrak{o}_v \not\subset \mathfrak{p}, \\ \ker(H^1(K_v, \mathcal{R}) \rightarrow H^1(K_v, \mathcal{R} \otimes \mathbb{Q}_p \text{Br}_\mathfrak{o}_v)) & \text{if } \mathfrak{o}_v \subset \mathfrak{p}, \end{cases}$$

where $I_v$ denotes the inertia subgroup of $G_K$. We next review the definition of the integral part of algebraic $K$-groups. Let $V$ be a proper smooth variety over the number field $K$. First fix a finite place $v$ of $K$. By de Jong’s alteration theorem [dJ], there is a proper generically finite morphism

$$\pi : V' \rightarrow V_v := V \otimes_K K_v$$

such that $V'$ has a projective regular model $X'$ with strict semistable reduction over the integer ring of some finite extension of $K_v$. The integral part $K_m(V_v)_{\mathfrak{o}_v} \subset K_m(V_v) \otimes \mathbb{Q}$ is the kernel of the composite map

$$K_m(V_v) \otimes \mathbb{Q} \xrightarrow{\pi^*} K_m(V') \otimes \mathbb{Q} \rightarrow \text{Image of } K_m(X') \otimes \mathbb{Q},$$

which is in fact independent of $X'$ ([Sch] §1). The integral part $K_m(V)_{\mathfrak{o}}$ is defined as

$$K_m(V)_{\mathfrak{o}} := \ker \left( K_m(V) \otimes \mathbb{Q} \rightarrow \prod_{v \text{ finite}} K_m(V_v) \otimes \mathbb{Q} \right),$$

$\text{Image of } K_m(X') \otimes \mathbb{Q}$.
where $v$ runs through all finite places of $K$. If $V$ admits a regular model which is proper flat over the integer ring of $K$, then $K_m(V)_\omega$ agrees with the image of the $K$-group of the model, i.e., the integral part considered by Beilinson. By these definitions of $H^i_\omega(K, V')$ and $K_{2r-i-1}(V)_\omega$, Theorem 1.1 is immediately reduced to Theorem 10.1 below, which is an analogue of Theorem 1.1 over local fields.

Let $\ell$ and $p$ be prime numbers. We change the setting slightly, and let $K$ be an $\ell$-adic local field, i.e., a finite extension of $\mathbb{Q}_\ell$. Let $O_K$ be the integer ring of $K$, and let $V$ be a proper smooth variety over $K$. Let $i$ and $r$ be integers with $2r \geq i + 1 \geq 1$.

There is a $p$-adic regulator map obtained from étale Chern character

$$\text{reg}_p : K_{2r-i-1}(V)_{O_K, \text{hom}} \to H^1(K, H^i(\mathbb{V}, \mathbb{Q}_p(r))),$$

where $\mathbb{V}$ denotes $V \otimes_K \mathbb{F}_p$, and $K_{2r-i-1}(V)_{O_K, \text{hom}}$ denotes the subspace of $K_{2r-i-1}(V)_{O_K}$ consisting of all elements which vanish in $H^{i+1}(\mathbb{V}, \mathbb{Q}_p(r))$ under the Chern character.

**Theorem 10.1** If $\ell = p$, assume that $r \leq p - 2$ and that Conjecture 8.11 holds for projective strict semistable varieties over $\mathbb{F}_p$ in degree $i$. Then $\text{Im}(\text{reg}_p)$ is contained in $H^1_\omega(K, H^i(\mathbb{V}, \mathbb{Q}_p(r)))$.

**Remark 10.2** When $\ell = p$, we need Conjecture 8.11 for the reduction of an alteration of $V$. When $\ell \neq p$, we do not need the monodromy-weight conjecture, but use Deligne’s proof of the Weil conjecture [D] to show that $\text{reg}_p$ is zero.

**Proof of Theorem 10.1.** We first reduce the problem to the case that $V$ has a regular model which is projective flat over $O_K$ with strict semistable reduction. By de Jong’s alteration theorem [dJ], there exists a proper generically finite morphism $\pi : V' \to V$ such that $V'$ has a projective regular model with strict semistable reduction over the integer ring $O_L$ of $L := \Gamma(V', \mathcal{O}_{V'})$. Then there is a commutative diagram

$$\begin{array}{ccc}
K_{2r-i-1}(V)_{O_K, \text{hom}} & \xrightarrow{\pi^*} & H^1(K, H^i(\mathbb{V}, \mathbb{Q}_p(r))) \\
\xrightarrow{\text{reg}_p} & & \xrightarrow{\pi^*} \\
K_{2r-i-1}(V')_{O_L, \text{hom}} & \xrightarrow{\text{reg}_p} & H^1(L, H^i(\mathbb{V}', \mathbb{Q}_p(r))) \\
\end{array}$$

where $\mathbb{V}'$ denotes $V' \otimes_L \mathbb{F}_p$, and the right (and the middle) vertical arrows are split injective by a standard argument using a corestriction map of Galois cohomology and a trace map of étale cohomology. By this diagram and the definition of $K_\omega(V)_{O_K}$, Theorem 10.1 for $V$ is reduced to that for $V'$. Thus we may assume that $V$ has a projective regular model $X$ with strict semistable reduction over $O_K$. Then the case $\ell \neq p$ follows from [Ne1] II Theorem 2.2 (cf. [D]). We prove the case $\ell = p$. Let $Y$ be the reduction of $X$. Put

$$H^{i+1}(X, \Omega_{\mathbb{Q}_p}(r)) := \mathbb{Q}_p \otimes_{\mathbb{Z}_p} \lim_{n \to \infty} H^{i+1}(X, \Omega_n(r)),$$
and let $K_{2r−i−1}(X)_{\text{hom}}$ (resp. $H^{i+1}(X, \mathbb{Q}_p(r))^0$) be the kernel of the composite map

$$K_{2r−i−1}(X) \to K_{2r−i−1}(V) \xrightarrow{ch} H^{i+1}(V, \mathbb{Q}_p(r)) \to H^{i+1}(\overline{V}, \mathbb{Q}_p(r))$$

(resp. $H^{i+1}(X, \mathbb{Q}_p(r)) \to H^{i+1}(V, \mathbb{Q}_p(r)) \to H^{i+1}(\overline{V}, \mathbb{Q}_p(r))$).

There is a commutative diagram by Corollary 5.7

$$
\begin{array}{ccc}
K_{2r−i−1}(X)_{\text{hom}} & \rightarrow & K_{2r−i−1}(V)_{\text{O_K, hom}} \\
\downarrow & & \downarrow ch \\
H^{i+1}(X, \mathbb{Q}_p(r))^0 & \rightarrow & H^1(K, H^i(\overline{V}, \mathbb{Q}_p(r)))
\end{array}
$$

The image of the bottom arrow is contained in $H^1(K, H^i(\overline{V}, \mathbb{Q}_p(r)))$ by Theorem 9.1 and Conjecture 8.11 for $D^i = D^i(Y)$, which implies Theorem 10.1. This completes the proof of Theorems 10.1 and 1.1. □

A Continuous crystalline cohomology

In this appendix, we formulate continuous versions of crystalline and syntomic cohomology of log schemes, combining the methods of Jannsen, Kato and Tsuji ([J], [K2], [K3], [Ts1], [Ts2]). The results of this appendix have been used in §9 of this paper. See [K3] for the general framework of log structures and log schemes.

Let $p$ be a prime number, and let $K$ be a complete discrete valuation field of characteristic 0 whose residue field $k$ is a perfect field of characteristic $p$. Let $O_K$ be the integer ring of $K$. Put $W := W(k)$, $W_n := W_n(k) (n \geq 1)$ and $K_0 := \text{Frac}(W)$. Let $X$ be a regular scheme which is projective flat over $O_K$ with semistable reduction, and put

$$
X_K := X \otimes_{O_K} K, \quad Y := X \otimes_{O_K} k \\
X_{\overline{K}} := X_K \otimes_K \overline{K}, \quad \overline{Y} := Y \otimes_k \overline{K} \quad \text{and} \quad \overline{X} := X \otimes_{O_K} \overline{O_K},
$$

where $\overline{O_K}$ denotes the integral closure of $O_K$ in $\overline{K}$. Let $M$ be the log structure on $X$ associated with the normal crossing divisor $Y$. We endow $\overline{X}$ with a log structure $\overline{M}$ as follows. For a finite field extension $L/K$, put $S_L := \text{Spec}(O_L)$, and let $M_{S_L}$ be the log structure on $S_L$ associated with its closed point. We denote $(S_K, M_{S_K})$ simply by $(S, M_S)$, and define $(X_{O_L}, M_{O_L})$ by base-change in the category of log schemes

$$
(X_{O_L}, M_{O_L}) := (X, M) \times_{(S, M_S)} (S_L, M_{S_L}). \quad (A.0.1)
$$

We then define the log structure $\overline{M}$ on $\overline{X}$ as that associated with the pre-log structure

$$
\lim_{K \subset L \subset \overline{K}} M_{O_L | \overline{X}}, \quad (A.0.2)
$$
where \( L \) runs through all finite field extensions of \( K \) contained in \( \overline{K} \), \( M_{O_L} \mid \overline{K} \) denotes the topological inverse image of \( M_{O_L} \) onto \( \overline{K} \), and the inductive limit is taken in the category of étale sheaves of monoids on \( \overline{X} \).

For a log scheme \((Z, M_Z)\) and an integer \( n \geq 1\), we define

\[
(Z_n, M_{Z_n}) := (Z, M_Z) \times_{\text{Spec}(\mathbb{Z})} \text{Spec}(\mathbb{Z}/p^n),
\]

where we regarded \( \text{Spec}(\mathbb{Z}) \) and \( \text{Spec}(\mathbb{Z}/p^n) \) as log schemes by endowing them with the trivial log structures and the fiber product is taken in the category of log schemes.

We denote by \( \text{Mod}(\mathbb{Z}) \) and \( \text{Spec}(\mathbb{Z}/p^n) \) as log schemes by endowing them with the trivial log structures and the fiber product is taken in the category of log schemes.

For a scheme \( T \), \( \text{Shv}(T, \mathbb{Z}/p^*) \) denotes the category of the projective systems \( \{F_n\}_{n \geq 1} \) of \( \mathbb{Z}/p^n \)-sheaves on \( T \) such that \( F_n \) is a \( \mathbb{Z}/p^n \)-module for each \( n \geq 1 \). For a pro-finite group \( G \), \( \text{Mod}(G, \mathbb{Z}/p^*) \) denotes the category of the projective systems \( \{F_n\}_{n \geq 1} \) of discrete \( G \)-modules such that \( F_n \) is a \( \mathbb{Z}/p^n \)-module.

For a scheme \( T \), \( \text{Shv}(T, \mathbb{Z}/p^*) \) denotes the category of the projective systems \( \{F_n\}_{n \geq 1} \) of étale sheaves on \( T \) such that \( F_n \) is a \( \mathbb{Z}/p^n \)-sheaf for each \( n \geq 1 \). For a pro-finite group \( G \) acting on \( T \), \( \text{Shv}(T, G, \mathbb{Z}/p^*) \) denotes the category of the projective systems \( \{F_n\}_{n \geq 1} \) of étale \( G \)-sheaves on \( T \) such that \( F_n \) is a \( \mathbb{Z}/p^n \)-sheaf for each \( n \geq 1 \). We write \( D(T, \mathbb{Z}/p^*) \) (resp. \( D(T, G, \mathbb{Z}/p^*) \)) for the derived category of \( \text{Shv}(T, \mathbb{Z}/p^*) \) (resp. \( \text{Shv}(T, G, \mathbb{Z}/p^*) \)).

For an additive category \( \mathcal{E} \), we write \( \mathbb{Q} \otimes \mathcal{E} \) for the \( \mathbb{Q} \)-tensor category of \( \mathcal{E} \), i.e., the category whose objects are the same as \( \mathcal{E} \) and such that for objects \( A, B \in \mathcal{E} \), the group of morphisms \( \text{Hom}_{\mathbb{Q} \otimes \mathcal{E}}(A, B) \) is given by \( \mathbb{Q} \otimes \text{Hom}_{\mathcal{E}}(A, B) \). We often write \( \mathbb{Q} \otimes A \) for \( A \in \mathcal{E} \) regarded as an object of \( \mathbb{Q} \otimes \mathcal{E} \) to avoid confusions.

### A.1 Crystalline complexes

We construct the following objects:

\[
\begin{align*}
\mathcal{E}(X^*_\alpha, M_\alpha)/W_n \otimes_{\mathcal{E}(X^*_\alpha, M_\alpha)/(\mathcal{E}_\alpha, M_{\mathcal{E}_\alpha})} & \in D^+(Y_{\text{et}}, \mathbb{Z}/p^*) \\
\mathcal{E}(X^*_\alpha, \mathcal{M}_\alpha)/W_n \otimes_{\mathcal{E}(X^*_\alpha, \mathcal{M}_\alpha)/(\mathcal{E}_\alpha, M_{\mathcal{E}_\alpha})} & \in D^+(Y_{\text{et}}, G_{K-\mathbb{Z}/p^*}),
\end{align*}
\]

where \( W_n \) means \( \text{Spec}(W_n) \) endowed with the trivial log structure for each \( n \geq 1 \). See Definition 9.6 for \( (\mathcal{E}_\alpha, M_{\mathcal{E}_\alpha}) \). To construct \( \mathcal{E}(X^*_\alpha, M_\alpha)/W_n \), we fix an étale hypercovering \((X^*, M^*) \hookrightarrow (X, M)\) and a closed immersion \((X^*, M^*) \hookrightarrow (Z^*, M_{Z^*})\) of simplicial fine log schemes over \( W \) such that \((Z^*, M_{Z^*})\) is smooth over \( W \) for each \( i \in \mathbb{N} \) (cf. [HK] (2.18)). Put

\[
Y^* := X^* \otimes_{\mathcal{O}_{X^*}} k,
\]

which is an étale hypercovering of \( Y \). For \( n \geq 1 \), let \( (\mathcal{D}_n^*, M_{\mathcal{D}_n^*}) \) be the PD-envelope of \((X^*_n, M_{X^*_n}) \hookrightarrow (Z^*_n, M_{Z^*_n})\) with respect to the canonical PD-structure on \((p) \subset W_n\) ([K3] Definition 5.4), and we define a complex \( \mathcal{E}(X^*_n, M_{X^*_n})/W_n \) of sheaves on \( Y_{\text{et}}^* \) as

\[
\begin{align*}
\mathcal{O}_{\mathcal{D}_n^*} & \xrightarrow{d} \mathcal{O}_{\mathcal{D}_n^*} \otimes \mathcal{O}_{\mathcal{D}_n^*} \omega_{(Z^*_n, M_{Z^*_n})}/W_n \xrightarrow{d} \cdots \xrightarrow{d} \mathcal{O}_{\mathcal{D}_n^*} \otimes \mathcal{O}_{\mathcal{D}_n^*} \omega_{(Z^*_n, M_{Z^*_n})}/W_n \xrightarrow{d} \cdots,
\end{align*}
\]

where the first term is placed in degree 0 and \( \omega_{(Z^*_n, M_{Z^*_n})}/W_n \) denotes the \( q \)-th differential module of \((Z^*_n, M_{Z^*_n})\) over \( W_n \) cf. [K3] (1.7). See loc. cit. Theorem 6.2 for \( d \).
Regarding this complex as a complex of projective systems (with respect to $n \geq 1$) of sheaves on $Y_{\text{et}}$, we obtain a complex $E_{(X^\bullet, M^\bullet)}/W^\bullet$ of objects of $\text{Shv}(Y_{\text{et}}^\bullet, \mathbb{Z}/p^\bullet)$. We then define

$$E_{(X^\bullet, M^\bullet)}(V, M_S) := R\theta_v(E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)) \in D^+(V_{\text{et}}),$$

where $\theta : \text{Shv}(Y_{\text{et}}^\bullet, \mathbb{Z}/p^\bullet) \to \text{Shv}(Y_{\text{et}}^\bullet, \mathbb{Z}/p^\bullet)$ denotes the natural morphism of topoi. The resulting object $E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)$ is independent of the choice of the pair $((X^\bullet, M^\bullet), (Z^\bullet, M_Z^\bullet))$ by a standard argument (cf. [K2] p. 212).

To construct $E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)$, we put $V := \text{Spec}(W[t])$, and define $M_V$ as the log structure on $V$ associated with the divisor $\{t = 0\}$. We regard $(X, M)$ as a log scheme over $(V, M_V)$ by the composite map

$$(X, M) \to (S, M_S) \to (V, M_V),$$

where the last map is given by $T \mapsto \pi$, the prime element of $O_K$ we fixed in Definition 9.6 to define $(\mathcal{E}_n, M_{\mathcal{E}_n})$. We then apply the same construction as for $E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)$ to the morphism $(X, M) \to (V, M_V)$, i.e. fix an étale hypercovering $(X^\bullet, M^\bullet) \to (X, M)$ and a closed immersion $\iota^* : (X^\bullet, M^\bullet) \to (Z^\bullet, M_Z^\bullet)$ of simplicial log schemes over $(V, M_V)$ such that $(Z^\bullet, M_Z^\bullet)$ is smooth over $W$ for each $i \in \mathbb{N}$. We define $E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)$ by replacing $\omega_{Z^\bullet, M_Z^\bullet}^\bullet(W^\bullet)$ with $\omega_{Z^\bullet, M_Z^\bullet}^\bullet(W^\bullet)$ in the above definition of $E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)$.

We construct $E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)$ as follows. Fix an étale hypercovering $(X^\bullet, M^\bullet) \to (X, M)$ and for each finite extension $L/K$ contained in $K$, fix a closed immersion $(X^\bullet_{\text{et}}, M^\bullet_{\text{et}}) \hookrightarrow (Z^\bullet_L, M_Z^\bullet_L)$ of simplicial fine log schemes over $W$ such that $(Z^\bullet_L, M_Z^\bullet_L)$ is smooth over $W$ for each $i \in \mathbb{N}$ and $L/K$, and such that for finite extensions $L'/L$ there are morphisms $\tau_{L'/L} : (Z^\bullet_{L'}, M_Z^\bullet_{L'}) \to (Z^\bullet_L, M_Z^\bullet_L)$ satisfying transitivity. For a finite extension $L/K$, let $k_L$ be the residue field of $L$, and put

$$Y^\bullet_L := X^\bullet \otimes_{O_K} k_L \quad \text{and} \quad Y^\bullet := X^\bullet \otimes_{O_K} K,$

which are étale hypercoverings of $Y_L := Y \otimes_k k_L$ and $Y$, respectively. We define a complex $E_{(X^\bullet_{\text{et}}, M^\bullet_{\text{et}})}(V, M_S^\bullet)$ on $Y^\bullet_L$ applying the same construction as for $E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)$ to the embedding $(X^\bullet_{\text{et}}, M^\bullet_{\text{et}}) \hookrightarrow (Z^\bullet_L, M_Z^\bullet_L)$, whose inverse image onto $Y^\bullet$ yields an inductive system of complexes of objects of $\text{Shv}(Y_{\text{et}}^\bullet, G_K \cdot \mathbb{Z}/p^\bullet)$ with respect to finite extensions $L/K$. We then define

$$E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet) := \lim_{\rightarrow} E_{(X^\bullet_{\text{et}}, M^\bullet_{\text{et}})}(V, M_S^\bullet),$$

where $\theta^\bullet : \text{Shv}(Y^\bullet_L, G_K \cdot \mathbb{Z}/p^\bullet) \to \text{Shv}(Y^\bullet_{\text{et}}, G_K \cdot \mathbb{Z}/p^\bullet)$ denotes the natural morphism of topoi. We construct $E_{(X^\bullet, M^\bullet)}(V, M_S^\bullet)$ applying similar arguments to the complexes $E_{(X^\bullet_{\text{et}}, M^\bullet_{\text{et}})}(V, M_S^\bullet)$.
We define the following objects of $\text{Mod}(\mathbb{Z}/p^*):$

\begin{align*}
H^i_{\text{crys}}((X_\bullet, M_\bullet)/W_\bullet) &:= R^i \Gamma(E_{(X_\bullet, M_\bullet)}/W_\bullet), \quad (A.1.1) \\
H^i_{\text{crys}}((X_\bullet, M_\bullet)/((\mathcal{E}_\bullet, M_{\mathcal{E}})_\bullet)) &:= R^i \Gamma(E_{(X_\bullet, M_\bullet)}/((\mathcal{E}_\bullet, M_{\mathcal{E}})_\bullet)), \quad (A.1.2) \\
H^i_{\text{crys}}((\overline{X}_\bullet, \overline{M}_\bullet)/W_\bullet) &:= R^i \mathcal{T}(E_{(\overline{X}_\bullet, \overline{M}_\bullet)}/W_\bullet), \quad (A.1.3) \\
H^i_{\text{crys}}((\overline{X}_\bullet, \overline{M}_\bullet)/((\mathcal{E}_\bullet, M_{\mathcal{E}})_\bullet)) &:= R^i \mathcal{T}(E_{(\overline{X}_\bullet, \overline{M}_\bullet)}/((\mathcal{E}_\bullet, M_{\mathcal{E}})_\bullet)), \quad (A.1.4)
\end{align*}

where $\Gamma$ and $\mathcal{T}$ denote the following left exact functors, respectively:

\begin{align*}
\Gamma &:= \Gamma(Y, -) : \text{Shv}(Y_{\text{et}}, \mathbb{Z}/p^*) \to \text{Mod}(\mathbb{Z}/p^*), \\
\mathcal{T} &:= \Gamma(Y, -) : \text{Shv}(Y_{\text{et}}, G_K\cdot\mathbb{Z}/p^*) \to \text{Mod}(G_K\cdot\mathbb{Z}/p^*). \quad (A.1.5)
\end{align*}

We define ‘naive’ crystalline cohomology groups

\begin{align*}
H^i_{\text{crys}}((X, M)/W), & \quad H^i_{\text{crys}}((X, M)/((\mathcal{E}, M_{\mathcal{E}}))), \\
H^i_{\text{crys}}((\overline{X}, \overline{M})/W) & \quad \text{and} \quad H^i_{\text{crys}}((\overline{X}, \overline{M})/((\mathcal{E}, M_{\mathcal{E}})))
\end{align*}

as the projective limit of (A.1.1)–(A.1.4), respectively.

**Remark A.1.6** Let $\pi_n : \text{Shv}(Y_{\text{et}}, \mathbb{Z}/p^n) \to \text{Shv}(Y_{\text{et}}, \mathbb{Z}/p^n)$ be the natural functor sending $\{\mathcal{F}_m\}_{m \geq 1}$ to $\mathcal{F}_n$. Since $\pi_n$ is exact, it extends to a triangulated functor $\pi_n : D^+(Y_{\text{et}}, \mathbb{Z}/p^n) \to D^+(Y_{\text{et}}, \mathbb{Z}/p^n)$, which sends $E_{(X_\bullet, M_\bullet)}/W_\bullet \mapsto E_{(X_\bullet, M_\bullet)}/W_\bullet$, an object computing the crystalline cohomology of $(X_\bullet, M_\bullet)/W_\bullet$, because $\pi_n$ is compatible with the gluing functor $\mathcal{T}$.

Moreover by [1] Proposition 1.1 (b), we have

\begin{align*}
H^i_{\text{crys}}((X_\bullet, M_\bullet)/W_\bullet) & \simeq \{H^i_{\text{crys}}((X_n, M_n)/W_n)\}_{n \geq 1}.
\end{align*}

We have similar facts for (A.1.2), (A.1.3) and (A.1.4) as well.

**A.2 Continuous crystalline cohomology**

We define continuous crystalline cohomology groups as follows:

\begin{align*}
H^i_{\text{cont-cr}}((X, M)/W) &:= R^i \left( \varprojlim \Gamma \right) (E_{(X_\bullet, M_\bullet)}/W_\bullet), \quad (A.2.1) \\
H^i_{\text{cont-cr}}((X, M)/((\mathcal{E}, M_{\mathcal{E}}))) &:= R^i \left( \varprojlim \Gamma \right) (E_{(X_\bullet, M_\bullet)}/((\mathcal{E}, M_{\mathcal{E}})_\bullet)), \quad (A.2.2)
\end{align*}

where $\Gamma$ is as in (A.1.5). Because $\Gamma$ has an exact left adjoint, it preserves injectives and there exists a spectral sequence

\[ E_{a,b}^2 = R^a \varprojlim \Gamma \left( H^b_{\text{crys}}((X_\bullet, M_\bullet)/W_\bullet) \right) \Rightarrow H_{\text{cont-cr}}^{a+b}((X, M)/W). \]

Because $R^a \varprojlim = 0$ for $a \geq 2$, this spectral sequence breaks up into short exact sequences

\[ 0 \to R^1 \varprojlim \Gamma \left( H^{i-1}_{\text{crys}} ((X_\bullet, M_\bullet)/W_\bullet) \right) \to H^i_{\text{cont-cr}}((X, M)/W) \to H^i_{\text{crys}}((X, M)/W) \to 0. \quad (A.2.3) \]

We have similar exact sequences for (A.2.2) by the same arguments.
There is a commutative diagram (see Definition 9.6 for $\mathcal{N}$)

\[
\begin{array}{c}
\lim \rightarrow & E(X, M^*)/W^n & \rightarrow & E(X, M^*)/(\mathcal{E}, M_{\mathcal{E}}) & \rightarrow & E(X, M^*)/W_n[1] \\
\end{array}
\]

in $D^+(Y_{\text{et}}, \mathbb{Z}/p^*)$. Consequently we obtain a long exact sequence

\[
\cdots \rightarrow H^{i}_{\text{cont-cr}}((X, M)/W) \rightarrow H^{i}_{\text{cont-cr}}((X, M)/(\mathcal{E}, M_{\mathcal{E}})) \rightarrow H^{i+1}_{\text{cont-cr}}((X, M)/W) \rightarrow \cdots.
\]

Proof. The assertion follows from the same arguments as in the proof of [K4] Lemma 4.2 with $\mathcal{F} = \mathcal{O}^{\text{cris}}$.

There is a commutative diagram (see Definition 9.6 for $\mathcal{N}$)

\[
\begin{array}{c}
H^{i}_{\text{cont-cr}}((X, M)/(\mathcal{E}, M_{\mathcal{E}})) & \rightarrow & H^{i}_{\text{cont-cr}}((X, M)/(\mathcal{E}, M_{\mathcal{E}})) \\
\downarrow & & \downarrow \\
H^{i}_{\text{cris}}((X, M)/(\mathcal{E}, M_{\mathcal{E}})) & \rightarrow & H^{i}_{\text{cris}}((X, M)/(\mathcal{E}, M_{\mathcal{E}})).
\end{array}
\]

We will show later that the vertical arrows are bijective in Corollary A.3.2 (1) below.

A.3 Comparison of projective systems of crystalline cohomology

We recall here the following comparison facts on projective systems of crystalline cohomology groups, which will be useful later. Note that for $A_* \in \text{Mod}(\mathbb{Z}/p^*)$, both $\lim A_*$ and $R^1 \lim A_*$ have finite exponents if $\mathbb{Q} \otimes (A_*) \simeq 0$ in $\mathbb{Q} \otimes \text{Mod}(\mathbb{Z}/p^*)$.

Proposition A.3.1 Let $i$ be a non-negative integer, and put $D^n_i := \mathbb{Q}(Y, \mathbb{W}_n \omega^n_\bullet)$.

Let $P_\bullet$ be the ring defined in [Ts1] §1.6, and put $R_{\mathcal{E}_n} := \Gamma(\mathcal{E}_n, \mathcal{O}_{\mathcal{E}_n})$.

1. There is an isomorphism in $\mathbb{Q} \otimes \text{Mod}(\mathbb{Z}/p^*)$

\[
\mathbb{Q} \otimes \{H^i_{\text{cris}}((X_*, M_*)/(\mathcal{E}, M_{\mathcal{E}}))\} \simeq \mathbb{Q} \otimes \{R_{\mathcal{E}_n} \otimes \mathbb{W}_n D^n_i\}_{n \geq 1}
\]

and $R^1 \lim H^i_{\text{cris}}((X_*, M_*)/(\mathcal{E}, M_{\mathcal{E}}))$ has a finite exponent.

2. There is an isomorphism in $\mathbb{Q} \otimes \text{Mod}(G_K \cdot \mathbb{Z}/p^*)$

\[
\mathbb{Q} \otimes \{H^i_{\text{cris}}((\overline{X_*, M_*)}/(\mathcal{E}, M_{\mathcal{E}}))\} \simeq \mathbb{Q} \otimes \{P_n \otimes \mathbb{W}_n D^n_i\}_{n \geq 1}
\]

and $R^1 \lim H^i_{\text{cris}}((\overline{X_*, M_*)}/(\mathcal{E}, M_{\mathcal{E}}))$ has a finite exponent.

3. There is an isomorphism in $\mathbb{Q} \otimes \text{Mod}(G_K \cdot \mathbb{Z}/p^*)$

\[
\mathbb{Q} \otimes \{H^i_{\text{cris}}((\overline{X_*, M_*)}/\mathbb{W}_n)\} \simeq \mathbb{Q} \otimes \{(P_n \otimes \mathbb{W}_n D^n_i)^N = 0\}_{n \geq 1}
\]

and $R^1 \lim H^i_{\text{cris}}((\overline{X_*, M_*)}/\mathbb{W}_n)$ has a finite exponent. Here $N$ acts on $P_n \otimes \mathbb{W}_n D^n_i$ by $N_{P_n} \otimes 1 + 1 \otimes N$ and $N_{P_n}$ denotes the monodromy operator on $P_n$, cf. [Ts1] p. 253.
Proof. (1) The first assertion follows from [HK] Lemma 5.2 (see also (*) in the proof of [Ts1] Proposition 4.4.6). To show the second assertion, we check that the projective system \( \{ R_{\mathfrak{d}_n} \otimes_{W_n} D_n \}_{n \geq 1} \) satisfies the Mittag-Leffler condition. Indeed, \( D_n \) is finitely generated over \( W_n \), and the projection \( R_{\mathfrak{d}_{n+1}} \to R_{\mathfrak{d}_n} \) is surjective by the presentation

\[
R_{\mathfrak{d}_n} = W[t, t^{\nu}/\nu! \ (\nu \geq 1)] \otimes_W W_n \quad (e := [K : K_0])
\]

obtained from the definition of \( (\mathfrak{d}_n, M_{\mathfrak{d}_n}) \) (see the proof of [Ts1] Proposition 4.4.6).

(2) The first assertion follows from that of (1) and [Ts1] Proposition 4.5.4. The second assertion follows from the fact that \( \{ P_n \otimes_{W_n} D_n \}_{n \geq 1} \) satisfies the Mittag-Leffler condition. Indeed, the natural projection \( P_{n+1} \to P_n \) is surjective (loc. cit. Lemma 1.6.7) and \( D_n \) is finitely generated over \( W_n \).

(3) See the proof of loc. cit. Proposition 4.5.6 for the first assertion. We show the second assertion. Note that \( P_n \) is flat over \( W_n \), because \( R_{\mathfrak{d}_n} \) is flat over \( W_n \) and \( P_n \) is flat over \( R_{\mathfrak{d}_n} \) by the above presentation of \( R_{\mathfrak{d}_n} \) and loc. cit. Proposition 4.1.5. Let \( \{ D_n \}_{n \geq 1} \) be a projective system of \( W \)-modules such that \( D_n \) is finitely generated over \( W_n \) for each \( n \), and let \( N : \{ D_n \}_{n \geq 1} \to \{ D_n \}_{n \geq 1} \) be a nilpotent \( W \)-endomorphism. Our task is to show that

\[
R^1 \lim \{ (P_n \otimes_{W_n} D_n)^{N=0} \}_{n \geq 1} = 0.
\]

Consider a short exact sequence of projective systems

\[
0 \to \{ (D_n)^{N=0} \}_{n \geq 1} \to \{ D_n \}_{n \geq 1} \to \{ N(D_n) \}_{n \geq 1} \to 0.
\]

Note that \( (D_n)^{N=0} \) and \( N(D_n) \) are finitely generated over \( W_n \) for each \( n \). Let \( b \) be the minimal integer for which \( N^b = 0 \) on \( \{ D_n \}_{n \geq 1} \). By [K4] Lemma 4.3 and the flatness of \( P_n \) over \( W_n \), we have a short exact sequence for each \( n \geq 1 \)

\[
0 \to (P_n \otimes_{W_n} (D_n)^{N=0})^{N=0} \to (P_n \otimes_{W_n} D_n)^{N=0} \to (P_n \otimes_{W_n} N(D_n))^{N=0} \to 0,
\]

which yields a short exact sequence of projective systems with respect to \( n \geq 1 \). Since \( N^{b-1} = 0 \) on \( \{ N(D_n) \}_{n \geq 1} \), we may assume \( b = 1 \) by induction on \( b \geq 1 \). Now let \( B_n \) be as in [Ts1] §1.6 and let \( A_{\text{cris}} \) be as in loc. cit. §1.1. Then we have isomorphisms

\[
(P_n \otimes_{W_n} D_n)^{N=0} \overset{(1)}{\cong} (P_n)^{N=0} \otimes_{W_n} D_n \overset{(2)}{\cong} B_n \otimes_{W_n} D_n \overset{(3)}{=} (A_{\text{cris}}/p^n) \otimes_{W_n} D_n,
\]

where (1) follows from the assumption \( b = 1 \) and the flatness of \( P_n \) over \( W_n \). The isomorphism (2) (resp. (3)) follows from loc. cit. Corollary 1.6.6 (resp. the definition of \( B_n \) in loc. cit. §1.6). Thus \( R^1 \lim \{ (P_n \otimes_{W_n} D_n)^{N=0} \}_{n \geq 1} \) is zero, and we obtain the assertion. \( \square \)

Corollary A.3.2. (1) For \( i \geq 0 \), we have

\[
H^i_{\text{cris}}((X, M)/(\mathfrak{d}, M_{\mathfrak{d}}))_{\mathbb{Q}_p} \simeq H^i_{\text{cont}-\text{cris}}((X, M)/(\mathfrak{d}, M_{\mathfrak{d}}))_{\mathbb{Q}_p}.
\]
(2) The torsion subgroups of $H^i_{\text{crys}}((X, M)/(\mathcal{E}, M_\mathcal{E}))$ and $H^i_{\text{crys}}((X, M)/W)$ have finite exponents for any $i \geq 0$.

**Proof.** (1) follows from Proposition A.3.1 (1) and the remark after (A.2.3). Since $\varprojlim_{n \geq 1} P_n$ is $p$-torsion free, the assertion (2) follows from the isomorphisms in Proposition A.3.1 (3). □

**Remark A.3.3** If $K$ is absolutely unramified, then we have

$$H^i_{\text{crys}}((X_n, M_n)/W_n) \simeq \mathbb{H}^i(Y, W_n \otimes \mathbb{F}_p),$$

which is finitely generated over $W_n$ by Theorem 8.8 (2) and (8.10.1) (see also [Hy2] (1.4.3), (2.4.2)). Consequently, the projective system $H^i_{\text{crys}}((X, M)/W)$ satisfies the Mittag-Leffler condition and we obtain a long exact sequence

$$\cdots \rightarrow H^i_{\text{crys}}((X, M)/W)_p \rightarrow H^i_{\text{crys}}((X, M)/(\mathcal{E}, M_\mathcal{E}))[p] \rightarrow H^{i+1}_{\text{crys}}((X, M)/W)_p \rightarrow \cdots,$$

which removes the assumption $(*$) in [La] p. 191. On the other hand, the author does not know if $H^i_{\text{crys}}((X, M)/W)$ satisfies the Mittag-Leffler condition even up to torsion, when $K$ is not absolutely unramified.

The following corollary has been used in the proof of Lemma 9.7 (1):

**Corollary A.3.4** In the following commutative diagram of canonical maps, the arrows (3) and (4) are injective:

$$
\begin{array}{ccc}
H^i_{\text{cont-cr}}((X, M)/W)_p & \xrightarrow{(1)} & H^i_{\text{crys}}((X, M)/(\mathcal{E}, M_\mathcal{E}))[p] \\
\downarrow (2) & & \downarrow (3) \\
H^i_{\text{crys}}((X, M)/W)_p & \xrightarrow{(4)} & H^i_{\text{crys}}((X, M)/(\mathcal{E}, M_\mathcal{E}))[p].
\end{array}
$$

In particular, the kernel of (1) agrees with that of (2).

**Proof.** The injectivity of (3) follows from Proposition A.3.1 (1) and (2) and the injectivity of the natural maps $R_{E_n} \rightarrow P_n$ for $n \geq 1$ ([Ts1] Proposition 4.1.5). The injectivity of (4) follows from Proposition A.3.1 (2) and (3). □

### A.4 Continuous-Galois crystalline cohomology

We define the continuous-Galois crystalline cohomology as follows:

$$H^i_{\text{cG-cr}}((X, M)/W) := R^i \left( \varprojlim_{n \geq 1} \Gamma_{\text{Gal}}(\overline{\mathbb{F}}_q) \right) \left( \mathcal{E}_{(X, M_n)} / W_n \right).$$
where $\overline{T}$ is as in (A.1.6), and $\Gamma_{\text{Gal}}$ denotes the functor taking $G_K$-invariant subgroups:

$$\Gamma_{\text{Gal}} := \Gamma(G_K, -) : \text{Mod}(G_K, \mathbb{Z}/p^*) \rightarrow \text{Mod}(\mathbb{Z}/p^*).$$

There is a natural map

$$H^i_{\text{cont-cris}}((X, M)/W) \rightarrow H^i_{\text{crys}}((X, M)/W) \quad (A.4.1)$$

by definition.

**Theorem A.4.2** There exists a Hochschild-Serre spectral sequence

$$E^2_{a,b} = H^a(K, H^b_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}) \Rightarrow H^{a+b}_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}.$$

**Proof.** Because $\overline{T}$ has an exact left adjoint functor, it preserves injectives and there exists a spectral sequence

$$E^2_{a,b} = H^a(K, H^b_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}) \Rightarrow H^{a+b}_{\text{crys}}((X, M)/W),$$

where $H^a(K, \{F_n\}_{n\geq 1})$ for a projective system $\{F_n\}_{n\geq 1}$ of discrete $G_K$-modules denotes the continuous Galois cohomology of $G_K$ in the sense of Jannsen [J] §2. We show that the canonical map

$$\alpha^{a,b} : H^a(K, H^b_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}) \rightarrow H^a(K, H^{b}_{\text{crys}}((X, M)/W))_{\mathbb{Q}_p}$$

(see the proof of loc. cit. Theorem 2.2) is bijective for $a, b \geq 0$. Put

$$L^b_n := \{p_n \otimes \mathbb{Z}_p \lim_{\leftarrow} L_n\} = B^+_{\text{st}} \otimes_{K_0} D^b_n,$$

$$T^b_n := H^b_{\text{crys}}((X, M)/W_n)$$

and $T^b := H^b_{\text{crys}}((X, M)/W)_{\mathbb{Q}_p}$.

Since the torsion part of $\lim_{\leftarrow} L^b_n$ has a finite exponent, the canonical map

$$\beta^{a,b} : H^a(K, L^b) \rightarrow H^a(K, L^b)_{\mathbb{Q}_p} \quad (A.4.3)$$

is bijective by loc. cit. Theorem 5.15 (c). We have a short exact sequence of objects in $\mathbb{Q} \otimes \text{Mod}(G_K, \mathbb{Z}/p^*)$

$$0 \rightarrow \mathbb{Q} \otimes (T^b_n) \rightarrow \mathbb{Q} \otimes (L^b_n) \rightarrow \mathbb{Q} \otimes (L^b) \rightarrow 0$$

and a short exact sequence of topological $G_K$-modules

$$0 \rightarrow T^b \rightarrow L^b \rightarrow L^b \rightarrow 0 \quad (A.4.4)$$

by Proposition A.3.1 (3) and [K4] Lemma 4.3. These exact sequences yield a commutative diagram with exact rows

$$\cdots \rightarrow H^{n-1}(K, L^b) \rightarrow H^n(K, T^b) \rightarrow H^n(K, L^b) \rightarrow H^n(K, L^b) \rightarrow \cdots$$

$$\downarrow \alpha^{n-1,b} \quad \downarrow \alpha^{n,b} \quad \downarrow \beta^{n-1,b} \quad \downarrow \beta^{n,b}$$

$$\cdots \rightarrow H^{n-1}(K, L^b)_{\mathbb{Q}_p} \rightarrow H^n(K, T^b_{\mathbb{Q}_p}) \rightarrow H^n(K, L^b)_{\mathbb{Q}_p} \rightarrow H^n(K, L^b)_{\mathbb{Q}_p} \rightarrow \cdots,$$

where the arrows $\beta^{n-1,b}$ and $\beta^{n,b}$ are bijective by (A.4.3). Hence the assertion follows from the five lemma. \qed
Corollary A.4.5 Put

\[ H^i_{\text{cont-cr}}((X, M)/W)^0_{Q_p} := \text{Ker}(H^i_{\text{cont-cr}}((X, M)/W)_{Q_p} \to H^{i+1}_{\text{crys}}((\mathcal{X}, \mathcal{M})/W)_{Q_p}). \]

Then the canonical homomorphism

\[ H^i_{\text{cont-cr}}((X, M)/W)^0_{Q_p} \to H^1(K, H^i_{\text{crys}}((\mathcal{X}, \mathcal{M})/W)_{Q_p}), \]

obtained from (A.4.1) and an edge homomorphism of the spectral sequence in Theorem A.4.2, fits into a commutative diagram

\[
\begin{array}{ccc}
H^i_{\text{crys}}((X, M)/(\mathcal{E}, M_\xi))_{Q_p} & \xrightarrow{\partial} & H^{i+1}_{\text{cont-cr}}((X, M)/W)^0_{Q_p} \\
\downarrow & & \downarrow \\
(B^+_{\text{st}} \otimes_{K_o} D^+)^{G_K} & \xrightarrow{(-1)^{i+1} \delta} & H^1(K, H^i_{\text{crys}}((\mathcal{X}, \mathcal{M})/W)_{Q_p}),
\end{array}
\]

where \( \partial \) is the connecting map induced by that in Proposition A.2.4 (see Corollary A.3.4 for the existence and the surjectivity), and \( \delta \) denotes the connecting map of continuous Galois cohomology associated with (A.4.4) with \( b = i \). The left vertical arrow is obtained from Proposition A.3.1 (2).

**Proof.** The assertion follows from Proposition A.3.1 (2) and a simple computation on boundary maps arising from the distinguished triangle

\[
\text{R}^T(E_{(\mathcal{X}, \mathcal{M})}/W) \xrightarrow{\text{can}} \text{R}^T(E_{(\mathcal{X}, \mathcal{M})}/(\mathcal{E}, M_\xi)) \\
\xrightarrow{\nu} \text{R}^T(E_{(\mathcal{X}, \mathcal{M})}/(\mathcal{E}, M_\xi)) \xrightarrow{\Phi^\wedge} \text{R}^T(E_{(\mathcal{X}, \mathcal{M})}/W)[1]
\]

in \( D^+(\text{Mod}(G_K, \mathbb{Z}/p^\bullet)) \) (a variant of Proposition A.2.4). The sign \((-1)^{i+1}\) in the diagram arises from the difference of the orientations between this distinguished triangle and (A.4.4), and the fact that the construction of connecting morphisms associated with short exact sequences of complexes commutes with the shift functor \([i]\) up to the sign \((-1)^i\). The details are straightforward and left to the reader. \( \square \)

### A.5 Syntomic complexes

We construct the following objects for \( r \geq 0 \):

\[
\mathcal{S}^\wedge(r)_{(X, M)} \in D^+(Y_{\text{et}}, \mathbb{Z}/p^\bullet) \quad \text{and} \quad \mathcal{S}^\wedge(r)_{(\mathcal{X}, \mathcal{M})} \in D^+(Y_{\text{et}}, G_K, \mathbb{Z}/p^\bullet),
\]

and the following objects for \( 0 \leq r \leq p - 1 \):

\[
\mathcal{S}(r)_{(X, M)} \in D^+(Y_{\text{et}}, \mathbb{Z}/p^\bullet) \quad \text{and} \quad \mathcal{S}(r)_{(\mathcal{X}, \mathcal{M})} \in D^+(Y_{\text{et}}, G_K, \mathbb{Z}/p^\bullet).
\]

**Definition A.5.1** Let \((T, M_T)\) be a log scheme over \( \mathbb{Z}_p \). A Frobenius endomorphism \( \varphi : (T, M_T) \to (T, M_T) \) is a morphism over \( \mathbb{Z}_p \) such that \( \varphi \otimes \mathbb{Z}/p : (T_1, M_{T_1}) \to (T_1, M_{T_1}) \) is the absolute Frobenius endomorphism in the sense of [K3] Definition 4.7.
To construct the objects $\mathcal{S}_r^\bullet (X, M)$ and $\mathcal{S}_r^\bullet (X, M)$, we fix an étale hypercovering $(X^*, M^*) \to (X, M)$ and a closed immersion $(X^*, M^*) \hookrightarrow (Z^*, M_{Z^*})$ of simplicial fine log schemes over $W$ such that $(Z^*, M_{Z^*})$ is smooth over $W$ and has a Frobenius endomorphism for each $i \in \mathbb{N}$. Let $n \geq 1$ be an integer, and let $(\mathcal{S}_n^*, M_{\mathcal{S}_n}^*)$ be the PD-envelope of $(X_n^*, M_n^*)$ in $(Z_n^*, M_{Z_n^*})$ with respect to the canonical PD-structure on $(p) \subset W_n$. For $i \geq 1$, let $\mathcal{J}^i \subset \mathcal{O}_{\mathcal{S}_n^*}$ be the $i$-th divided power of the ideal $\mathcal{J} = \text{Ker}(\mathcal{O}_{\mathcal{S}_n^*} \to \mathcal{O}_{X_n^*})$. For $i \leq 0$, put $\mathcal{J}^i := \mathcal{O}_{\mathcal{S}_n^*}$. Let $\mathcal{J}^{[r]}_{(X_n^*, M_n^*)}/W_n$ be the complex of sheaves on $Y_{et}^*$

$$\mathcal{J}^{[r]} \xrightarrow{d_i} \mathcal{J}^{[r-i]} \otimes_{\mathcal{O}_{Z_n^*}} \omega_{(Z_n^*, M_{Z_n^*})}/W_n \xrightarrow{d_1} \cdots \xrightarrow{d_i} \mathcal{J}^{[r-q]} \otimes_{\mathcal{O}_{Z_n^*}} \omega_{(Z_n^*, M_{Z_n^*})}/W_n \xrightarrow{d_1} \cdots ,$$

where $\mathcal{J}^{[r]}$ is placed in degree 0. See [Ts2] Corollary 1.10 for $d$. The complex $\mathcal{E}_{(X_n^*, M_n^*)}/W_n$ we considered in §A.1 agrees with $\mathcal{J}^{[r]}_{(X_n^*, M_n^*)}/W_n$. We define a complex $\mathcal{S}_n^\bullet (r)(X^*, M^*)$ on $Y_{et}^*$ as the mapping fiber of the homomorphism

$$p^r - \varphi_n^r : \mathcal{J}^{[r]}_{(X_n^*, M_n^*)} \to \mathcal{E}_{(X_n^*, M_n^*)}$$

For $0 \leq r \leq p - 1$, the Frobenius endomorphism on $(Z_{n+1}^*, M_{Z_{n+1}^*})$ induces a homomorphism of complexes

$$f_r := p^r - \varphi_{n+r}^r : \mathcal{J}^{[r]}_{(X_n^*, M_n^*)}/W_n \to \mathcal{E}_{(X_n^*, M_n^*)}/W_n$$

(cf. [Ts2] p. 540). We define a complex $\mathcal{S}_n^\bullet (r)(X^*, M^*)$ on $Y^*_{et}$ as the mapping fiber of the homomorphism

$$1 - f_r : \mathcal{J}^{[r]}_{(X_n^*, M_n^*)} \to \mathcal{E}_{(X_n^*, M_n^*)}$$

Regarding $\mathcal{S}_n^\bullet (r)(X^*, M^*)$ and $\mathcal{S}_n^\bullet (r)(X^*, M^*)$ as complexes of projective systems (on $n \geq 1$) of sheaves on $Y_{et}^*$, we define

$$\mathcal{S}_r^\bullet (X, M) := R\theta_*(\mathcal{S}_r^\bullet (r)(X^*, M^*)) \quad (r \geq 0),$$

$$\mathcal{S}_r^\bullet (X, M) := R\theta_*(\mathcal{S}_r^\bullet (r)(X^*, M^*)) \quad (0 \leq r \leq p - 1),$$

where $\theta : \text{Shv}(Y_{et}^*, \mathbb{Z}/p^*) \to \text{Shv}(Y_{et}^*, \mathbb{Z}/p^*)$ denotes the natural morphism of topoi. The resulting objects are independent of the pair $((X^*, M^*), (Z^*, M_{Z^*}))$ (cf. [K2] p. 212).

We construct $\mathcal{S}_r^\bullet (r)(X_{\mathcal{S}_n^*})$ for $r \geq 0$ and $\mathcal{S}_r^\bullet (r)(X_{\mathcal{S}_n^*})$ for $0 \leq r \leq p - 1$ as follows. Fix an étale hypercovering $(X^*, M^*) \to (X, M)$ and for each finite extension $L/K$ contained in $\mathcal{R}$, fix a closed immersion $(X_{O_L^*, M_{O_L}}^*) \hookrightarrow (Z_{L}^*, M_{Z_L^*})$ of simplicial fine log schemes over $W$ such that $(Z_L^*, M_{Z_L^*})$ is smooth over $W$ for each $i \in \mathbb{N}$ and $L/K$, such that $Z_L^*$ has a Frobenius endomorphism for each $i \in \mathbb{N}$ and $L/K$, and such that for finite extensions $L'/L$ there are morphisms $\tau_{L'/L} : (Z_{L^*}, M_{Z_{L}}) \to (Z_{L'}^*, M_{Z_{L}^*})$ which satisfy transitivity and compatibility with Frobenius morphisms. For a finite extension $L/K$, we define $Y^*_{L}$ in a similar way as in §A.1. We define complexes
which satisfy

\[ \text{Let } a \text{ with } a \in \mathbb{Z} \text{ satisfy} \]

\( \text{For } \theta \in \mathbb{Z}, \text{ we define} \)

\[ \text{Theorem A.5.6 ([Ts1] \S 3.1, [K4] Theorem 5.4) For } r \geq 0, \text{ there exists a} \]

\[ \text{compatible with product structures} \]

\[ \text{If } r \leq p - 2, \text{ then } \eta^*_r \text{ factors through an isomorphism} \]

\[ \text{Documenta Mathematica 18 (2013) 177–247} \]
Proof. We define \( \eta^*_p \) applying the arguments in [Ts1] §3.1 in the category of \( \mathbb{Z}/p^\bullet \)-sheaves. If \( r \leq p - 2 \), then we have \( \mathbb{Z}/p^\bullet(r)^* = \mu_{p^r}^\infty \) and \( \eta^*_p \) factors through a morphism \( \eta^*_p : \mathcal{A}(r)(\mathcal{X}, \mathcal{M}) \to \mathcal{T} H^\infty_{\text{cont}} \mu_{p^r}^\infty \) by the construction of \( \eta^*_p \) (loc. cit. (3.1.11)). The morphism \( \eta^*_p \) induces an isomorphism as claimed, because \( \mathcal{A}(r)(\mathcal{X}, \mathcal{M}) \) is concentrated in \([0, r]\) and \( \eta^*_p \) induces isomorphisms on the \( q \)-th cohomology objects with \( 0 \leq q \leq r \) by [K4] Theorem 5.4.

We define

\[
H^i_{\text{syn}}((X, M), \mathcal{A}(r)) := R^i \Gamma(\mathcal{A}(r)(X, M)), \\
H^i_{\text{syn}}((X, M), \mathcal{Z}_p(r)) := \varprojlim H^i_{\text{syn}}((X, M), \mathcal{A}(r)), \\
H^i_{\text{syn}}((X, M), \mathcal{F}_p(r)) := Q_p \otimes \mathbb{Z}_p H^i_{\text{syn}}((X, M), \mathcal{A}(r)), \\
H^i_{\text{syn}}((X, M), \mathcal{F}_p(r)) := Q_p \otimes \mathbb{Z}_p \varprojlim R^i \mathcal{T}(\mathcal{A}(r)(\mathcal{X}, \mathcal{M})).
\]

where \( \Gamma \) and \( \mathcal{T} \) are as in (A.1.5) and (A.1.6), respectively.

A.6 CONTINUOUS(-GAOIS) SYNTOMIC COHOMOLOGY

We assume \( r \leq p - 2 \) in what follows. For \( i \geq 0 \), we define the continuous syntomic cohomology as follows:

\[
H^i_{\text{cont-syn}}((X, M), \mathcal{Z}_p(r)) := R^i \left( \varprojlim \Gamma \right) (\mathcal{A}(r)(X, M)).
\]

Similarly, we define the continuous-Galois syntomic cohomology as follows:

\[
H^i_{\text{G-cont-syn}}((X, M), \mathcal{F}_p(r)) := R^i \left( \varprojlim \Gamma_{\text{Gal}} \right) (\mathcal{A}(r)(\mathcal{X}, \mathcal{M})).
\]

We put

\[
H^i_{\text{cont-syn}}((X, M), \mathcal{F}_p(r)) := Q_p \otimes \mathbb{Z}_p H^i_{\text{cont-syn}}((X, M), \mathcal{Z}_p(r)).
\]

PROPOSITION A.6.1 Let \( i \geq 0 \) be an integer.

1. Let \( \eta : H^i_{\text{syn}}((X, M), \mathcal{F}_p(r)) \to H^i(X, M, \mathcal{A}(r)) \) be as in Lemma 9.5. Then the kernel of the composite map

\[
H^i_{\text{cont-syn}}((X, M), \mathcal{F}_p(r)) \to H^i_{\text{syn}}((X, M), \mathcal{F}_p(r)) \xrightarrow{\eta} H^i(X, M, \mathcal{A}(r))
\]

agrees with that of the composite map

\[
H^i_{\text{cont-syn}}((X, M), \mathcal{F}_p(r)) \to H^i_{\text{syn}}((X, M), \mathcal{F}_p(r)) \to H^i_{\text{crys}}((X, \mathcal{M})/W)_{\mathbb{Q}_p},
\]

which we denote by \( H^i_{\text{cont-syn}}((X, M), \mathcal{F}_p(r))^{(0)} \), in what follows.

2. If \( K \) is a \( p \)-adic local field (i.e., \( k \) is finite), then we have

\[
H^i_{\text{cont-syn}}((X, M), \mathcal{Z}_p(r)) \xrightarrow{\sim} H^i_{\text{syn}}((X, M), \mathcal{Z}_p(r)).
\]

In particular, we have the following canonical map in this case:

\[
H^i_{\text{syn}}((X, M), \mathcal{Z}_p(r)) \to H^i_{\text{cont-cr}}((X, M)/W).
\]
(3) If \( r \geq d := \dim(X_K) \), then we have
\[
H_\text{crys}^{i}( (X, M) , \mathcal{Z}_p(r) ) \rightarrow H_{\text{cont}}^{i}(X_K, \mathbb{Z}_p(r)),
\]
where \( H_{\text{cont}}^{*}(X_K, \mathbb{Z}_p(r)) \) denotes the continuous étale cohomology \([J]\).

**Proof.** (1) The assertion immediately follows from Lemma 9.5.
(2) There is a short exact sequence analogous to (A.2.3)
\[
0 \rightarrow R^{1}\lim_{\rightarrow} H_{\text{syn}}^{i-1}( (X, M), \mathcal{Z}_p(r) ) \rightarrow H_{\text{cont}}^{i}( (X, M), \mathcal{Z}_p(r) ) \rightarrow H_{\text{syn}}^{i}( (X, M), \mathcal{Z}_p(r) ) \rightarrow 0.
\]
Since \( X \) is proper over \( k \) and \( k \) is finite by assumption, \( H_{\text{syn}}^{i-1}( (X, M), \mathcal{Z}_p(r) ) \) is finite for any \( i \), \( n \geq 1 \) by \([Ts1]\) Proposition 2.4.1. See also the remark in loc. cit., p. 263. The assertion follows from these facts.
(3) Since \( r \geq d \) by assumption, (A.5.7) implies \( \mathcal{Z}_p(r) \simeq \mathcal{T}' R_{\gamma}^{*} \mu_p^{\otimes r} \). Hence the assertion follows from the isomorphisms in \( D^{+}(\text{Mod}(G_{K-\mathbb{Z}/p}) \big) \)
\[
R\Gamma(X_{\mathbb{F}}, \mu_p^{\otimes r}) = R\Gamma(X, R_{\gamma}^{*} \mu_p^{\otimes r}) \rightarrow R\Gamma(Y, \tau^{*} R_{\gamma}^{*} \mu_p^{\otimes r}), \quad \text{(A.6.2)}
\]
where the last isomorphism is a consequence of the proper base-change theorem for the usual étale cohomology. \( \square \)

For \( i \geq 0 \), put
\[
H_{\text{syn}}^{i}( (X, M), \mathcal{Z}_p(r) ) := \text{Ker}(\eta : H_{\text{syn}}^{i}( (X, M), \mathcal{Z}_p(r) ) \rightarrow H_{\text{syn}}^{i}(X_K, \mathbb{Z}_p(r))),
\]
\[
H_{\text{cont}}^{i}( (X, M), \mathcal{Z}_p(r) ) := \text{Ker}(\eta : H_{\text{cont}}^{i}( (X, M), \mathcal{Z}_p(r) ) \rightarrow H_{\text{cont}}^{i}(X_K, \mathbb{Z}_p(r))).
\]
We have \( H_{\text{cont}}^{i}(X_K, \mathbb{Z}_p(r)) = H_{\text{cont}}^{i}(X_K, \mathbb{Q}_p(r)) \) when \( K \) is a \( p \)-adic local field. The following corollary is a consequence of Proposition A.6.1 (1), (2) and the covariant functoriality of Hochschild-Serre spectral sequences in coefficients (see also the diagram in the proof of Theorem A.6.7 below).

**Corollary A.6.3** Assume that \( K \) is a \( p \)-adic local field, and let \( e \) be the composite map
\[
e : H_{\text{syn}}^{i+1}( (X, M), \mathcal{Z}_p(r) ) \rightarrow H_{\text{cont}}^{i+1}(X_K, \mathbb{Q}_p(r)) \rightarrow H^{i}(K, H^{1}(X_{\mathbb{F}}, \mathbb{Q}_p(r))),
\]
where the last map is an edge homomorphism of the Hochschild-Serre spectral sequence (9.0.1). If \( r \geq d \), then \( e \) fits into a commutative diagram
\[
\begin{array}{ccc}
H_{\text{syn}}^{i+1}( (X, M), \mathcal{Z}_p(r) ) & \rightarrow & H_{\text{cont}}^{i+1}( (X, M)/W)_{\mathbb{Q}_p} \\
\downarrow e & & \downarrow \beta^{i,r} \\
H^{i}(K, H^{1}(X_{\mathbb{F}}, \mathbb{Q}_p(r))) & \rightarrow & H^{i}(K, H_{\text{syn}}^{1}(X_{\mathbb{F}}, (X, M)/W)_{\mathbb{Q}_p}),
\end{array}
\]
where the top arrow is obtained from Proposition A.6.1 (1) and (2), the right vertical arrow is the map in Corollary A.4.5 and \( \beta^{i,r} \) is as in the proof of Theorem 9.9.
We next construct a commutative diagram in $\mathbb{Q} \otimes D^+(\text{Mod}(G_K, \mathbb{Z}/p^\bullet))$ assuming $r < d$

$$
\begin{array}{ccc}
\mathbb{Q} \otimes R\Gamma(\overline{Y}, \mathcal{A}_r(r)(\overline{X_\bullet \mathcal{M}_\bullet})) & \xrightarrow{f^r} & \mathbb{Q} \otimes R\Gamma(X_{\overline{\Gamma}}, \mu^{\overline{\bullet}}) \\
\downarrow & & \downarrow g^r \\
\mathbb{Q} \otimes R\Gamma(X_{\overline{\Gamma}}, \mu^{\overline{\bullet}}) & \xrightarrow{h^r} & \mathbb{Q} \otimes (R\Gamma(\overline{Y}, E_{(\overline{X_\bullet \mathcal{M}_\bullet})/W_\bullet}) \otimes \mathbb{Z}/p^*(r-d)'),
\end{array}
$$

which is a key ingredient of the commutative diagram (A) of §9 for the case $r < d$. See (A.5.5) for the isomorphism (1) and the product of crystalline complexes. We define $h^r$ as the composite morphism

$$
\begin{align*}
\mathbb{Q} \otimes R\Gamma(\overline{Y}, \mathcal{A}_r(r)(\overline{X_\bullet \mathcal{M}_\bullet})) & \xrightarrow{(A.5.3)} \mathbb{Q} \otimes R\Gamma(\overline{Y}, E_{(\overline{X_\bullet \mathcal{M}_\bullet})/W_\bullet}) \\
\mathbb{Q} \otimes R\Gamma(\overline{Y}, E_{(\overline{X_\bullet \mathcal{M}_\bullet})/W_\bullet}) & \xrightarrow{(A.5.2)} \mathbb{Q} \otimes \{ R\Gamma(\overline{Y}, E_{(\overline{X_\bullet \mathcal{M}_\bullet})/W_\bullet}) \otimes \mathbb{Z}/p^*(d-r') \otimes \mathbb{Z}/p^*(r-d') \} \\
\mathbb{Q} \otimes \{ R\Gamma(\overline{Y}, E_{(\overline{X_\bullet \mathcal{M}_\bullet})/W_\bullet}) \otimes \mathbb{Z}/p^*(d-r') \otimes \mathbb{Z}/p^*(r-d') \} & \rightarrow \mathbb{Q} \otimes \{ R\Gamma(\overline{Y}, E_{(\overline{X_\bullet \mathcal{M}_\bullet})/W_\bullet}) \otimes \mathbb{Z}/p^*(r-d') \}.
\end{align*}
$$

where the last arrow is induced by the $p^{d-r}$-times of the composite map

$$
\begin{align*}
\mathbb{Q} \otimes \mathbb{Z}/p^*(d-r') & \xrightarrow{(1)} \mathbb{Q} \otimes H^0_{\text{syn}}((\overline{X_\bullet \mathcal{M}_\bullet}), \mathcal{A}_r(d-r)) \\
\mathbb{Q} \otimes H^0_{\text{syn}}((\overline{X_\bullet \mathcal{M}_\bullet}), \mathcal{A}_r(d-r)) & \xrightarrow{(A.5.2)} \mathbb{Q} \otimes H^0_{\text{crys}}((\overline{X_\bullet \mathcal{M}_\bullet})/W_\bullet)
\end{align*}
$$

(see (A.6.6) below for the isomorphism (1)) and the product of crystalline complexes. We define $f^r$ as the morphism induced by (A.5.7) and the isomorphisms in (A.6.2). To define $g^r$, we need an isomorphism

$$
\tilde{j}^d : \mathbb{Q} \otimes R\Gamma(\overline{Y}, \mathcal{A}_r(d)(\overline{X_\bullet \mathcal{M}_\bullet})) \simeq \mathbb{Q} \otimes R\Gamma(X_{\overline{\Gamma}}, \mathbb{Z}/p^*(d'))
$$

induced by $\tilde{n}_d^\circ$ in Theorem A.5.6 (cf. [Ts1] Theorem 3.3.2 (1)). We define

$$
\tilde{h}^d : \mathbb{Q} \otimes R\Gamma(\overline{Y}, \mathcal{A}_r(d)(\overline{X_\bullet \mathcal{M}_\bullet})) \rightarrow \mathbb{Q} \otimes R\Gamma(\overline{Y}, E_{(\overline{X_\bullet \mathcal{M}_\bullet})/W_\bullet})
$$

in the same way as $h^r$ (using (A.5.2) instead of (A.5.3)) and define $g^d := p^d \cdot (\tilde{h}^d \circ (\tilde{j}^d)^{-1})$. Finally we define $g^r$ as the composite of the natural map

$$
R\Gamma(X_{\overline{\Gamma}}, \mu^{\overline{\bullet}}) \rightarrow R\Gamma(X_{\overline{\Gamma}}, \mathbb{Z}/p^*(d')) \otimes \mathbb{Z}/p^*(r-d')
$$

and $g^d \circ \text{id}$. The above diagram is commutative by the definition of $g^r$ and the compatibility of $\eta_\delta^\circ$ with products (cf. [Ts1] §3.1). Now we prove

**Theorem A.6.7** Assume that $K$ is a $p$-adic local field (i.e., $k$ is finite). Then the diagram (9.9.2) commutes for $r < d$. 

*Documenta Mathematica* 18 (2013) 177–247
Proof. We first note the isomorphisms

\[ H^{i+1}_{\text{cont-syn}}((X, M), \mathcal{F}_{p^*}(r)) \xrightarrow{\sim} H^{i+1}_{\text{syn}}((X, M), \mathcal{F}_{p^*}(r)), \]
\[ H^{i+1}_{\text{cont}}(X_K, \mathbb{Q}_p(r)) \xrightarrow{\sim} H^{i+1}(X_K, \mathbb{Q}_p(r)) \]

by the assumption that \( k \) is finite (cf. Proposition A.6.1 (2)). For integers \( m \geq 0 \) and \( s < 0 \), put

\[ H^m_{\text{crys}}((X, M)/W; s) := R^m(\lim_{\to} \Gamma_{g_0}) \{ RF(\mathcal{Y}, \mathbb{Z}/p^{s+1}) \otimes \mathbb{Z}/p^s(s') \}, \]
\[ H^m_{\text{crys}}((X, M)/W; s)_{\mathbb{Q}_p} := \text{Ker}(H^m_{\text{crys}}((X, M)/W; s)_{\mathbb{Q}_p} \to H^m_{\text{crys}}((X, \bar{M})/W)_{\mathbb{Q}_p}(s)). \]

By (A.6.4), there is a commutative diagram

\[
\begin{array}{ccc}
H^{i+1}_{\text{syn}}((\bar{X}, \bar{M}), \mathcal{F}_{p^*}(r)) & \xrightarrow{f^*} & H^{i+1}(X_{\mathbb{Q}_p}(r))_{\mathbb{Q}_p} \\
\downarrow{h^*} & & \downarrow{g^*} \\
H^{i+1}(X_{K'}, \mathbb{Q}_p(r))_{\mathbb{Q}_p} & \xrightarrow{\gamma^*} & H^{i+1}_{\text{cont}}((X, \bar{M})/W)_{\mathbb{Q}_p}(r - d),
\end{array}
\]

where the bottom arrow is the same as \( \beta^{i+1,r} \). By this diagram we obtain the arrows \((g^*)^0\) and \((h^*)^0\) in the following commutative diagram:

\[
\begin{array}{ccc}
H^{i+1}_{\text{syn}}((X, M), \mathcal{F}_{p^*}(r))^0 & \xrightarrow{(f^*)^0} & H^{i+1}_{\text{cont}}(X, \mathbb{Q}_p(r))^0 \\
\downarrow{(h^*)^0} & & \downarrow{(g^*)^0} \\
H^{i+1}(X_{K'}, \mathbb{Q}_p(r))^0 & \xrightarrow{\gamma^*} & H^{i+1}_{\text{crys}}((X, \bar{M})/W; r - d)^0_{\mathbb{Q}_p} \\
\downarrow{\text{edge}} & & \downarrow{\text{edge}} \\
H^1(K, H^i(X_{\mathbb{Q}_p}, \mathbb{Z}/p^*)_{\mathbb{Q}_p}) & \xrightarrow{g^*} & H^1(K, H^i_{\text{crys}}((\bar{X}, \bar{M})/W; r - d))^0_{\mathbb{Q}_p} \\
\downarrow{\text{edge}} & & \downarrow{\text{edge}} \\
H^1(K, H^i(X_{\mathbb{Q}_p}, \mathbb{Q}_p(r))) & \xrightarrow{\beta^i,r} & H^1(K, H^i_{\text{crys}}((\bar{X}, \bar{M})/W)_{\mathbb{Q}_p}(r - d)),
\end{array}
\]

where the top triangle is induced by (A.6.4) and the central square commutes by the functoriality of Hochschild-Serre spectral sequences. The arrows \( c \) and \( c' \) are canonical maps, and the bottom square commutes by the definitions of \( \beta^{i,r} \) and \( g^* \). The map \( c \) is bijective by [J] Theorem 5.15 (c) and the finiteness of \( H^i(X_{\mathbb{Q}_p}, \mathbb{Z}/p^*) \) for \( n \geq 1 \). The bijectivity of \( c' \) is obtained from a similar argument as for the proof of Theorem A.4.2. Moreover it is easy to see that the composite of the left column agrees with the left vertical arrow of (9.9.2), and that the composite of \( h^* \) and the right column agrees with \( \gamma^{i,r} \) in (9.9.2). The commutativity in question follows from these facts. \( \square \)
References


Cycle Class and $p$-adic Regulator


K. Sato
Department of Mathematics
Chuo University
1-13-27 Kasuga, Bunkyo-ku
112-8552 Tokyo, Japan
kanetomo_sato@hotmail.com
Factorial Cluster Algebras

Christof Geiß, Bernard Leclerc, and Jan Schröer

Received: October 13, 2011
Communicated by Edward Frenkel

Abstract. We show that cluster algebras do not contain non-trivial units and that all cluster variables are irreducible elements. Both statements follow from Fomin and Zelevinsky’s Laurent phenomenon. As an application we give a criterion for a cluster algebra to be a factorial algebra. This can be used to construct cluster algebras, which are isomorphic to polynomial rings. We also study various kinds of upper bounds for cluster algebras, and we prove that factorial cluster algebras coincide with their upper bounds.

2010 Mathematics Subject Classification: Primary 13F60; Secondary 13F15, 17B37

Contents

1. Introduction and main results 250
2. Invertible elements in cluster algebras 255
3. Irreducibility of cluster variables 257
4. Factorial cluster algebras 258
5. The divisibility group of a cluster algebra 260
6. Examples of non-factorial cluster algebras 261
7. Examples of factorial cluster algebras 264
8. Applications 271
References 273
1. Introduction and main results

1.1. Introduction. The introduction of cluster algebras by Fomin and Zelevinsky \cite{FZ1} triggered an extensive theory. Most results deal with the combinatorics of seed and quiver mutations, with various categorifications of cluster algebras, and with cluster phenomena occurring in various areas of mathematics, like representation theory of finite-dimensional algebras, quantum groups and Lie theory, Calabi-Yau categories, non-commutative Donaldson-Thomas invariants, Poisson geometry, discrete dynamical systems and algebraic combinatorics.

On the other hand, there are not many results on cluster algebras themselves. As a subalgebra of a field, any cluster algebra $A$ is obviously an integral domain. It is also easy to show that its field of fractions $\text{Frac}(A)$ is isomorphic to a field $K(x_1, \ldots, x_m)$ of rational functions. Several classes of cluster algebras are known to be finitely generated, e.g. acyclic cluster algebras \cite[Corollary 1.21]{BFZ} and also a class of cluster algebras arising from Lie theory \cite[Theorem 3.2]{GLS2}. Berenstein, Fomin and Zelevinsky gave an example of a cluster algebra which is not finitely generated. (One applies \cite[Theorem 1.24]{BFZ} to the example mentioned in \cite[Proposition 1.26]{BFZ}.) Only very little is known on further ring theoretic properties of an arbitrary cluster algebra $A$. Here are some basic questions we would like to address:

- Which elements in $A$ are invertible, irreducible or prime?
- When is $A$ a factorial ring?
- When is $A$ a polynomial ring?

In this paper, we work with cluster algebras of geometric type.

1.2. Definition of a cluster algebra. In this section we repeat Fomin and Zelevinsky’s definition of a cluster algebra.

A matrix $A = (a_{ij}) \in M_{n,n}(\mathbb{Z})$ is skew-symmetrizable (resp. symmetrizable) if there exists a diagonal matrix $D = \text{Diag}(d_1, \ldots, d_n) \in M_{n,n}(\mathbb{Z})$ with positive diagonal entries $d_1, \ldots, d_n$ such that $DA$ is skew-symmetric (resp. symmetric), i.e. $d_ia_{ij} = -d_ja_{ji}$ (resp. $d_ia_{ij} = d_ja_{ji}$) for all $i, j$.

Let $m, n$ and $p$ be integers with

$$m \geq p \geq n \geq 1 \quad \text{and} \quad m > 1.$$ 

Let $B = (b_{ij}) \in M_{m,n}(\mathbb{Z})$ be an $(m \times n)$-matrix with integer entries. By $B^\circ \in M_{n,n}(\mathbb{Z})$ we denote the principal part of $B$, which is obtained from $B$ by deleting the last $m - n$ rows.

Let $\Delta(B)$ be the graph with vertices $1, \ldots, m$ and an edge between $i$ and $j$ provided $b_{ij}$ or $b_{ji}$ is non-zero. We call $B$ connected if the graph $\Delta(B)$ is connected.
Throughout, we assume that $K$ is a field of characteristic 0 or $K = \mathbb{Z}$. Let $\mathcal{F} := K(X_1, \ldots, X_m)$ be the field of rational functions in $m$ variables.

A **seed** of $\mathcal{F}$ is a pair $(x, B)$ such that the following hold:

(i) $B \in M_{m,n}(\mathbb{Z})$,
(ii) $B$ is connected,
(iii) $B^\circ$ is skew-symmetrizable,
(iv) $x = (x_1, \ldots, x_m)$ is an $m$-tuple of elements in $\mathcal{F}$ such that $x_1, \ldots, x_m$ are algebraically independent over $K$.

For a seed $(x, B)$, the matrix $B$ is the **exchange matrix** of $(x, B)$. We say that $B$ has **maximal rank** if $\text{rank}(B) = n$.

Given a seed $(x, B)$ and some $1 \leq k \leq n$ we define the **mutation** of $(x, B)$ at $k$ as

$$
\mu_k(x, B) := (x', B'),
$$

where $B' = (b'_{ij})$ is defined as

$$
b'_{ij} := \begin{cases} 
-b_{ij} & \text{if } i = k \text{ or } j = k, \\
b_{ij} + \frac{|b_{ik}|b_{kj} + b_{ik}|b_{kj}|}{2} & \text{otherwise}, 
\end{cases}
$$

and $x' = (x'_1, \ldots, x'_m)$ is defined as

$$
x'_s := \begin{cases} 
x_k^{-1} \prod_{b_{ik} > 0} x_{ki}^{b_{ik}} + x_k^{-1} \prod_{b_{ik} < 0} x_{ki}^{-b_{ik}} & \text{if } s = k, \\
x_s & \text{otherwise}. 
\end{cases}
$$

The equality

$$
x_kx'_k = \prod_{b_{ik} > 0} x_{ki}^{b_{ik}} + \prod_{b_{ik} < 0} x_{ki}^{-b_{ik}}
$$

is called an **exchange relation**. We write

$$
\mu(x, B)(x_k) := x'_k
$$

and

$$
\mu_k(B) := B'.
$$

It is easy to check that $(x', B')$ is again a seed. Furthermore, we have $\mu_k\mu_k(x, B) = (x, B)$.

Two seeds $(x, B)$ and $(y, C)$ are **mutation equivalent** if there exists a sequence $(i_1, \ldots, i_t)$ with $1 \leq i_j \leq n$ for all $j$ such that

$$
\mu_{i_t} \cdots \mu_{i_2} \mu_{i_1}(x, B) = (y, C).
$$

In this case, we write $(y, C) \sim (x, B)$. This yields an equivalence relation on all seeds of $\mathcal{F}$. (By definition $(x, B)$ is also mutation equivalent to itself.)

For a seed $(x, B)$ of $\mathcal{F}$ let

$$
\mathcal{X}(x, B) := \bigcup_{(y, C) \sim (x, B)} \{y_1, \ldots, y_n\},
$$

**Factorial Cluster Algebras**

251

Documenta Mathematica 18 (2013) 249–274
where the union is over all seeds \((y, C)\) with \((y, C) \sim (x, B)\). By definition, the cluster algebra \(A(x, B)\) associated to \((x, B)\) is the \(L\)-subalgebra of \(F\) generated by \(X_{(x,B)}\), where

\[
L := K[x_{n+1}^{\pm 1}, \ldots, x_p^{\pm 1}, x_{p+1}, \ldots, x_m]
\]

is the localization of the polynomial ring \(K[x_{n+1}, \ldots, x_m]\) at \(x_{n+1} \cdots x_p\). (For \(p = n\) we set \(x_{n+1} \cdots x_p := 1\).) Thus \(A(x, B)\) is the \(K\)-subalgebra of \(F\) generated by

\[
\{x_{n+1}^{\pm 1}, \ldots, x_p^{\pm 1}, x_{p+1}, \ldots, x_m\} \cup X_{(x,B)}.
\]

The elements of \(X_{(x,B)}\) are the cluster variables of \(A(x, B)\).

We call \((y, C)\) a seed of \(A(x, B)\) if \((y, C) \sim (x, B)\). In this case, for any \(1 \leq k \leq n\) we call \((y_k, \mu_{(y,C)}(y_k))\) an exchange pair of \(A(x, B)\). Furthermore, the \(m\)-tuple \(y\) is a cluster of \(A(x, B)\), and monomials of the form \(y_1^{a_1} y_2^{a_2} \cdots y_m^{a_m}\) with \(a_i \geq 0\) for all \(i\) are called cluster monomials of \(A(x, B)\).

Note that for any cluster \(y\) of \(A(x, B)\) we have \(y_i = x_i\) for all \(n + 1 \leq i \leq m\). These \(m - n\) elements are the coefficients of \(A(x, B)\). There are no invertible coefficients if \(p = n\).

Clearly, for any two seeds of the form \((x, B)\) and \((y, B)\), there is an algebra isomorphism \(\eta: A(x, B) \to A(y, B)\) with \(\eta(x_i) = y_i\) for all \(1 \leq i \leq m\), which respects the exchange relations. Furthermore, if \((x, B)\) and \((y, C)\) are mutation equivalent seeds, then \(A(x, B) = A(y, C)\) and we have \(K(x_1, \ldots, x_m) = K(y_1, \ldots, y_m)\).

1.3. TRIVIAL CLUSTER ALGEBRAS AND CONNECTEDNESS OF EXCHANGE MATRICES. Note that we always assume \(m > 1\). For \(m = 1\) we would get the trivial cluster algebra \(A(x, B)\) with exactly two cluster variables, namely \(x_1\) and \(x_1' := \mu_{(x,B)}(x_1) = x_1^{-1}(1 + 1)\). In particular, for \(K \neq \mathbb{Z}\), both cluster variables are invertible in \(A(x, B)\), and \(A(x, B)\) is just the Laurent polynomial ring \(K[x_1^{\pm 1}]\).

Furthermore, for any seed \((x, B)\) of \(F\) the exchange matrix \(B\) is by definition connected. For non-connected \(B\) one could write \(A(x, B)\) as a product \(A(x_1, B_1) \times A(x_2, B_2)\) of smaller cluster algebras and study the factors \(A(x_i, B_i)\) separately. The connectedness assumption also ensures that there are no exchange relations of the form \(x_1 x_1' = 1 + 1\).

1.4. THE LAURENT PHENOMENON. It follows by induction from the exchange relations that for any cluster \(y\) of \(A(x, B)\), any cluster variable \(z\) of \(A(x, B)\) is of the form

\[
z = \frac{f}{g},
\]

where \(f, g \in \mathbb{N}[y_1, \ldots, y_m]\) are integer polynomials in the cluster variables \(y_1, \ldots, y_m\) with non-negative coefficients. For any seed \((x, B)\) of \(F\) let

\[
\mathcal{L}_x := K[x_1^{\pm 1}, \ldots, x_n^{\pm 1}, x_{n+1}^{\pm 1}, \ldots, x_p^{\pm 1}, x_{p+1}, \ldots, x_m]
\]
be the localization of $K[x_1,\ldots,x_m]$ at $x_1 x_2 \cdots x_p$, and let

$$\mathcal{L}_{x, Z} := \mathbb{Z}[x_1^{\pm 1}, \ldots, x_n^{\pm 1}, x_{n+1}, \ldots, x_m]$$

be the localization of $\mathbb{Z}[x_1,\ldots,x_m]$ at $x_1 x_2 \cdots x_n$. We consider $\mathcal{L}_x$ and $\mathcal{L}_{x,Z}$ as subrings of the field $F$. The following remarkable result, known as the \textit{Laurent phenomenon}, is due to Fomin and Zelevinsky and is our key tool to derive some ring theoretic properties of cluster algebras.

\textbf{Theorem 1.1} ([FZ1, Theorem 3.1], [FZ2, Proposition 11.2]). For each seed $(x, B)$ of $F$ we have

$$A(x, B) \subseteq \overline{A}(x, B) := \bigcap_{(y, C) \sim (x, B)} \mathcal{L}_y$$

and

$$X(x, B) \subset \bigcap_{(y, C) \sim (x, B)} \mathcal{L}_{y,Z}.$$ 

The algebra $\overline{A}(x, B)$ is called the \textit{upper cluster algebra} associated to $(x, B)$, compare [BFZ, Section 1].

1.5. Upper bounds. For a seed $(x, B)$ and $1 \leq k \leq n$ let $(x_k, B_k) := \mu_k(x, B)$. Berenstein, Fomin and Zelevinsky [BFZ] called

$$U(x, B) := \mathcal{L}_x \cap \bigcap_{k=1}^n \mathcal{L}_{x_k}$$

the \textit{upper bound} of $A(x, B)$. They prove the following:

\textbf{Theorem 1.2} ([BFZ, Corollary 1.9]). Let $(x, B)$ and $(y, C)$ be mutation equivalent seeds of $F$. If $B$ has maximal rank and $p = m$, then $U(x, B) = U(y, C)$. In particular, we have $\overline{A}(x, B) = U(x, B)$.

For clusters $y$ and $z$ of $A(x, B)$ define

$$U(y, z) := \mathcal{L}_y \cap \mathcal{L}_z.$$ 

1.6. Acyclic cluster algebras. Let $(x, B)$ be a seed of $F$ with $B = (b_{ij})$. Let $\Sigma(B)$ be the quiver with vertices $1,\ldots,n$, and arrows $i \to j$ for all $1 \leq i, j \leq n$ with $b_{ij} > 0$, compare [BFZ, Section 1.4]. So $\Sigma(B)$ encodes the sign-pattern of the principal part $B^\circ$ of $B$.

The seed $(x, B)$ and $B$ are called acyclic if $\Sigma(B)$ does not contain any oriented cycle. The cluster algebra $A(x, B)$ is \textit{acyclic} if there exists an acyclic seed $(y, C)$ with $(y, C) \sim (x, B)$.
1.7. Skew-symmetric exchange matrices and quivers. Let $B = (b_{ij})$ be a matrix in $M_{m,n}(\mathbb{Z})$ such that $B^\tau$ is skew-symmetric. Let $\Gamma(B)$ be the quiver with vertices 1, ..., $m$ and $b_{ij}$ arrows $i \rightarrow j$ if $b_{ij} > 0$, and $-b_{ij}$ arrows $j \rightarrow i$ if $b_{ij} < 0$. Thus given $\Gamma(B)$, we can recover $B$. In the skew-symmetric case one often works with quivers and their mutations instead of exchange matrices.

1.8. Main results. For a ring $R$ with 1, let $R^\times$ be the set of invertible elements in $R$. Non-zero rings without zero divisors are called integral domains. A non-invertible element $a$ in an integral domain $R$ is irreducible if it cannot be written as a product $a = bc$ with $b, c \in R$ both non-invertible. Cluster algebras are integral domains, since they are by definition subrings of fields.

Theorem 1.3. For any seed $(x, B)$ of $\mathcal{F}$ the following hold:

(i) We have $A(x, B) = \{ \lambda x_{n+1}^{a_{n+1}} \cdots x_p^{a_p} \mid \lambda \in K^\times, a_i \in \mathbb{Z} \}$.

(ii) Any cluster variable in $A(x, B)$ is irreducible.

For elements $a, b$ in an integral domain $R$ we write $a | b$ if there exists some $c \in R$ with $b = ac$. A non-invertible element $a$ in a commutative ring $R$ is prime if whenever $a | bc$ for some $b, c \in R$, then $a | b$ or $a | c$. Every prime element is irreducible, but the converse is not true in general. Non-zero elements $a, b \in R$ are associate if there is some unit $c \in R^\times$ with $a = bc$. An integral domain $R$ is factorial if the following hold:

(i) Every non-zero non-invertible element $r \in R$ can be written as a product $r = a_1 \cdots a_s$ of irreducible elements $a_i \in R$.

(ii) If $a_1 \cdots a_s = b_1 \cdots b_t$ with $a_i, b_j \in R$ irreducible for all $i$ and $j$, then $s = t$ and there is a bijection $\pi: \{1, \ldots, s\} \rightarrow \{1, \ldots, t\}$ such that $a_i$ and $b_{\pi(i)}$ are associate for all $1 \leq i \leq s$.

For example, any polynomial ring is factorial. In a factorial ring, all irreducible elements are prime.

Two clusters $y$ and $z$ of a cluster algebra $A(x, B)$ are disjoint if $\{y_1, \ldots, y_n\} \cap \{z_1, \ldots, z_n\} = \emptyset$.

The next result gives a useful criterion when a cluster algebra is a factorial ring.

Theorem 1.4. Let $y$ and $z$ be disjoint clusters of $A(x, B)$. If there is a subalgebra $U$ of $A(x, B)$, such that $U$ is factorial and

$\{y_1, \ldots, y_n, z_1, \ldots, z_n, x_{n+1}^{\pm 1}, \ldots, x_p^{\pm 1}, x_{p+1}, \ldots, x_m\} \subset U$,

then

$U = A(x, B) = U(y, z)$.

In particular, $A(x, B)$ is factorial and all cluster variables are prime.

We obtain the following corollary on upper bounds of factorial cluster algebras.
Corollary 1.5. Assume that $\mathcal{A}(x, B)$ is factorial.

(i) If $y$ and $z$ are disjoint clusters of $\mathcal{A}(x, B)$, then $\mathcal{A}(x, B) = U(y, z)$.

(ii) For any $(y, C) \sim (x, B)$ we have $\mathcal{A}(x, B) = U(y, C)$.

In Section 7 we apply the above results to show that many cluster algebras are polynomial rings. In Section 8 we discuss some further applications concerning the dual of Lusztig’s semicanonical basis and monoidal categorifications of cluster algebras.

1.9. Factoriality and maximal rank. In Section 6.1 we give examples of cluster algebras $\mathcal{A}(x, B)$, which are not factorial. In these examples, $B$ does not have maximal rank.

After we presented our results at the Abel Symposium in Balestrand in June 2011, Zelevinsky asked the following question:

Problem 1.6. Suppose $(x, B)$ is a seed of $\mathcal{F}$ such that $B$ has maximal rank. Does it follow that $\mathcal{A}(x, B)$ is factorial?

After we circulated a first version of this article, Philipp Lampe [La] discovered an example of a non-factorial cluster algebra $\mathcal{A}(x, B)$ with $B$ having maximal rank. With his permission, we explain a generalization of his example in Section 6.2.

2. Invertible elements in cluster algebras

In this section we prove Theorem 1.3(i), classifying the invertible elements of cluster algebras.

The following lemma is straightforward and well-known.

Lemma 2.1. For any seed $(x, B)$ of $\mathcal{F}$ we have

$$L_x^\times = \{\lambda x_1^{a_1} \cdots x_p^{a_p} | \lambda \in K^\times, a_i \in \mathbb{Z}\}.$$  

Theorem 2.2. For any seed $(x, B)$ of $\mathcal{F}$ we have

$$\mathcal{A}(x, B)^\times = \{\lambda x_{n+1}^{a_{n+1}} \cdots x_p^{a_p} | \lambda \in K^\times, a_i \in \mathbb{Z}\}.$$  

Proof. Let $u$ be an invertible element in $\mathcal{A}(x, B)$, and let $(y, C)$ be any seed of $\mathcal{A}(x, B)$. By the Laurent phenomenon Theorem 1.1 we know that $\mathcal{A}(x, B) \subseteq L_y$. It follows that $u$ is also invertible in $L_y$. Thus by Lemma 2.1 there are $a_1, \ldots, a_p \in \mathbb{Z}$ and $\lambda \in K^\times$ such that $u = \lambda M$, where

$$M = y_1^{a_1} \cdots y_k^{a_k} \cdots y_p^{a_p}.$$  

If all $a_i$ with $1 \leq i \leq n$ are zero, we are done. To get a contradiction, assume that there is some $1 \leq k \leq n$ with $a_k \neq 0$. 

Documenta Mathematica 18 (2013) 249–274
Let $y_k^*: =\mu_{(y,C)}(y_k)$. Again the Laurent phenomenon yields $b_1, \ldots, b_p \in \mathbb{Z}$ and $\nu \in K^\times$ such that

$$u = \nu y_1^{b_1} \cdots y_{k-1}^{b_{k-1}} (y_k^{*})^{b_k} y_{k+1}^{b_{k+1}} \cdots y_p^{b_p}.$$ 

Without loss of generality let $b_k \geq 0$. (Otherwise we can work with $u^{-1}$ instead of $u$.) If $b_k = 0$, we get

$$\lambda y_1^{a_1} \cdots y_k^{a_k} \cdots y_p^{a_p} = \nu y_1^{b_1} \cdots y_{k-1}^{b_{k-1}} y_{k+1}^{b_{k+1}} \cdots y_p^{b_p},$$

where $\lambda, \nu \in K^\times$. This is a contradiction, because $a_k \neq 0$ and $y_1, \ldots, y_m$ are algebraically independent, and therefore Laurent monomials in $y_1, \ldots, y_m$ are linearly independent in $\mathcal{F}$.

Next, assume that $b_k > 0$. By definition we have

$$y_k^* = M_1 + M_2$$

with

$$M_1 = y_k^{-1} \prod_{c_{ik} > 0} y_i^{c_{ik}} \quad \text{and} \quad M_2 = y_k^{-1} \prod_{c_{ik} < 0} y_i^{-c_{ik}},$$

where the products run over the positive, respectively negative, entries in the $k$th column of the matrix $C$.

Thus we get an equality of the form

$$(2) \quad u = \lambda M = \nu (y_1^{b_1} \cdots y_{k-1}^{b_{k-1}}) (M_1 + M_2) y_k^{b_k} (y_{k+1}^{b_{k+1}} \cdots y_p^{b_p}).$$

We know that $M_1 \neq M_2$. (Here we use that $m > 1$ and that exchange matrices are by definition connected. Otherwise, one could get exchange relations of the form $x_k x_k^i = 1 + 1$.) Thus the right-hand side of Equation $(2)$ is a non-trivial linear combination of $b_k + 1 \geq 2$ pairwise different Laurent monomials in $y_1, \ldots, y_m$. This is again a contradiction, since $y_1, \ldots, y_m$ are algebraically independent. \hfill \Box

**Corollary 2.3.** For any seed $(x, B)$ of $\mathcal{F}$ the following hold:

(i) Let $y$ and $z$ be non-zero elements in $\mathcal{A}(x, B)$. Then $y$ and $z$ are associate if and only if there exist $a_{n+1}, \ldots, a_p \in \mathbb{Z}$ and $\lambda \in K^\times$ with

$$y = \lambda x_1^{a_{n+1}} \cdots x_p^{a_p} z.$$ 

(ii) Let $y$ and $z$ be cluster variables of $\mathcal{A}(x, B)$. Then $y$ and $z$ are associate if and only if $y = z$.

**Proof.** Part (i) follows directly from Theorem 2.2. To prove (ii), let $y$ and $z$ be clusters of $\mathcal{A}(x, B)$. Assume $y_i$ and $z_j$ are associate for some $1 \leq i, j \leq n$. By (i) there are $a_{n+1}, \ldots, a_p \in \mathbb{Z}$ and $\lambda \in K^\times$ with $y_i = \lambda x_{n+1}^{a_{n+1}} \cdots x_p^{a_p} z_j$. By
Theorem 1.1 we know that there exist $b_1, \ldots, b_n \in \mathbb{Z}$ and a polynomial $f$ in $\mathbb{Z}[z_1, \ldots, z_m]$ with

$$y_i = \frac{f}{z_1^{b_1} \cdots z_n^{b_n}}$$

and $f$ is not divisible by any $z_1, \ldots, z_n$. The polynomial $f$ and $b_1, \ldots, b_n$ are uniquely determined by $y_i$. It follows that $\lambda \in \mathbb{Z}$ and $a_{n+1}, \ldots, a_p \geq 0$. But we also have $z_j = \lambda^{-1}x_{n+1}^{-a_{n+1}} \cdots x_p^{-a_p}y_i$. Reversing the role of $y_i$ and $z_j$ we get $-a_{n+1}, \ldots, -a_p \geq 0$ and $\lambda = \frac{y_i}{z_j}$. This implies $y_i = z_j$ or $-y_i = z_j$. By the remark at the beginning of Section 1.3 we know that $z_j = \frac{f}{g}$ for some $f, g \in \mathbb{N}[y_1, \ldots, y_m]$. Assume that $-y_i = z_j$. We get $z_j = -y_i = f/g$ and therefore $f + y_ig = 0$. This is a contradiction to the algebraic independence of $y_1, \ldots, y_m$. Thus we proved (ii).

We thank Giovanni Cerulli Irelli for helping us with the final step of the proof of Corollary 2.3(ii).

Two clusters $y$ and $z$ of a cluster algebra $A(x, B)$ are non-associate if there are no $1 \leq i, j \leq n$ such that $y_i$ and $z_j$ are associate.

**Corollary 2.4.** For clusters $y$ and $z$ of $A(x, B)$ the following are equivalent:

(i) The clusters $y$ and $z$ are non-associate.

(ii) The clusters $y$ and $z$ are disjoint.

**Proof.** Non-associate clusters are obviously disjoint. The converse follows directly from Corollary 2.3(ii).

3. Irreducibility of cluster variables

In this section we prove Theorem 1.3(ii). The proof is very similar to the proof of Theorem 2.2.

**Theorem 3.1.** Let $(x, B)$ be a seed of $F$. Then any cluster variable in $A(x, B)$ is irreducible.

**Proof.** Let $(y, C)$ be any seed of $A(x, B)$. We know from Theorem 2.2 that the cluster variables of $A(x, B)$ are non-invertible in $A(x, B)$.

Assume that $y_k$ is not irreducible for some $1 \leq k \leq n$. Thus $y_k = y_k' y_k''$ for some non-invertible elements $y_k'$ and $y_k''$ in $A(x, B)$. Since $y_k$ is invertible in $L_y$, we know that $y_k'$ and $y_k''$ are both invertible in $L_y$. Thus by Lemma 2.1 there are $a_s, b_i \in \mathbb{Z}$ and $\lambda, \lambda' \in K^\times$ with

$$y_k' = \lambda' y_1^{a_1} \cdots y_s^{a_s} \cdots y_p^{a_p}$$

and

$$y_k'' = \lambda'' y_1^{b_1} \cdots y_s^{b_s} \cdots y_p^{b_p}.$$ 

Since $y_k = y_k' y_k''$, we get $a_s + b_s = 0$ for all $s \neq k$ and $a_k + b_k = 1.$

Documenta Mathematica 18 (2013) 249–274
Assume that \( a_s = 0 \) for all \( 1 \leq s \leq n \) with \( s \neq k \). Then \( y'_k = \lambda' y_k^{a_k} y_{a_k+1}^{a+1} \cdots y_p^{a_p} \) and \( y''_k = \lambda'' y_k^{b_k} y_{b_k+1}^{b+1} \cdots y_p^{b_p} \). If \( a_k \leq 0 \), then \( y'_k \) is invertible in \( \mathcal{A}(x, B) \), and if \( a_k > 0 \), then \( y''_k \) is invertible in \( \mathcal{A}(x, B) \). In both cases we get a contradiction.

Next assume \( a_s \neq 0 \) for some \( 1 \leq s \leq n \) with \( s \neq k \). Let \( y^*_s := \mu_{(y, C)}(y_s) \). Thus we have

\[
y^*_s = M_1 + M_2
\]

with

\[
M_1 = y_s^{-1} \prod_{c_i > 0} y_i^{c_i} \quad \text{and} \quad M_2 = y_s^{-1} \prod_{c_i < 0} y_i^{-c_i},
\]

where the products run over the positive, respectively negative, entries in the \( s \)th column of the matrix \( C \).

Since \( s \neq k \), we see that \( y_k \) and therefore also \( y'_k \) and \( y''_k \) are invertible in \( \mathcal{L}_{\mu_{s}, (y, C)} \). Thus by Lemma 2.1 there are \( c_i, d_i \in \mathbb{Z} \) and \( \nu', \nu'' \in K^\times \) with

\[
y'_k = \nu' y_1^{c_1} \cdots y_{s-1}^{c_{s-1}} (y_s)^{c_s} y_{s+1}^{c_{s+1}} \cdots y_p^{c_p}
\]

and

\[
y''_k = \nu'' y_1^{d_1} \cdots y_{s-1}^{d_{s-1}} (y_s)^{d_s} y_{s+1}^{d_{s+1}} \cdots y_p^{d_p}.
\]

Note that \( c_s + d_s = 0 \). Without loss of generality we assume that \( c_s \geq 0 \). (If \( c_s < 0 \), we continue to work with \( y''_k \) instead of \( y'_k \).) If \( c_s = 0 \), we get

\[
y'_k = \lambda' y_1^{c_1} \cdots y_{s-1}^{c_{s-1}} y_s^{c_s} y_{s+1}^{c_{s+1}} \cdots y_p^{c_p} = \nu' y_1^{c_1} \cdots y_s^0 \cdots y_p^{c_p}.
\]

This is a contradiction, since \( a_s \neq 0 \) and \( y_1, \ldots, y_m \) are algebraically independent. If \( c_s > 0 \), then

\[
y'_k = \lambda' y_1^{c_1} \cdots y_{s-1}^{c_{s-1}} y_s^{c_s} y_{s+1}^{c_{s+1}} \cdots y_p^{c_p} = \nu' y_1^{c_1} \cdots y_{s-1}^{c_{s-1}} (y_s)^{c_s} y_{s+1}^{c_{s+1}} \cdots y_p^{c_p} = \nu' y_1^{c_1} \cdots y_{s-1}^{c_{s-1}} (M_1 + M_2)^{c_s} y_{s+1}^{c_{s+1}} \cdots y_p^{c_p}.
\]

We know that \( M_1 \neq M_2 \). Thus the Laurent monomial \( y_k \) is a non-trivial linear combination of \( c_s + 1 \geq 2 \) pairwise different Laurent monomials in \( y_1, \ldots, y_m \), a contradiction.

Note that the coefficients \( x_{p+1}, \ldots, x_m \) of \( \mathcal{A}(x, B) \) are obviously irreducible in \( \mathcal{L}_x \). Since \( \mathcal{A}(x, B) \subseteq \mathcal{L}_x \), they are also irreducible in \( \mathcal{A}(x, B) \).

4. Factorial cluster algebras

4.1. A factoriality criterion. This section contains the proofs of Theorem 4.4 and Corollary 4.5.

Theorem 4.1. Let \( y \) and \( z \) be disjoint clusters of \( \mathcal{A}(x, B) \), and let \( U \) be a factorial subalgebra of \( \mathcal{A}(x, B) \) such that

\[
\{y_1, \ldots, y_n, z_1, \ldots, z_m, x_{n+1}^{\pm1}, \ldots, x_p^{\pm1}, x_{p+1}, \ldots, x_m \} \subset U.
\]
Then we have

\[ U = \mathcal{A}(x, B) = U(y, z). \]

Proof. Let \( u \in U(y, z) = \mathcal{L}_y \cap \mathcal{L}_z \). Thus we have

\[ u = \frac{f}{y_1^{a_1} y_2^{a_2} \cdots y_p^{a_p}} = \frac{g}{z_1^{b_1} z_2^{b_2} \cdots z_p^{b_p}}, \]

where \( f \) is a polynomial in \( y_1, \ldots, y_m \), and \( g \) is a polynomial in \( z_1, \ldots, z_m \), and \( a_i, b_i \geq 0 \) for all \( 1 \leq i \leq p \). By the Laurent phenomenon it is enough to show that \( u \in U \).

Since \( y_i, z_i \in U \) for all \( 1 \leq i \leq m \), we get the identity

\[ f z_1^{b_1} z_2^{b_2} \cdots z_n^{b_n} \cdots z_{n+1}^{b_{n+1}} \cdots z_p^{b_p} = g y_1^{a_1} y_2^{a_2} \cdots y_n^{a_n} y_{n+1}^{a_{n+1}} \cdots y_p^{a_p} \]

in \( U \).

By Theorem 4.1 the cluster variables \( y_i \) and \( z_i \) with \( 1 \leq i \leq n \) are irreducible in \( \mathcal{A}(x, B) \). In particular, they are irreducible in the subalgebra \( U \) of \( \mathcal{A}(x, B) \). The elements \( y_n^{a_n+1} \cdots y_p^{a_p} \) and \( z_{n+1}^{b_{n+1}} \cdots z_p^{b_p} \) are units in \( U \). (Recall that \( x_i = y_i = z_i \) for all \( n+1 \leq i \leq m \).)

The clusters \( y \) and \( z \) are disjoint. Now Corollary 2.4 implies that the elements \( y_i \) and \( z_j \) are non-associate for all \( 1 \leq i, j \leq n \). Thus, by the factoriality of \( U \), the monomial \( y_1^{a_1} y_2^{a_2} \cdots y_n^{a_n} \) divides \( f \) in \( U \). In other words there is some \( h \in U \) with \( f = h y_1^{a_1} y_2^{a_2} \cdots y_n^{a_n} \). It follows that

\[ u = \frac{f}{y_1^{a_1} y_2^{a_2} \cdots y_p^{a_p}} = \frac{h y_1^{a_1} y_2^{a_2} \cdots y_n^{a_n}}{y_1^{a_1} y_2^{a_2} \cdots y_p^{a_p}} = \frac{h}{y_n^{a_n+1} \cdots y_p^{a_p}} = hy_n^{a_n+1} \cdots y_p^{a_p}. \]

Since \( h \in U \) and \( y_n^{a_n+1} \cdots y_p^{a_p} \in U \), we get \( u \in U \). This finishes the proof. \( \square \)

Corollary 4.2. Assume that \( \mathcal{A}(x, B) \) is factorial.

(i) If \( y \) and \( z \) are disjoint clusters of \( \mathcal{A}(x, B) \), then \( \mathcal{A}(x, B) = U(y, z) \).

(ii) For any \( (y, C) \sim (x, B) \) we have \( \mathcal{A}(x, B) = U(y, C) \).

Proof. Part (i) follows directly from Theorem 4.1. To prove part (ii), assume \( (y, C) \sim (x, B) \) and let \( u \in U(y, C) \). For \( 1 \leq k \leq n \) let \( (y_k, C_k) := \mu_k(y, C) \) and \( y_k^* := \mu_{(y, C)}(y_k) \). We get

\[ u = \frac{f}{y_1^{a_1} \cdots y_k^{a_k} \cdots y_p^{a_p}} = \frac{f_k}{y_1^{b_1} \cdots (y_k^*)^{b_k} \cdots y_p^{b_p}} \]

for a polynomial \( f \) in \( y_1, \ldots, y_k, \ldots, y_m \), a polynomial \( f_k \) in \( y_1, \ldots, y_k^*, \ldots, y_m \), and \( a_i, b_i \geq 0 \). This yields an equality

\[ f y_1^{b_1} \cdots (y_k^*)^{b_k} \cdots y_p^{b_p} = f_k y_1^{a_1} \cdots y_k^{a_k} \cdots y_p^{a_p} \]

in \( \mathcal{A}(x, B) \). Now we argue similarly as in the proof of Theorem 4.1. The cluster variables \( y_1, \ldots, y_n, y_1^*, \ldots, y_n^* \) are obviously pairwise different. Now
For \( k \leq 1 \) and \( y \) that follows that \( y \in A \) the existence of disjoint clusters in \( y \). Since \( u \in A(x, B) \), Equation 1 implies that \( y^{y_0} \) divides \( f \) in \( A(x, B) \). Since this holds for all \( 1 \leq k \leq n \), we get that \( y_{y_0}^1 \cdots y_{y_0}^{n_0} \) divides \( f \) in \( A(x, B) \). It follows that \( u \in A(x, B) \).

4.2. EXISTENCE OF DISJOINT CLUSTERS. One assumption of Theorem 4.1 is the existence of disjoint clusters in \( A(x, B) \). We can prove this under a mild assumption. But it should be true in general.

**Proposition 4.3.** Assume that the cluster monomials of \( A(x, B) \) are linearly independent. Let \( (y, C) \) be a seed of \( A(x, B) \), and let 

\[
(z, D) := \mu_n \cdots \mu_2 \mu_1 (y, C).
\]

Then the clusters \( y \) and \( z \) are disjoint.

**Proof.** Set \( (y[0], C[0]) := (y, C) \), and for \( 1 \leq k \leq n \) let \( (y[k], C[k]) := \mu_k (y[k-1], C[k-1]) \) and \( (y_1[k], \ldots, y_m[k]) := y[k] \). We claim that 

\[
\{y_1[k], \ldots, y_k[k]\} \cap \{y_1, \ldots, y_n\} = \emptyset.
\]

For \( k = 1 \) this is straightforward. Thus let \( k \geq 2 \) and assume that our claim is true for \( k - 1 \). To get a contradiction, assume that \( y_k[k] = y_j \) for some \( 1 \leq j \leq n \). By the induction assumption we know that \( \{y_1[k], \ldots, y_{k-1}[k]\} \cap \{y_1, \ldots, y_n\} = \emptyset \), since \( y_i[k] = y_i[k-1] \) for all \( 1 \leq i \leq k - 1 \).

We have \( y[k] = (y_1[k], \ldots, y_k[k], y_{k+1}, \ldots, y_m) \).

Since \( y_1[k], \ldots, y_k[k], y_{k+1}, \ldots, y_m \) are algebraically independent and \( y_k[k] \neq y_k \), we get \( 1 \leq j \leq k - 1 \). Since \( (y[j], C[j]) = \mu_j (y[j-1], C[j-1]) \), it follows that \( (y[j-1], y_j[j]) \) is an exchange pair of \( A(x, B) \). Next, observe that \( y_k[k] = y_j = y_j[j-1] \) and \( y_j[k] = y_j[j] \). Thus \( y_j[j-1] \) and \( y_j[j] \) are both contained in \( \{y_1[k], \ldots, y_m[k]\} \), and therefore \( y_j[j-1]y_j[j] \) is a cluster monomial. The corresponding exchange relation gives a contradiction to the linear independence of cluster monomials.

Fomin and Zelevinsky [FZ3, Conjecture 4.16] conjecture that the cluster monomials of \( A(x, B) \) are always linearly independent. Under the assumptions that \( B \) has maximal rank and that \( B^2 \) is skew-symmetric, the conjecture follows from [DWZ, Theorem 1.7].

5. THE DIVISIBILITY GROUP OF A CLUSTER ALGEBRA

Let \( R \) be an integral domain, and let \( \text{Frac}(R) \) be the field of fractions of \( R \). Set \( \text{Frac}(R)^* := \text{Frac}(R) \setminus \{0\} \). The abelian group 

\[
G(R) := (\text{Frac}(R)^*/R^x, \cdot)
\]

is the divisibility group of \( R \).
For $g, h \in \text{Frac}(R)^*$ let $g \leq h$ provided $hg^{-1} \in R$. This relation is reflexive and transitive and it induces a partial ordering on $G(R)$.

Let $I$ be a set. The abelian group $(\mathbb{Z}(I), +)$ is equipped with the following partial ordering: We set $(x_i)_{i \in I} \leq (y_i)_{i \in I}$ if $x_i \leq y_i$ for all $i$. (By definition, the elements in $\mathbb{Z}(I)$ are tuples $(x_i)_{i \in I}$ of integers $x_i$ such that only finitely many $x_i$ are non-zero.)

There is the following well-known criterion for the factoriality of $R$, see for example [C, Section 2].

**Proposition 5.1.** For an integral domain $R$ the following are equivalent:

(i) $R$ is factorial.

(ii) There is a set $I$ and a group isomorphism

$$\phi: G(R) \to \mathbb{Z}(I)$$

such that for all $g, h \in G(R)$ we have $g \leq h$ if and only if $\phi(g) \leq \phi(h)$.

Not all cluster algebras $A(\mathbf{x}, B)$ are factorial, but at least one part of the above factoriality criterion is satisfied:

**Proposition 5.2.** For any seed $(\mathbf{x}, B)$ of $\mathcal{F}$ the divisibility group $G(A(\mathbf{x}, B))$ is isomorphic to $\mathbb{Z}(I)$, where

$I := \{ f \in K[x_1, \ldots, x_m] \mid f \text{ is irreducible and } f \not\equiv x_i \text{ for } n+1 \leq i \leq p \}/K^*$

is the set of irreducible polynomials unequal to any $x_{n+1}, \ldots, x_p$ in $K[x_1, \ldots, x_m]$ up to non-zero scalar multiples.

**Proof.** By the Laurent phenomenon and the definition of a seed we get

$$\text{Frac}(A(\mathbf{x}, B)) = \text{Frac}(\mathcal{Z}_\mathbf{x}) = K(x_1, \ldots, x_m).$$

Furthermore, by Theorem 2.2 we have

$$A(\mathbf{x}, B)^\times = \{ \lambda x_{n+1}^{a_{n+1}} \cdots x_p^{a_p} \mid \lambda \in K^*, a_i \in \mathbb{Z} \}.$$  

Any element in $K(x_1, \ldots, x_m)$ is of the form $f_1 \cdots f_m g_1^{-1} \cdots g_i^{-1}$ with $f_j, g_j$ irreducible in $K[x_1, \ldots, x_m]$. Using that the polynomial ring $K[x_1, \ldots, x_m]$ is factorial, and working modulo $A(\mathbf{x}, B)^\times$ yields the result. \qed

6. Examples of non-factorial cluster algebras

6.1. For a matrix $A \in M_{m,n}(\mathbb{Z})$ and $1 \leq i \leq n$ let $c_i(A)$ be the $i$th column of $A$.

**Proposition 6.1.** Let $(\mathbf{x}, B)$ be a seed of $\mathcal{F}$. Assume that $c_k(B) = c_s(B)$ or $c_k(B) = -c_s(B)$ for some $k \neq s$ with $b_{ks} = 0$. Then $A(\mathbf{x}, B)$ is not factorial.
Proof. Define \((y, C) := \mu_k(x, B)\) and \((z, D) := \mu_s(y, C)\). We get
\[ y_k = z_k = x_k^{-1}(M_1 + M_2), \]
where
\[ M_1 := \prod_{b_{ik} > 0} x_i^{b_{ik}} \quad \text{and} \quad M_2 := \prod_{b_{ik} < 0} x_i^{-b_{ik}}. \]
By the mutation rule, we have \(c_k(C) = -c_k(B)\), and since \(b_{ks} = 0\), we get \(c_s(C) = c_s(B)\). Since \(c_s(B) = c_k(B)\) or \(c_s(B) = -c_k(B)\), this implies that
\[ z_s = x_s^{-1}(M_1 + M_2). \]
The cluster variables \(x_k, x_s, z_k, z_s\) are pairwise different. Thus they are pairwise non-associate by Corollary 2.3(ii), and by Theorem 3.1 they are irreducible in \(A(x, B)\). Obviously, we have
\[ x_kz_k = x_sz_s. \]
Thus \(A(x, B)\) is not factorial. 

To give a concrete example of a cluster algebra, which is not factorial, assume \(m = n = p = 3\), and let \(B \in M_{m,n}(\mathbb{Z})\) be the matrix
\[
B = \begin{pmatrix}
0 & -1 & 0 \\
1 & 0 & -1 \\
0 & 1 & 0 
\end{pmatrix}.
\]
The matrix \(B\) obviously satisfies the assumptions of Proposition 6.1. Note that \(B = B^\circ\) is skew-symmetric, and that \(\Gamma(B)\) is the quiver
\[
3 \rightarrow 2 \rightarrow 1.
\]
Thus \(A(x, B)\) is a cluster algebra of Dynkin type \(A_3\). (Cluster algebras with finitely many cluster variables are classified via Dynkin types, for details see [FZ2].)

Define \((z, D) := \mu_3 \mu_1(x, B)\). We get \(z_1 = x_1^{-1}(1 + x_2)\), \(z_3 = x_3^{-1}(1 + x_2)\) and therefore \(x_1z_1 = x_3z_3\).

Clearly, the cluster variables \(x_1, x_3, z_1, z_3\) are pairwise different. Using Corollary 2.3(ii) we get that \(x_1, x_3, z_1, z_3\) are pairwise non-associate, and by Theorem 3.1 they are irreducible. Thus \(A(x, B)\) is not factorial.

6.2. The next example is due to Philipp Lampe. It gives a negative answer to Zelevinsky’s Question 1.6.

Proposition 6.2 (La). Let \(K = \mathbb{C}\), \(m = n = 2\) and
\[
B = \begin{pmatrix}
0 & -2 \\
2 & 0 
\end{pmatrix}.
\]
Then \(A(x, B)\) is not factorial.
The proof of the following result is a straightforward generalization of Lampe’s proof of Proposition 6.3.

**Proposition 6.3.** Let $(x, B)$ be a seed of $\mathcal{F}$. Assume that there exists some $1 \leq k \leq n$ such that the polynomial $X^d + Y^d$ is not irreducible in $K[X,Y]$, where $d := \gcd(b_{1k}, \ldots, b_{nk})$ is the greatest common divisor of $b_{1k}, \ldots, b_{nk}$. Then $\mathcal{A}(x, B)$ is not factorial.

**Proof.** Let $X^d + Y^d = f_1 \cdots f_t$, where the $f_j$ are irreducible polynomials in $K[X,Y]$. Since $X^d + Y^d$ is not irreducible in $K[X,Y]$, we have $t \geq 2$. Let $y_k := \mu_{(x,B)}(x_k)$. The corresponding exchange relation is

$$x_k y_k = \prod_{b_{ik} > 0} x_i^{b_{ik}} + \prod_{b_{ik} < 0} x_i^{-b_{ik}} = M^d + N^d = \prod_{j=1}^t f_j(M,N),$$

where

$$M := \prod_{b_{ik} > 0} x_i^{b_{ik}/d} \quad \text{and} \quad N := \prod_{b_{ik} < 0} x_i^{-b_{ik}/d}.$$

Clearly, each $f_j(M,N)$ is contained in $\mathcal{A}(x, B)$. To get a contradiction, assume that $\mathcal{A}(x, B)$ is factorial. By Theorem 2.2 none of the elements $f_j(M,N)$ is invertible in $\mathcal{A}(x, B)$. Since $\mathcal{A}(x, B)$ is factorial, each $f_j(M,N)$ is equal to a product $f_{1j} \cdots f_{aj,j}$, where the $f_{ij}$ are irreducible in $\mathcal{A}(x, B)$ and $a_j \geq 1$. By Theorem 3.1 the cluster variables $x_k$ and $y_k$ are irreducible in $\mathcal{A}(x, B)$. It follows that $a_1 + \cdots + a_t = 2$, since $\mathcal{A}(x, B)$ is factorial. This implies $t = 2$ and $a_1 = a_2 = 1$. In particular, $f_1(M,N)$ and $f_2(M,N)$ are irreducible in $\mathcal{A}(x, B)$, and we have $x_k y_k = f_1(M,N)f_2(M,N)$. For $j = 1, 2$ the elements $x_k$ and $f_j(M,N)$ cannot be associate, since $f_j(M,N)$ is just a $K$-linear combination of monomials in $\{x_1, \ldots, x_m\} \setminus \{x_k\}$. (Here we use Corollary 2.3(i) and the fact that $b_{kk} = 0$.) This is a contradiction to the factoriality of $\mathcal{A}(x, B)$. $\square$

Note that a polynomial of the form $X^d + Y^d$ is irreducible if and only if $X^d + 1$ is irreducible.

**Corollary 6.4.** Let $K = \mathbb{C}$, $m = n = 2$ and

$$B = \begin{pmatrix} 0 & -c \\ d & 0 \end{pmatrix}$$

with $c \geq 1$ and $d \geq 2$. Then $\mathcal{A}(x, B)$ is not factorial.

**Proof.** For $k = 1$ the assumptions of Proposition 6.3 hold. (We have $\gcd(0, d) = d$, and the polynomial $X^d + 1$ is not irreducible in $\mathbb{C}[X]$.) $\square$

**Corollary 6.5.** Let $m = n = 2$ and

$$B = \begin{pmatrix} 0 & -c \\ d & 0 \end{pmatrix}$$

with $c \geq 1$ and $d \geq 3$ an odd number. Then $\mathcal{A}(x, B)$ is not factorial.
Proof. For \( k = 1 \) the assumptions of Proposition 6.3 hold. (We have \( \gcd(0,d) = d \), and for odd \( d \) we have
\[
X^d + 1 = (X + 1) \left( \sum_{j=0}^{d-1} (-1)^j X^j \right).
\]
Thus \( X^d + 1 \) is not irreducible in \( K[X] \). \( \square \)

7. Examples of factorial cluster algebras

7.1. Cluster algebras of Dynkin type \( A \) as polynomial rings. Assume \( m = n + 1 = p + 1 \), and let \( B \in M_{m,n}(\mathbb{Z}) \) be the matrix
\[
B = \begin{pmatrix}
0 & -1 & 0 & -1 & \\
1 & 0 & -1 & \\
& 1 & 0 & -1 & \\
& & \\
& & & & \\
& & & & \\
& & & & \\
& & & & 1
\end{pmatrix}.
\]
Obviously, \( B^o \) is skew-symmetric, \( \Gamma(B) \) is the quiver
\[
\begin{aligned}
m &\longrightarrow \cdots \longrightarrow 2 \longrightarrow 1,
\end{aligned}
\]
and \( \mathcal{A}(x, B) \) is a cluster algebra of Dynkin type \( \mathbb{A}_n \). Note that \( \mathcal{A}(x, B) \) has exactly one coefficient, and that this coefficient is non-invertible.

Let \( (x[0], B[0]) := (x, B) \). For each \( 1 \leq i \leq m - 1 \) we define inductively a seed by
\[
(x[i], B[i]) := \mu_{m-i} \cdots \mu_2 \mu_1 (x[i-1], B[i-1]).
\]
For \( 0 \leq i \leq m - 1 \) set \( (x_1[i], \ldots, x_m[i]) := x[i] \).

For simplicity we define \( x_0[i] := 1 \) and \( x_{-1}[i] := 0 \) for all \( i \).

Lemma 7.1. For \( 0 \leq i \leq m - 2, 1 \leq k \leq m - 1 - i \) and \( 0 \leq j \leq i \) we have
\[
\mu_{(x[i], B[i])}(x_k[i]) = \frac{x_{k-1}[i] + x_{k+1}[i]}{x_k[i]} = \frac{x_{k-1+i}[i-j] + x_{k+1+i}[i-j]}{x_{k+j}[i-j]}.
\]

Proof. The first equality follows from the definition of \( (x[i], B[i]) \) and the mutation rule. The second equality is proved by induction on \( i \). \( \square \)

Corollary 7.2. For \( 0 \leq i \leq m - 2 \) we have
\[
\begin{aligned}
x_{i+2} &= x_1[i+1]x_{i+1} - x_i, \\
x_{i+1}[1] &= x_1[i+1]x_i[1] - x_{i-1}[1].
\end{aligned}
\]
Proof. Equation (5) follows from (4) for \( k = 1 \) and \( j = i \). The case \( k = 1 \) and \( j = i - 1 \) yields Equation (6).

\[\]

**Proposition 7.3.** The elements \( x_1[0], x_1[1], \ldots, x_1[m - 1] \) are algebraically independent and

\[ K[x_1[0], x_1[1], \ldots, x_1[m - 1]] = A(x, B). \]

In particular, the cluster algebra \( A(x, B) \) is a polynomial ring in \( m \) variables.

Proof. It follows from Equation (5) that

\[ x_1[i] \in K(x_1, \ldots, x_{i+1}) \setminus K(x_1, \ldots, x_i) \]

for all \( 1 \leq i \leq m - 1 \). Since \( x_1, \ldots, x_m \) are algebraically independent, this implies that \( x_1[0], x_1[1], \ldots, x_1[m - 1] \) are algebraically independent as well. Thus

\[ U := K[x_1[0], x_1[1], \ldots, x_1[m - 1]] \]

is a polynomial ring in \( m \) variables. In particular, \( U \) is factorial. Equation (5) implies that \( x_1, \ldots, x_m \in U \), and Equation (6) yields that \( x_1[1], \ldots, x_m[1] \in U \). Clearly, the clusters \( x \) and \( x[1] \) are disjoint. Thus the assumptions of Theorem 4.1 are satisfied, and we get \( U = A(x, B) \).

The cluster algebra \( A(x, B) \) as defined above has been studied by several people. It is related to a \( T \)-system of Dynkin type \( A_1 \) with a certain boundary condition, see [DK]. Furthermore, for \( K = \mathbb{C} \) the cluster algebra \( A(x, B) \) is naturally isomorphic to the complexified Grothendieck ring of the category \( C_n \) of finite-dimensional modules of level \( n \) over the quantum loop algebra of Dynkin type \( A_1 \), see [HL, N2]. It is well known, that \( A(x, B) \) is a polynomial ring. We just wanted to demonstrate how to use Theorem 4.1 in practise.

### 7.2. Acyclic cluster algebras as polynomial rings

Let \( C = (c_{ij}) \in M_{n,n}(\mathbb{Z}) \) be a generalized Cartan matrix, i.e. \( C \) is symmetrizable, \( c_{ii} = 2 \) for all \( i \) and \( c_{ij} \leq 0 \) for all \( i \neq j \).

Assume that \( m = 2n = 2p \), and let \( (x, B) \) be a seed of \( \mathcal{F} \), where \( B = (b_{ij}) \in M_{2n,n}(\mathbb{Z}) \) is defined as follows: For \( 1 \leq i \leq 2n \) and \( 1 \leq j \leq n \) let

\[
\begin{aligned}
  b_{ij} := &\begin{cases}
    0 & \text{if } i = j, \\
    -c_{ij} & \text{if } 1 \leq i < j \leq n, \\
    c_{ij} & \text{if } 1 \leq j < i \leq n, \\
    1 & \text{if } i = n + j, \\
    c_{i-n,j} & \text{if } n + 1 \leq i \leq 2n \text{ and } i - n < j, \\
    0 & \text{if } n + 1 \leq i \leq 2n \text{ and } i - n > j.
  \end{cases}
\end{aligned}
\]
Thus we have

\[
B = \begin{pmatrix}
0 & b_{12} & b_{13} & \cdots & b_{1n} \\
b_{21} & 0 & b_{23} & \cdots & b_{2n} \\
b_{31} & b_{32} & 0 & \ddots & : \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
b_{n1} & b_{n2} & \cdots & b_{n,n-1} & 0 \\
1 & -b_{12} & -b_{13} & \cdots & -b_{1n} \\
0 & 1 & -b_{23} & \cdots & -b_{2n} \\
0 & 0 & 1 & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

Clearly, \((x, B)\) is an acyclic seed. Namely, if \(i \rightarrow j\) is an arrow in \(\Sigma(B)\), then \(i < j\). Up to simultaneous reordering of columns and rows, each acyclic skew-symmetrizable matrix in \(M_{n,n}(\mathbb{Z})\) is of the form \(B^\circ\) with \(B\) defined as above. Note that \(A(x, B)\) has exactly \(n\) coefficients, and that all these coefficients are non-invertible.

For \(1 \leq i \leq n\) let

\[
(x[1], B[1]) := \mu_n \cdots \mu_2 \mu_1 (x, B)
\]

and \((x_1[1], \ldots, x_{2n}[1]) := x[1]\). Let \(B_0 := B\), and for \(1 \leq i \leq n\) let \(B_i := \mu_i(B_{i-1})\). Thus we have \(B_n = B[1]\). It is easy to work out the matrices \(B_i\) explicitly: The matrix \(B_i\) is obtained from \(B_{i-1}\) by changing the sign in the \(i\)th row and the \(i\)th column of the principal part \(B^\circ_{i-1}\). Furthermore, the \((n+i)\)th row of \(B_{i-1}\) gets replaced by

\[
(0, \ldots, 0, 1, -b_{i,i+1}, -b_{i,i+2}, \ldots, -b_{in})
\]

of \(B_{i-1}\) gets replaced by

\[
(-b_{1i}, -b_{2i}, \ldots, -b_{i,i-1}, -1, 0, \ldots, 0).
\]

If we write \(N_+\) (resp. \(N_-\)) for the upper (resp. lower) triangular part of \(B^\circ\), we get

\[
B = \begin{pmatrix}
1 & N_+ \\
\cdots & \cdots \\
N_- & 1 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & 1
\end{pmatrix} \quad \text{and} \quad B[1] = \begin{pmatrix}
1 & N_+ \\
\cdots & \cdots \\
N_- & 1 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
-N_- & -1 & \cdots & \cdots & \cdots
\end{pmatrix}.
\]

In particular, the principal part \(B^\circ\) of \(B\) is equal to the principal part \(B[1]^\circ\) of \(B[1]\).
Now the definition of seed mutation yields

\[(7)\quad x_k[1] = x_k^{-1}\left(x_{n+k} + \prod_{i=1}^{k-1} x_i[1]^{b_{ik}} \prod_{i=k+1}^{n} x_i^{-b_{ik}}\right)\]

for \(1 \leq k \leq n\).

**Proposition 7.4.** The elements \(x_1, \ldots, x_n, x_1[1], \ldots, x_n[1]\) are algebraically independent and

\[K[x_1, \ldots, x_n, x_1[1], \ldots, x_n[1]] = A(x, B).\]

In particular, the cluster algebra \(A(x, B)\) is a polynomial ring in \(2n\) variables.

**Proof.** By Equation (7) and induction we have

\[x_k[1] \in K[x_1, \ldots, x_{n+k}] \setminus K[x_1, \ldots, x_{n+k-1}]\]

for all \(1 \leq k \leq n\). It follows that \(x_1, \ldots, x_n, x_1[1], \ldots, x_n[1]\) are algebraically independent, and that the clusters \(x\) and \(x[1]\) are disjoint. Let \(U := K[x_1, \ldots, x_n, x_1[1], \ldots, x_n[1]]\).

Thus \(U\) is a polynomial ring in \(2n\) variables. In particular, \(U\) is factorial. It follows from Equation (7) that

\[(8)\quad x_{n+k} = x_k[1]x_k - \prod_{i=1}^{k-1} x_i[1]^{b_{ik}} \prod_{i=k+1}^{n} x_i^{-b_{ik}}.\]

This implies \(x_{n+k} \in U\) for all \(1 \leq k \leq n\). Thus the assumptions of Theorem 4.1 are satisfied, and we can conclude that \(U = A(x, B)\). \(\square\)

**Proposition 7.4** is a special case of a much more general result proved in [GLS2]. But the proof presented here is new and more elementary.

Next, we compare the basis

\[P_{GLS} := \{x[a] := x_1^{a_1} \cdots x_n^{a_n} x_1[1]^{a_{n+1}} \cdots x_n[1]^{a_{2n}} \mid a = (a_1, \ldots, a_{2n}) \in \mathbb{N}^{2n}\}\]

of \(A(x, B)\) resulting from Proposition 7.4 with a basis constructed by Berenstein, Fomin and Zelevinsky [BFZ]. For \(1 \leq k \leq n\) let

\[(9)\quad x'_k := \mu(x, B)(x_k) = x_k^{-1}\left(x_{n+k} \prod_{i=1}^{k-1} x_i^{b_{ik}} + \prod_{i=k+1}^{n} x_i^{-b_{ik}} \prod_{i=1}^{k-1} x_{n+i}^{-b_{ik}}\right)\]

and set

\[P_{BFZ} := \{x'[a] := x_1^{a_1} \cdots x_n^{a_n} x_1[1]^{a_{n+1}} \cdots x_n[1]^{a_{2n}} (x'_1)^{a_{2n+1}} \cdots (x'_n)^{a_{2n}} \mid a = (a_1, \ldots, a_{3n}) \in \mathbb{N}^{3n}, a_k a_{2n+k} = 0 \text{ for } 1 \leq k \leq n\}.\]

**Proposition 7.5** ([BFZ Corollary 1.21]). The set \(P_{BFZ}\) is a basis of \(A(x, B)\).
Note that the basis $P_{\text{GLS}}$ is constructed by using cluster variables from two seeds, namely $(x, B)$ and $\mu_n \cdots \mu_1(x, B)$, whereas $P_{\text{BFZ}}$ uses cluster variables from $n + 1$ seeds, namely $(x, B)$ and $\mu_k(x, B)$, where $1 \leq k \leq n$.

Now we insert Equation (8) into Equation (9) and obtain

\begin{equation}
(10) \quad x_k x'_k = \left( x_k[1] - \prod_{i=1}^{k-1} x_i[1]^{b_{ik}} \prod_{i=k+1}^{n} x_i^{b_{ik}} \right) \prod_{i=1}^{k-1} x_i^{b_{ik}} + \\
\prod_{i=k+1}^{n} x_i^{b_{ik}} \prod_{i=1}^{k-1} \left( x_i[1] - \prod_{j=1}^{i-1} x_j[1]^{b_{ij}} \prod_{j=i+1}^{n} x_j^{-b_{ij}} \right) b_{ik}.
\end{equation}

Then we observe that the right-hand side of Equation (10) is divisible by $x_k$ and that $x'_k$ is a polynomial in $x_1, \ldots, x_n, x_1[1], \ldots, x_n[1]$. Thus we can express every element of the basis $P_{\text{BFZ}}$ explicitly as a linear combination of vectors from the basis $P_{\text{GLS}}$.

One could use Equation (10) to get an alternative proof of Proposition 7.4 as pointed out by Zelevinsky [Z]. Vice versa, using Proposition 7.4 yields another proof that $P_{\text{BFZ}}$ is a basis.

As an illustration, for $n = 3$ the matrices $B_i$ look as follows:

\begin{align*}
B_0 &= \begin{pmatrix}
0 & b_{12} & b_{13} \\
b_{21} & 0 & b_{23} \\
b_{31} & b_{32} & 0 \\
1 & -b_{12} & -b_{13} \\
0 & 1 & -b_{23} \\
0 & 0 & 1
\end{pmatrix}, & B_1 &= \begin{pmatrix}
0 & -b_{12} & -b_{13} \\
b_{21} & 0 & b_{23} \\
b_{31} & b_{32} & 0 \\
-1 & 0 & 0 \\
0 & 1 & -b_{23} \\
0 & 0 & 1
\end{pmatrix}, \\
B_2 &= \begin{pmatrix}
0 & b_{12} & -b_{13} \\
b_{21} & 0 & -b_{23} \\
-b_{31} & -b_{32} & 0 \\
-1 & 0 & 0 \\
-b_{21} & -1 & 0 \\
0 & 0 & 1
\end{pmatrix}, & B_3 &= \begin{pmatrix}
0 & b_{12} & b_{13} \\
b_{21} & 0 & b_{23} \\
b_{31} & b_{32} & 0 \\
-1 & 0 & 0 \\
-b_{21} & -1 & 0 \\
-b_{31} & -b_{32} & -1
\end{pmatrix}.
\end{align*}

For example, for

\begin{equation}
B = B_0 = \begin{pmatrix}
0 & 2 & 0 \\
-2 & 0 & 1 \\
0 & -1 & 0 \\
1 & -2 & 0 \\
0 & 1 & -1 \\
0 & 0 & 1
\end{pmatrix}
\end{equation}
the quivers $\Gamma(B_0)$ and $\Gamma(B_3)$ look as follows:

$\Gamma(B_0)$: 

```
\[ \begin{array}{c}
5 & 4 & 1 & 2 & 3 & 6 \\
\searrow & \nearrow & \downarrow & \nearrow & \searrow & \downarrow \\
\end{array} \]
```

$\Gamma(B_3)$: 

```
\[ \begin{array}{c}
5 & 4 & 1 & 2 & 3 & 6 \\
\searrow & \nearrow & \downarrow & \nearrow & \searrow & \downarrow \\
\end{array} \]
```

The cluster algebra $A(x, B)$ is a polynomial ring in the 6 variables $x_1, x_2, x_3$ and

- $x_1[1] = \frac{x_2^2 + x_4}{x_1}$,
- $x_2[1] = \frac{x_1^2 x_3 + 2 x_2^3 x_4 + x_3 x_2^2 + x_1^3 x_2}{x_1^2 x_2}$,
- $x_3[1] = \frac{x_2^3 x_3 + 2 x_2^3 x_4 + x_3 x_2^2 + x_1^3 x_3 + x_1^2 x_2 x_6}{x_1^2 x_2 x_3}$.

7.3. Cluster algebras arising in Lie theory as polynomial rings.

The next class of examples can be seen as a fusion of the examples discussed in Sections 7.1 and 7.2. In the following we use the same notation as in [GLS 2].

Let $C \in M_{n,n}(\mathbb{Z})$ be a symmetric generalized Cartan matrix, and let $g$ be the associated Kac-Moody Lie algebra over $K = \mathbb{C}$ with triangular decomposition $g = n^- \oplus h \oplus n$, see [K].

Let $U(n)^*_gr$ be the graded dual of the enveloping algebra $U(n)$ of $n$. To each element $w$ in the Weyl group $W$ of $g$ one can associate a subalgebra $R(C_w)$ of $U(n)^*_gr$ and a cluster algebra $A(C_w)$, see [GLS 2]. Here $C_w$ denotes a Frobenius category associated to $w$, see [BIRS, GLS 2].

In [GLS 2] we constructed a natural algebra isomorphism

$$A(C_w) \rightarrow R(C_w).$$

This yields a cluster algebra structure on $R(C_w)$.

Let $i = (i_1, \ldots, i_r)$ be a reduced expression of $w$. In [GLS 2] we studied two cluster-tilting modules $V_i = V_{i_1} \oplus \cdots \oplus V_{i_r}$ and $T_i = T_{i_1} \oplus \cdots \oplus T_{i_r}$ in $C_w$, which are associated to $i$. These modules yield two disjoint clusters $(\delta_{V_1}, \ldots, \delta_{V_r})$ and $(\delta_{T_1}, \ldots, \delta_{T_r})$ of $R(C_w)$. The exchanges matrices are of size $r \times (r - n)$. In contrast to our conventions in this article, the $n$ coefficients are $\delta_{V_k} = \delta_{T_k}$ with $k^+ = r + 1$, where $k^+$ is defined as in [GLS 2], and none of these coefficients is invertible. Furthermore, we studied a module $M_i = M_{i_1} \oplus \cdots \oplus M_{i_r}$ in $C_w$, which yields cluster variables $\delta_{M_1}, \ldots, \delta_{M_r}$ of $R(C_w)$. (These do not form a cluster.) Using methods from Lie theory we obtained the following result.

**Theorem 7.6 ([GLS 2, Theorem 3.2]).** The cluster algebra $R(C_w)$ is a polynomial ring in the variables $\delta_{M_1}, \ldots, \delta_{M_r}$.
To obtain an alternative proof of Theorem 7.6, one can proceed as follows:

(i) Show that the cluster variables $\delta_{M_1}, \ldots, \delta_{M_r}$ are algebraically independent.

(ii) Show that for $1 \leq k \leq r$ the cluster variables $\delta_{V_k}$ and $\delta_{T_k}$ are polynomials in $\delta_{M_1}, \ldots, \delta_{M_r}$.

(iii) Apply Theorem 4.1.

Part (i) can be done easily using induction and the mutation sequence in [GLS2, Section 13]. Part (ii) is not at all straightforward.

Let us give a concrete example illustrating Theorem 7.6. Let $\mathfrak{g}$ be the Kac-Moody Lie algebra associated to the generalized Cartan matrix

$$C = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix},$$

and let $i = (2, 1, 2, 1, 2, 1, 2, 1)$. Then $\mathcal{A}(C_w) = \mathcal{A}(x, B_i)$, where $r = n + 2 = 8$, $x_7$ and $x_8$ are the (non-invertible) coefficients, and

$$B_i = \begin{pmatrix} 0 & 2 & -1 \\ -2 & 0 & 2 & -1 \\ 1 & -2 & 0 & 2 & -1 \\ 1 & -2 & 0 & 2 \\ 1 & -2 & 0 & \end{pmatrix}.$$

The principal part $B_i^\circ$ of $B_i$ is skew-symmetric, and the quiver $\Gamma(B_i)$ looks as follows:

![Quiver Diagram](image)

Define

$$(x[0], B[0]) := (x, B_i),$$

$$(x[1], B[1]) := \mu_5 \mu_3 \mu_1 (x[0], B[0]), \quad (x[2], B[2]) := \mu_6 \mu_4 \mu_2 (x[1], B[1]),$$

$$(x[3], B[3]) := \mu_3 \mu_1 (x[2], B[2]), \quad (x[4], B[4]) := \mu_4 \mu_2 (x[3], B[3]),$$

$$(x[5], B[5]) := \mu_1 (x[4], B[4]), \quad (x[6], B[6]) := \mu_2 (x[5], B[5]),$$

and for $0 \leq k \leq 6$ let $(x[k], \ldots, x[k]) := x[k]$.

Under the isomorphism $\mathcal{A}(C_w) \to \mathcal{R}(C_w)$ the cluster $x[0]$ of $\mathcal{A}(C_w) = \mathcal{A}(x, B_i)$ corresponds to the cluster $(\delta_{V_1}, \ldots, \delta_{V_8})$ of $\mathcal{R}(C_w)$, the cluster $x[6]$ corresponds...
to \((\delta_{T_1}, \ldots, \delta_{T_8})\), and we have
\[
\begin{align*}
x_1[0] &\mapsto \delta_{M_1}, \\
x_2[0] &\mapsto \delta_{M_2}, \\
x_1[2] &\mapsto \delta_{M_3}, \\
x_2[2] &\mapsto \delta_{M_4}, \\
x_1[4] &\mapsto \delta_{M_5}, \\
x_2[4] &\mapsto \delta_{M_6}, \\
x_1[6] &\mapsto \delta_{M_7}, \\
x_2[6] &\mapsto \delta_{M_8}.
\end{align*}
\]
By Theorem 7.6 we know that the cluster algebra \(A(x, B_1)\) is a polynomial ring in the variables \(x_1[0], x_2[0], x_1[2], x_2[2], x_1[4], x_2[4], x_1[6], x_2[6]\).

8. Applications

8.1. Prime elements in the dual semicanonical basis. As in Section 7.3 let \(C \in M_{n,n}(\mathbb{Z})\) be a symmetric generalized Cartan matrix, and let \(\mathfrak{g} = \mathfrak{n}_- \oplus \mathfrak{h} \oplus \mathfrak{n}\) be the associated Lie algebra.

As before let \(W\) be the Weyl group of \(\mathfrak{g}\). To \(C\) one can also associate a preprojective algebra \(\Lambda\) over \(\mathbb{C}\), see for example [GLS2, R].

Lusztig [Lu] realized the universal enveloping algebra \(U(\mathfrak{n})\) of \(\mathfrak{n}\) as an algebra of constructible functions on the varieties \(\Lambda_d\) of nilpotent \(\Lambda\)-modules with dimension vector \(d \in \mathbb{N}^n\). He also constructed the semicanonical basis \(S\) of \(U(\mathfrak{n})\). The elements of \(S\) are naturally parametrized by the irreducible components of the varieties \(\Lambda_d\).

An irreducible component \(Z\) of \(\Lambda_d\) is called indecomposable if it contains a Zariski dense subset of indecomposable \(\Lambda\)-modules, and \(Z\) is rigid if it contains a rigid \(\Lambda\)-module \(M\), i.e. \(M\) is a module with \(\text{Ext}^1_{\Lambda}(M, M) = 0\).

Let \(S^*\) be the dual semicanonical basis of the graded dual \(U(\mathfrak{n})^*_{\text{gr}}\) of \(U(\mathfrak{n})\). The elements \(\rho_{Z}\) in \(S^*\) are also parametrized by irreducible components \(Z\) of the varieties \(\Lambda_d\). We call \(\rho_{Z}\) indecomposable (resp. rigid) if \(Z\) is indecomposable (resp. rigid). An element \(b \in S^*\) is called primitive if it cannot be written as a product \(b = b_1 b_2\) with \(b_1, b_2 \in S^* \setminus \{1\}\).

**Theorem 8.1** ([GLS1 Theorem 1.1]). If \(\rho_{Z}\) is primitive, then \(Z\) is indecomposable.

**Theorem 8.2** ([GLS2 Theorem 3.1]). For \(w \in W\) all cluster monomials of the cluster algebra \(\mathcal{R}(C_w)\) belong to the dual semicanonical basis \(S^*\) of \(U(\mathfrak{n})^*_{\text{gr}}\). More precisely, we have
\[
\{\text{cluster variables of } \mathcal{R}(C_w)\} \subseteq \{\rho_{Z} \in S^* \mid Z \text{ is indecomposable and rigid}\},
\]
\[
\{\text{cluster monomials of } \mathcal{R}(C_w)\} \subseteq \{\rho_{Z} \in S^* \mid Z \text{ is rigid}\}.
\]
Combining Theorems 8.1, 7.6 and 8.2 we obtain a partial converse of Theorem 8.1.

**Theorem 8.3.** The cluster variables in \(\mathcal{R}(C_w)\) are prime, and they are primitive elements of \(S^*\).

**Conjecture 8.4.** If \(\rho_{Z} \in S^*\) is indecomposable and rigid, then \(\rho_{Z}\) is prime in \(U(\mathfrak{n})^*_{\text{gr}}\).
8.2. Monoidal categorifications of cluster algebras. Let $\mathcal{C}$ be an abelian tensor category with unit object $I$. We assume that $\mathcal{C}$ is a Krull-Schmidt category, and that all objects in $\mathcal{C}$ are of finite length. Let $\mathcal{M}(\mathcal{C}) := K_0(\mathcal{C})$ be the Grothendieck ring of $\mathcal{C}$. The class of an object $M \in \mathcal{C}$ is denoted by $[M]$. The addition in $\mathcal{M}(\mathcal{C})$ is given by $[M] + [N] := [M \oplus N]$ and the multiplication is defined by $[M][N] := [M \otimes N]$. We assume that $[M \otimes N] = [N \otimes M]$.

The concept of a monoidal categorification of a cluster algebra was introduced in [HL]. Combining Proposition 8.5 with Theorem 3.1 we get the following result.

A monoidal categorification of a cluster algebra $\mathcal{A}(x, B)$ is an algebra isomorphism

$$\Phi : \mathcal{A}(x, B) \to \mathcal{M}_K(\mathcal{C}),$$

where $\mathcal{C}$ is a tensor category as above, such that each cluster monomial $y = y_1^{a_1} \cdots y_m^{a_m}$ of $\mathcal{A}(x, B)$ is mapped to a class $[S_y]$ of some simple object $S_y \in \mathcal{C}$. In particular, we have

$$[S_y] = [S_{y_1}]^{a_1} \cdots [S_{y_m}]^{a_m} = [S_{y_1}^{\otimes a_1} \otimes \cdots \otimes S_{y_m}^{\otimes a_m}].$$

For an object $M \in \mathcal{C}$ let $x_M$ be the element in $\mathcal{A}(x, B)$ with $\Phi(x_M) = [M]$.

The concept of a monoidal categorification of a cluster algebra was introduced in [HL] Definition 2.1. But note that our definition uses weaker conditions than in [HL].

An object $M \in \mathcal{C}$ is called invertible if $[M]$ is invertible in $\mathcal{M}_K(\mathcal{C})$. An object $M \in \mathcal{C}$ is primitive if there are no non-invertible objects $M_1$ and $M_2$ in $\mathcal{C}$ with $M \cong M_1 \otimes M_2$.

**Proposition 8.5.** Let $\Phi : \mathcal{A}(x, B) \to \mathcal{M}_K(\mathcal{C})$ be a monoidal categorification of a cluster algebra $\mathcal{A}(x, B)$. Then the following hold:

(i) The invertible elements in $\mathcal{M}_K(\mathcal{C})$ are

$$\mathcal{M}_K(\mathcal{C})^\times = \{ \lambda |I| | S_{x_1}^{a_1} \otimes \cdots \otimes S_{x_p}^{a_p} \mid \lambda \in K^\times, a_i \in \mathbb{Z} \}.$$

(ii) Let $M$ be an object in $\mathcal{C}$ such that the element $x_M$ is irreducible in $\mathcal{A}(x, B)$. Then $M$ is primitive.

**Proof.** Part (i) follows directly from Theorem 2.2. To prove (ii), assume that $M$ is not primitive. Thus there are non-invertible objects $M_1$ and $M_2$ in $\mathcal{C}$ with $M \cong M_1 \otimes M_2$. Thus in $\mathcal{M}_K(\mathcal{C})$ we have $[M] = [M_1][M_2]$. Since $\Phi$ is an algebra isomorphism, we get $x_M = x_{M_1}x_{M_2}$ with $x_{M_1}$ and $x_{M_2}$ non-invertible in $\mathcal{A}(x, B)$. Since $x_M$ is irreducible, we have a contradiction.

Combining Proposition 8.5 with Theorem 3.1 we get the following result.

**Documenta Mathematica** 18 (2013) 249–274
Corollary 8.6. Let $\Phi : A(x, B) \to \mathcal{M}_K(C)$ be a monoidal categorification of a cluster algebra $A(x, B)$. For each cluster variable $y$ of $A(x, B)$, the simple object $S_y$ is primitive.

Examples of monoidal categorifications of cluster algebras can be found in [HL][NI], see also [Le].

Acknowledgements. We thank Giovanni Cerulli Irelli, Sergey Fomin, Daniel Labardini Fragoso, Philipp Lampe and Andrei Zelevinsky for helpful discussions. We thank Giovanni Cerulli Irelli for carefully reading several preliminary versions of this article. The first author likes to thank the Max-Planck Institute for Mathematics in Bonn for a one year research stay in 2010/2011. The third author thanks the Sonderforschungsbereich/Transregio SFB 45 for financial support, and all three authors thank the Hausdorff Institute for Mathematics in Bonn for support and hospitality.

References


[La] P. Lampe, Email from October 8th, 2011.


---

Christof Geiß
Instituto de Matemáticas
Universidad Nacional Autónoma de México
Ciudad Universitaria
04510 México D.F.
christol@math.unam.mx

Bernard Leclerc
Institut de Mathématiques
Université de Caen
LMNO
CNRS UMR 6139
14032 Caen cedex
bernard.leclerc@unicaen.fr

Jan Schröer
Mathematisches Institut
Universität Bonn
Endenicher Allee 60
53115 Bonn
Germany
schoer@math.uni-bonn.de
Large Parallel Volumes of Finite
and Compact Sets in d-Dimensional Euclidean Space

Jürgen Kampf and Markus Kiderlen

Received: January 10, 2012
Revised: June 18, 2012
Communicated by Lars Hesselholt

Abstract. The \( r \)-parallel volume \( V(C_r) \) of a compact subset \( C \) in
\( d \)-dimensional Euclidean space is the volume of the set \( C_r \) of all points
of Euclidean distance at most \( r > 0 \) from \( C \). According to Steiner’s
formula, \( V(C_r) \) is a polynomial in \( r \) when \( C \) is convex. For finite
sets \( C \) satisfying a certain geometric condition, a Laurent expansion
of \( V(C_r) \) for large \( r \) is obtained. The dependence of the coefficients
on the geometry of \( C \) is explicitly given by so-called intrinsic power
volumes of \( C \). In the planar case such an expansion holds for all finite
sets \( C \). Finally, when \( C \) is a compact set in arbitrary dimension, it is
shown that the difference of large \( r \)-parallel volumes of \( C \) and of its
convex hull behaves like \( cr^{d-3} \), where \( c \) is an intrinsic power volume
of \( C \).

2000 Mathematics Subject Classification: Primary 52A39; Secondary
52B11, 41A10.
Keywords and Phrases: Large parallel sets, Laurent expansion of par-
allel volume, Steiner formula, intrinsic power volume.

1 Introduction

In 1840, Jakob Steiner [10] showed that the \( r \)-parallel volume of certain compact
convex sets in \( \mathbb{R}^d \) are polynomials in \( r \geq 0 \) when \( d = 2 \) or \( d = 3 \). The
generalization to arbitrary dimensions is now known as the Steiner formula

\[
V_d(C_r) = \sum_{i=0}^{d} \kappa_i V_{d-i}(C)r^i, \quad r \geq 0.
\]

Both authors were supported by the Villum Foundation in the framework of the Centre
for Stochastic Geometry and Advanced Bioimaging, CSGB.
Here $C$ is any compact convex subset of $\mathbb{R}^d$, $C_r$ is the (outer) $r$-parallel set of $C$ consisting of all points of distance at most $r$ from $C$, $V_d$ is the Lebesgue measure, $\kappa_i$ denotes the volume of the $i$-dimensional Euclidean unit ball and $V_i(C)$ are the intrinsic volumes of $C$, which are defined by this relation. A possible proof of (1) uses the fact that the intrinsic volumes can be given explicitly in the case where the set $C$ is a convex polytope $P$:

$$V_i(P) = \sum_{F \in \mathcal{F}_i(P)} \gamma(F, P)V_i(F), \quad (2)$$

$i = 0, \ldots, d$. Here $\mathcal{F}_i(P)$ is the family of $i$-dimensional faces of $P$ and $\gamma(F, P)$ is the outer angle of $P$ at $F$ defined in Equation (13) below. There are numerous generalizations of the Steiner formula, for instance local versions [8, Chapter 4] and even Steiner-type results for arbitrary closed sets $C$; see [3]. Often, Steiner-type formulas are only valid for sufficiently small $r$, e.g. in the case of sets of positive reach [1]. In the present work we focus on $r$-parallel volumes for large $r$ and determine a Laurent series expansion or at least its leading coefficients. We discuss mainly the parallel volume of finite sets, which turns out to be already nontrivial, but Theorem 2, below, deals with compact sets $C \subseteq \mathbb{R}^d$. A first result for large parallel volumes has been obtained in [4] for compact sets $C \subseteq \mathbb{R}^d$, where it is shown that the volume of $C_r$ is close to the volume of its convex hull $\text{conv} C_r = (\text{conv} C)_r$. In fact, there is a constant $c = c(C)$ such that

$$0 \leq V_d(\text{conv} C_r) - V_d(C_r) \leq cr^{d-3} \quad (3)$$

for all sufficiently large $r$. In [4] an example set $C$ was given where this volume difference behaves like $r^{d-3}$, so the exponent here is best possible. Independently, a weaker version of (3) was shown in [7, Lemma 2], where $C$ was assumed to be an at most two-dimensional subset of the set $[0,1]^d$, $d \geq 3$, and the volume difference was shown to converge to zero faster than $r^{d-2}$. In [7] this was used to obtain an (incomplete) collection of asymptotic Miles-type formulas for the specific intrinsic volumes of stationary digitized Boolean models of balls. It was the original motivation for the present research to complete and generalize these results. One might even ask for similar asymptotic formulas for the specific intrinsic volumes of digitized standard random sets. (Standard random sets are stationary, a.s. locally polyconvex random sets, satisfying a certain integrability condition; see [9, Definition 9.2.1]). While the Boolean model case only requires (truncated) Laurent expansions for the volume of the set $C_r$, being the Minkowski sum of $C$ with the $r$-scaled Euclidean unit ball $B^d$, results for general standard random sets would require such expansions for the volume of the Minkowski sum of $C$ with an arbitrary $r$-scaled polyconvex set. In [6, Corollary 2.2] the first two leading coefficients of such a Laurent expansion are determined and [6, §5] discusses its application to the theory of random sets, including the calculation of the one-sided derivative of the contact distribution function at zero. However, higher order expansions in this general setting appear not to be known.

Documenta Mathematica 18 (2013) 275–295
Other applications of our expansions appear to be in reach. For example, in [4] formula (3) was used to examine the expected value of the parallel volume of Brownian paths, when the time is small, and to relate analytical properties of \( r \mapsto W(rK) \) to geometric properties of \( K \), where \( W \) denotes the Wills functional and \( K \) is a fixed compact set.

In the next section, the main results will be stated starting with the definition of functionals of finite subsets of \( \mathbb{R}^d \) capturing all geometric properties that are relevant when considering large parallel volumes. An optimal lower bound for the constant \( c \) in (3) will be given in Theorem 2. For finite sets \( C \), we obtain an explicit Laurent expansion in the case \( d = 2 \) in Proposition 1 and, under an additional condition on \( C \), in Theorem 5 for all \( d \geq 2 \). Section 3 provides proofs of the main results and discusses in particular the properties of the coefficients in the Laurent expansion.

2 Main results

The \( r \)-parallel volumes of one-dimensional compact sets are trivially affine functions for sufficiently large \( r \). We therefore assume \( d \geq 2 \) throughout the following. In the spirit of (2) define

\[
V_{i}^{(m)}(C) = \sum_{F \in \mathcal{F}_{i}(\text{conv} C)} \gamma(F, \text{conv} C) \int_{F} \frac{d(C \cap F, x)^{m-i}}{m!} \, dx,
\]

where \( i = 0, \ldots, d, m \geq i \), and \( C \) is a finite subset of \( \mathbb{R}^d \). We call the functionals \( V_{i}^{(m)} \) the intrinsic power volumes. Here, integration is understood with respect to \( i \)-dimensional Lebesgue measure in the affine hull of \( F \), and \( d(C \cap F, x) \) is the smallest Euclidean distance between \( x \) and a point in \( C \cap F \). Due to (2), we have

\[
V_{i}^{(i)}(C) = V_{i}(\text{conv} C)
\]

for all \( i = 0, \ldots, d \). The functionals \( V_{i}^{(m)} \), defined on the family of finite subsets of \( \mathbb{R}^d \), share many properties with usual intrinsic volumes (see Lemma 9) and are in particular independent of the dimension of the embedding space. The functional \( V_{i}^{(m)}(C) \), where the sum in (4) extends over all edges of \( \text{conv} C \), is given more explicitly in (17). In particular, if \( C \) is the set of vertices of a convex polytope,

\[
V_{i}^{(m)}(C) = \frac{1}{m!} \sum_{F \in \mathcal{F}_{i}(\text{conv} C)} \gamma(F, \text{conv} C) V_{i}^{m}(F)
\]

\( m \geq 1 \). In the case of planar finite sets, these (and the classical intrinsic volumes) are the only intrinsic power volumes that occur in a Laurent expansion of large parallel volumes.

Documenta Mathematica 18 (2013) 275–295
Proposition 1. Let $C \subseteq \mathbb{R}^2$ be a finite set and put $K = \text{conv} \ C$. Then

$$V_2(K_r) - V_2(C_r) = 2 \sum_{n=1}^{\infty} \frac{(2n-3)!!}{(2n)!!} V_1^{(2n+1)}(C) r^{-(2n-1)}$$

for all sufficiently large $r$. (The definition of the double factorial $n!!$ is recalled in Equation (12) below.)

Proposition 1 and Corollary 6 follow directly from Theorem 5 below. All other results in this section will be shown in Section 3. The leading coefficient of the Laurent expansion in Proposition 1 depends on $C$ through the intrinsic power volume $V_1^{(3)}(C)$. An analogue statement holds when $C$ is only assumed to be compact, and in all dimensions $d \geq 2$. To define $V_1^{(m)}(C)$, $m \geq 1$, also for compact $C \subseteq \mathbb{R}^d$, we choose for any $x \in \mathbb{R}^d$ two points $p_x^C$ and $q_x^C$ in $C$ with $x$ in the line segment $[p_x^C, q_x^C]$ and $C \cap [p_x^C, q_x^C] = \{p_x^C, q_x^C\}$, whenever such points exist and are unique up to permutation. Otherwise we put $p_x^C = q_x^C = x$. We then define

$$V_1^{(m)}(C) = \frac{1}{\kappa_{d-1}} \int_{\mathbb{R}^d} d\{p_x^C, q_x^C\} dx^{m-1} C_1(\text{conv} C, dx),$$

where $C_1(\text{conv} C, \cdot)$ is the first curvature measure of $\text{conv} C$. If $C$ is finite, then the definitions (4) and (6) coincide as will be shown in Remark 7.

Theorem 2. If $K$ is the convex hull of a compact set $C \subseteq \mathbb{R}^d$, then

$$\lim_{r \to \infty} \frac{V_d(K_r) - V_d(C_r)}{r^{d-3}} = \frac{\omega_{d-1}}{2} V_1^{(3)}(C).$$

Here $\omega_i$ is the surface area of the $(i-1)$-dimensional unit sphere in $\mathbb{R}^i$.

A natural question is whether the speed of convergence in Theorem 2 can be determined: is there an $\alpha > 0$ such that, for any compact $C \subseteq \mathbb{R}^d$, there is a constant $c = c(C)$ with

$$\left| \frac{V_d(K_r) - V_d(C_r)}{r^{d-3}} - \frac{\omega_{d-1}}{2} V_1^{(3)}(C) \right| \leq cr^{-\alpha}$$

for all sufficiently large $r$? The following proposition (with $f(r) = r^{-\alpha/2}$) shows that, already in $\mathbb{R}^2$, such a stability result cannot hold.

Proposition 3. Let $f : (0, \infty) \to (0, \infty)$ be a continuous bijective map with $\lim_{r \to \infty} f(r) = 0$. Then there is a compact set $C \subseteq \mathbb{R}^2$ and a number $c > 0$ such that

$$cf(r) \leq \left| \frac{V_2(K_r) - V_2(C_r)}{r^{-1}} - V_1^{(3)}(C) \right|$$

for all sufficiently large $r$. As usual, we have put $K = \text{conv} C$. 

Documenta Mathematica 18 (2013) 275–295
However, if the class of compact sets is replaced by the smaller class of all finite sets, a stability result with $\alpha = 1$ can be obtained. (That this speed of convergence is optimal for this class when $d \geq 3$ follows from Theorem 5 below.)

**Theorem 4.** Let $C \subseteq \mathbb{R}^d$ be finite and put $K = \text{conv} C$. Then there is a constant $c = c(C)$ such that

$$\left| \frac{V_d(K_r) - V_d(C_r)}{r^{d-3}} - \frac{\omega_{d-1}}{2} V_1^{(3)}(C) \right| \leq \frac{c}{r},$$

for sufficiently large $r$.

In the remainder of this section, $C$ will always be a finite subset of $\mathbb{R}^d$. Under the following condition on the facets (these are the $(d-1)$-dimensional faces) of $\text{conv} C$, we can also obtain a Laurent series expansion of infinite order in higher dimensions, generalizing Proposition 1.

**Condition (A).** For all facets $G$ of the polytope $K = \text{conv} C$ and all faces $F$ of $G$ we have

$$d(C \cap F, x) = d(C \cap G, x) \quad \text{for all } x \in F. \quad (9)$$

Clearly (9) holds whenever $F$ is a singleton or $F = G$. Hence Condition (A) must only be checked for faces $F$ of dimension between 1 and $d - 2$, and is in particular satisfied for all finite sets $C$ in $\mathbb{R}^2$. In $\mathbb{R}^3$, only edges $F \in \mathcal{F}_1(K)$ must be considered. In particular, if $C$ is the vertex set of a simplicial polytope in $\mathbb{R}^3$, Condition (A) is violated if and only if at least one facet of $\text{conv} C$ is a triangle with a strictly obtuse angle.

For instance, Condition (A) is fulfilled if $C \subseteq \mathbb{R}^d$ is the set of vertices of a rectangular cuboid. If $C$ is the set of the vertices of the standard simplex

$$S_d = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : \sum_{i=1}^d x_i \leq 1, x_1, \ldots, x_d \geq 0 \}$$

in $\mathbb{R}^d$, Condition (A) is satisfied if and only if $d \in \{2, 3\}$. This can be shown as follows. We have already seen that Condition (A) is trivially fulfilled when $d = 2$, and it also holds for $d = 3$ as the triangles forming the facets of $S_3$ do not contain a strictly obtuse angle. For $d \geq 4$ the point $x = (1/3, 1/3, 1/3, 0, \ldots, 0) \in \mathbb{R}^d$ is contained in the face

$$F = \{(y_1, y_2, y_3, 0, \ldots, 0) \in \mathbb{R}^d : \sum_{i=1}^3 y_i = 1, y_1, y_2, y_3 \geq 0 \}$$

of the facet $G = S_{d-1} \times \{0\}$. Hence (9) is violated, as

$$d(C \cap F, x) = \sqrt{\frac{2}{3}} > \sqrt{\frac{1}{3}} = \|x - o\| = d(C \cap G, x).$$
If \( C \subseteq \mathbb{R}^d \) satisfies Condition (A), a Laurent expansion for large parallel volumes with explicit coefficients can be shown.

**Theorem 5.** Let \( K \) be the convex hull of a finite set \( C \subseteq \mathbb{R}^d \) that satisfies Condition (A). Then

\[
V_d(K_r) - V_d(C_r) = \sum_{n=3-d}^{\infty} a_n(C)r^{-n}
\]

for all sufficiently large \( r \) with coefficients

\[
a_n(C) = \min_{2(n+d-i)}(d-1,n+d-2) \left(\frac{-1}{(n+d-i)/2}\right) \left(\frac{-1}{n+2}\right) \kappa_{d-(n+2)}V_i^{(n+2)}(C)
\]

which vanish for all even positive \( n \).

As Condition (A) is satisfied for all planar finite sets, Proposition 1 is a direct consequence of Theorem 5. Indeed, in the case when \( n > 0 \) is odd and \( d = 2 \) the sum in the definition of \( a_n(C) \) consists only of one summand, namely

\[
a_n(C) = \left(\frac{-1}{n+3/2}\right) \kappa_2V_1^{(n+2)}(C) = 2 \frac{(n-2)!!}{(n+1)!!} V_1^{(n+2)}(C).
\]

Also for \( d = 3 \) the representation in Theorem 5 simplifies considerably. We get similarly as above

\[
a_0(C) = \left(\frac{-1}{1}\right) \kappa_2V_1^{(3)}(C) = \pi V_1^{(3)}(C),
\]

and

\[
a_n(C) = \left(\frac{-1}{n+3/2}\right) \kappa_2V_2^{(n+3)}(C) = 2 \frac{(n-2)!!}{(n+1)!!} V_2^{(n+3)}(C)
\]

for positive odd \( n \).

**Corollary 6.** Let \( K \) be the convex hull of a finite set \( C \subseteq \mathbb{R}^3 \) that satisfies Condition (A). Then, for all sufficient large \( r \),

\[
V_3(K_r) - V_3(C_r) = \pi V_1^{(3)}(C) + 2 \sum_{n=1}^{\infty} \frac{(2n-3)!!}{(2n)!!} V_2^{(2n+2)}(C) r^{-(2n-1)}.
\]

This section concludes by an example where Corollary 6 is applied to the vertex set

\[
C^d = \text{vert}([0,1]^d) = \bigcup_{F \in \mathcal{F}_d([0,1]^d)} F
\]
of the unit cube $[0,1]^d \subseteq \mathbb{R}^d$ for $d = 3$. It will be shown in Subsection 3.6 below, that (10) then becomes

$$V_3([0,1]^3) - V_3(C_r^3) = \frac{\pi}{4} + \sum_{n=1}^{\infty} b_n r^{-(2n-1)}$$

with

$$b_n = \frac{6}{(2n+1)(2n-1)(n+1)4^n} \sum_{i=0}^{n} \frac{(2i-1)!!}{i!}.$$ 

3 Proofs and auxiliary results

3.1 Notations and the intrinsic power volumes

The following subsection summarizes the notation and discusses basic properties of the intrinsic power volumes. For the reader’s convenience we repeat the most important concepts already named in the previous sections.

The Euclidean norm in $\mathbb{R}^d$ is denoted by $\| \cdot \|$, the closed Euclidean unit ball by $B^d$ and its boundary, the $(d-1)$-dimensional unit sphere by $S^{d-1}$. We write $A^C$, diam $A$, bd $A$, and relint $A$ for the complement, the diameter, the topological boundary, and the relative interior of a set $A \subseteq \mathbb{R}^d$, respectively. Let $d(A,x) = \inf_{y \in A} \|x-y\|$ be the distance of $x \in \mathbb{R}^d$ to $A \subseteq \mathbb{R}^d$. Let $[x,y]$ be the line segment with endpoints $x,y \in \mathbb{R}^d$. We denote the $m$-dimensional Hausdorff measure by $\mathcal{H}^m$ and use the volume of the Euclidean unit ball $\kappa_d = \mathcal{H}^d(B^d)$ and the surface area of the $(d-1)$-dimensional unit sphere $\omega_d = \mathcal{H}^{d-1}(S^{d-1})$. Often, we will write $V_d = \mathcal{H}^d$ for the usual Lebesgue measure. The double factorial of an integer $n \geq -1$ is defined to be

$$n!! = \begin{cases} n \cdot (n-2) \cdots 2, & \text{if } n > 0 \text{ is even}, \\ n \cdot (n-2) \cdots 1, & \text{if } n > 0 \text{ is odd}, \end{cases}$$

and the usual convention $(-1)!! = 0!! = 1$.

We recall some basic notions from convex geometry; see [8] for details. A convex body is a nonempty compact convex set in $\mathbb{R}^d$. For a convex body $K$ let $p(K,x)$ be the metric projection of $x \in \mathbb{R}^d$ onto $K$, that is the point in $K$ closest to $x$. Then $d(K,x) = \|x-p(K,x)\|$ is the distance between $x$ and $K$. If $x \not\in K$,

$$u(K,x) = \frac{x-p(K,x)}{d(K,x)} \in S^{d-1}$$

is the negative projection direction. A convex subset $F$ of a convex body $K$ is called face of $K$ if for any two points $x,y \in K$ with $x,y \in F$ we have $x,y \in F$. Let $\mathcal{F}_i(K)$ be the family of all $i$-dimensional faces of $K$, $i = 0, \ldots, d$. We denote the normal cone of $K$ at $F \in \mathcal{F}_i(K)$ by $N(K,F)$ and put $n(K,F) =$
Then \( n(K, F) = \{ u(K, x) \in S^{d-1} : x \notin K, p(K, x) \in \text{relint } F \} \).

For a polytope \( K \)

\[
\gamma(F, K) = \frac{\mathcal{H}^{d-i-1}(n(K, F))}{\omega_{d-i}}
\]

is the exterior angle of \( K \) at \( F \in \mathcal{F}(K) \), \( i = 0, \ldots, d-1 \). For completeness we set \( \gamma(K, K) = 1 \) if \( K \) is full-dimensional. We recall that the support measures \( \Theta_0(K, \cdot), \ldots, \Theta_{d-1}(K, \cdot) \) of \( K \) are the measures on \( \mathbb{R}^d \times S^{d-1} \) satisfying

\[
\mathcal{H}^d(\{ x \in \mathbb{R}^d \setminus K : d(K, x) \leq r, (p(K, x), u(K, x)) \in \eta \}) = \sum_{m=0}^{d-1} \binom{d-1}{m} r^{d-m} \omega_d \Theta_m(K, \eta)
\]

for all \( r \geq 0 \) and all Borel sets \( \eta \subseteq \mathbb{R}^d \times S^{d-1} \). More generally, for any measurable function \( f \geq 0 \) on \( \mathbb{R}^d \) we have

\[
\int_{\mathbb{R}^d \setminus K} f(x) \, dx = \sum_{m=0}^{d-1} \binom{d-1}{m} \int_{\mathbb{R}^d} \int_0^\infty f(x + su) \times
\]

\[
\times s^{d-m-1} \, ds \, \Theta_m(K, d(x, u)).
\]

The support measures are concentrated on the (generalized) normal bundle

\[
\mathcal{N}(K) = \{(p(K, y), u(K, y)) : y \notin K \} \subseteq (\text{bd } K) \times S^{d-1}.
\]

Their total mass is

\[
\Theta_m(K, \mathbb{R}^d \times S^{d-1}) = \omega_d \omega_m \binom{d-1}{m}^{-1} V_m(K).
\]

The area measures of \( K \) are the projections of the support measures to their second component:

\[
S_m(K, \omega) = \Theta_m(K, \mathbb{R}^d \times \omega)
\]

where \( \omega \) is a Borel set in \( S^{d-1} \). The curvature measures are their projections on the first component given by

\[
C_m(K, \beta) = \Theta_m(K, \beta \times S^{d-1})
\]

for Borel sets \( \beta \subseteq \mathbb{R}^d \). If \( K \) is a polytope, then

\[
\binom{d-1}{m} \Theta_m(K, \eta) = \sum_{F \in \mathcal{F}(K)} \int_F \int_{n(K, F)} 1_\eta(x, u) \, d\mathcal{H}^{d-m-1}(u) \, d\mathcal{H}^m(x)
\]

for any Borel set \( \eta \subseteq \mathbb{R}^d \times S^{d-1} \); see [8, (4.2.2)].
Remark 7. If $C$ is finite, then (16) implies
\[ C_1(\text{conv } C, \cdot) = \kappa_{d-1} \sum_{F \in F_1(\text{conv } C)} \gamma(F, \text{conv } C) \mathcal{H}^1(F \cap \cdot), \]
and thus, (6) coincides with the definition (4) when $i = 1$.

For a convex body $K \subseteq \mathbb{R}^d$ we call
\[ F_K(u) = \{ x \in K : \langle x, u \rangle = h_K(u) \} \]
the support set of $K$ in direction $u \in S^{d-1}$, where $h_K : S^{d-1} \to \mathbb{R}$ is the support function of $K$ and $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product.

Lemma 8. Let $K \subseteq \mathbb{R}^d$ be a convex body and $m \in \{0, \ldots, d-2\}$. Set
\[ A_m = \{ u \in S^{d-1} : \dim F_K(u) \geq m + 1 \}. \]
Then $S_m(K, A_m) = 0$.

Proof. As a straight-forward generalization of [8, Theorem 2.2.9], we obtain
\[ \mathcal{H}^{d-m-1}(A_m) = 0. \]
By [8, Theorem 4.6.5] there is a constant $a$ such that
\[ S_m(K, \omega) \leq a \mathcal{H}^{d-m-1}(\omega) \]
for each $(\mathcal{H}^{d-m-1}, d-m-1)$-rectifiable set $\omega \subseteq S^{d-1}$, where we just want to mention that zero sets are always rectifiable and refer to [2, 3.2.14] for a complete definition. So, $S_m(K, A_m) = 0$.

We now summarize properties of the intrinsic power volumes as functionals on the family $\mathcal{E}$ of finite subsets $C$ of $\mathbb{R}^d$. In order to get a simplified expression for $V_i^{(m)}(C), m \geq 1$, we define the set
\[ F_1^*(C) = \{ [x,y] : x \neq y, [x,y] \cap C = \{ x,y \}, \]
there is $e \in F_1(\text{conv } C)$ with $[x,y] \subseteq e \}
of all refined edges of $\text{conv } C$, where every edge in $F_1(\text{conv } C)$ is partitioned into line segments such that exactly their endpoints are in $C$. Note that if $C = F_0(K)$ for a convex polytope $K$, then $F_1^*(C) = F_1(\text{conv } C)$, but in general this does not hold. For $C \in E$ and $e \in F_1^*(C)$ we put $N(\text{conv } C, e) = N(\text{conv } C, \tilde{e})$ and $\gamma(e, \text{conv } C) = \gamma(\tilde{e}, \text{conv } C)$, where $\tilde{e}$ is the unique edge of $\text{conv } C$ with $e \subseteq \tilde{e}$.

Lemma 9 (Properties of $V_i^{(m)}$). Let $i = 1, \ldots, d$, and $m \geq i$ be given.

(a) $V_i^{(i)}(C) = V_i(\text{conv } C)$ for any $C \in \mathcal{E}$. 

Documenta Mathematica 18 (2013) 275–295
(b) $V^{(m)}_i$ is homogeneous of degree $m$:

$$V^{(m)}_i(\alpha C) = \alpha^m V^{(m)}_i(C)$$

for all $\alpha \geq 0$ and all $C \in \mathcal{E}$.

(c) $V^{(m)}_i$ is motion invariant.

(d) $V^{(m)}_i$ is independent of the embedding space: Let $C$ be a finite set in $\mathbb{R}^{d'} \subseteq \mathbb{R}^d$ for some $d' < d$. Then $V^{(m)}_i(C)$ considered as a subset of $\mathbb{R}^{d'}$ coincides with $V^{(m)}_i(C)$ for the (lower-dimensional) subset $C$ of $\mathbb{R}^d$.

(e) Simplified expression for $V^{(m)}_1$: For finite $C \subseteq \mathbb{R}^d$ we have

$$V^{(m)}_1(C) = \sum_{F \in \mathcal{F}_1(\text{conv } C)} \gamma(F, \text{conv } C) V_1(F)^m.$$

Thus, $V^{(m)}_i$, as a functional on $\mathcal{E}$, has similar properties as the intrinsic volume on the family of convex bodies. However, in contrast to the latter, $V^{(m)}_i$ is in general not a valuation, i.e. it is not additive, and it is not continuous with respect to the Hausdorff metric. The properties (a)-(d) of Lemma 9 hold for arbitrary compact sets $C$ if $i = 1$.

3.2 Proof of Theorem 2

The proof of Theorem 2 is divided into a sequence of lemmas. Several times, the following analytical lemma will be needed. It can be shown using a Taylor expansion of order two.

**Lemma 10.** Let $n \in \mathbb{N}$, $0 \leq a \leq r$, and $f_n(r) = r^n - \sqrt{r^2 - a^2}$. Then

$$\frac{1}{2} a^2 r^{-1} \leq f_n(r) \leq \frac{1}{2} a^2 r^{-1} + \frac{\sqrt{2}}{4} a^4 r^{-3}, \quad \text{if } n = 1, \text{ and } a \leq \frac{r}{2},$$

$$\frac{3}{2} a^2 r - \frac{3\sqrt{2}}{8} a^4 r^{-1} \leq f_n(r) \leq \frac{3}{2} a^2 r, \quad \text{if } n = 3, \text{ and } a \leq \frac{r}{2},$$

$$\frac{n}{2} a^2 r^{n-2} - \frac{n(n-2)}{8} a^4 r^{n-4} \leq f_n(r) \leq \frac{n}{2} a^2 r^{n-2}, \quad \text{if } n \geq 4.$$

In short,

$$|f_n(r) - \frac{n}{2} a^2 r^{n-2}| \leq c_n a^4 r^{n-4} \quad (18)$$

where $c_2 = 0$ and $c_n = n(n-2)/8$ for $n \geq 4$. Inequality (18) also holds for $n = 1$ with $c_1 = \sqrt{2}/4$ and for $n = 3$ with $c_3 = 3\sqrt{2}/8$ if $0 \leq a \leq r/2$. In particular, the inequality $f_n(r) \leq na^2 r^{n-2}$, which is elementary for $n = 1$, holds for all $n$. 
For compact $C \subseteq \mathbb{R}^d$ we now show that the difference between the parallel volume of $K = \text{conv} \ C$ and the parallel volume of $C$ is approximately

$$I_C(r) = (d - 1) \int_{\mathbb{R}^d \times S^{d-1}} \int_0^\infty 1_{K \setminus C}(x + su) s^{d-2} ds \Theta_1(K, d(x, u)).$$

**Lemma 11.** Let $C \subseteq \mathbb{R}^d$ be a compact set and put $K = \text{conv} \ C$. Then there is a constant $c = c(C)$ with

$$0 \leq V_d(K_r) - V_d(C_r) - I_C(r) \leq c \cdot r^{d-4}$$

for all $r \geq \text{diam} \ C$.

**Proof.** Without loss of generality, we may assume that the diameter of $C$ is positive. If $x \in K_r \setminus C_r$, then due to $\text{ext} \ K := \bigcup_{F \in F_0(K)} F \subseteq C$, we have $p(K, x) / \notin \text{ext} \ K,$ (20) and an application of the Pythagorean theorem implies

$$d(K, x) > \sqrt{r^2 - (\text{diam} \ C)^2},$$

whenever $r \geq \text{diam} \ C$; see [5, Example 3.3 and Lemma 3.5]. Since $r \geq \text{diam} \ C$ implies $(K_r \setminus C_r) \cap K = 0$ and (20) together with Lemma 8 implies $\Theta_0(K, \{(p(K, x), u(K, x)) \mid x \in K_r \setminus C_r\}) = 0$, we get from (14) with $f = 1_{K_r \setminus C_r}$ that

$$V_d(K_r) - V_d(C_r) = V_d(K_r \setminus C_r)$$

$$= \sum_{m=1}^{d-1} \left( \frac{d-1}{m} \right) \int_{\mathbb{R}^d \times S^{d-1}} \int_0^\infty 1_{K_r \setminus C_r}(x + su) \times$$

$$\times s^{d-m-1} ds \Theta_m(K, d(x, u)).$$

(22)

The term on the right hand side of (22) corresponding to $m = 1$ is $I_C(r)$. We now consider the summands of the right hand side of (22) for which $m \geq 2$. For $d = 2$ no such summands exist. Since all these summands are non-negative for $d \geq 3$, the left inequality of the assertion is shown. The right inequality follows from the fact that – due to equations (21) and (15) and Lemma 10 – the right hand side of (22), without the summand for $m = 1$, is bounded from above by

$$\sum_{m=2}^{d-1} \left( \frac{d-1}{m} \right) \int_{\mathbb{R}^d \times S^{d-1}} \int_r^\infty s^{d-m-1} ds \Theta_m(K, d(x, u))$$

$$= \sum_{m=2}^{d-1} \omega_{d-m} \frac{1}{d-m} \left( r^{d-m} - (r^2 - (\text{diam} \ C)^2)^{(d-m)/2} \right) V_m(K)$$

$$\leq \sum_{m=2}^{d-1} \omega_{d-m}(\text{diam} \ C)^2 r^{d-m-2} V_m(K)$$

$$\leq cr^{d-4},$$

where $c = c(C)$.

**Documenta Mathematica 18 (2013) 275–295**
where \( c = \sum_{m=2}^{d-1} \omega_{d-m}(\text{diam } C)^{4-m} V_m(K) \).

An upper bound for \( I_C(r) \) is obtained easily.

**Lemma 12.** Let \( C \subseteq \mathbb{R}^d \) be a compact set and put \( K = \text{conv } C \). If \( r \geq 2(\text{diam } C) \), then

\[
I_C(r) \leq \frac{\omega_{d-1}}{2} V_1^{(3)}(C) r^{d-3}
\]

for \( d \geq 3 \) and

\[
I_C(r) \leq V_1^{(3)}(C) r^{-1} + \frac{\omega_2}{2} V_1^{(5)}(C) r^{-3}
\]

for \( d = 2 \).

**Proof.** Recall \( A_1 := \{ u \in S^{d-1} \mid \dim F_K(u) \geq 2 \} \). For any \( (x,u) \in ((bd K \setminus C) \times A_1) \cap N(K) \) the set \( F_K(u) \) is at most 1-dimensional due to \( x \notin A_1 \) and at least 1-dimensional due to \( x \notin C \). Thus we have \( p_x^C \neq x \) and \( q_x^C \neq x \), where \( p_x^C \) and \( q_x^C \) were defined before formula (6). Now

\[
1_{K \setminus C} (x+su) = 1_{\{ p_x^C \neq x \}} (x+su) \leq 1_{\{ p_x^C \neq x \}} (x+su)
\]

holds for all \((x,u) \in ((bd K \setminus C) \cap N(K) \) by Lemma 8. Setting \( a_x = d(\{ p_x^C \neq x \}, x) \), we get

\[
I_C(r) \leq (d-1) \int_{\mathbb{R}^d \times S^{d-1}} \int_0^\infty 1_{\{ p_x^C \neq x \}} (x+su) s^{d-2} ds \Theta_1(K, d(x,u))
\]

\[
= (d-1) \int_{\mathbb{R}^d \times S^{d-1}} \int_0^r \sqrt{r^2 - a_x^{d-1}} s^{d-2} ds \Theta_1(K, d(x,u))
\]

\[
= \int_{\mathbb{R}^d} \left( r^{d-1} - \sqrt{r^2 - a_x^{d-1}} \right) C_1(K, dx).
\]

Lemma 10 with \( n = d-1 \) and \( a = a_x \) yields

\[
I_C(r) \leq \frac{d-1}{2} \int_{\mathbb{R}^d} a_x^2 C_1(K, dx) r^{d-3}
\]

for \( d \geq 3 \), and

\[
I_C(r) \leq \frac{1}{2} \int_{\mathbb{R}^d} a_x^2 C_1(K, dx) r^{-1} + \frac{\omega_2}{2} \int_{\mathbb{R}^d} a_x^4 C_1(K, dx) r^{-3}
\]

for \( d = 2 \). In view of (6) this shows the assertion.

We now derive the corresponding asymptotic lower bound for \( I_C(r)^{-d/(d-3)} \).

**Lemma 13.** For any compact \( C \subseteq \mathbb{R}^d \) we have

\[
\liminf_{r \to \infty} \frac{I_C(r)}{r^{d-3}} \geq \frac{\omega_{d-1}}{2} V_1^{(3)}(C).
\]
Proof. Let $K$ be the convex hull of $C$. For $x \in \mathbb{R}^d$ set $\tau_x = (q^C_x - p^C_x) / \|q^C_x - p^C_x\|$ if $p^C_x \neq q^C_x$, and set $\tau_x = 0$, otherwise. The following arguments do not depend on the orientation of $\tau_x$. For $(x, u) \in \mathcal{N}(K)$, $\epsilon \geq 0$ and $\delta > 0$ we denote the indicator of the following event by $\zeta(x, u, \epsilon, \delta)$: for any $y \in C$ the implication

$$\langle x - y, u \rangle \leq \epsilon \implies \langle y, \tau_x \rangle < \langle p^C_x, \tau_x \rangle + \epsilon \text{ or } \langle y, \tau_x \rangle > \langle q^C_x, \tau_x \rangle - \delta$$

holds. Note that if $p^C_x = q^C_x$ (and in particular if $x \in C$), we have $\zeta(x, u, \epsilon, \delta) = 1$.

Fix $(x, u) \in \mathcal{N}(K)$ with $x \notin C$, and numbers $\delta > 0, \epsilon, s \geq 0$. Put

$$a_x = d((p^C_x, q^C_x), x) \leq \frac{\text{diam } C}{2}$$

and $r_x = \max\{(\text{diam } C)^2/(4\epsilon), (\text{diam } C)/2\}$. In order to find a lower bound for $I_C(r)$, we will first show the inequality

$$\zeta(x, u, \epsilon, \delta) \mathbf{1}_{\{\sqrt{r^2 - ((a_x - \delta)^+)^2} \leq s\}} \leq \mathbf{1}_{\{d(C, x + su) \geq r\}}, \quad (24)$$

for all $r \geq r_x$. So assume $\zeta(x, u, \epsilon, \delta) = 1$ and $\sqrt{r^2 - ((a_x - \delta)^+)^2} \leq s$. Let $y \in C$. If $\langle x - y, u \rangle \leq \epsilon$, then $\zeta(x, u, \epsilon, \delta) = 1$ implies

$$d(y, x + su)^2 \geq s^2 + \langle x - y, \tau_x \rangle^2 \geq s^2 + ((a_x - \delta)^+)^2 \geq r^2.$$ 

If $\langle x - y, u \rangle > \epsilon$, then

$$d(y, x + su)^2 \geq \langle x - y, u \rangle^2 + s^2 \geq \epsilon^2 + \sqrt{r^2 - ((a_x - \delta)^+)^2}^2 \geq r^2$$

for all $r \geq r_x$ by Lemma 10. So $d(C, x + su) \geq r$ for all $r \geq r_x$, which completes the proof of (24).

Now let $\epsilon, \delta > 0$. For all $r \geq r_x$ Lemma 8 and inequality (24) imply

$$I_C(r) = (d - 1) \int_{(bdK) \times A^C_1} \int_0^r \mathbf{1}_{\{d(C, x + su) \geq r\}} s^{d-2} ds \Theta_1(K, d(x, u))$$

$$\geq (d - 1) \int_{(bdK) \times A^C_1} \int_0^r \mathbf{1}_{\{d(C, x + su) \geq r\}} \zeta(x, u, \epsilon, \delta) s^{d-2} ds \Theta_1(K, d(x, u))$$

$$= \int_{(bdK) \times A^C_1} \zeta(x, u, \epsilon, \delta) \left(r^{d-1} - \sqrt{r^2 - ((a_x - \delta)^+)^2}^{d-1}\right) \Theta_1(K, d(x, u)).$$

Since

$$\lim_{r \to \infty} \frac{r^{d-1} - \sqrt{r^2 - ((a_x - \delta)^+)^2}^{d-1}}{r^{d-3}} = \frac{d-1}{2}((a_x - \delta)^+)$$

due to Lemma 10, Fatou’s lemma gives

$$\liminf_{r \to \infty} \frac{I_C(r)}{r^{d-3}} \geq \frac{d-1}{2} \int_{(bdK) \times A^C_1} \zeta(x, u, \epsilon, \delta) ((a_x - \delta)^+) \Theta_1(K, d(x, u)).$$
Now we first let $\epsilon \to 0$ using $\lim_{\epsilon \to 0} \zeta(x, u, \epsilon, \delta) = \zeta(x, u, 0, \delta)$, and the monotone convergence theorem. Since $\zeta(x, u, 0, \delta)((a_x - \delta)^+)^2 \leq (\text{diam } C)^2$, we can use the dominated convergence theorem to let $\delta \to 0$, and get

$$\lim_{r \to \infty} \inf_{r \to \infty} \frac{1}{r^{d-3}} \int_{(\text{bd } K) \times \mathcal{A}^C_1} \zeta(x, u, 0, \delta) a_x^2 \Theta_1(K, d(x, u)).$$

For all $(x, u) \in (\text{bd } K) \times \mathcal{A}^C_1$, we have $\lim_{\delta \to 0} \zeta(x, u, 0, \delta) = 1$, and (23) follows using (6) and Lemma 8.

Theorem 2 now follows directly from Lemmas 11, 12, and 13.

3.3 Proof of Proposition 3

Without loss of generality, may assume that $f(r) \geq \frac{1}{4}$ holds for all $r \in (0, \infty)$. Put $g : [0, 1] \to (-\infty, 0], x \mapsto -x/(6f^{-1}(x))$, and let $S = g(1) = 1$,

$$C = \{(x, g(x)) : x \in [0, 1]\} \cup \{(x, g(2 - x)) : x \in [1, 2]\} \cup \{(0, S), (2, S)\}$$

and

$$C^0 = \{(0, 0), (2, 0), (0, S), (2, S)\}.$$

For any $r \in \mathbb{R}$, large enough that $f(r) \leq 1/3$, we have $r \geq \frac{1}{f(r)} \geq 3$ and

$$V_2(C_r) - V_2(C_r^0) \geq \int_0^1 \left( \max\{y \in \mathbb{R} : (x, y) \in C_r\} - \max\{y \in \mathbb{R} : (x, y) \in C_r^0\} \right) dx$$

$$\geq \int_0^1 \left( \max\{y \in \mathbb{R} : \|\langle x, y \rangle - (f(r), g(f(r)))\| \leq r\} - \max\{y \in \mathbb{R} : (x, y) \in C_r^0\} \right) dx$$

$$= \int_0^1 \left( \sqrt{r^2 - (x - f(r))^2} + g(f(r)) - \sqrt{r^2 - x^2} \right) dx$$

$$\geq \int_0^1 \left( \frac{2xf(r) - f(r)^2}{2\sqrt{r^2 - (x - f(r))^2}} + g(f(r)) \right) dx$$

by using Lemma 10 with $r, a,$ and $n$ replaced by $\sqrt{r^2 - (x - f(r))^2}$, $\sqrt{2xf(r) - f(r)^2}$, and 1, respectively. Since $g(f(r)) = -\frac{f(r)}{6r}$, this integral can be estimated from below by

$$\int_0^1 \frac{2xf(r) - f(r)^2}{2r} - \frac{f(r)}{6r} dx = \frac{f(r)}{2r} - \frac{f(r)^2}{2r} - \frac{f(r)}{6r} \geq \frac{f(r)}{6r}.$$
Observing that $K = \text{conv} C = \text{conv} C^0$ and $V_1^{(3)}(C) = V_1^{(3)}(C^0)$ we conclude from Proposition 1 that there is a constant $c_1 \geq 0$ with

$$V_1^{(3)}(C) \frac{1}{r} - (V_2(K_r) - V_2(C_r))$$

$$= - (V_2((\text{conv} C^0)_r) - V_2(C^0_r)) + (V_2(C_r) - V_2(C^0_r))$$

$$\geq -\frac{c_1}{r^3} + \frac{f(r)}{6r}$$

$$\geq \frac{f(r)}{12r}$$

for all sufficiently large $r$. Hence

$$V_1^{(3)}(C) - \frac{V_2(K_r) - V_2(C_r)}{r^{-1}} \geq \frac{f(r)}{12}$$

for all sufficiently large $r$ and Proposition 3 is shown.

Remark 14. In order to show a statement analogous to Proposition 3 in higher dimensions, one can consider bodies of revolution.

3.4 Proof of Theorem 4

Theorem 4 is a consequence of Lemmas 11, 12 and the following result.

Lemma 15. For any finite set $C \subseteq \mathbb{R}^d$ there is a constant $c = c(C) > 0$ such that

$$I_C(r) \geq \frac{\omega_{d-1}}{2} V_1^{(3)}(C)r^{d-3} - cr^{d-4}$$

for all sufficiently large $r$.

Proof. We have

$$I_C(r) = (d - 1) \int_{(bd K) \times S^{d-1}} \int_0^r 1_{(C_r)}(x + su)s^{d-2}ds \Theta_1(K,d(x,u)).$$

As

$$1_{(C_r)}(x + su) = 1_{(p^C, q^C)}(x + su)(1 - 1_{(C \setminus (p^C, q^C))}, (x + su))$$

$$\geq 1_{(p^C, q^C)}(x + su)(1 - \sum_{y \in C} 1_{[0,r]}(\|x + su - y\|)),$$

and

$$1_{(p^C, q^C)}(x + su) = 1_{(\sqrt{r^2 - a^2} \to \infty)}(s)$$

(with $a_x = d((p^C_x, q^C_x), x))$, we get

$$I_C(r) \geq \int_{(bd K) \times S^{d-1}} \left( r^{d-1} - \sqrt{r^2 - a_x^2}^{d-1} \right) \Theta_1(K,d(x,u)) - \sum_{y \in C} J_C(r,y).$$

(26)
Here,

$$J_C(r,y) = (d-1) \int_{(bd\ K)\times S^{d-1}} \int_0^r \mathbf{1}_{(p\notin aF)}(x) \xi(x + sy) \times$$

$$\times \mathbf{1}_{[0,r]}(||x + sy - y||) s^{d-2} ds \Theta_1(K, d(x,y)).$$

(27)

By Lemma 16 below, $J_C(r,y) = O(r^{d-4})$ for all $y \in C$, as $r \to \infty$, and thus the second term on the right hand side of (26) is $O(r^{d-4})$. Lemma 10 with $n = d - 1$ and $a = a_d$ together with (6) shows that the first term on the right hand side of (26) is bounded from below by

$$\frac{\omega_{d-1}}{2} V_1^{(3)}(C) r^{d-3} - c_{d-1} \kappa_{d-1} V_1^{(5)}(C) r^{d-5}$$

for all sufficiently large $r$. Hence (25) follows from (26).

**Lemma 16.** Let $C \subseteq \mathbb{R}^d$ be finite and $y \in C$. Then there is a constant $c = c(C,y) > 0$ such that

$$J_C(r,y) \leq cr^{d-4}$$

for all sufficiently large $r$, where $J_C(r,y)$ is defined by (27).

**Proof.** From (27) and (16) we get

$$J_C(r,y) = \sum_{e \in \mathcal{E}_1(K)} \int_{n(K,e)} \int_0^r \int_0^{\eta_K} \zeta(x + sy) s^{d-2} ds dH^{d-2}(u) dH_1(x).$$

where $\zeta(z) = \mathbf{1}_{(relbd\ e)}(z) \mathbf{1}_{[0,r]}(||z - y||)$. Note that the relative boundary relbd\ e of e consists just of the two endpoints of e. Spherical coordinates in hyperplanes orthogonal to e and Fubini’s theorem give

$$J_C(r,y) = \sum_{e \in \mathcal{E}_1(K)} \int_{\mathbb{R}^d} \mathbf{1}_e(p(K,z)) \mathbf{1}_n(K,e)(u(K,z)) \zeta(z) dz.$$

(28)

Fix $e \in \mathcal{E}_1(K)$ and let $x_1$ and $x_2$ be its endpoints. We assume without loss of generality that e contains the origin. Let g be the affine hull of e and let L be the affine hull of e and y. We may assume $y \notin g$, since we cannot have $y \in \text{relint} e$ and we have $\mathbf{1}_e(p(K,z)) \mathbf{1}_{n(K,e)}(u(K,z)) \zeta(z) = 0$ for all $z \in \mathbb{R}^d$ if $y \in g \setminus \text{relint} e$. Let $H^+$ be the closed half space containing e in its boundary with normal vector $y - p(g,y)$, such that $y \notin H^+$. Finally let $V_y$ be the Voronoi cell of y with respect to the set $\{x_1, x_2, y\}$. The planar set

$$T = L \cap V_y \cap H^+ \cap (e + g^{-})$$

is either empty or a bounded triangle. For nonempty T let $\delta > 0$ be the maximal distance from a point of T to $\{x_1, x_2\}$. If $T = \emptyset$ put $\delta = 0$. If we can show that

$$\mathbf{1}_e(p(K,z)) \mathbf{1}_{n(K,e)}(u(K,z)) \zeta(z) \leq \mathbf{1}_T(z) |L| \mathbf{1}_{(\text{relbd} e)}(||z||)$$

(29)
holds for all $r > \delta$ and $z \in \mathbb{R}^d$, then
\[
\int_{\mathbb{R}^d} 1_e(p(K, z))1_{u(K, z)}(\zeta(z)) \, dz \\
\leq \int_{L^+} \int_{L} 1_{T(y_1)}1_{(\sqrt{r^2 - \delta^2}, r]}(\|y_2\|) \, dy_1 \, dy_2 \\
= V_2(T) \cdot \kappa_{d-2}(\sqrt{r^2 - \delta^2} - \delta^{d-2}) \\
\leq (d-2)\kappa_{d-2}V_2(T)\delta^2r^{d-4},
\]
for all $r > \delta$. The last inequality is evident in the case $d = 2$, and follows for $d \geq 3$ from Lemma 10 with $n = d - 2$ and $a = \delta$. Bounding all summands in (28) in such a way shows the assertion.

It remains to prove (29). Assume that the left hand side of (29) is one. Then $p(K, z) \in e$, $u(K, z) \in u(K, e)$, $\|z - x_1\| > r$, $\|z - x_2\| > r$, and $\|z - y\| \leq r$. The last three inequalities imply $z \in V_y$ and $z|L \in V_y|L = V_y \cap L$. The convexity of $K$ and $y \in K$ imply $\langle z - p(K, z), y - p(K, z) \rangle \leq 0$. Since both $z - p(K, z)$ and $y - p(g, y)$ are perpendicular to $g$, this gives $z \in H^+$. Finally,
\[
z \in e + N(K, e) \subseteq e + g^\perp
\]
gives $z|L \in T$. As $\|z|L^\perp\| \leq \|z - y\| \leq r$ and
\[
r^2 < d(x_1, x_2, z)^2 = d(x_1, x_2, z|L)^2 + d(z|L, z)^2 \leq \delta^2 + \|z|L^\perp\|^2,
\]
we have $\sqrt{r^2 - \delta^2} < \|z|L^\perp\| \leq r$, and (29) is shown.

### 3.5 Proof of Theorem 5

We first show a key observation: If Condition (A) holds, then, for sufficiently large $r$, the part of the difference set $K_r \setminus C_r$ that is projected on a face $F$ is independent of the points of $C$ outside $F$.

**Lemma 17.** Let $C \subseteq \mathbb{R}^d$ be a finite set satisfying Condition (A). If $K = \text{conv} C$, $m \in \{0, \ldots, d - 1\}$, and $F \in \mathcal{F}_m(K)$ then
\[
(K_r \setminus C_r) \cap (F + N(K, F)) = (F \setminus (C \cap F)_r) \cap (F + N(K, F)) \tag{30}
\]
for all sufficiently large $r$.

**Proof.** Since
\[
K_r \cap (F + N(K, F)) = F_r \cap (F + N(K, F))
\]
and $(C \cap F)_r \subseteq C_r$, the set on the left-hand side is contained in the set on the right-hand side. To show the opposite inclusion, let $V_y$ be the Voronoi cell of $y \in C$ with respect to $C$, let $S_y = V_y \cap (F + N(K, F))$ be the set of all the points in $V_y$ with metric projection in $F$, and define
\[
r_0 = \min\{r \geq 0 : S_y \subseteq (C \cap F)_r \text{ for all } y \in C \text{ with bounded } S_y\}.
\]
Let \( r > r_0 \) and assume

\[
x \in (F_r \setminus (C \cap F)_r) \cap (F + N(K, F)).
\]

Then, clearly, \( x \in K_r \). Moreover, we have \( \| x - y \| > r \) for all \( y \in C \cap F \). As \( \{ V_y : y \in C \} \) covers \( \mathbb{R}^d \), there is a \( y \in C \) with \( x \in V_y \cap (F + N(K, F)) = S_y \). The definition of \( r_0 \) and \( x \notin (C \cap F)_r \) imply that the closed convex set \( S_y \) is unbounded. Hence, there is a ray with direction \( \langle v, y \rangle \) contained in \( S_y \). It follows that \( v \in N(K, F) \) and, as the ray is contained in \( V_y \), that \( \langle y, v \rangle \geq h_K(v) \). Hence \( F \) and \( y \) are contained in a supporting hyperplane of \( K \) (with normal \( v \)), and thus they are contained in some facet \( G \) of \( K \). As \( x' = p(K, x) \in F \), Condition (A) implies that there is a point \( y' \in C \cap F \) with \( \| y - y' \| \geq \| x' \| \) and thus

\[
d(C, x)^2 = \| y - x \|^2 \geq \| y - y' \|^2 + \| x' - x \|^2 \\
\geq \| y' - x' \|^2 + \| x' - x \|^2 = \| y' - x \|^2 > r^2,
\]

where the first equality is due to \( x \in V_y \), the first inequality follows from \( \langle y, x - x' \rangle \leq \langle x', x - x' \rangle \) (see [8, Lemma 1.3.1]) and the last inequality is due to \( x \notin (C \cap F)_r \). Hence \( x \notin C_r \), which completes the proof of (30).

We now prove Theorem 5. Let \( C \subseteq \mathbb{R}^d \) be a finite set that satisfies Condition (A), and set \( K = \text{conv} \, C \). Assume \( r > \text{diam} \, C \). Due to (14) with \( f = 1_{K_r \setminus C_r} \), (20), and (16) we have

\[
V_d(K_r) - V_d(C_r) = \sum_{m=1}^{d-1} \sum_{F \in F_m(K)} I_{F,m}
\]

with

\[
I_{F,m} = \int_{F} \int_{n(K,F)} \int_{0}^{\infty} 1_{K \setminus C_f}(x + su)s^{d-m-1}ds \, dH^{d-m-1}(u) \, dH^m(x),
\]

\( m \in \{1, \ldots, d-1\}, F \in F_m(K) \). For all sufficiently large \( r \), Lemma 17 implies

\[
I_{F,m} = \int_{F} \int_{n(K,F)} \int_{0}^{\infty} 1_{F_r \setminus (C \cap F)_r}(x + su)s^{d-m-1}ds \, dH^{d-m-1}(u) \, dH^m(x)
\]

\[
= \int_{F} \int_{n(K,F)} \int_{0}^{\sqrt{r^2 - d(C \cap F; x)^2}} s^{d-m-1}ds \, dH^{d-m-1}(u) \, dH^m(x)
\]

\[
= \frac{\omega_{d-m}}{d-m} \gamma(F,K) \int_{F} \left( r^{d-m} - \sqrt{r^2 - d(C \cap F; x)^2} \right)^{d-m} \, dH^m(x).
\]

Put \( a_{F,x} = d(C \cap F, x) \). The binomial series

\[
r^{d-m} - \sqrt{r^2 - a_{F,x}^2} = r^{d-m} \sum_{k=1}^{\infty} (-1)^{k+1} \binom{(d-m)/2}{k} \left( \frac{a_{F,x}}{r} \right)^{2k}
\]

\[\text{Documenta Mathematica 18 (2013) 275–295}\]
converges absolutely as \( r > \text{diam} \, C \geq a_{F,x} \). Hence

\[
I_{F,m} = \sum_{k=1}^{\infty} (-1)^{k+1} \left( \frac{(d-m)/2}{k} \right)^{\kappa_{d-m} \gamma(F,K)} \int_{F} a_{F,x}^{2k} \, d\mathcal{H}^{m}(x) \, r^{d-m-2k} = \sum_{n \geq \frac{2}{d}} \sum_{m \leq 2(n+d-m)} (-1)^{(n+d-m+2)/2} \left( \frac{(d-m)/2}{n+d-m} \right) \times \\
\times \kappa_{d-m} \gamma(F,K) \int_{F} a_{F,x}^{n+d-m} \, d\mathcal{H}^{m}(x) \, r^{-n}.
\]

Substitution into (31) and definition (4) gives

\[
V_{d}(K_{r}) - V_{d}(C_{r}) = \sum_{m=1}^{d-1} \sum_{n \geq \frac{2}{d}} \sum_{m \leq 2(n+d-m)} (-1)^{(n+d-m+2)/2} \left( \frac{(d-m)/2}{n+d-m} \right) \times \\
\times \kappa_{d-m} V_{m(n+d)}(C) \, r^{-n} = \sum_{n=3-d}^{\infty} \sum_{m \geq 2(n+d-m)} (-1)^{(n+d-m+2)/2} \left( \frac{(d-m)/2}{n+d-m} \right) \times \\
\times \kappa_{d-m} V_{m(n+d)}(C) \, r^{-n}.
\]

If \( n \) is positive and even, the inner sum only involves integers \( m \) with the same parity as \( d \), and \( (d-m)/2 < (n+d-m)/2 \) are both integers implying \( (d-m)/2 = 0 \). Hence all coefficients of \( r^{-n} \) vanish for even \( n > 0 \).

This concludes the proof of Theorem 5.

### 3.6 The example of the unit cube.

We show (11). In arbitrary dimension \( d \) the number of \( i \)-dimensional faces of \([0,1]^{d}\) is

\[
\# \mathcal{F}_{i}([0,1]^{d}) = 2^{d-i} \binom{d}{d-i}.
\]

Using orthogonal projections and symmetry we get

\[
\gamma(F, [0,1]^{d}) = \gamma([0,1]^{d-1}) = (\# \mathcal{F}_{0}([0,1]^{d-1}))^{-1} = 2^{-(d-i)}
\]

for any \( F \in \mathcal{F}_{i}([0,1]^{d}) \), \( i = 0, \ldots, d-1 \). Thus definition (4) and a symmetry argument give

\[
V_{i}^{(m+i)}(C^{d}) = \binom{d}{d-i} \int_{[0,1]} d(C^{i}, x)^{m} dx \times \\
= 2^{i} \binom{d}{d-i} \int_{[0,1]} \|x\|^{m} dx \times \\
= 2^{-m} \binom{d}{d-i} \int_{[0,1]} \|x\|^{m} dx.
\]
This implies
\[ V_1^{(m+1)}(C^d) = \frac{d}{(m+1)2^m} \] (32)
for \( m \geq 0 \), and, introducing polar coordinates,
\[ V_2^{(m+2)}(C^d) = 2^{-m} \left( \frac{d}{d-2} \right) 2 \int_0^{\pi/4} \int_0^{1/\cos \varphi} \| r \cdot (\cos \varphi, \sin \varphi) \|^m r \, dr \, d\varphi \]
\[ = \frac{d(d-1)}{(m+2)2^m} \int_0^{\pi/4} \cos^{-(m+2)}(\varphi) \, d\varphi. \]

We put \( d_n := \int_0^{\pi/4} \cos^{-n}(\varphi) \, d\varphi \). Integrating \( \int_0^{\pi/4} \cos^{-(n+1)}(\varphi) \cos(\varphi) \, d\varphi \) by parts, we obtain the recurrence relation
\[ (n+1)d_{n+2} = 2^{n/2} + nd_n, \quad n \geq 0, \]
with starting value \( d_0 = \frac{\pi}{4} \). Induction gives
\[ d_{2m} = 2^{m-1} \frac{(m-1)!}{(2m-1)!!} \sum_{i=0}^{m-1} \frac{(2i-1)!!}{i!}, \]
and we arrive at
\[ V_2^{(2m+2)}(C^3) = \frac{3}{(m+1)4^m} d_{2m+2} \]
\[ = \frac{3(m!)}{(2m+1)!! (m+1)2^m} \sum_{i=0}^{m} \frac{(2i-1)!!}{i!}. \] (33)

As \( C^3 \) satisfies Condition (A), (11) follows by substituting (32) with \( d = 3 \), \( m = 2 \), and (33) into (10).

References


TRANSPORTATION-COST INEQUALITIES
ON PATH SPACE OVER MANIFOLDS WITH BOUNDARY

FENG-YU WANG

Received: May 9, 2011
Revised: October 11, 2011
Communicated by Friedrich Götze

ABSTRACT. Let \( L = \Delta + Z \) for a \( C^1 \) vector field \( Z \) on a complete Riemannian manifold possibly with a boundary. A number of transportation-cost inequalities on the path space for the (reflecting) \( L \)-diffusion process are proved to be equivalent to the curvature condition \( \text{Ric} - \nabla Z \geq -K \) and the convexity of the boundary (if exists). These inequalities are new even for manifolds without boundary, and are partly extended to non-convex manifolds by using a conformal change of metric which makes the boundary from non-convex to convex.

2010 Mathematics Subject Classification: 60J60, 58G60.
Keywords and Phrases: Transportation-cost inequality, curvature, second fundamental form, path space.

1 INTRODUCTION

In 1996 Talagrand [13] found that the \( L^2 \)-Wasserstein distance to the standard Guassian measure can be dominated by the square root of twice relative entropy. This inequality is called (Talagrand) transportation-cost inequality, and has been extended to distributions on finite- and infinite-dimensional spaces. In particular, this inequality was established on the path space of diffusion processes with respect to several different distances (i.e. cost functions); see e.g. [7] for the study on the Wiener space with the Cameron-Martin distance, [17, 15] on the path space of diffusions with the \( L^2 \)-distance, [18] on the Riemannian path space with intrinsic distance induced by the Malliavin gradient.

---

1Supported in Part By NNSFC(11131003), SRFDP and the Fundamental Research Funds for the Central Universities.
operator, and \([B, 27]\) on the path space of diffusions with the uniform distance. The main purpose of this paper is to investigate the Talagrand inequality on the path space of reflecting diffusion process, for which both the curvature and the second fundamental form of the boundary will take important roles.

Let \(M\) be a connected complete Riemannian manifold possibly with a boundary \(\partial M\). Let \(L = \Delta + Z\) for a \(C^1\) vector field \(Z\) on \(M\). Let \(X_t\) be the (reflecting if \(\partial M \neq \emptyset\)) diffusion process generated by \(L\) with initial distribution \(\mu \in \mathcal{P}(M)\), where \(\mathcal{P}(M)\) is the set of all probability measures on \(M\). Assume that \(X_t\) is non-explosive, which is the case if \(\partial M\) is convex and the curvature condition

\[
\text{Ric} - \nabla Z \geq -K
\]

holds for some constant \(K \in \mathbb{R}\). In this case, for any \(T > 0\), the distribution \(\Pi_T^\mu\) of \(X_{[0,T]} := \{X_t : t \in [0,T]\}\) is a probability measure on the (free) path space

\[
M^T := C([0,T]; M).
\]

When \(\mu = \delta_o\), the Dirac measure at point \(o \in M\), we simply denote \(\Pi_T^\delta_o = \Pi_T^o\).

For any nonnegative measurable function \(F\) on \(M_T\) such that \(\Pi_T^\mu(\{F\}) = 1\), one has

\[
\mu_T^\mu(dx) := \Pi_T^\mu(F) \mu(dx) \in \mathcal{P}(M).
\]

Let \(\rho\) be the Riemannian distance on \(M\); i.e. for \(x, y \in M\), \(\rho(x, y)\) is the length of the shortest curve on \(M\) linking \(x\) and \(y\). Then \(M^T\) is a Polish space under the uniform distance

\[
\rho_\infty(\gamma, \eta) = \sup_{t \in [0,T]} \rho(\gamma_t, \eta_t), \quad \gamma, \eta \in M^T.
\]

Let \(W_{2, \rho_\infty}\) be the \(L^2\)-Wasserstein distance (or \(L^2\)-transportation cost) induced by \(\rho_\infty\). In general, for any \(p \geq 1\) and for two probability measures \(\Pi_1, \Pi_2\) on \(M^T\),

\[
W_{p, \rho_\infty}(\Pi_1, \Pi_2) := \inf_{\pi \in \mathcal{C}(\Pi_1, \Pi_2)} \left\{ \int_{M_T \times M_T} \rho_\infty(\gamma, \eta)^p \pi(d\gamma, d\eta) \right\}^{1/p}
\]

is the \(L^p\)-Wasserstein distance (or \(L^p\)-transportation cost) of \(\Pi_1\) and \(\Pi_2\) induced by the uniform norm, where \(\mathcal{C}(\Pi_1, \Pi_2)\) is the set of all couplings for \(\Pi_1\) and \(\Pi_2\).

Before moving on, let us recall the Talagrand transportation-cost inequality established in \([B]\) on the path space over Riemannian manifolds without boundary. Let \(\partial M = \emptyset\) and \(\rho_o = \rho(o, \cdot)\). If

\[
|Z| \leq \psi \circ \rho_o
\]

holds for some constant \(K \in \mathbb{R}\). In this case, for any \(T > 0\), the distribution \(\Pi_T^\mu\) of \(X_{[0,T]} := \{X_t : t \in [0,T]\}\) is a probability measure on the (free) path space

\[
M^T := C([0,T]; M).
\]

When \(\mu = \delta_o\), the Dirac measure at point \(o \in M\), we simply denote \(\Pi_T^\delta_o = \Pi_T^o\).

For any nonnegative measurable function \(F\) on \(M_T\) such that \(\Pi_T^\mu(\{F\}) = 1\), one has

\[
\mu_T^\mu(dx) := \Pi_T^\mu(F) \mu(dx) \in \mathcal{P}(M).
\]

Let \(\rho\) be the Riemannian distance on \(M\); i.e. for \(x, y \in M\), \(\rho(x, y)\) is the length of the shortest curve on \(M\) linking \(x\) and \(y\). Then \(M^T\) is a Polish space under the uniform distance

\[
\rho_\infty(\gamma, \eta) = \sup_{t \in [0,T]} \rho(\gamma_t, \eta_t), \quad \gamma, \eta \in M^T.
\]

Let \(W_{2, \rho_\infty}\) be the \(L^2\)-Wasserstein distance (or \(L^2\)-transportation cost) induced by \(\rho_\infty\). In general, for any \(p \geq 1\) and for two probability measures \(\Pi_1, \Pi_2\) on \(M^T\),

\[
W_{p, \rho_\infty}(\Pi_1, \Pi_2) := \inf_{\pi \in \mathcal{C}(\Pi_1, \Pi_2)} \left\{ \int_{M_T \times M_T} \rho_\infty(\gamma, \eta)^p \pi(d\gamma, d\eta) \right\}^{1/p}
\]

is the \(L^p\)-Wasserstein distance (or \(L^p\)-transportation cost) of \(\Pi_1\) and \(\Pi_2\) induced by the uniform norm, where \(\mathcal{C}(\Pi_1, \Pi_2)\) is the set of all couplings for \(\Pi_1\) and \(\Pi_2\).

Before moving on, let us recall the Talagrand transportation-cost inequality established in \([B]\) on the path space over Riemannian manifolds without boundary. Let \(\partial M = \emptyset\) and \(\rho_o = \rho(o, \cdot)\). If

\[
|Z| \leq \psi \circ \rho_o
\]

holds for some constant \(K \in \mathbb{R}\). In this case, for any \(T > 0\), the distribution \(\Pi_T^\mu\) of \(X_{[0,T]} := \{X_t : t \in [0,T]\}\) is a probability measure on the (free) path space

\[
M^T := C([0,T]; M).
\]

When \(\mu = \delta_o\), the Dirac measure at point \(o \in M\), we simply denote \(\Pi_T^\delta_o = \Pi_T^o\).

For any nonnegative measurable function \(F\) on \(M_T\) such that \(\Pi_T^\mu(\{F\}) = 1\), one has

\[
\mu_T^\mu(dx) := \Pi_T^\mu(F) \mu(dx) \in \mathcal{P}(M).
\]

Let \(\rho\) be the Riemannian distance on \(M\); i.e. for \(x, y \in M\), \(\rho(x, y)\) is the length of the shortest curve on \(M\) linking \(x\) and \(y\). Then \(M^T\) is a Polish space under the uniform distance

\[
\rho_\infty(\gamma, \eta) = \sup_{t \in [0,T]} \rho(\gamma_t, \eta_t), \quad \gamma, \eta \in M^T.
\]

Let \(W_{2, \rho_\infty}\) be the \(L^2\)-Wasserstein distance (or \(L^2\)-transportation cost) induced by \(\rho_\infty\). In general, for any \(p \geq 1\) and for two probability measures \(\Pi_1, \Pi_2\) on \(M^T\),

\[
W_{p, \rho_\infty}(\Pi_1, \Pi_2) := \inf_{\pi \in \mathcal{C}(\Pi_1, \Pi_2)} \left\{ \int_{M_T \times M_T} \rho_\infty(\gamma, \eta)^p \pi(d\gamma, d\eta) \right\}^{1/p}
\]

is the \(L^p\)-Wasserstein distance (or \(L^p\)-transportation cost) of \(\Pi_1\) and \(\Pi_2\) induced by the uniform norm, where \(\mathcal{C}(\Pi_1, \Pi_2)\) is the set of all couplings for \(\Pi_1\) and \(\Pi_2\).

Before moving on, let us recall the Talagrand transportation-cost inequality established in \([B]\) on the path space over Riemannian manifolds without boundary. Let \(\partial M = \emptyset\) and \(\rho_o = \rho(o, \cdot)\). If

\[
|Z| \leq \psi \circ \rho_o
\]
holds for some positive function $\psi$ such that $\int_0^\infty \frac{1}{\psi(s)} \, ds = \infty$, then (see [6, Theorem 1.1])

$$W_{2,\rho_\infty}(F\Pi^T, \Pi^T_\infty) \leq \frac{2}{K}(e^{2KT} - 1)\Pi^T_\infty(F\log F), \quad F \geq 0, \Pi^T_\infty(F) = 1.$$  \hspace{1cm} (1.4)

According to [12, 4, 18], the log-Sobolev inequality for a smooth elliptic diffusion implies the Talagrand transportation-cost inequality with the intrinsic distance. So, (1.4) was proved in [6] by using a known damped log-Sobolev inequality on the path space and finite-dimensional approximations. To ensure the smoothness of the approximating diffusions, one needs the boundedness of curvature. To get rid of this condition, a sequence of new metric approximating the original one were constructed in [6], which satisfy (1.1) and have bounded curvatures. In this way (1.4) was established without using curvature upper bounds. But to realize this approximation argument, the technical condition (1.3) with $\int_0^\infty \frac{1}{\psi(s)} \, ds = \infty$ was adopted.

In this paper we adopt a different argument developed in [27] for diffusions on $\mathbb{R}^d$ by using the martingale representation theorem and Girsanov transformations, so that this technical condition was avoided. Furthermore, we present a number of cost inequalities which are equivalent to the convexity of $\partial M$ (if exists) and the curvature condition (1.1).

When $\partial M \neq \emptyset$, let $N$ be the inward unit normal vector field of $\partial M$. Then the second fundamental form of $\partial M$ is defined by

$$\llcorner(U, V) = -\langle \nabla U N, V \rangle, \quad U, V \in T\partial M,$$

where $T\partial M$ is the tangent space of $\partial M$. If $\llcorner \geq 0$, i.e. $\llcorner(U, U) \geq 0$ for all $U \in T\partial M$, we call $M$ (or $\partial M$) convex.

**Theorem 1.1.** Let $P_T(o, \cdot)$ be the distribution of $X_T$ with $X_0 = o$, and let $P_T$ be the corresponding semigroup. The following statements are equivalent to each other:

1. $\partial M$ is either convex or empty, and (1.1) holds.
2. For any $T > 0$, $\mu \in \mathcal{P}(M)$ and nonnegative $F$ with $\Pi^T_\mu(F) = 1$,

$$W_{2,\rho_\infty}(F\Pi^T_\mu, \Pi^T_\mu, F) \leq \frac{2}{K}(e^{2KT} - 1)\Pi^T_\mu(F\log F)$$

holds, where $\Pi^T_\mu \in \mathcal{P}(M)$ is fixed by (1.2).
3. (1.4) holds for any $o \in M$ and $T > 0$.
4. For any $o \in M$ and $T > 0$,

$$W_{2,\rho}(P_T(o, \cdot), fP_T(o, \cdot)) \leq \frac{2}{K}(e^{2KT} - 1)P_T(f\log f)(o),$$

$\quad f \geq 0, P_T f(o) = 1$. 

**Documenta Mathematica** 18 (2013) 297–322
For any $T > 0$, $\mu, \nu \in \mathcal{P}(M)$, and $p \geq 1$,

$$W_{p,\rho}(\Pi_T^\mu, \Pi_T^\nu) \leq e^{KT}W_{p,\rho}(\mu, \nu),$$

where $W_{p,\rho}$ is the $L^p$-Wasserstein distance for probability measures on $M$ induced by $\rho$.

For any $x, y \in M$ and $T > 0$,

$$W_{2,\rho}(P_T(x, \cdot), P_T(y, \cdot)) \leq e^{KT}\rho(x, y).$$

For any $T > 0$, $\mu \in \mathcal{P}(M)$, and $F \geq 0$ with $\Pi_T^\mu(F) = 1$,

$$W_{2,\rho\infty}(F\Pi_T^\mu, \Pi_T^\mu) \leq \left\{\frac{2}{K}(e^{2KT} - 1)\Pi_T^\mu(F \log F)\right\}^{1/2} + e^{KT}W_{2,\rho}(\mu^T, \mu).$$

(8) For any $\mu \in \mathcal{P}(M)$ and $C \geq 0$ such that $W_{2,\rho}(f\mu, \mu)^2 \leq C\mu(f \log f)$, $f \geq 0$, $\mu(f) = 1$,

there holds

$$W_{2,\rho\infty}(F\Pi_T^\mu, \Pi_T^\mu)^2 \leq \left(\frac{2}{K}(e^{2KT} - 1) + e^{KT}\sqrt{C}\right)^2 \Pi_T^\mu(F \log F),$$

$$F \geq 0, \Pi_T^\mu(F) = 1.$$
3 we prove Theorem 1.1 by using results in Section 2, the martingale representation and Girsanov transformation for (reflecting) diffusions on (convex) manifolds.

To establish transportation-cost inequalities on the path space for non-convex manifolds, we shall adopt a conformal change of metric \( \langle \cdot, \cdot \rangle' = f^{-2} \langle \cdot, \cdot \rangle \) such that \( \partial M \) is convex under the new metric (see [21, Lemma 2.1]). Let \( \Delta' \) be the Laplacian induced by the new metric, we have (see [21, Lemma 2.2])

\[
L = f^{-2} \left\{ \Delta' + f^2 Z + \frac{d - 2}{2} \nabla f^2 \right\}. \tag{1.5}
\]

According to this fact, we will modify our arguments in Section 4 to study the reflecting diffusion process generated by \( L := \psi^2 (\Delta + Z) \) for a smooth function \( \psi \) on a convex manifold, then extend Theorem 1.1 in Section 5 to the non-convex setting.

2 Formulae for the second fundamental form and applications

When \( M \) is compact, the following formula on \( \partial M \) has been found in [22]:

\[
\lim_{t \to 0} \frac{\| \nabla f \|^2}{\sqrt{t}} \log \left( \frac{\| \nabla P_t f \|}{(P_t \| \nabla f \|^p)^{1/p}} \right) = -\frac{2}{\sqrt{\pi}} I(\nabla f, \nabla f), \quad p \geq 1, \tag{2.1}
\]

where \( f \) is a smooth function satisfying the Neumann boundary condition. When \( M \) is non-compact, some technical problems appear in the original proof when e.g. a dominated convergence is used. To fix these problems, we shall stop the process in a compact domain, so that we shall first study the behavior of hitting times.

Recall that the reflecting \( L \)-diffusion process can be constructed by solving the SDE

\[
dX_t = \sqrt{2} \Phi_t \circ dB_t + Z(X_t)dt + N(X_t)dl_t, \tag{2.2}
\]

where \( \Phi_t \) is the horizontal lift of \( X_t \) onto the frame bundle \( O(M) \), \( B_t \) is the \( d \)-dimensional Brownian motion.

By the It\( ô \) formula, for any \( f \in C^2(M) \) we have

\[
df(X_t) = \sqrt{2} (\nabla f(X_t), \Phi_t \circ dB_t) + Lf(X_t)dt + Nf(X_t)dl_t, \tag{2.3}
\]

where \( Nf = \langle N, \nabla f \rangle \). For any \( R > 0 \), let

\[
\tau_R = \inf\{ t \geq 0 : \rho(X_0, X_t) \geq R \}.
\]

Proposition 2.1. Let \( R > 0 \) and \( X_0 = o \in M \) be fixed. Then there exist two constants \( c_1, c_2 > 0 \) such that

\[
P(\tau_R \leq t) \leq c_1 e^{-c_2/t}, \quad t > 0.
\]
Proof. This result is well known on manifolds without boundary (cf. [2, Lemma 2.3]), and the proof works also when \( \partial M \) is convex. As in the present case the boundary is not necessarily convex, we shall follow [21] to make the boundary convex under a conformal change of metric. Since
\[
B_R := \{ x \in M : \rho(o, x) \leq R \}
\]
is compact, there exists a constant \( \sigma > 0 \) such that \( I \geq -\sigma \) holds on \( \partial M \cap B_R \).
Let \( f \geq 1 \) be smooth such that
\[
N \log f \geq \sigma \quad \text{on} \quad \partial M \cap B_R.
\]
Such a function can be constructed by using the distance function \( \rho_o \) to the boundary \( \partial M \). Since \( \rho_o \) is smooth in a neighborhood of the boundary, there exists a constant \( r_0 > 0 \) such that \( \rho_o \) is smooth on \( \{ x \in B_{2R} : \rho_o(x) \leq r_0 \} \).
Let \( h \in C^\infty([0, \infty)) \) such that \( h' \geq 0, h(0) = 1, h'(r) = 0 \) for \( r \geq r_0 \). Then \( h \circ \rho_o \) is smooth on \( B_{2R} \) and \( N \log h \circ \rho_o |_{\partial M \cap B_{2R}} = \sigma \). Thus, it suffices to take smooth \( f \geq 1 \) such that \( f = h \circ \rho_o \) on \( B_R \).
By [21] Lemma 2.1 and (2.4), \( \partial M \) is convex in \( B_R \) under the new metric
\[
\langle \cdot, \cdot \rangle' = f^{-2} \langle \cdot, \cdot \rangle,
\]
where \( \langle \cdot, \cdot \rangle \) is the original metric. Let \( \Delta' \) be the Laplacian induced by the new metric. We have (see [21] Lemma 2.2)
\[
L = f^{-2} (\Delta' + Z')
\]
for some \( C^1 \)-vector field \( Z' \). Let \( \tilde{\rho}_o \) be the Riemannian distance to \( o \) induced by the new metric. By the Laplacian comparison theorem,
\[
L \tilde{\rho}_o^2 \leq c \quad \text{on} \quad B_R
\]
holds for some constant \( c > 0 \) outside the cut-locus induced by \( \langle \cdot, \cdot \rangle' \). Since \( \partial M \) is convex on \( B_R \) and \( N \) is still the inward normal vector under the new metric, we have
\[
N \tilde{\rho}_o \leq 0 \quad \text{on} \quad \partial M \cap B_R.
\]
Therefore, by using Kendall’s Itô formula for the distance (cf. [3] for \( f = 1 \), [21] implies
\[
d \tilde{\rho}_o^2(X_t) \leq 2\sqrt{2} f^{-2}(X_t) \tilde{\rho}_o(X_t) dt + c dt, \quad t \leq \tau R,
\]
or equivalently,
\[
t \mapsto 2\sqrt{2} \int_0^{t \wedge \tau R} f^{-2}(X_s) \tilde{\rho}_o(X_s) ds + c(t \wedge \tau R) - \tilde{\rho}_o^2(X_{t \wedge \tau R})
\]
is an increasing process, where \( b_t \) is some one-dimensional Brownian motion. Since \( f^{-2} \leq 1 \), this implies that for any \( \delta > 0 \), the process

\[ \text{Documenta Mathematica 18 (2013) 297–322} \]
\[ Z_s := \exp \left[ \frac{\delta_s}{t} \rho_o(X_s) - \frac{\delta c s}{t} - \frac{4 \delta^2 t}{t^2} \int_{0}^{s} \rho_o(X_u) du \right], \quad s \leq \tau_R \]

is a super martingale. Therefore, letting \( C > 1 \) be a constant such that \( f \leq C \) on \( B_R \) and thus, \( \rho_o \geq \tilde{\rho}_o \geq C^{-1} \rho_o \) holds on \( B_R \), we obtain

\[
P(\tau_R \leq t) = P \left( \max_{s \in [0,t]} \rho_o(X_{s \wedge \tau_R}) \geq R \right) \leq P \left( \frac{R}{C} \geq \max_{s \in [0,t]} \tilde{\rho}_o(X_{s \wedge \tau_R}) \geq \frac{R}{C} \right)
\]

\[
\leq P \left( \max_{s \in [0,t]} Z_{s \wedge \tau_R} \geq \exp \left[ \frac{\delta R^2}{tC^2} - \delta c - \frac{4 \delta^2 R^2}{t} \right] \right)
\]

\[
\leq \exp \left[ c\delta - \frac{R^2}{tC^2} (\delta - 4C^2 \delta^2) \right], \quad \delta > 0.
\]

The proof is then completed by taking e.g. \( \delta = 1/(8C^2) \).

**Proposition 2.2.** Let \( X_0 = o \in \partial M \). Then for any \( R > 0 \),

\[
\limsup_{t \to 0} \frac{1}{t} \left| \mathbb{E} \left[ l_{t \wedge \tau_R} - 2 \sqrt{t/\pi} \right] \right| \leq \infty.
\]

**Proof.** Repeating the proof of [22, Lemma 2.2] by using \( t \wedge \tau_R \) in place of \( t \), we obtain

\[
\mathbb{E} l_{t \wedge \tau_R} \leq ct, \quad t \in [0,1] \tag{2.6}
\]

for some constant \( c > 0 \). Let \( r_0 > 0 \) be such that \( \rho_o \) is smooth on \( \{ \rho_o \leq r_0 \} \cap B_R \).

Let

\[
\tau = \inf \{ t \geq 0 : \rho_o(X_t) \geq r_0 \}.
\]

By the Itô formula we have

\[
d\rho_o(X_t) = \sqrt{2} dB_t + L \rho_o(X_t) dt + \frac{dt}{2}, \quad t \leq \tau \wedge \tau_R, \tag{2.7}
\]

where, as before, \( b_t \) is some one-dimensional Brownian motion. By the proof of [22, Theorem 2.1] using \( \tau \wedge \tau_R \) in place of \( \tau \), we have, instead of (2.4) in [22],

\[
\mathbb{E} (\rho_o(X_{t \wedge \tau \wedge \tau_R}) - \sqrt{2} |b_{t \wedge \tau \wedge \tau_R}|)^2 \leq c_1 t^2, \quad t \in [0,1] \tag{2.8}
\]

for some constant \( c_1 > 0 \), where \( \tilde{b}_t \) is some one-dimensional Brownian motion. Due to (2.7),

\[
\left| \mathbb{E} l_{t \wedge \tau \wedge \tau_R} - \mathbb{E} \rho_o(X_{t \wedge \tau \wedge \tau_R}) \right| \leq c_2 t
\]

holds for some constant \( c_2 > 0 \). Combining this with (2.8) we arrive at

\[
\left| \mathbb{E} l_{t \wedge \tau \wedge \tau_R} - \sqrt{2} \mathbb{E} |\tilde{b}_{t \wedge \tau \wedge \tau_R}| \right| \leq c_3 t, \quad t \in [0,1]
\]
for some constant $c_3 > 0$. Since $\mathbb{E}|\hat{b}_t| = \sqrt{2t/\pi}$ and $\mathbb{E}|\hat{b}_t|^2 = t$, this and (2.3) imply

$$\left|\mathbb{E}I_{t\wedge\tau_R} - \frac{2\sqrt{t}}{\sqrt{\pi}}\right| = \left|\mathbb{E}I_{t\wedge\tau_R} - \sqrt{2}\mathbb{E}|\hat{b}_t|\right|$$

$$\leq c_3 t + \mathbb{E}I_{1_{t\geq\tau\wedge\tau_R}}(t_{\wedge\tau_R} + \sqrt{2}|\hat{b}_t|)$$

$$\leq c_3 t + c_4 \sqrt{\log(t)}(t \geq \tau \wedge \tau_R), \quad t \in [0, 1].$$

Moreover, noting that

$$\mathbb{P}(\tau \wedge \tau_R \leq t, \tau_R > \tau) \leq \mathbb{P}\left(\max_{s \in [0, t]} \rho_0(X_{s\wedge\tau \wedge \tau_R}) \geq r_0\right),$$

by using $\tau \wedge \tau_R$ to replace $\tau$ in the proof of [22, Proposition A.2], we conclude that

$$\mathbb{P}(\tau \wedge \tau_R \leq t, \tau_R > \tau) \leq c_5 \exp[-r_0^2/(16t)], \quad t > 0$$

holds for some constant $c_5 > 0$. Combining this with Proposition 2.1, we obtain

$$\mathbb{P}(t \geq \tau \wedge \tau_R) \leq c_6 e^{-c_7/t}, \quad t > 0$$

for some constants $c_6, c_7 > 0$. Therefore, the proof is completed by (2.9).

**Theorem 2.3.** Let $f \in C^\infty(M)$ with $Nf|_{\partial M} = 0$. 

1. For any $p \geq 1$ and $R > 0$,

$$\lim_{t \to 0} \frac{\|\nabla f\|^2}{\sqrt{t}} \log \left(\frac{\mathbb{E}(|\nabla f|^p(X_{t\wedge\tau_R}))^{1/p}}{|\nabla f|}\right) = \frac{2}{\sqrt{\pi}} \|\nabla f, \nabla f\| (2.10)$$

holds at points on $\partial M$ such that $|\nabla f| > 0$.

2. Assume that for any $g \in C^\infty_0(M)$ the function $|\nabla Pg|$ is bounded on $[0, 1] \times M$. If moreover $f$ has a compact support, then (2.7) holds points on $\partial M$ such that $|\nabla f| > 0$.

**Proof.** (2.10) follows immediately from the proof of [22, Theorem 1.2] by using Proposition 2.2 in place of [22, Theorem 2.1], and using $t \wedge \tau_R$ in place of $t$. Next, let $f \in C^\infty_0(M)$. By the assumption of (2) and that $L^f \in C^\infty_0(M)$, $|\nabla PLf|$ is bounded on $[0, 1] \times M$. So, the proof of [22, (3.1)] implies that

$$\lim_{t \to 0} \frac{\|\nabla f\|^2}{\sqrt{t}} \log \left(\frac{\mathbb{E}(|\nabla f|^p(X_{t\wedge\tau_R}))^{1/p}}{|\nabla f|}\right) = -\lim_{t \to 0} \frac{\|\nabla f\|^2}{\sqrt{t}} \log \left(\frac{\mathbb{E}(P_t|\nabla f|^p)^{1/p}}{|\nabla f|}\right). (2.11)$$

Since by Proposition 2.1 there exist two constant $c_1, c_2 > 0$ such that

$$|P_t|\nabla f|^p - \mathbb{E}|\nabla f|^p(X_{t\wedge\tau_R})| \leq ||\nabla f||_\infty \mathbb{P}(t > \tau_R) \leq c_1 e^{-c_2/t}, \quad t > 0,$$

we conclude that (2.11) follows from (2.11) and (2.10).
As an application of (2.10), the following result provides equivalent semigroup log-Sobolev/Poincaré inequalities for Theorem 1.1(1).

**Theorem 2.4.** Each of the following statements is equivalent to Theorem 1.1(1):

(9) For any $T > 0$ and $f \in C_b(M)$,

$$P_T f^2 \log f^2 \leq (P_T f^2) \log P_T f^2 + \frac{e^{2KT} - 1}{2K} \|\nabla f\|^2.$$  

(10) For any $T > 0$ and $f \in C_b(M)$,

$$P_T f^2 \leq (P_T f)^2 + \frac{e^{2KT} - 1}{K} \|\nabla f\|^2.$$

**Proof.** According to e.g. [16, Lemma 3.1], which holds also for the non-symmetric case, Theorem 1.1(1) implies the semigroup log-Sobolev inequality (9). It is well known that the log-Sobolev inequality implies the Poincaré inequality. So, (10) follows from (9). Hence, it remains to show that (10) implies Theorem 1.1(1). Below we shall prove the convexity of $\partial M$ and the curvature condition (1.1) respectively.

(a) Let $\partial M \neq \emptyset$. For any $o \in \partial M$ and non-trivial $U \in T_o \partial M$, we aim to show that $I(U, U) \geq 0$. Let $f \in C_0^\infty(M)$ such that $Nf|_{\partial M} = 0$ and $\nabla f(o) = U$. Let $X_0 = o$ and $\tau_1 = \inf\{t \geq 0 : \rho(o, X_t) \geq 1\}$.

Since $f$ and $f^2$ satisfies the Neumann boundary condition, we have

$$\mathbb{E} f(X_{t \wedge \tau_1}) = f(o) + \mathbb{E} \int_0^{t \wedge \tau_1} Lf(X_s)ds,$$

$$\mathbb{E} f^2(X_{t \wedge \tau_1}) = f^2(o) + 2\mathbb{E} \int_0^{t \wedge \tau_1} (fLf)(X_s)ds + 2\mathbb{E} \int_0^{t \wedge \tau_1} |\nabla f|^2(X_s)ds.$$

So,

$$\mathbb{E} f^2(X_{t \wedge \tau_1}) - \{\mathbb{E} f(X_{t \wedge \tau_1})\}^2 = 2\mathbb{E} \int_0^{t \wedge \tau_1} \{f(X_s) - f(X_0)\} Lf(X_s)ds$$

$$- \left( \mathbb{E} \int_0^{t \wedge \tau_1} Lf(X_s)ds \right)^2 + 2\mathbb{E} \int_0^{t \wedge \tau_1} |\nabla f|^2(X_s)ds. \quad (2.12)$$

Since $Lf$ is bounded on $B_1 := \{x : \rho(o, x) \leq 1\}$, we have

$$\left( \mathbb{E} \int_0^{t \wedge \tau_1} Lf(X_s)ds \right)^2 \leq ct^2 \quad (2.13)$$

**Documenta Mathematica 18 (2013) 297–322**
for some $c > 0$. Moreover, due to Proposition 2.1,

$$P(\tau_1 \leq t) \leq c_1 e^{-c_2 / t}, \quad t > 0 \quad (2.14)$$

holds for some constants $c_1, c_2 > 0$. Thus,

$$|P_t f^2(o) - (P_t f)^2(o) - (E f^2(X_{t \wedge \tau_1}) - \{E f(X_{t \wedge \tau_1})\}^2)| = o(t^2),$$

$$\mathbb{E} \int_0^{t \wedge \tau_1} |\nabla f|^2(X_s) ds = t|\nabla f(o)|^2 + \int_0^t \mathbb{E} \{ |\nabla f|^2(X_{s \wedge \tau_1}) - |\nabla f(o)|^2 \} ds + o(t^2), \quad (2.15)$$

where and in what follows, $o(s)$ stands for a function of $s > 0$ such that

$$\lim_{s \to 0} o(s) / s = 0.$$

Similarly, applying the Itô formula to $\{ f(X_s) - f(o) \} Lf(X_s)$, we obtain (note that $Nf|_{\partial M} = 0$)

$$\mathbb{E} \int_0^{t \wedge \tau_1} \{ f(X_s) - f(o) \} Lf(X_s) ds$$

$$= o(t^2) + \int_0^t \mathbb{E} \{ (f(X_{s \wedge \tau_1}) - f(o)) Lf(X_{s \wedge \tau_1}) \} ds$$

$$= o(t^2) + \mathbb{E} \int_0^t ds \int_0^{s \wedge \tau_1} L((f - f(o)) Lf)(X_r) dr$$

$$+ \mathbb{E} \int_0^t ds \int_0^{s \wedge \tau_1} \{ (f - f(o)) NLf \}(X_r) dl_r. \quad (2.16)$$

Noting that

$$f(X_r) - f(o) = \sqrt{2} \int_0^r \langle \nabla f(X_u), \Phi_u \circ dB_u \rangle + \int_0^r Lf(X_u) du, \quad u \leq \tau_1,$$

and that

$$\mathbb{E} \sup_{r \in [0, t]} \left( \int_0^r \langle \nabla f(X_u), \Phi_u \circ dB_u \rangle \right)^2 \leq c_2 t, \quad t \in [0, 1]$$

holds for some constant $c_2 > 0$, we obtain from (2.16) and (2.6) that

$$\mathbb{E} \int_0^{t \wedge \tau_1} \{ f(X_s) - f(o) \} Lf(X_s) ds \leq c_3 t^2, \quad t \in [0, 1] \quad (2.17)$$

holds for some constant $c_3 > 0$. Finally, by Theorem 2.3(1), we have

$$\mathbb{E} |\nabla f|^2(X_{t \wedge \tau_1}) = |\nabla f|^2(o) + \frac{4 \sqrt{t}}{\sqrt{\pi}} \lVert \nabla f \rVert + o(t^{1/2}) \quad (2.18)$$
Combining this with (2.12), (2.13), (2.15) and (2.17), and noting that $U = \nabla f(o)$, we conclude that
\[ P tf^2(o) - (P tf)^2(o) = 2 t |\nabla f(o)|^2 + \frac{16 t^{3/2}}{3\sqrt{\pi}} \mathbb{I}(U, U) + o(t^{3/2}). \]  
(2.19)

Finally, (2.18) and (2.14) imply that
\[ \frac{e^{2Kt}}{K} P tf^2(o) = 2 t |\nabla f(o)|^2 + \frac{8t^{3/2}}{\sqrt{\pi}} \mathbb{I}(U, U) + o(t^{3/2}). \]

Since $\frac{16}{3} < 8$, combining this with (10) and (2.19) we conclude that $\mathbb{I}(U, U) \geq 0$.

(b) Let $X_o = o \in M \setminus \partial M$, we aim to show that $\text{Ric} - \nabla Z \geq - K$ holds on $T_o M$.

Let $R > 0$ such that $B_R \cap \partial M = \emptyset$. Since $l_t$ increases only when $X_t \in \partial M$, $l_t = 0$ for $t \leq \tau_R$. Hence, due to Proposition 2.1 for any $f \in C^\infty_o(M)$,

\[ P tf^2(o) - (P tf)^2(o) = o(t^2) + \mathbb{E} f^2(X_{t \wedge \tau_R}) - \left( \mathbb{E} f(X_{t \wedge \tau_R}) \right)^2. \]

By the continuity of $s \mapsto \mathbb{L} f(X_{s \wedge \tau_R})$, we have
\[ \left( \int_0^t \mathbb{E} f(X_{s \wedge \tau_R}) ds \right)^2 = (L f)^2(o) t^2 + o(t^2). \]

Similarly, it is easy to see that
\[ \mathbb{E} f^2(X_{s \wedge \tau_R}) - 2 f(o) \mathbb{E} L f(X_{s \wedge \tau_R}) \]
\[ = L f^2(o) - 2 f(o) L f(o) + s \{ L f^2 - 2 f LL f \}(0) + o(s) \]
\[ = 2 |\nabla f|^2(o) + 2 s \{ |\nabla f|^2 + (L f)^2(o) + 2 \langle \nabla f, \nabla L f \rangle(o) \} + o(s). \]

Combining this with (2.20) and (2.21) we obtain
\[ P tf^2(o) - (P tf)^2(o) = 2 t |\nabla f(o)|^2 + t^2 \{ L |\nabla f|^2 + 2 \langle \nabla f, \nabla L f \rangle \}(o) + o(t^2). \]

(2.22)

Finally, by Proposition 2.1 and noting that $l_s = 0$ for $s \leq \tau_R$, we have
\[ P t |\nabla f|^2(o) = o(t^2) + \mathbb{E} |\nabla f|^2(X_{t \wedge \tau_R}) = |\nabla f|^2(o) + t L |\nabla f|^2(o) + o(t). \]

Combining this with (10) and (2.22), we conclude that
\[ \frac{1}{2} L |\nabla f|^2(o) - \langle \nabla f, \nabla L f \rangle(o) \geq - K |\nabla f|(o), \quad f \in C^\infty_o(M). \]

This completes the proof by the Bochner-Weitzenböck formula.
3 Proof of Theorem 1.1

By taking \( \mu = \delta_\alpha \), we have \( \mu_T^T = \Pi_\alpha^T(F) \delta_\alpha = \delta_\alpha \). So, (3) follows from each of (2), (7) and (8). Next, (4) follows from (3) by taking \( F(X_{[0,T]}) = f(X_T) \), and (5) implies (6) by taking \( p = 2 \) and \( \mu = \delta_x, \nu = \delta_y \). Moreover, it is clear that (8) follows from (7) while (7) is implied by (2) and (5). So, it suffices to prove that (1) \( \Rightarrow \) (3) \( \Rightarrow \) (2), (4) \( \Rightarrow \) (1) \( \Rightarrow \) (6) \( \Rightarrow \) (5) and (6) \( \Rightarrow \) (1), where "\( \Rightarrow \)" stands for "implies".

(a) (1) \( \Rightarrow \) (3). We shall only consider the case where \( \partial M \) is non-empty and convex. For the case without boundary, the following argument works well by taking \( \lambda_t = 0 \) and \( N = 0 \). The idea of the proof comes from [27], where elliptic diffusions on \( \mathbb{R}^d \) were concerned. Let \( B_t \) be the \( d \)-dimensional Brownian motion on the naturally filtered probability space \( (\Omega, \mathcal{F}_t, P) \). Let \( \{ X_t : t \geq 0 \} \) solve (2.2) with \( X_0 = o \).

Next, let \( F \) be a positive bounded measurable function on \( M_T \) such that \( \inf F > 0 \) and \( \Pi_\mu^T(F) = 1 \). Then

\[
 m_t := E_P(F(X_{[0,T]}))|\mathcal{F}_t) \quad \text{and} \quad L_t := \int_0^t \frac{dm_s}{m_s}, \quad t \in [0,T]
\]

are square-integrable \( \mathcal{F}_t \)-martingales under \( P \), where \( E_P \) is the expectation taken for the probability measure \( P \). Obviously, we have

\[
 m_t = e^{L_t - \frac{1}{2} \langle L \rangle_t}, \quad t \in [0,T]. \tag{3.1}
\]

Since \( \mathcal{F}_t \) is the natural filtration of \( B_t \), by the martingale representation theorem (cf. [8, Theorem 6.6]), there exists a unique \( \mathcal{F}_t \)-predictable process \( \beta_t \) on \( \mathbb{R}^d \) such that

\[
 L_t = \int_0^t \langle \beta_s, dB_s \rangle, \quad t \in [0,T]. \tag{3.2}
\]

Let \( dQ = F(X_{[0,T]})d\mathbb{P} \). Since \( E_P F(X_{[0,T]}) = \Pi_\mu^T(F) = 1 \), \( Q \) is a probability measure on \( \Omega \). Due to (3.1) and (3.2) we have

\[
 F(X_{[0,T]}) = m_T = e^{\int_0^T \langle \beta_s, dB_s \rangle - \frac{1}{2} \int_0^T \| \beta_s \|^2 ds}.
\]

Moreover, by the Girsanov theorem,

\[
 \tilde{B}_t := B_t - \int_0^t \beta_s ds, \quad t \in [0,T] \tag{3.3}
\]

is a \( d \)-dimensional Brownian motion under the probability measure \( Q \).

Let \( Y_t \) solve the SDE

\[
 dY_t = \sqrt{2} P_{X_t,Y_t} \Phi_t \circ dB_t + Z(Y_t)dt + N(Y_t)d\tilde{l}_t, \quad Y_0 = o, \tag{3.4}
\]

where \( P_{X_t,Y_t} \) is the parallel displacement along the minimal geodesic from \( X_t \) to \( Y_t \) and \( \tilde{l}_t \) is the local time of \( Y_t \) on \( \partial M \). As explained in e.g. [10] Section
Finally, since (3.1) and (3.2) yield
\[ E \parallel \beta \parallel^2 = E\left(\parallel \beta_s \parallel^2 \right), \]
we conclude that under \( Q \) the distribution of \( X_{[0,T]} \) coincides with \( F\Pi_T^\rho \). Therefore,
\[ W_{2,\infty}(F\Pi_T^\rho, \Pi_T^\rho) \leq E_Q(\rho(0,T), Y_{[0,T]})^2 = E_Q \max_{t\in[0,T]} \rho(X_t, Y_t)^2. \]
By the convexity of \( \partial M \) we have
\[ \langle N(x), \nabla \rho(y, \cdot)(x) \rangle = \langle N(x), \nabla \rho(\cdot, y)(x) \rangle \leq 0, \quad x \in \partial M. \]
Combining this with the Itô formula for \( (X_t, Y_t) \) given by (3.1) and (3.2), we obtain from (3.4) that
\[ d \rho(X_t, Y_t) \leq K \rho(X_t, Y_t) dt + \sqrt{2} \langle \dot{\Phi}_t, \nabla \rho(\cdot, Y_t)(X_t) \rangle dt \]
\[ \leq \left( K \rho(X_t, Y_t) + \sqrt{2} \parallel \beta_t \parallel \right) dt, \]
see e.g. [15] Lemmas 2.1 and 2.2. Since we are using the coupling by parallel displacement instead of the mirror reflection, the martingale part here disappears (cf. Theorem 2 and (2.5) in [9]). Since \( X_0 = Y_0 \), this implies
\[ \rho(X_t, Y_t)^2 \leq e^{2Kt} \left( \sqrt{2} \int_0^t e^{-Ks} \parallel \beta_s \parallel ds \right)^2 \leq \frac{e^{2Kt} - 1}{K} \int_0^t \parallel \beta_s \parallel^2 ds, \quad t \in [0,T]. \]
Therefore,
\[ E_Q \max_{t\in[0,T]} \rho(X_t, Y_t)^2 \leq \frac{e^{2KT} - 1}{K} \int_0^T E_Q \parallel \beta_s \parallel^2 ds. \]
It is clear that
\[ E_Q \parallel \beta_s \parallel^2 = E\left(\parallel \beta_s \parallel^2 \right) = E\left(\parallel \beta_s \parallel^2 E\left(\parallel \beta_s \parallel^2 \right) \right) = E\left(\parallel \beta_s \parallel^2 \right), \quad s \in [0,T]. \]
\[ d(m)_t = m_t^2 d(L)_t = m_t^2 \| \beta_t \|^2 dt, \]

we have

\[
d m_t \log m_t = (1 + \log m_t) d m_t + \frac{d(m)_t}{2m_t}
= (1 + \log m_t) d m_t + \frac{m_t}{2} \| \beta_t \|^2 dt.
\]

As \( m_t \) is a \( \mathbb{P} \)-martingale, combining this with (3.8) we obtain

\[
\int_0^T \mathbb{E} \| \beta_s \|^2 ds = 2 \mathbb{E}_P F(X_{[0,T]}) \log F(X_{[0,T]}). \tag{3.9}
\]

Therefore, (1.4) follows from (3.6), (3.7) and (3.9).

(b) (3) \Rightarrow (2). By (3), for each \( x \in M \), there exists \( \pi_x \in C(F \Pi^T_x, \Pi^T_x) \) such that

\[
\int_M \rho \infty (\gamma, \eta)^2 \pi_x(d\gamma, d\eta) \leq \frac{2}{K} (e^{2KT} - 1) \Pi^T_x \left( \frac{F}{\Pi^T_x(F)} \log \frac{F}{\Pi^T_x(F)} \right). \tag{3.10}
\]

If \( x \mapsto \pi_x(G) \) is measurable for bounded continuous functions \( G \) on \( M^T \times M^T \), then

\[
\pi := \int_M \pi_x \mu^T_F(dx) \in C(\mu^T_F, \mu^T_F)
\]
is well defined and by (3.10)

\[
\int_{M^T \times M^T} \rho^2 \infty d\pi \leq \frac{2}{K} (e^{2KT} - 1) \int_M \Pi^T_x \left( F \log \frac{F}{\Pi^T_x(F)} \right) \mu(dx)
\leq \frac{2}{K} (e^{2KT} - 1) \Pi^T_{\mu} (F \log F).
\]

This implies the inequality in (2).

To confirm the measurability of \( x \mapsto \pi_x \), we first consider discrete \( \mu \), i.e. \( \mu = \sum_{n=1}^\infty \varepsilon_n \delta_{x_n} \) for some \( \{x_n\} \subset M \) and \( \varepsilon_n \geq 0 \) with \( \sum_{n=1}^\infty \varepsilon_n = 1 \). In this case

\[
\pi_x = \sum_{n=1}^\infty 1_{\{x=x_n\}} \pi_{x_n}, \quad \mu\text{-a.e.}
\]

which is measurable in \( x \) and \( \pi = \sum_{n=1}^\infty \mu^T_F (\{x_n\}) \pi_{x_n} \). Hence, the inequality in (2) holds. Then, for general \( \mu \), the desired inequality can be derived by
approximating \( \mu \) with discrete distributions in a standard way, see (b) in the proof of [\ref{6}, Theorem 4.1].

(c) (4) \( \Rightarrow \) (1). According to [\ref{12}, Section 7] (see also [\ref{4}, Section 4.1]), by first applying the transportation-cost inequality in (3) to \( 1 - \varepsilon + \varepsilon f \) in place of \( f \), then letting \( \varepsilon \to 0 \), we obtain the Poincaré inequality

\[
PT f^2 \leq \frac{e^{2KT}}{K} PT|\nabla f|^2 + (PT f)^2, \quad f \in C^1_b(M), T > 0.
\]  

(3.11)

Thus, the proof is finished by Theorem 2.4.

(d) (1) \( \Rightarrow \) (6). Let \( X_t \) solve (2.2) with \( X_0 = x \) and \( Y_t \) solve

\[
dY_t = \sqrt{2} \, d_{X_t,Y_t} \Phi_t \circ dB_t + Z(Y_t) \, dt - N(Y_t) \, d\tilde{t}, \quad Y_0 = y,
\]  

(3.12)

where \( \tilde{t} \) is the local time of \( Y_t \) on \( \partial M \). Since \( \partial M \) is convex and (1.1) holds, as explained in (a), we have

\[
d\rho(X_t,Y_t) \leq K \rho(X_t,Y_t) \, dt.
\]

Thus, \( \rho_\infty(X,Y) \leq e^{KT} \rho(x,y) \). This implies (6).

(e) (6) \( \Rightarrow \) (5). By (6), for any \( x,y \in M \), there exists \( \pi_{x,y} \in C(\Pi^T_x, \Pi^T_y) \) such that

\[
\int_{M^T \times M^T} \rho_\infty^p \, d\pi_{x,y} \leq e^{KT} \rho(x,y)^p.
\]

As explained in (b), we assume that \( \mu \) and \( \nu \) are discrete, so that for any \( \pi^0 \in (\mu, \nu) \), \( \pi_{x,y} \) has a \( \pi^0 \)-version measurable in \( (x,y) \). Thus,

\[
\pi := \int_{M \times M} \pi_{x,y} \pi^0(dx,dy) \in C(\Pi^T_\mu, \Pi^T_\nu)
\]

satisfies

\[
\int_{M^T \times M^T} \rho_\infty^p \, d\pi \leq e^{KT} \int_{M^T \times M^T} \rho(x,y)^p \, d\pi^0(d\pi_{x,y}).
\]

This implies the desired inequality in (5).

(f) (6) \( \Rightarrow \) (1). Let \( T > 0 \) be fixed. For any \( x,y \in M \), let \( \pi_{x,y} \in C(PT(x,\cdot), PT(y,\cdot)) \) be the optimal coupling for \( W_{2,\rho} \), i.e.

\[
W_{2,\rho}(PT(x,\cdot), PT(y,\cdot))^2 = \int_{M \times M} \rho^2 \, d\pi_{x,y}.
\]  

(3.13)
Then for any \( f \in C^2_b(M) \), (6) implies
\[
\frac{|Pt f(x) - Pt f(y)|}{\rho(x, y)} \leq \int_{M \times M} \frac{|f(z_1) - f(z_2)|}{\rho(z_1, z_2)^2} \cdot \pi_{x,y}(dz_1, dz_2) \rho(x, y)
\]
\[
\leq W_{2, \rho}(Pt(x, \cdot), Pt(y, \cdot)) \left\{ \int_{M \times M} \frac{(f(z_1) - f(z_2))^2}{\rho(z_1, z_2)^2} \pi_{x,y}(dz_1, dz_2) \right\}^{1/2}
\]
\[
\leq e^{KT} \left\{ \int_{M \times M} \frac{(f(z_1) - f(z_2))^2}{\rho(z_1, z_2)^2} \pi_{x,y}(dz_1, dz_2) \right\}^{1/2}.
\]
(3.14)

Noting that \( f \in C^2_b(M) \) implies
\[
|f(z_1) - f(z_2)|^2 \leq \rho(z_1, z_2)^2|\nabla f|^2(z_1) + c\rho(z_1, z_2)^3
\]
for some constant \( c > 0 \), by (6) and (3.13) we obtain
\[
\int_{M \times M} \frac{(f(z_1) - f(z_2))^2}{\rho(z_1, z_2)^2} \pi_{x,y}(dz_1, dz_2) \leq Pt|\nabla f|^2(x) + c e^{KT} \rho(x, y).
\]
Therefore, letting \( y \to x \) in (3.14) we arrive at
\[
|\nabla Pt f(x)| \leq e^{KT} (Pt|\nabla f|^2(x))^{1/2}.
\]
By a standard argument of Bakry and Emery, this implies the Poincaré inequality (3.11). Thus, (1) holds according to Theorem 2.4.

4 The case with a diffusion coefficient

Let \( \psi > 0 \) be a smooth function on \( M \), and let \( \Pi_{\mu, \psi}^T \) be the distribution of the (reflecting if \( \partial M \neq \emptyset \)) diffusion process generated by \( L_{\psi} := \psi^2(\Delta + Z) \) on time interval \([0, T]\) with initial distribution \( \mu \), and let \( \Pi_{x,\psi}^T = \Pi_{\delta_x, \psi}^T \) for \( x \in M \). Moreover, for \( F \geq 0 \) with \( \Pi_{\mu, \psi}^T(F) = 0 \), let
\[
\mu_{T,\psi}^T(dx) = \Pi_{x,\psi}^T(F) \mu(dx).
\]

Theorem 4.1. Assume that \( \partial M \) is either empty or convex and let (1.1) hold.
Let \( \psi \in C^\infty_b(M) \) be strictly positive. Let
\[
K_{\psi} = K^+ \|\psi\|_\infty^2 + 2\|Z\|_\infty \|\nabla \psi\|_\infty \|\psi\|_\infty + (d - 1) \|\nabla \psi\|_\infty^2.
\]
Then
\[
W_{2, \rho_{\psi}}(F\Pi_{\mu, \psi}^T, \Pi_{\mu_{T, \psi}^T, \psi}^T)^2 \leq 2C(T, \psi)\Pi_{\mu, \psi}^T(F \log F),
\]
\( \mu \in \mathcal{P}(M), \ F \geq 0, \ \Pi_{\mu, \psi}^T(F) = 1 \)
holds for

\[ C(T, \psi) := \inf_{R > 0} \left\{ (1 + R^{-1})\|\psi\|_\infty^2 \frac{e^{2K_T^{-1}} - 1}{K_\psi} \exp \left[ 2(1 + R)\|\nabla \psi\|_\infty^2 \frac{e^{2K_T^{-1}} - 1}{K_\psi} \right] \right\}. \]

**Proof.** As explained in (a) of the proof of Theorem 1.1, we shall only consider the case that \( \partial M \) is non-empty and convex. According to the proof of “\((3) \Rightarrow (2)\)”, it suffices to prove for \( \mu = \delta_o, o \in M \). In this case the desired inequality reduces to

\[ W_{2, \rho, \infty}^\ast (F \Pi_{o, \psi}^T, \Pi_{o, \psi}^T) \leq 2C(T, \psi)\Pi_{o, \psi}^T(F \log F), \quad F \geq 0, \Pi_{o, \psi}^T(F) = 1. \]  

(4.1)

Since the diffusion coefficient is non-constant, it is convenient to adopt the Itô differential \( d_I \) for the Girsanov transformation. So, the reflecting diffusion process generated by \( L_\psi := \psi^2(\nabla + Z) \) can be constructed by solving the Itô SDE

\[ d_I X_t = \sqrt{2} \psi(X_t) \Phi_t dB_t + \psi^2(X_t) Z(X_t) dt + N(X_t) dl_t, \]  

(4.2)

where \( X_0 = o \) and \( B_t \) is the \( d \)-dimensional Brownian motion with natural filtration \( \mathcal{F}_t \). Let \( \beta_t, Q \) and \( \tilde{B}_t \) be fixed in the proof of Theorem 1.1. Then

\[ d_I X_t = \sqrt{2} \psi(X_t) \Phi_t d\tilde{B}_t + \left\{ \psi^2(X_t) Z(X_t) + \sqrt{2} \psi(X_t) \Phi_t \beta_t \right\} dt + N(X_t) dl_t, \]  

(4.3)

Let \( Y_t \) solve

\[ d_I Y_t = \sqrt{2} \psi(Y_t) \Phi_t d\tilde{B}_t + \psi^2(Y_t) Z(Y_t) dt + N(Y_t) dl_t, \quad Y_0 = o, \]  

(4.4)

where \( \tilde{l}_t \) is the local time of \( Y_t \) on \( \partial M \). As in (a) of the proof of Theorem 1.1 under \( Q \), the distributions of \( Y_{[0, T]} \) and \( X_{[0, T]} \) are \( \Pi_{o, \psi}^T \) and \( F \Pi_{o, \psi}^T \) respectively. So,

\[ W_{2, \rho, \infty}^\ast (F \Pi_{o, \psi}^T, \Pi_{o, \psi}^T)^2 \leq \mathbb{E}_Q \max_{t \in [0, T]} \rho(X_t, Y_t)^2. \]  

(4.5)

Noting that due to the convexity of \( \partial M \)

\[ \langle N(x), \nabla \rho(y, \cdot)(x) \rangle = \langle N(x), \nabla \rho(\cdot, y)(x) \rangle \leq 0, \quad x \in \partial M, \]
by (4.3), (4.4) and the Itô formula, we obtain

\[
\begin{align*}
d\rho(X_t, Y_t) &\leq \sqrt{2} \left\{ \psi(X_t)(\nabla \rho(\cdot, Y_t)(X_t), \Phi_t d\dot{B}_t) \\
&\quad + \psi(Y_t)(\nabla \rho(X_t, \cdot)(Y_t), P_{X_t, Y_t} \Phi_t d\dot{B}_t) \right\} \\
&\quad + \left\{ \sum_{i=1}^{d-1} U_i^2 \rho(X_t, Y_t) + \langle \psi(X_t)^2 Z(X_t) + \sqrt{2} \psi(X_t) \Phi_t \beta_t, \nabla \rho(\cdot, Y_t)(X_t) \rangle \\
&\quad + \psi(Y_t)^2 (Z(Y_t), \nabla \rho(X_t, \cdot)(Y_t)) \right\} dt,
\end{align*}
\]

(4.6)

where \( \{U_i\}_{i=1}^{d-1} \) are vector fields on \( M \times M \) such that \( \nabla U_i(X_t, Y_t) = 0 \) and

\[ U_i(X_t, Y_t) = \psi(X_t)V_i + \psi(Y_t)P_{X_t, Y_t}V_i, \quad 1 \leq i \leq d - 1 \]

for \( \{V_i\}_{i=1}^d \) an OBN of \( T_X M \) with \( V_d = \nabla \rho(\cdot, Y_t)(X_t) \).

In order to calculate \( U_i^2 \rho(X_t, Y_t) \), we adopt the second variational formula for the distance. Let \( \rho_t = \rho(X_t, Y_t) \) and let \( \{J_i\}_{i=1}^{d-1} \) be Jacobi fields along the minimal geodesic \( \gamma : [0, \rho_t] \to M \) from \( X_t \) to \( Y_t \) such that \( J_i(0) = \psi(X_t)V_i \) and \( J_i(\rho_t) = \psi(Y_t)P_{X_t, Y_t}V_i, 1 \leq i \leq d - 1 \). Note that the existence of \( \gamma \) is ensured by the convexity of \( \partial M \). Then, by the second variational formula and noting that \( \nabla U_i(X_t, Y_t) = 0 \), we have

\[
I := \sum_{i=1}^{d-1} U_i^2 \rho(X_t, Y_t) = \sum_{i=1}^{d-1} \int_0^{\rho_t} \left\{ |\nabla \gamma J_i|^2 - \langle \mathcal{R}(\gamma, J_i) J_i, \gamma \rangle \right\}(s) ds,
\]

(4.7)

where \( \mathcal{R} \) is the curvature tensor. Let

\[
\tilde{J}_i(s) = \left( \frac{s}{\rho_t} \psi(Y_t) + \frac{\rho_t - s}{\rho_t} \psi(X_t) \right) P_{\gamma(0), \gamma(s)} V_i, \quad 1 \leq i \leq d - 1.
\]

We have \( \tilde{J}_i(0) = J_i(0) \) and \( \tilde{J}_i(\rho_t) = J_i(\rho_t), 1 \leq i \leq i - 1 \). By the index lemma,

\[
I \leq \sum_{i=1}^{d-1} \int_0^{\rho_t} \left\{ |\nabla \gamma \tilde{J}_i|^2 - \langle \mathcal{R}(\gamma, \tilde{J}_i) \tilde{J}_i, \gamma \rangle \right\}(s) ds
\]

\[
\leq (d - 1) \|\nabla \psi\|^2_{L^\infty} \rho_t - \frac{1}{\rho_t^2} \int_0^{\rho_t} \left\{ s \psi(Y_t) + (\rho_t - s) \psi(X_t) \right\}^2 \text{Ric}(\gamma(s), \gamma(s)) ds.
\]

(4.8)
Moreover, 

\[ \psi(X_t)^2 \langle Z(X_t), \nabla \rho(\cdot, Y_t)(X_t) \rangle + \psi(Y_t)^2 \langle Z(Y_t), \nabla \rho(X_t, \cdot)(Y_t) \rangle \]

\[ = \frac{1}{\rho_t^2} \int_0^{\rho_t} d\langle \psi(Y_t) + (\rho_t - s)\psi(X_t) \rangle (Z(\gamma(s)), \dot{\gamma}(s))ds \]

\[ = \frac{1}{\rho_t^2} \int_0^{\rho_t} (s\psi(Y_t) + (\rho_t - s)\psi(X_t))^2 (\nabla \dot{\gamma} Z \circ \gamma, \dot{\gamma}(s))ds \]

\[ + 2\int_0^{\rho_t} (\psi(Y_t) - \psi(X_t)) (s\psi(Y_t) + (\rho_t - s)\psi(X_t))ds \]

\[ \leq 1 \rho_t^2 \int_0^{\rho_t} (s\psi(Y_t) + (\rho_t - s)\psi(X_t))^2 (\nabla \dot{\gamma} Z \circ \gamma, \dot{\gamma}(s))ds \]

\[ + 2\|Z\|_\infty \|\psi\|_\infty \|\nabla \psi\|_\infty \rho_t. \] (4.9)

Finally, we have

\[ \langle \nabla \rho(X_t, \cdot)(Y_t), P_{X_t, Y_t} \Phi_t d\tilde{B}_t \rangle = \langle P_{Y_t, X_t} \nabla \rho(X_t, \cdot)(Y_t), \Phi_t d\tilde{B}_t \rangle = \]

\[ = -\langle \nabla \rho(\cdot, Y_t)(X_t), \Phi_t d\tilde{B}_t \rangle. \]

Combining this with (4.6), (4.7), (4.8) and (4.9), we arrive at

\[ d\rho(X_t, Y_t) \leq \sqrt{2} (\psi(X_t) - \psi(Y_t)) (\nabla \rho(\cdot, Y_t)(X_t), \Phi_t d\tilde{B}_t) \]

\[ + K_\psi \rho(X_t, Y_t) dt + \sqrt{2} \|\psi\|_\infty \|\beta_t\| dt. \]

Then

\[ M_t := \sqrt{2} \int_0^t e^{-K_\psi s} (\psi(X_s) - \psi(Y_s)) (\nabla \rho(\cdot, Y_s)(X_s), \Phi_s d\tilde{B}_s) \]

is a \( \mathbb{Q} \)-martingale such that

\[ \rho(X_t, Y_t) \leq e^{K_\psi t} M_t + \sqrt{2} e^{K_\psi t} \int_0^t e^{-K_\psi s} \|\psi\|_\infty \|\beta_s\| ds, \quad t \in [0, T]. \] (4.10)

So, by the Doob inequality we obtain
\[ h_t := \mathbb{E}_Q \max_{s \in [0,t]} \rho(X_s, Y_s)^2 \]
\[ \leq (1 + R)e^{2K_y t} \mathbb{E}_Q \max_{s \in [0,t]} M_s^2 ds \]
\[ + 2\|\psi\|_\infty^2 (1 + R^{-1}) e^{2K_y t} \mathbb{E}_Q \left( \int_0^t e^{-K_y s} \|\beta_s\| ds \right)^2 \]
\[ \leq 4(1 + R)e^{2K_y t} \mathbb{E}_Q M_t^2 + (1 + R^{-1}) \|\psi\|_\infty^2 \frac{e^{2K_y t} - 1}{K_y} \int_0^t \mathbb{E}_Q \|\beta_s\|^2 ds \]
\[ \leq 4(1 + R)\|\nabla \psi\|_\infty^2 e^{2K_y t} \int_0^t e^{-2K_y s} h_s ds \]
\[ + (1 + R^{-1}) \|\psi\|_\infty^2 \frac{e^{2K_y t} - 1}{K_y} \int_0^t \mathbb{E}_Q \|\beta_s\|^2 ds \]
for any \( R > 0 \). Since \( e^{-2K_y s} \) is decreasing in \( s \) while \( h_s \) is increasing in \( s \), by the FKG inequality we have
\[ \int_0^t e^{-2K_y s} h_s ds \leq \left( \frac{1}{t} \int_0^t e^{-2K_y s} ds \right) \int_0^t h_s ds \leq \frac{1}{2K_y t} \int_0^t h_s ds. \]

Therefore,
\[ h_t \leq 2(1 + R)\|\nabla \psi\|_\infty^2 \frac{e^{2K_y t} - 1}{K_y} \int_0^t h_s ds + (1 + R^{-1}) \|\psi\|_\infty^2 \frac{e^{2K_y t} - 1}{K_y} \int_0^t \mathbb{E}_Q \|\beta_s\|^2 ds \]
holds for \( t \in [0,T] \). Since \( h_0 = 0 \), this implies that
\[ \mathbb{E}_Q \max_{t \in [0,T]} \rho(X_t, Y_t)^2 = h_T \]
\[ \leq (1 + R^{-1}) \|\psi\|_\infty^2 \frac{e^{2K_y T} - 1}{K_y} \exp \left( 2(1 + R)\|\nabla \psi\|_\infty^2 \frac{e^{2K_y T} - 1}{K_y} \right) \int_0^T \mathbb{E}_Q \|\beta_s\|^2 ds. \]

Combining this with (4.5) and (3.9), we complete the proof. \( \square \)

**Theorem 4.2.** In the situation of Theorem 4.1,
\[ W_{2,\rho_*} (\Pi_{\mu,\psi}^T, \Pi_{\nu,\psi}^T) \leq 2\rho(K_y + \|\nabla \psi\|_\infty^2)T W_{2,\rho}(\mu, \nu), \quad \mu, \nu \in \mathcal{P}(M), T > 0. \]

**Proof.** As explained in the proof of “(6) \Rightarrow (5)”, we only consider \( \mu = \delta_x \) and \( \nu = \delta_y \). Let \( X_t \) solve (4.2) with \( X_0 = x \), and let \( Y_t \) solve, instead of (4.4),
\[ dY_t = \sqrt{2} \psi(Y_t) P_{X_t,Y_t} \Phi t d\tilde{B}_t + \psi^2(Y_t) Z(Y_t) dt + N(Y_t) d\tilde{U}_t, \quad Y_0 = y. \]
Then, repeating the proof of Theorem 4.1 we have, instead of (4.10),
\[ \rho(X_t, Y_t) \leq e^{K_y t}(M_t + \rho(x, y)), \quad t \geq 0 \quad (4.11) \]
holds for
\[ M_t := \sqrt{2} \int_0^t e^{-K_{\psi,s}(\psi(X_s) - \psi(Y_s))} \langle \nabla \rho(\cdot, Y_s)(X_s), \Phi_s dB_s \rangle. \]

So,
\[ \mathbb{E}\rho(X_1, Y_2)^2 \leq e^{2K_{\psi,t}} \mathbb{E} \left\{ \rho(x, y)^2 + 2\| \nabla \psi \|_{X_c}^2 \int_0^t e^{-2K_{\psi,s}} \mathbb{E}\rho(X_s, Y_s)^2 \right\}, \]
which implies
\[ \mathbb{E}\rho(X_1, Y_2)^2 \leq e^{2(K_{\psi} + \| \nabla \psi \|_{X_c}) t} \rho(x, y)^2. \]
Combining this with (1.1) and the Doob inequality, we arrive at
\[
W_{2,\rho_{\infty}}(\Pi_{\rho_{\psi},\phi}^T, \Pi_{\rho_{\psi},\phi}^T)^2 \leq \mathbb{E} \max_{t \in [0,T]} \rho(X_t, Y_t)^2 \leq e^{2K_{\psi,T}} \mathbb{E} \max_{t \in [0,T]} (M_t + \rho(x, y))^2 \\
\leq 4e^{2K_{\psi,T}} \mathbb{E} (M_T + \rho(x, y))^2 = 4e^{2K_{\psi,T}} \left( \mathbb{E} (M_T^2 + \rho(x, y)^2) \right) \\
= 4e^{2K_{\psi,T}} \left( \rho(x, y)^2 + 2\| \nabla \psi \|_{X_c}^2 \int_0^T e^{-2K_{\psi,t}} \mathbb{E}\rho(X_t, Y_t)^2 dt \right) \\
\leq 4e^{2(K_{\psi} + \| \nabla \psi \|_{X_c}) T} \rho(x, y)^2.
\]
This implies the desired inequality for \( \mu = \delta_x \) and \( \nu = \delta_y \).

5 Extensions to non-convex manifolds

As explained in the end of Section 1, combining Theorem 1.1 with a proper conformal change of metric, we are able to establish the following transportation-cost inequality on a class of manifolds with non-convex boundary.

**Theorem 5.1.** Let \( \partial M \neq \emptyset \) with \( \| \geq -\sigma \) for some constant \( \sigma > 0 \), and let (1.1) hold for some \( K \in \mathbb{R} \). Then for any \( f \in C_0^\infty(M) \) with \( f \geq 1 \) and \( N \log f \mid_{\partial M} \geq \sigma \), and for any \( \mu \in \mathcal{P}(M) \),
\[
W_{2,\rho_{\infty}}(\Pi_{\rho_{\mu},\phi}^T, \Pi_{\rho_{\mu},\phi}^T)^2 \leq 2\| f \|_{X_c}^2 c(T, f) \Pi_{\rho_{\mu}}^T(F \log F), \quad F \geq 0, \Pi_{\rho_{\mu}}^T(F) = 1
\] holds for
\[
c(T, f) = \inf_{R > 0} \left\{ \left( 1 + R^{-1} \right) \frac{e^{2K_{\psi,T}} - 1}{\kappa_f} \exp \left[ 2(1 + R) \| \nabla f \|_{X_c} \frac{e^{2K_{\psi,T}} - 1}{\kappa_f} \right] \right\},
\]
where
\[
\kappa_f = 5\| f \|_{X_c} \| \nabla f \|_{X_c} \| Z \|_{X_c} + \left\{ 3d - 5 + (d - 3)^+ \right\} \| \nabla f \|_{X_c}^2 + \| (Kf^2 - f \Delta f)^+ \|_{X_c}.
\]
In particular,
\[
W_{2,\rho_{\infty}}(\Pi_{\rho_{\mu},\phi}^T, \Pi_{\rho_{\mu},\phi}^T)^2 \leq 2\| f \|_{X_c}^2 c(T, f) \Pi_{\rho_{\mu}}^T(F \log F), \quad o \in M, F \geq 0, \Pi_{\rho_{\mu}}^T(F) = 1.
\]
Proof. Let $f \in C^\infty_c(M)$ such that $f \geq 1$. Since $1 \geq -\sigma$ and $N \log f|_{\partial M} \geq \sigma$, by [21, Lemma 2.1] the boundary $\partial M$ is convex under the new metric

$$\langle \cdot, \cdot \rangle' = f^{-2}\langle \cdot, \cdot \rangle.$$ 

Let $\Delta'$ and $\nabla'$ be induced by the new metric. Then (see formula (2.2) in [14])

$$L = f^{-2}(\Delta' + Z'), \quad Z' := f^2Z + \frac{d-2}{2} \nabla f^2.$$

Let $\text{Ric}'$ be the Ricci curvature induced by the new metric, we have (cf. formula (3.2) in [6])

$$\text{Ric}' = \text{Ric} + (d-2)f^{-1}\text{Hess}_f + \left(f^{-1}\Delta f - (d-3)|\nabla \log f|^2\right)\langle \cdot, \cdot \rangle. \quad (5.1)$$

Since the Levi-Civita connection induced by $\langle \cdot, \cdot \rangle'$ satisfies (cf. [3, Theorem 1.59(a)])

$$\nabla'_U V = \nabla_U V - (U, \nabla \log f)V - (V, \nabla \log f)U + \langle U, V \rangle \nabla \log f, \quad U, V \in TM,$$

we have

$$\langle \nabla'_U Z', U \rangle' = f^{-2}\left\{ \langle \nabla_U Z', U \rangle - \langle Z', \nabla \log f \rangle|U|^2 \right\}$$

$$= 2\langle U, \nabla \log f \rangle \langle Z, U \rangle + \langle \nabla_U Z, U \rangle + \frac{d-2}{2f^2} \text{Hess}_f(U, U)$$

$$- \langle Z, \nabla \log f \rangle|U|^2 - \frac{d-2}{2} \langle |\nabla \log f|^2, \nabla \log f \rangle|U|^2$$

$$\leq \langle \nabla_U Z, U \rangle + 3|\nabla \log f| \cdot |Z| \cdot |U|^2 + (d-2)f^{-1}\text{Hess}_f(U, U).$$

Combining this with (5.1), we obtain

$$\text{Ric}'(U, U) - \langle \nabla_U Z', U \rangle'$$

$$\geq \text{Ric}(U, U) - \langle \nabla_U Z, U \rangle + \left\{ f^{-1}\Delta f - (d-3)|\nabla \log f| - 3|Z| \cdot |\nabla \log f| \right\}|U|^2$$

$$\geq -K'(U, U)', \quad U \in TM,$$

where

$$K' = \sup_M\{Kf^2 - f\Delta f + (d-3)|\nabla f|^2 + 3|Z|f|\nabla f| \}. \quad (5.2)$$

Noting that $f \geq 1$, we have

$$\sqrt{\langle Z', Z' \rangle'} = f^{-1}|f^2Z + (d-2)f\nabla f| \leq \|f\|_\infty\|Z\|_\infty + (d-2)\|\nabla f\|_\infty,$$

$$\sqrt{\langle \nabla' f^{-1}, \nabla' f^{-1} \rangle'} = f|\nabla f^{-1}| \leq \|\nabla f\|_\infty. \quad (5.3)$$
Letting $K_\psi$ be defined in Theorem 4.1 for the manifold $(M, \langle \cdot, \cdot \rangle)$ and $L = \psi^2(\Delta' + Z')$ with $\psi = f^{-1}$, we deduce from $f \geq 1$, (5.2) and (5.3) that

$$K_\psi \leq \kappa_f.$$ 

Therefore, $C(T, \psi) \leq c(T, f)$ and thus, Theorem 4.1 implies

$$W_2, \rho'_{\infty}(\Pi^T_{\mu}, \Pi^T_{\nu}) \leq 2\|f\|_{\infty}e^{(\kappa_f + \|\nabla f^{-1}\|_{\infty})^T}W_2, \rho_{\infty}(\mu, \nu), \quad \mu, \nu \in \mathcal{P}(M), T > 0.$$ 

As a consequence of Theorems 5.1 and 5.2, we present below an explicit transportation-cost inequalities for a class of non-convex manifolds.

**Corollary 5.3.** Assume that (1.1) holds for some $K \geq 0$ and the injectivity radius $i_{\partial M}$ of $\partial M$ is strictly positive. Let $\sigma \geq 0$ and $\gamma, k, > 0$ be such that $-\sigma \leq I \leq \gamma$ and $\text{Sect}_M \leq k$. Let

$$0 < r \leq \min \left\{ i_{\partial M}, \frac{1}{\sqrt{k}} \arcsin \left( \frac{\sqrt{k}}{\sqrt{k + \gamma^2}} \right) \right\}.$$ 

(i) The transportation-cost inequality

$$W_2, \rho_{\infty}(F\Pi^T_{\mu}, \Pi^T_{\nu}) \leq (2 + rd\sigma)^2 \frac{e^{2\theta T} - 1}{\theta} \exp \left[ \frac{4(e^{2\theta T} - 1)}{\theta} \right] \Pi^T_{\mu}(F \log F)$$

holds for all $\mu \in \mathcal{P}(M)$ and $F \geq 0$ with $\Pi^T_{\mu}(F) = 1$, where

$$\theta = K \left( 1 + rd\sigma + \frac{r^2d^2\sigma^2}{4} + \frac{d\sigma}{r} \left( 3d - 5 + (d-3)^+ + \frac{d^2}{2} \right) \sigma^2 + 5\|Z\|_{\infty}\sigma \left( 1 + \frac{rd\sigma}{2} \right) \right).$$

In particular,

$$W_2, \rho_{\infty}(F\Pi^T_{\nu}, \Pi^T_{\mu}) \leq (2 + rd\sigma)^2 \frac{e^{2\theta T} - 1}{\theta} \exp \left[ \frac{4(e^{2\theta T} - 1)}{\theta} \right] \Pi^T_{\mu}(F \log F).$$
holds for all $F \geq 0$ with $\Pi^T_\mu (F) = 1$.

(ii) For any $T > 0$ and $\mu, \nu \in P(M)$, 

$$W_{2,\rho^\infty}(\Pi^T_\mu, \Pi^T_\nu) \leq (2 + \sigma r d) e^{(\theta + \sigma^2) T} W_{2,\rho}(\mu, \nu).$$

Proof. Let 

$$h(s) = \cos (\sqrt{k} s) - \frac{\gamma}{\sqrt{k}} \sin (\sqrt{k} s), \quad s \geq 0.$$ 

Then $h$ is the unique solution to the equation 

$$h'' + kh = 0, \quad h(0) = 1, h'(0) = -\gamma.$$ 

Up to an approximation argument presented in the proof of [20, Theorem 1.1], we may apply Theorem 5.1 to 

$$f = 1 + \sigma \varphi \circ \rho_{\partial M},$$ 

where $\rho_{\partial}$ is the Riemannian distance to $\partial M$, which is smooth on $\{ \rho_{\partial M} < i_{\partial M} \}$, and 

$$\alpha = (1 - h(r))^{1-d} \int_0^r (h(s) - h(r))^{d-1} ds,$$ 

$$\varphi(s) = \frac{1}{\alpha} \int_0^s (h(t) - h(r))^{1-d} dt \int_{t \wedge r}^r (h(u) - h(r))^{d-1} du, \quad s \geq 0.$$ 

We have $\varphi(0) = 1, 0 \leq \varphi' \leq \varphi'(0) = 1$. Moreover, as observed in [20, Proof of Theorem 1.1], 

$$\alpha \geq \frac{r}{d}, \quad \varphi(r) \leq \frac{r^2}{2\alpha} \leq \frac{dr}{2}, \quad \Delta \varphi \circ \rho_{\partial M} \geq -\frac{1}{\alpha} \geq -\frac{d}{r},$$ 

So, 

$$\|f\|_\infty \leq 1 + \sigma \varphi(r) \leq 1 + \frac{rd\sigma}{2}, \quad \|
abla f\|_\infty \leq \varphi'(0) = \sigma, \quad \Delta f \geq -\frac{\sigma d}{r}. \quad (5.4)$$ 

Noting that (recall that $K \geq 0$) 

$$\sup(K f^2) \leq K \left(1 + r d\sigma + \frac{r^2 d^2 \sigma^2}{4}\right).$$ 

from (5.4), we conclude that $\kappa_f \leq \theta$. So, (i) follows from (1.5) and (5.1) for $R = 1$, and (ii) follows from Theorem 4.2 and (5.4).

Acknowledgement. The author would like to thank the referee for very helpful comments and corrections.
References


Framed Sheaves over Treefolds
and Symmetric Obstruction Theories

Dragos Oprea

Received: September 20, 2012
Communicated by Edward Frenkel

Abstract. We note that open moduli spaces of sheaves over local Calabi-Yau surface geometries framed along the divisor at infinity admit symmetric perfect obstruction theories. We calculate the corresponding Donaldson-Thomas weighted Euler characteristics (as well as the topological Euler characteristics). Furthermore, for blowup geometries, we discuss the contribution of exceptional curves.

2010 Mathematics Subject Classification: 14N35, 14D20
Keywords and Phrases: Moduli spaces of sheaves, Donaldson-Thomas invariants

1 Introduction

Moduli spaces of sheaves over threefolds admit virtual fundamental classes in a lot of examples, yielding Donaldson-Thomas invariants [T]. The rank 1 case is particularly interesting, bearing connections with virtual curve counts [MNOP].

In this note, we study open moduli spaces of higher rank sheaves over local Calabi-Yau surface geometries, framed along the divisor at infinity. We prove that the moduli spaces admit symmetric perfect obstruction theories, and in this context, we compute the ensuing Donaldson-Thomas Euler characteristics. In addition, we find the topological Euler characteristics of the compactified moduli spaces of framed modules. We also discuss a “blowup” formula. Finally, we point out other geometries which can be studied by the same methods.

This way, we extend previously known results in two directions 1.

1 The search for such generalizations motivated our interest this topic.
(i) First, there is quite a bit of literature on moduli spaces of framed sheaves over surfaces. An exhaustive survey is not our intention here, but we refer the reader to [BPT], [N] for calculations which we partially carry out in the higher dimensional setting, and also for a more comprehensive bibliography.

(ii) Second, as suggested above, we partially generalize to higher rank results about the Hilbert scheme of points over threefolds. For these, the Donaldson-Thomas Euler characteristics, in the form needed here, were calculated in [BF].

2 Framed sheaves over local Calabi-Yau surface geometries

We now detail the discussion. Let $S$ be a smooth complex projective surface, and let $X^o$ denote the total space of the canonical bundle $K_S \to S$. We are concerned with moduli spaces of sheaves over the open Calabi-Yau threefold $X^o$. The noncompact geometry does not allow for a good moduli space of semistable sheaves. Instead, we will consider the compact threefold $\pi: X = \mathbb{P}(K_S + \mathcal{O}_S) \to S$. This comes equipped with two divisors $S_\infty = \mathbb{P}(K_S + 0)$ and $S_0 = \mathbb{P}(0 + \mathcal{O}_S)$ corresponding to the summands $K_S$ and $\mathcal{O}_S$. Clearly, $X \setminus S_\infty = X^o$. We form the moduli space $\mathfrak{M}_n$ of semistable framed modules $(E, \phi)$ of rank $r$ with

$$c_1(E) = c_2(E) = 0, \quad \chi(F) = N := r\chi(\mathcal{O}_S) - n,$$

with a non-zero framing over $S_\infty$:

$$\phi: E \to \mathcal{O}_{S_\infty} \otimes \mathbb{C}^r.$$

The moduli space $\mathfrak{M}_n$ was constructed in [HL]. Semistability was defined with respect to a polynomial $\delta$ of degree $\leq 2$ with positive leading coefficient, as well as an ample divisor $H$ on $X$. We will pick $H = \pi^*H_0 + \epsilon c_1(\mathcal{O}_S(1))$, for an ample divisor $H_0$ on $S$ and a sufficiently small rational $\epsilon > 0$. By definition, $(E, \phi)$ is semistable provided that

(i) for all proper subsheaves $F$ of $E$, the Hilbert polynomials satisfy

$$P_F - \delta \leq \frac{\text{rk}F}{\text{rk}E}(P_E - \delta);$$
(ii) if $F$ is contained in the kernel of $\phi$, then

$$P_F \leq \frac{\text{rk} F}{\text{rk} E} (P_E - \delta).$$

Semistable framed modules admit Harder-Narasimhan filtrations, yielding the notion of $S$-equivalence. There is a projective moduli space $\mathcal{M}_n$ of $S$-equivalence classes of framed modules, cf. [HL].

We will consider the open subset

$$\mathcal{M}_n^o \hookrightarrow \mathcal{M}_n$$

corresponding to what are called framed sheaves in [L], [N]. These are stable framed modules $(E, \phi)$ such that

(iii) $E$ is torsion free, locally free near $S_\infty$, and $\phi$ is an isomorphism along $S_\infty$.

Over curves and certain surfaces and for special framings, the stability conditions (i) and (ii) are automatic for the framed sheaves of (iii), cf. [BM], but for threefolds stability is not yet known to follow on general grounds.

**Example** We describe the moduli space in perhaps the simplest example, that of the Hilbert polynomial

$$P_E = P_{\mathcal{O}_X} - \ell.$$ 

Intuitively, in this case we should get the Hilbert scheme of points. This is not entirely obvious because framed modules are not required to be torsion free and because of the stability condition. The exact description will be determined by comparing $\delta$ to the polynomial

$$\Delta = \chi(mH|_{S_\infty}) = \frac{m^2 H^2}{2} + \text{l.o.t.}$$

To avoid strictly semistables, we assume that $\Delta - \delta$ is not a constant $a$ with $0 \leq a \leq \ell$.

The sheaves $E$ in the moduli space have rank 1 and can be written in the form

$$0 \to T \to E \to E^\circ \to 0$$

where $T$ is torsion and $E^\circ$ is torsion free. In fact,

$$E^\circ = I_Z \otimes L$$

for some line bundle $L \to X$, and some subscheme $Z$ of dimension at most 1. Now, by stability, or using Lemma 1.2 of [HL], the kernel $K$ of the restricted framing

$$\phi_\infty : T \to \mathcal{O}_{S_\infty}$$
must satisfy
\[ p_K \leq 0 \implies K = 0. \]
Therefore \( \phi_\infty \) gives an inclusion of the torsion module \( T \) into the framing \( \mathcal{O}_{S_\infty} \), showing that
\[ T = 0 \text{ or } T = i_*(I_W \otimes M), \]
for some line bundle \( M \) over \( i: S_\infty \to X \), and a subscheme \( W \subset S_\infty \) of dimension zero.

In the first case, since \( c_1(E) = c_2(E) = 0 \), we must have \( c_1(L) = 0 \) and \( Z \) is zero dimensional. We claim that \( L = \mathcal{O}_X \). Indeed, \( L \) restricts trivially to the fibers of \( X \to S \), hence it must be a pullback
\[ L = \pi^* N \]
of a degree zero line bundle \( N \) on \( S \). The framing condition implies that there must exist a nonzero morphism \( N \to \mathcal{O}_S \), hence \( N \) must be trivial. Therefore, up to isomorphisms, the only framed modules are
\[(E, i): I_Z \to \mathcal{O}_{S_\infty} \]
for zero dimensional subschemes \( Z \) of length \( \ell \). We analyze semistability. The kernel of \( i \) takes the form \( I_U(-S_\infty) \) for some zero dimensional scheme \( U \). Thus, we must have
\[ \chi(I_U(mH - S_\infty)) \leq \chi(mH) - \ell - \delta \iff \ell - \ell(U) \leq \Delta - \delta. \]
If \( \Delta - \delta \) has negative leading term, the inequality cannot be satisfied. If \( \Delta - \delta \) has positive leading term, then the inequality is automatic and stability follows. Hence, the moduli space is either empty, or isomorphic to \( X[\ell] \). \(^2\)

We claim the second case cannot occur under our assumptions. If it did, then
\[ 0 \to i_*(I_W \otimes M) \to E \to L \otimes I_Z \to 0. \]
Calculating the Chern class \( c_1(E) = 0 \), we find
\[ L = \mathcal{O}(-S_\infty) \otimes \pi^* N, \]
for some degree 0 line bundle \( N \) over \( S \). Therefore
\[ 0 \to i_*(I_W \otimes M) \to E \to I_Z(-S_\infty) \otimes \pi^* N \to 0, \ \phi: E \to \mathcal{O}_{S_\infty}. \]

\(^2\)We also remark here that if \( \Delta - \delta = a \) for some \( a \in \{0, \ldots, \ell\} \), then there are strictly semistable framed modules. Indeed, pairs of subschemes \( Z^0 \) of \( X^0 \) and \( Z_\infty \) of \( S_\infty \) with \( \ell(Z_\infty) = a \) yield the strictly semistable framed modules \( (I_Z(-S_\infty), 0) \oplus (i_* I_{Z_\infty/S_\infty}, \ell) \).
We already argued above that the restriction $\phi_\infty$ of $\phi$ to the torsion module $i_*(I_W \otimes M)$ must be injective. Since $\phi_\infty \neq 0$, there must exist a non-zero morphism $M \to \mathcal{O}_S$, hence $M'$ must be effective. Since $c_2(E) = 0$, we have

$$\iota_* c_1(M) = [Z].$$

Therefore, $M$ is trivial and $Z$ is of dimension zero. The Hilbert polynomial gives

$$\ell(Z) + \ell(W) = \ell.$$

Furthermore, up to scalars, $\phi_\infty$ must be the natural inclusion. Semistability implies that

$$P_{i_* I_W} - \delta \leq 0 \implies \Delta - \delta \leq \ell(W).$$

Since $\phi_\infty$ is the natural inclusion, the image of $\phi$ in $\mathcal{O}_{S_\infty}$ must be the ideal sheaf $I_U$ of a scheme $U \subset W$. Let $K$ be the kernel of $\phi$. We have

$$P_K = P_E - P_{i_* I_U} = P_E - (\Delta - \ell(U)).$$

Semistability implies

$$P_K \leq P_E - \delta \implies \Delta - \delta \geq \ell(U).$$

Therefore, we conclude that

$$\Delta - \delta = a$$

for some constant $a$ such that

$$\ell(W) \geq a \geq \ell(U).$$

In particular $0 \leq a \leq \ell$ which contradicts our assumption.

To summarize, when $\Delta - \delta$ is not equal to a constant between $0$ and $\ell$, we obtain the following description of the moduli space:

(a) if (the leading term of) $\Delta - \delta < 0$, then we get $\emptyset$;

(b) if (the leading term of) $\Delta - \delta > 0$, then the moduli space is the Hilbert scheme $X[^\ell].$

3 Obstruction theory

We note now that the obstruction theory of framed sheaves is symmetric. To this end, we assume that $\delta$ is good, i.e. it satisfies the following conditions:

- if $\deg \delta = 0$, then $\delta > (r - 1)n$;

- if $\deg \delta = 2$, the quadratic term of $\delta$ is sufficiently small compared to that of $\Delta$. 

Documenta Mathematica 18 (2013) 323–342
In particular, any $\delta$ of degree 1 is good for all $n$.

**Theorem 1.** When $\delta$ is good, the moduli space $\mathcal{M}_n^0$ admits a symmetric perfect obstruction theory at the stable points $(E, \phi)$.

**Proof.** The deformation theory of stable framed sheaves was worked out in [HL], [S]. Write $$(\mathcal{E}, \Phi) \rightarrow \mathcal{M}_n^0 \times X$$
for the universal family, which exists by [HL], and let $p$ and $q$ be the natural projections. The complex
$$\mathbf{F} = R^p_\ast R\mathcal{H}om(\mathcal{E}(-S_\infty), \mathcal{E} \otimes q^\ast K_X)[2]$$
is an obstruction theory over $\mathcal{M}_n^0$. The obstruction theory is symmetric in the sense that there is a symmetric isomorphism
$$\mathbf{F} \rightarrow \mathbf{F}^\vee[1].$$
This is a consequence of Grothendieck duality and of the crucial observation that
$$K_X = \mathcal{O}(-2S_\infty).$$
The calculation of the canonical bundle standardly follows from the Euler sequence of the projective bundle $X$.

The obstruction theory is perfect with amplitude contained in $[-1, 0]$. Indeed, the amplitude is clearly contained in $[-2, 1]$. By symmetry, it suffices to explain that the degree $-2$ term is zero. In turn, this is implied by the vanishing
$$\text{Hom}(E, E(-S_\infty)) = 0$$
which holds for all sheaves $E$ in $\mathcal{M}_n^0$. Indeed, assuming there is a non-zero morphism
$$E \rightarrow E(-S_\infty),$$
we let $K$ and $I$ denote its kernel and image, and write $r_K$ and $r_I$ for their ranks. We have $r_I > 0$. By stability
$$P_K - \delta \leq \frac{r_K}{r}(P_E - \delta), \quad P_I(S_\infty) - \delta \leq \frac{r_I}{r}(P_E - \delta).$$
Considering the quadratic terms of these inequalities, we obtain
$$c_1(K) \cdot H^2 \leq \delta_0 \left(1 - \frac{r_K}{r}\right), \quad c_1(I) \cdot H^2 + r_I S_\infty \cdot H^2 \leq \delta_0 \left(1 - \frac{r_I}{r}\right),$$
where $\delta_0$ is half the leading term of $\delta$. We also have
$$c_1(K) + c_1(I) = 0.$$\[1\]
Adding, we obtain
$$r_I S_\infty \cdot H^2 \leq \delta_0$$
which is impossible when the leading term $\delta_0 < S_\infty \cdot H^2$ is sufficiently small. This completes the proof.  \[QED\]
Example. We determine the obstruction theory for the previous example. We consider case (b) corresponding to $\Delta - \delta > 0$, $\Delta - \delta$ does not equal a constant $a$ with $0 \leq a \leq \ell$. The tangent space at the ideal sheaf $I_Z$ was found in [HL] to be

$$T_Z \mathcal{M} = \text{Ext}^1(I_Z, [I_Z \to \mathcal{O}_{S_\infty}]).$$

This can be calculated from the exact triangle

$$[I_Z \to \mathcal{O}_{S_\infty}] \to [\mathcal{O}_X \to \mathcal{O}_{S_\infty}] \cong \mathcal{O}(-S_\infty) \to \mathcal{O}_Z.$$

We have

$$\text{Ext}^0(I_Z, \mathcal{O}_X(-S_\infty)) = \text{Ext}^3(\mathcal{O}_X, I_Z(-S_\infty))^\vee = H^3(I_Z(-S_\infty))^\vee = H^3(\mathcal{O}_X(-S_\infty))^\vee = 0,$$

and similarly for $\text{Ext}^1$. From the exact triangle, we obtain

$$0 \to \text{Ext}^0(I_Z, \mathcal{O}_Z) \cong T_Z X^{[\ell]} \to T_Z \mathcal{M} \to 0.$$

In particular, this agrees with the identification

$$\mathcal{M} \cong X^{[\ell]}.$$

Thus, by symmetry, the obstruction theory of $\mathcal{M}$ coincides with the usual obstruction theory for the Hilbert scheme only along the open part $(X^\circ)^{[\ell]}$.

4 Calculations

Symmetric perfect obstruction theories have associated Behrend functions [B]. In particular, the open moduli space

$$\mathcal{M}^\circ_n \hookrightarrow \mathcal{M}_n$$

is endowed with a constructible function

$$\nu : \mathcal{M}^\circ_n \to \mathbb{Z}.$$

We will calculate the Donaldson-Thomas weighted Euler characteristic

$$\tilde{\chi}(\mathcal{M}_n^\circ) = \sum_k k \chi(\nu^{-1}(k)).$$

Since the obstruction theory is not perfect symmetric over the boundary, these weighted Euler characteristics do not calculate intersection theoretic Donaldson-Thomas invariants of $\mathcal{M}_n$.  

**Documenta Mathematica 18 (2013) 323–342**
4.1 Virtual localization

Our computation is via equivariant localization. The following result was proved in [BF] for torus actions with isolated fixed points, and in [LQ] in arbitrary generality. Let $\mathcal{M}$ be a moduli space admitting a $\mathbb{C}^*$-action compatible with the symmetric perfect obstruction theory. Then the fixed point set $\mathcal{M}^{\mathbb{C}^*}$ also inherits a symmetric perfect obstruction theory. Furthermore, the Behrend functions of $\mathcal{M}$ and $\mathcal{M}^{\mathbb{C}^*}$ at torus fixed points $p$ are related by

$$\nu_{\mathcal{M}}(p) = (-1)^{\epsilon_p} \nu_{\mathcal{M}^{\mathbb{C}^*}}(p),$$

where $\epsilon_p$ is given by the difference in the dimension of the Zariski tangent spaces

$$\epsilon_p = \dim T_p \mathcal{M} - \dim T_p \mathcal{M}^{\mathbb{C}^*}.$$

This observation is used in [LQ] as follows. The torus acts on the subscheme

$$\{ p \in \mathcal{M} \setminus \mathcal{M}^{\mathbb{C}^*} : \nu_{\mathcal{M}}(p) = k \}$$

with no fixed points, hence its Euler characteristic must be zero, cf. [LY]. Therefore,

$$\chi(\{ p \in \mathcal{M} : \nu_{\mathcal{M}}(p) = k \}) = \chi(\{ p \in \mathcal{M}^{\mathbb{C}^*} : \nu_{\mathcal{M}}(p) = k \})$$

yielding

$$\tilde{\chi}(\mathcal{M}) = \sum_k k \chi(\{ p \in \mathcal{M}^{\mathbb{C}^*} : \nu_{\mathcal{M}}(p) = k(-1)^{\epsilon_p} \}).$$

We will apply these remarks to the action of $\mathbb{C}^*$ on $\mathcal{M}_n$ induced by the scaling action in the fibers of the projective bundle $X \to S$ and the scaling action on the framing coming from a generic embedding

$$\mathbb{C}^* \hookrightarrow \text{GL}_r.$$

We will find the torus fixed points in $\mathcal{M}_n$.

**Lemma 2.** Assume $\delta$ is good. The $\mathbb{C}^*$-fixed framed modules in $\mathcal{M}_n$ take the form

$$E = \bigoplus_{i=1}^r I_{Z_i}$$

where $Z_i$ are zero dimensional subschemes of $X$ invariant under the action of the torus, of total length $n$. The framing $\phi$ is the natural composition $E \hookrightarrow \mathcal{O}_X^r \to \mathcal{O}_{S,\infty}^r$.

**Proof.** We first prove that all invariant framed modules are torsion free. Indeed, the torsion module $T$ of $E$ is $\mathbb{C}^*$-fixed. By stability, the framing $\phi$ gives a $\mathbb{C}^*$-invariant injection

$$\phi_\infty : T \hookrightarrow \mathcal{O}_{S,\infty} \otimes \mathbb{C}^r.$$
Therefore, the torsion module splits
\[ T = \bigoplus_{j=1}^r i_*(I_{W_j} \otimes M_j), \]
for zero dimensional subschemes \( W_j \) of \( S_\infty \) and line bundles \( M_j \) over \( S_\infty \). Again by stability applied to the torsion submodule \( T \) we find
\[ p_T - \delta \leq 0 \implies \ell \cdot H^2 \cdot S_\infty \leq \delta_0 \]
where \( \delta_0 \) is half the quadratic coefficient of \( \delta \). By assumption, we may take \( \delta_0 < H^2 \cdot S_\infty \), implying that \( \ell = 0 \) and showing that the torsion module vanishes.

Now, since \( E \) is torsion free and \( \mathbf{C}^*_\text{-}\text{invariant} \), the argument of [BPT] shows that
\[ E = \bigoplus_{i=1}^r I_{Z_i} \otimes L_i \]
where \( L_i \) are line bundles over \( X \) and \( Z_i \) are subschemes of dimension at most 1. The subschemes \( Z_i \) must be torus invariant. Since \( c_1(E) = c_2(E) = 0 \), we find
\[ \sum c_1(L_i) = 0 \]  \hspace{1cm} (1)
and furthermore
\[ \sum_{i=1}^r c_1(L_i)^2 = 2 \sum_{i=1}^r [Z_i]. \]  \hspace{1cm} (2)

Since the framed module \( (E, \phi) \) is semistable, for all submodules \( F \) of \( E \) of positive rank we must have
\[ \frac{c_1(F) \cdot H^2 - \delta_0}{\text{rk } F} \leq \frac{c_1(E) \cdot H^2 - \delta_0}{\text{rk } E}. \]
Taking \( F = I_{Z_i} \otimes L_i \) we find
\[ c_1(L_i) \cdot H^2 \leq \delta_0 \left(1 - \frac{1}{r}\right). \]

Now, since \( \delta_0 \) is sufficiently small compared to the denominator of the rational divisor \( H \), we conclude
\[ c_1(L_i) \cdot H^2 \leq 0. \]
In fact, \( c_1(L_i) \cdot H^2 = 0 \) for all \( i \), because of (1). We argue that \( L_i \) are trivial and \( Z_i \) are zero dimensional.

Write
\[ c_1(L_i) = \pi^* D_i + d_i \zeta, \]
where \( D_i \) are divisors on the surface \( S \) and \( \zeta = c_1(O_p(1)) \). We calculate
\[ \sum_i c_1(L_i)^2 = \pi^* \left( \sum_i D_i^2 \right) + 2 \sum_i d_i (\pi^* D_i \cdot \zeta) + (\sum_i d_i^2) \zeta^2 = 2 \sum_i [Z_i]. \]  \hspace{1cm} (3)
Set
\[ M = 2 \sum_i d_i D_i - \left( \sum_i d_i^2 \right) K_S. \]
Using
\[ \zeta^2 + K_S \cdot \zeta = 0, \]
we conclude from (3) that
\[ \pi^* \left( \sum_i D_i^2 \right) + \pi^* M \cdot \zeta = 2 \sum_i [Z_i]. \] (4)

Pushing (4) forward under \( \pi \) we find
\[ M = 2 \sum_i \pi_* [Z_i]. \]

As a consequence, \( M \) is effective. The requirement that the slopes of \( L_i \) are trivial translates into the condition
\[ (\pi^* D_i + d_i \zeta)(\pi^* H_0 + \epsilon \zeta)^2 = 0 \]
which rewrites as
\[ (D_i - d_i K_S) \cdot \Sigma = -d_i \]
where
\[ \Sigma = \frac{\epsilon(2H_0 - \epsilon K_S)}{H_0^2} \]
is an ample rational curve class on \( S \) for small \( \epsilon \). Write
\[ F_i = D_i - d_i K_S \]
so that
\[ F_i \cdot \Sigma = -d_i. \]

Since
\[ M = 2 \sum_i d_i F_i + \left( \sum_i d_i^2 \right) K_S \]
is effective, its intersection with \( \Sigma \) must be positive. This gives
\[ -2 \sum_i d_i^2 + \left( \sum_i d_i^2 \right) K_S \cdot \Sigma \geq 0. \]
For small \( \epsilon \) we have \( K_S \cdot \Sigma < 2 \). We conclude from here that \( d_i = 0 \) for all \( i \).
Therefore \( M = 0 \), and by (4) we must have
\[ \pi^* \left( \sum_i D_i^2 \right) = 2 \sum_i [Z_i]. \]
is effective. Note that the left hand side is supported on fibers. Therefore,
\[ \sum_i D_i^2 \geq 0. \] (5)

We moreover proved
\[ F_i \cdot \Sigma = 0 \implies D_i \cdot \Sigma = 0. \]

Since \( \Sigma \) is ample, by Hodge index theorem we have
\[ D_i^2 \leq 0, \]
with equality only if \( D_i \) is numerically equivalent to 0. In fact equality must occur because of (5). This yields \( c_1(L_i) = 0. \) In turn,
\[ L_i = \pi^*N_i \]
for some line bundles \( N_i \to S \) of first Chern class 0. Furthermore, from (3) we find \([Z_i] = 0\) hence \( Z_i \) must be zero dimensional.

Thus
\[ E = \bigoplus_{i=1}^{r} I_{Z_i} \otimes \pi^*N_i, \]
where
\[ \sum \ell(Z_i) = n. \]

Clearly, \( Z_i \) must be torus invariant and \( \phi = \bigoplus \phi_i \) where
\[ \phi_i : I_{Z_i} \otimes \pi^*N_i \to \mathcal{O}_{S_{\infty}}. \]

We next claim that \( \phi_i \neq 0 \) for all \( i \). Indeed, if \( \phi_i = 0 \) for some \( i \), then \( I_{Z_i} \otimes \pi^*N_i \) is in the kernel of \( \phi \), yielding by stability
\[ \chi(I_{Z_i}(mH) \otimes \pi^*N_i) \leq \frac{1}{r} \left( \sum_j \chi(I_{Z_j}(mH) \otimes \pi^*N_j) - \delta \right). \]

This gives
\[ n \geq \ell(Z_i) \geq \frac{n}{r} + \frac{\delta}{r}. \]

This is a contradiction since \( \delta > (r-1)n \). Therefore \( \phi_i \neq 0 \), showing that there exists a non-zero morphism \( N_i \to \mathcal{O}_S. \) Therefore \( N_i \) must be trivial, completing the proof.

Over the open moduli space \( \mathcal{M}^\circ \), the same result holds without any restrictions on \( \delta \):
Lemma 2A. The $C^*$-fixed framed sheaves $E$ in $\mathcal{M}_n^s$ must split

$$E = \bigoplus_{i=1}^r I_{Z_i}$$

where $Z_i$ are zero dimensional subschemes of $X^\circ$ invariant under the action of the torus, of total length $n$.

Proof. By assumption $E$ is torsion free, hence

$$E = \bigoplus_{i=1}^r I_{Z_i} \otimes L_i.$$  

Since the framing is an isomorphism, we conclude $Z_i$ is contained in $X^\circ$ and $L_i$ is trivial on $S_\infty$. Hence,

$$L_i = \mathcal{O}(d_i S_0)$$

for some integers $d_i$. We claim that $d_i = 0$ for all $i$. This in turn implies that $Z_i$ are zero dimensional by using $c_2(E) = 0$.

Assume first that the quadratic term of $\delta$ is sufficiently small. This case is already covered by Lemma 2, but a simpler argument is possible over $\mathcal{M}_n^s$; we record it here for future reference. Indeed, the stability condition applied to $L_i \otimes I_{Z_i}$ gives

$$d_i S_0 \cdot H^2 = c_1(L_i) \cdot H^2 \leq \delta_0 \left(1 - \frac{1}{r}\right) \implies d_i \leq 0.$$  

Since $c_1(E) = 0$, we have $\sum_i d_i = 0$. Hence $d_i = 0$ for all $i$, as claimed.

We now give the general argument. Using that $c_1(E) = c_2(E) = 0$, we find

$$\left(\sum_i d_i^2\right) S_0^2 = 2 \sum_i [Z_i].$$

Assume not all $d_i$ are equal to 0. Since the $Z_i$’s are torus invariant and disjoint from $S_\infty$, their cohomology classes are supported on the surface $S_0$. Using that

$$S_0^2 = (K_S^2) f + \zeta \cdot \pi^* K_S,$$

from equation (6) we find $K_S^2 = 0$. From here, pushing forward under $\pi$, we conclude

$$\left(\sum_i d_i^2\right) K_S = 2 \sum_i \pi_* [Z_i]$$

which is effective. Hence

$$K_S \cdot H_0 \geq 0.$$  

Now, $I_{Z_i}(d_i S_0 - S_\infty)$ is contained in the kernel of $\phi$. Hence by stability

$$(d_i S_0 - S_\infty) \cdot H^2 \leq -\delta_0 \frac{r}{r}.$$
Pick an index $i$ such that $d_i \geq 1$. The above inequality implies

$$(S_0 - S_\infty) \cdot H^2 \leq -\frac{\delta_0}{r} \implies \pi^* K_S \cdot H^2 \leq -\frac{\delta_0}{r}.$$  

However,

$$\pi^* K_S \cdot H^2 = \epsilon \cdot (2H_0 - \epsilon K_S) K_S = 2 \epsilon \cdot K_S \cdot H_0 \geq 0.$$  

Therefore, $K_S \cdot H_0 = \delta_0 = 0$. Since the quadratic term of $\delta$ is 0, the previous paragraph applies, showing that in fact all $d_i = 0$. 

**Lemma 3.** If $\delta$ is good, all torus fixed framed sheaves $E$ in $\mathcal{M}_n$ described above are stable.

**Proof.** Let $F$ be a subsheaf of $E = \oplus I_{Z_i}$ of rank $r'$. Since $F$ is a subsheaf of $O_X$, by Gieseker semistability we have

$$P_F \leq r' \chi(mH) < \frac{r'}{r} P_E + \frac{r - r'}{r} \delta,$$

at least when $r' \neq r$, using that $\delta > (r - 1)n$. When $r' = r$, induction on $r$ yields the claim. For the inductive step, consider the non-zero map $F \to I_{Z_i}$, and write $F'$ for the kernel. Then, apply the induction hypothesis to $F'$ which is contained in $\oplus_{i=1}^{r-1} I_{Z_i}$.

Next, assume $F$ is in the kernel of $\phi$. The kernel of $\phi$ is contained in $O_X(-S_\infty)^r$ (and it is isomorphic to $\oplus_{j} I_{Z_j}(-S_\infty)$ for $E$ in $\mathcal{M}_n^2$). By Gieseker-semistability, we have

$$P_F \leq r' \chi(mH - S_\infty) < \frac{r'}{r} (r \chi(mH) - n - \delta) = \frac{r'}{r} (P_E - \delta),$$

using that $\delta$ is good.

**Lemma 4.** For all torus fixed sheaves $E$ in $\mathcal{M}_n$, we have

$$\dim T_E \mathcal{M}_n \equiv nr \mod 2.$$  

**Proof.** Since $E = \oplus I_{Z_i}$ is stable, the tangent space is calculated in [HL]:

$$T_E \mathcal{M}_n = \text{Ext}^1(E, E(-S_\infty)) = \sum_{i,j} \text{Ext}^1(I_{Z_i}, I_{Z_j}(-S_\infty)).$$

We consider first the contributions of terms corresponding to pairs of indices $(i, j)$ and $(j, i)$ for $i \neq j$:

$$\text{Ext}^1(I_{Z_i}, I_{Z_j}(-S_\infty)) + \text{Ext}^1(I_{Z_j}, I_{Z_i}(-S_\infty))$$
\[ = \text{Ext}^1(I_{Z_i}, I_{Z_j}(-S_\infty)) + \text{Ext}^2(I_{Z_i}, I_{Z_j}(-S_\infty)) \]

by Serre duality. Now, considering the above expression modulo 2 we obtain

\[ \chi(I_{Z_i}, I_{Z_j}(-S_\infty)) + \text{Ext}^0(I_{Z_i}, I_{Z_j}(-S_\infty)) + \text{Ext}^3(I_{Z_i}, I_{Z_j}(-S_\infty)). \]

Next, it is easily seen that \( \text{Ext}^0 \) vanishes, and same for \( \text{Ext}^3 \) by duality. Thus, we are left with

\[ \chi(I_{Z_i}, I_{Z_j}(-S_\infty)) = \chi(O_{X}, O_{Z_i}(-S_\infty)) - \chi(O_{Z_i}, O_{Z_j}) - \chi(O_{Z_j}, O_{X}(-S_\infty)) = \ell(Z_i) + \ell(Z_j) \mod 2. \]

We consider now the terms with \( i = j \):

\[ \text{Ext}^1(I_{Z_i}, I_{Z_i}(-S_\infty)). \]

This term was already worked out in the deformation theory of Example 1. We obtained

\[ \text{Ext}^1(I_{Z_i}, I_{Z_i}(-S_\infty)) = \text{Ext}^0(I_{Z_i}, O_{Z_i}) \equiv \ell(Z_i) \mod 2, \]

where for the last congruence we used [BF] or [MNOP]. The lemma follows by collecting the above facts.

We can now put together the calculation of Lemma 4 and the remarks about Behrend functions in Subsection 4.1 to calculate the Donaldson-Thomas Euler characteristic of \( \mathcal{M}^n \). We write \( \mathcal{X}_\ell \) for the subset of the Hilbert scheme of points in \( X^\ell \) which parametrizes torus fixed \( Z \)'s of length \( \ell \). For each partition \( \ell \) into \( r \) parts \( (\ell_1, \ldots, \ell_r) \) with \( \ell_1 + \ldots + \ell_r = n \) we write

\[ \mathcal{X}^\ell = \mathcal{X}^{\ell_1} \times \ldots \times \mathcal{X}^{\ell_r}. \]

Then, \( \mathcal{X}^\ell \) are the \( C^* \)-fixed loci of \( \mathcal{M}^n \). With the convention that

\[ \vec{Z} = (Z_1, \ldots, Z_r) \]

represents an \( r \)-tuple of schemes in \( \mathcal{X}^\ell \), we calculate

\[ \bar{\chi}(\mathcal{M}^n) = \]

\[ = \sum_{\vec{\ell}} \sum_k k \chi(\{ \vec{Z} \in \mathcal{X}^{\ell} : \nu_{\mathcal{X}^{\ell}}(\vec{Z}) = k(-1)^{\ell - \dim \mathcal{T}_{\mathcal{X}^{\ell}}} \}) \]

\[ = (-1)^{(r-1)n} \sum_{\vec{\ell}} \prod_{i=1}^r k_i \chi(\{ Z_i \in \mathcal{X}^{\ell_i} : \nu_{\mathcal{X}^{\ell_i}}(Z_i) = k_i(-1)^{\ell_i - \dim \mathcal{T}_{\mathcal{X}^{\ell_i}}} \}) \]

\[ = (-1)^{(r-1)n} \prod_{i=1}^r \left( \sum_{k} k \chi(\{ Z : \nu_{\mathcal{X}^{\ell_i}}(Z) = k(-1)^{\ell_i - \dim \mathcal{T}_{\mathcal{X}^{\ell_i}}} \}) \right) \]
By applying these results when \( r = 1 \), and using the identification of the rank 1 moduli space with the Hilbert scheme worked out in Example 1, we obtain
\[
\check{\chi}((X^\circ)^{|f|}) = \sum_{k} k\chi(\{Z : \nu X^\circ(Z) = k(-1)^{\ell - \dim T_Z X^\circ}\}).
\]
This yields
\[
\check{\chi}(\mathfrak{M}^\circ_n) = (-1)^{(r-1)n}\sum_{\ell} \check{\chi}((X^\circ)^{|f|_\ell}) \cdots \check{\chi}((X^\circ)^{|f_r|}).
\]
We form the generating series
\[
\sum_n q^n \check{\chi}(\mathfrak{M}^\circ_n) = \left(\sum_{\ell} ((-1)^{(r-1)}q)^{\ell} \check{\chi}((X^\circ)^{|f|_\ell})\right)^r.
\]
Now, from [BF] we lift the calculation
\[
\sum_{\ell} q^{\ell} \check{\chi}((X^\circ)^{|f|}) = M(-q)^{r e(X^\circ)} = M(-q)^{r e(S)};
\]
where \( M(q) \) is the MacMahon function
\[
M(q) = \prod_{k=1}^{\infty} (1 - q^k)^{-k}.
\]
To summarize, for \( \delta \) good, we proved

**Theorem 5.** The following equality holds
\[
\sum_{n \geq 0} q^n \check{\chi}(\mathfrak{M}^\circ_n) = M((-1)^r q)^{r e(S)}.
\] (7)

We are unable to define (and calculate) the virtual motive \([\mathfrak{M}^\circ_n]^{vir}\), as it is done in rank 1 in [BBS] and for surfaces in [N]. This question may deserve further study.

### 4.2 Topological Euler characteristics

Lemma 1 also allows us to calculate the topological Euler characteristics of the compact spaces \( \mathcal{M}_n \) via the localization results of [LY]:
\[
e(\mathfrak{M}_n) = e(\mathfrak{M}_n^C).
\]
The same calculation as above shows that
\[
\sum q^n e(\mathfrak{M}_n) = \left(\sum q^n e(\mathcal{I}_n)\right)^r = M(q)^{r e(X)}
\] (8)
where \( \mathcal{I}_n \cong X_n^{|n|} \) denotes the rank 1 moduli space. The series needed here
\[
\sum q^n e(\mathcal{I}_n) = M(q)^{e(X)}
\]
is computed in [C]. The answer we found is valid whenever \( \delta \) is good.
4.3 Blow-up surfaces

A slightly more complicated example arises by considering blow-up geometries. Indeed, assume that the surface $S$ contains a $(-1)$-curve $C$. Then, $C \hookrightarrow S_0$ is super-rigid in $X$:

$$N_{C/X} = N_{C/S} \oplus N_{S_0/X}|_C = \mathcal{O}_C(-1) \oplus \mathcal{O}_C(-1).$$

We consider the moduli space $\mathfrak{M}_{n,k}$ of rank $r$ modules over $X$ framed at infinity, with numerics

$$c_1 = 0, \quad c_2 = k[C], \quad \chi = r\chi(\mathcal{O}_X) - n.$$ 

In order not to worry about stability, we assume that $\delta$ is good of degree 2. By the argument in the first two paragraphs of Lemma 2A, the torus fixed sheaves in $\mathfrak{M}_{n,k}$ take the form

$$E = \oplus_i I_{Z_i},$$

where $Z_i$ may have at most 1 dimensional components contained in $X^o$. Furthermore,

$$\sum_i [Z_i] = k[C], \quad \sum_i \chi(\mathcal{O}_{Z_i}) = n.$$ 

In fact, $[Z_i] = k_i[C]$ for non-negative integers $k_i$ adding up to $k$. Indeed, after projecting to the blowdown surface $X \to S \to \bar{S}$

the classes of the effective curves $Z_i$ add up to 0, hence they must be trivial. This shows that the components of $Z_i$ are supported on the fibers of $X \to S$ or are contained in the Hirzebruch surface $F = \mathbb{P}(K_S|_C \oplus \mathcal{O}_C) \to C$.

In fact, by torus invariance, all components of $Z_i$ must be supported over fibers or over the zero section $C \hookrightarrow S_0$. Since

$$\sum_i [Z_i] = k[C]$$

contains no fiber classes, or alternatively since the framing must be an isomorphism along $S_{\infty}$, we conclude that $Z_i$ has no support over fibers, hence $[Z_i] = k_i[C]$ as claimed.

We carry out the computation of the Donaldson-Thomas Euler characteristics. In the new setting, for all $\mathbb{C}^*$-fixed sheaves $E$ in $\mathfrak{M}_{n,k}$ we have

$$\dim T_E \mathfrak{M}_{n,k} \equiv rn - k \mod 2.$$
The proof follows that of Lemma 4. The only change is the calculation
\[ \dim \text{Ext}^1(I_{Z_i}, I_{Z_i}(-S_\infty)) \equiv \chi(O_{Z_{i}}) - k_i \mod 2. \]

To this end, consider the exact sequence
\[ 0 \rightarrow I_{Z_i}(-S_\infty) \rightarrow I_{Z_i} \rightarrow O_{S_\infty} \rightarrow 0. \]
Since the map \( \text{Ext}^0(I_{Z_i}, I_{Z_i}) \rightarrow \text{Ext}^0(I_{Z_i}, O_{S_\infty}) \) is an isomorphism, we obtain the exact sequence
\[ 0 \rightarrow \dim \text{Ext}^1(I_{Z_i}, I_{Z_i}(-S_\infty)) \rightarrow \text{Ext}^1(I_{Z_i}, I_{Z_i}) \rightarrow \text{Ext}^1(I_{Z_i}, O_{S_\infty}). \]

To find the last group, we use the local to global spectral sequence
\[ H^p(\text{Ext}^q(I_{Z_i}, O_{S_\infty})) \Rightarrow \text{Ext}^{p+q}(I_{Z_i}, O_{S_\infty}). \]
The terms with \( q \geq 1 \) vanish since \( Z_i \) avoids \( S_\infty \), while the \( q = 0 \) terms equal \( H^p(O_S) \). Therefore,
\[ \text{Ext}^1(I_{Z_i}, O_{S_\infty}) = H^1(O_S). \]
From the exact sequence, we conclude
\[ \text{Ext}^1(I_{Z_i}, I_{Z_i}(-S_\infty)) = \text{Ext}^1(I_{Z_i}, I_{Z_i})_0. \]
The dimension of the last vector space was found in [BB] using Theorem 2 of [MNOP]. The answer is
\[ \text{Ext}^1(I_{Z_i}, I_{Z_i}(-S_\infty)) \equiv \chi(O_{Z_{i}}) - k_i \mod 2 \]
as claimed above.

We form the generating series
\[ \sum_{n,k} \chi(\mathcal{M}^n_{n,k}) q^n v^k = \left( \sum_{n,k} ((-1)^{r-1} q)^n v^k \chi(\mathcal{M}^0_{n,k}) \right)^r \]
where \( \mathcal{M}^0_{n,k} \) denotes the rank 1 framed moduli space. This is isomorphic to the Hilbert scheme. The rank 1 Donaldson-Thomas invariants of super-rigid curves were calculated in [BB]:
\[ \sum_{n,k} q^n v^k \chi(\mathcal{M}^0_{n,k}) = M(-q)^{e(X^\infty)} \cdot \prod_{m=1}^{\infty} (1 - (-q)^{m} v)^{m}. \]
Therefore, we obtain
\[ \sum_{n,k} \chi(\mathcal{M}^0_{n,k}) q^n v^k = M((-1)^r q)^{e(S)} \cdot \prod_{m=1}^{\infty} (1 - ((-1)^r q)^{m} v)^{mr}. \]
4.4 Other geometries

There are other geometries for which the above methods apply. We discuss some of them here. Most straightforwardly, assuming $S_\infty$ is a smooth framing divisor with

$$K_X = -2S_\infty,$$

then our techniques yield

$$\sum_n q^n \tilde{\chi}(\mathcal{M}_n^\infty) = M((-1)^r q)^{\text{re}(X^\infty)}.$$  \hfill (10)

In order to make the proof of Lemma 2A work, we need to assume for instance that the restriction

$$\text{Pic}(X) \cap (H^2)^\perp \to \text{Pic}(S_\infty)$$

is injective. If $X$ is Fano of index 2, this requirement is satisfied by the Lefschetz hyperplane theorem applied to the ample class $S_\infty$. Examples pertinent to this setting include, among others:

- $X$ is a cubic in $\mathbb{P}^4$ or a $(2,2)$ complete intersection in $\mathbb{P}^5$, and $S_\infty$ is a hyperplane section;
- $X$ is a double cover of $\mathbb{P}^3$ branched along a quartic, and $S_\infty$ is the pullback of a hyperplane.

Fano threefolds of index higher than 2 also yield symmetric perfect obstruction theories. This can be checked directly using the well-known classification:

- $X$ is a quadric in $\mathbb{P}^4$ or $X = \mathbb{P}^3$, and $S_\infty$ is a hyperplane section.

In index 3, more examples arise from the curve geometry:

- $X = \mathbb{P}(\mathcal{O}_C + E) \to C$, with $E \to C$ any rank 2 bundle of determinant $\det E = K_C$, and $S_\infty$ the divisor at infinity.

Since the same argument works in all cases above, let us only discuss the rank $r$ sheaves over $\mathbb{P}^3$ framed along the plane at infinity $\mathbb{P}^2 \hookrightarrow \mathbb{P}^3$. Along $\mathcal{M}_n^\infty$, the obstruction theory is symmetric since

$$T_k \mathcal{M}_n^\infty = \text{Ext}^1(E, E(-1)) = \text{Ext}^1(E, E(-3)) = \text{Ext}^2(E, E(-1))^\vee = \text{ob}_E^\infty.$$  

The second isomorphism follows from the short exact sequences

$$0 \to E(-k-1) \to E(-k) \to \mathcal{O}_{\mathbb{P}^2}(-k) \to 0$$

for $k = 1$ and $k = 2$, and the vanishings

$$\text{Ext}^0(E, \mathcal{O}_{\mathbb{P}^2}(-k)) = \text{Ext}^1(E, \mathcal{O}_{\mathbb{P}^2}(-k)) = 0 \text{ for } k = 1, 2.$$  

The first vanishing is clear. The second follows from the local to global spectral sequence:

$$E_2^{p,q} = H^p(\text{Ext}^q(E, \mathcal{O}_{\mathbb{P}^2}(-k))) \to \text{Ext}^{p+q}(E, \mathcal{O}_{\mathbb{P}^2}(-k))$$

with vanishing $E_2$ terms when $p + q = 1$. This proves the claim about the obstruction theory. Equation (10) still holds by the same methods.
5 Acknowledgements

The author gratefully acknowledges correspondence with Balázs Szendrői, as well as support from the NSF via grant DMS 1001486 and from the Sloan Foundation.

References


ON THE GENERALIZED SEMI-RELATIVISTIC SCHröDINGER-POISSON SYSTEM IN $\mathbb{R}^n$

WALID ABOU SALEM, THOMAS CHEN, AND VITALI VOUGALTER

Received: December 21, 2012
Revised: March 23, 2013

Communicated by Heinz Siedentop

ABSTRACT. The Cauchy problem for the semi-relativistic Schrödinger-Poisson system of equations is studied in $\mathbb{R}^n$, $n \geq 1$, for a wide class of nonlocal interactions. Furthermore, the asymptotic behavior of the solution as the mass tends to infinity is rigorously discussed, and compared with solutions to the non-relativistic Schrödinger-Poisson system.

2010 Mathematics Subject Classification: 82D10, 82C10
Keywords and Phrases: Schrödinger-Poisson system, mean-field dynamics, long-range interaction, functional spaces, density matrices, Cauchy problem, global existence, infinite mass limit.

1 INTRODUCTION

1.1 MOTIVATION AND HEURISTIC DISCUSSION

In this article, we study the global Cauchy problem for the semi-relativistic Schrödinger-Poisson system in $\mathbb{R}^n$, $n \geq 1$, for a wide class of nonlocal interactions, both in the attractive and repulsive cases. This system is relevant to the description of many-body semi-relativistic quantum particles in the mean-field limit. We consider a system of $N$ semi-relativistic quantum particles in $\mathbb{R}^n$, $n \geq 1$ with long-range two-body interactions $g \sum_{1\leq i<j\leq N} \frac{1}{|x_i-x_j|^\gamma}$, with $0 < \gamma \leq 1$ if $n \geq 2$, and $0 < \gamma < 1$ if $n = 1$, and with $g \in \mathbb{R}$. In the mean-field
limit, one can formally show that the density matrix that describes the mixed state of the system satisfies the Hartree-von Neumann equation

\[
\begin{cases}
i \partial_t \rho(t) = [H_m + w(t) \ast n(t), \rho(t)], & x \in \mathbb{R}^n, \ n \geq 1, \ t \geq 0 \\
H_m = \sqrt{m^2 - \Delta} - m, \ w(t) = g \frac{1}{|x|^n}, \ n(t, x) = \rho(t, x, x), \ \rho(0) = \rho_0
\end{cases}
\]

where \( \Delta \) stands for the \( n \)-dimensional Laplacian, \( \ast \) stands for convolution in \( \mathbb{R}^n \), and \( m \geq 0 \) is the mass.\(^1\) Since \( \rho_0 \) is a positive, self-adjoint trace-class operator acting on \( L^2(\mathbb{R}^n) \), its kernel can be decomposed with respect to an orthonormal basis of \( L^2(\mathbb{R}^n) \),

\[
\rho_0(x, y) = \sum_{k \in \mathbb{N}} \lambda_k \psi_k(x) \overline{\psi_k(y)}
\]

where \( \{\psi_k\}_{k \in \mathbb{N}} \) denotes an orthonormal basis of \( L^2(\mathbb{R}^n) \). Furthermore,

\[
\Delta := \{\lambda_k\}_{k \in \mathbb{N}} \in l^1, \ \lambda_k \geq 0, \ \sum_k \lambda_k = 1.
\]

We will show that there exists a one-parameter family of complete orthonormal bases of \( L^2(\mathbb{R}^n) \), \( \{\psi_k(t)\}_{k \in \mathbb{N}} \), for \( t \in \mathbb{R}_+ \), such that the kernel of the solution \( \rho(t) \) to (1.1) can be represented as

\[
\rho(t, x, y) = \sum_{k \in \mathbb{N}} \lambda_k \psi_k(t, x) \overline{\psi_k(t, y)}.
\]

Substituting (1.3) in (1.1), the one-parameter family of orthonormal vectors \( \{\psi_k(t)\}_{k \in \mathbb{N}} \) is seen to satisfy the semi-relativistic Schrödinger-Poisson system

\[
\frac{i}{\partial t} \psi_k = H_m \psi_k + V \psi_k, \ k \in \mathbb{N}
\]

\[
V[\Psi] = w \ast n[\Psi], \ \Psi := \{\psi_k\}_{k=1}^{\infty},
\]

\[
n[\Psi(x, t)] = \sum_{k=1}^{\infty} \lambda_k |\psi_k|^2.
\]

The purpose of this note is to show global well-posedness of (1.4) in a suitable Banach space (to be specified below), and to study the asymptotics of the solution as the mass \( m \) tends to \( \infty \), which we compare to solutions to the non-relativistic Schrödinger-Poisson system, see [11]. The semi-relativistic Schrödinger-Poisson system of equations in a finite domain of \( \mathbb{R}^3 \) and with repulsive Coulomb interactions has been studied recently in [1, 2]. Here, we generalize the result of [1] in several ways. First, the problem is studied in \( \mathbb{R}^n, \ n \geq 1.\)

\(^1\)The rigorous derivation of the semi-relativistic Hartree-von Neumann equation is a topic of future work, see [3, 4] for a derivation of this system of equations in the non-relativistic case.
Semi-relativistic Schrödinger-Poisson system in $\mathbb{R}^n$

Second, we consider a wide class of nonlocal interactions in both the attractive and repulsive cases, and which includes the repulsive Coulomb case in three spatial dimensions. Third, in the infinite mass limit $m \to \infty$, we prove that solutions to the semi-relativistic, and to the non-relativistic Schrödinger-Poisson systems become indistinguishable; the latter has been studied extensively, see for example [5, 8] and references therein. In the special case when the initial density matrix is a pure state $\rho_0 = |\psi_0\rangle\langle\psi_0|$, the Schrödinger-Poisson system becomes a single Hartree equation

$$i\partial_t \psi = (\sqrt{m^2 - \Delta} - m)\psi + (w_\gamma \ast |\psi|^2)\psi, \quad \psi(0) = \psi_0.$$ 

In that sense, our analysis generalizes the results of [10, 7] to the effective dynamics of a mixed state of a semi-relativistic system.

The organization of this paper is as follows. In Subsection 1.3 we state our main results. We prove local and global well-posedness in Section 2. Finally, in Section 3, we discuss the asymptotic behavior of the solutions as the mass tends to infinity. For the benefit of a general reader, we recall some useful results about fractional integration and fractional Leibniz rule in Appendix A.

1.2 Notation

- $A \lessapprox B$ means that there exists a positive constant $C$ independent mass $m$ such that $A \leq C B$.
- $L^p$ stands for the standard Lebesgue space. Furthermore, $L^p B = L^p(I; B)$. $\langle \cdot, \cdot \rangle_{L^2}$ denotes the $L^2(\mathbb{R}^n)$ inner product. We will often use the abbreviated notation $L^p_T$ for $L^p([0,T])$, in the situation where $[0,T]$ denotes a time interval.
- $l^1 = \{ \{ a_l \}_{l \in \mathbb{N}} \mid \sum_{l \geq 1} |a_l| < \infty \}$. 
- $W^{s,p} = (-\Delta + 1)^{-s} L^p$, the standard (complex) Sobolev space. When $p = 2$, $W^{s,2} = H^s$. $H^s$ denotes the homogeneous Sobolev space with norm $\|\psi\|_{H^s} = (\langle \psi, (-\Delta)^s \psi \rangle_{L^2})^{1/2}$.
- For fixed $\lambda \in l^1$, $\lambda_k \geq 0$, and for sequences of functions $\Phi := \{ \phi_k \}_{k \in \mathbb{N}}$ and $\Psi := \{ \psi_k \}_{k \in \mathbb{N}}$, we define the inner product

$$\langle \Phi, \Psi \rangle_{L^2} := \sum_{k \geq 1} \lambda_k \langle \phi_k, \psi_k \rangle_{L^2},$$

which induces the norm

$$\|\Phi\|_{L^2} = \left( \sum_{k \geq 1} \lambda_k \|\phi_k\|_{L^2}^2 \right)^{1/2}.$$ 

The corresponding Hilbert space is $L^2$. 

Documenta Mathematica 18 (2013) 343–357
For fixed $\lambda \in l^1$, $\lambda_k \geq 0$,

$$\mathcal{H}^s = \{ \Psi = \{ \psi_k \}_{k \in \mathbb{N}} \mid \psi_k \in H^s, \sum_{k \geq 1} \lambda_k \| \psi_k \|^2_{H^s} < \infty \}$$

is a Banach space with norm $\| \Psi \|_{\mathcal{H}^s} = (\sum_{k \geq 1} \lambda_k \| \psi_k \|^2_{H^s})^{\frac{1}{2}}$.

For fixed $\lambda \in l^1$, $\lambda_k \geq 0$,

$$\dot{\mathcal{H}}^s = \{ \dot{\Psi} = \{ \dot{\psi}_k \}_{k \in \mathbb{N}} \mid \dot{\psi}_k \in \dot{H}^s, \sum_{k \geq 1} \lambda_k \| \dot{\psi}_k \|^2_{\dot{H}^s} < \infty \}$$

is a Banach space with norm $\| \dot{\Psi} \|_{\dot{\mathcal{H}}^s} = (\sum_{k \geq 1} \lambda_k \| \dot{\psi}_k \|^2_{\dot{H}^s})^{\frac{1}{2}}$.

1.3 Statement of main results

For $s > 0$, we define the state space for the Schrödinger-Poisson system by

$$S^s := \{(\Psi, \Delta) \mid \Psi = \{ \psi_k \}_{k=1}^{\infty} \text{ is a complete orthonormal system in } L^2(\mathbb{R}^n), \Delta = \{ \lambda_k \}_{k \in \mathbb{N}} \in l^1, \lambda_k \geq 0 \}.$$ 

The following is our first main result about the global Cauchy problem.

**Theorem 1.1.** Consider the system of equations (1.4)-(1.6), with $m \geq 0$, with $0 < \gamma \leq 1$ if $n \geq 2$, and $0 < \gamma < 1$ if $n = 1$. Suppose that $(\Psi(0), \Delta) \in S^s$, $s \geq 1/2$. If $g \geq 0$, or $g < 0$ with $\| \Psi(0) \|_{L^2}$ small enough, then there is a unique mild solution $(\Psi, \Delta) \in C([0, \infty), S^s)$.

**Remark 1.2.** $\Delta$ is time-independent, and hence the evolution can be thought as that of $\Psi \in \mathcal{H}^s$.

**Remark 1.3.** Local well-posedness requires less regularity, in particular, $s \geq \gamma/2$, see Proposition 2.2 in Section 2.1. On the other hand, in order to enhance local to global well-posedness, energy conservation is used, and consequently, $s \geq \frac{1}{2}$ is assumed to ensure finiteness of the energy.

**Remark 1.4.** It follows from the proof of local well-posedness (Proposition 2.2 in Section 2.1) that there exists a positive time $T$ independent of $m \geq 0$ such that $\| \Psi \|_{L^\infty_T \mathcal{H}^s} \leq C \| \Psi(0) \|_{\mathcal{H}^s}$, $s \geq \gamma/2$, where $C > 0$ is independent of $m$.

**Remark 1.5.** The solution is continuous in the mass $m$. In particular, as $m \searrow 0$, and for $T > 0$ fixed, $\Psi \rightarrow \Psi(0)$ strongly in $L^\infty_T(\mathcal{H}^s)$, $s \geq 1/2$, where $\Psi(0)$ satisfies (1.4)-(1.6) with $m = 0$ and initial condition $\Psi(0)$, see Proposition 2.6 in Sect. 2.
The second result is about the infinite mass limit. Let $\Gamma$ satisfy the nonrelativistic Schrödinger-Poisson system of equations

$$i\frac{\partial \psi_k}{\partial t} = -\frac{1}{2m} \Delta \psi_k + V(\psi_k), \quad k \in \mathbb{N}$$

$$V[\Psi] = w_\gamma * n[\Psi], \quad \Psi := \{\psi_k\}_{k=1}^\infty, \quad n[\Psi(x, t)] = \sum_{k=1}^\infty \lambda_k |\psi_k|^2,$$

with initial condition $\Psi(0) = \{\psi_k(0)\}_{k \in \mathbb{N}}$.

**Theorem 1.6.** Suppose that the hypotheses of Theorem 1.1 hold. Then there exists $\tau > 0$ such that $\Psi - \Gamma \to 0$ in $L^\infty(\mathcal{H}^s)$, $s \geq \gamma/2$, as $m \to \infty$.

In other words, when the mass tends to infinity, the solutions of the semi-relativistic, and of the non-relativistic Schrödinger-Poisson systems of equations asymptotically become indistinguishable.

**Remark 1.7.** The proof of Theorem 1.6 relies on local well-posedness, and this is why the result holds for $s \geq \gamma/2$.

2 **Well-posedness**

2.1 **Local well-posedness**

In what follows, we fix $\lambda \in L^1$, $\lambda_i \geq 0$, $i \in \mathbb{N}$. We start by showing that the nonlinearity $V[\Psi] \Psi$ is locally Lipschitz.

**Lemma 2.1.** For $\Psi, \Phi \in \mathcal{H}^s$, $s \geq \gamma/2$,

$$\|V[\Psi] \Psi - V[\Phi] \Phi\|_{\mathcal{H}^s} \lesssim (\|\Psi\|_{\mathcal{H}^s}^2 + \|\Phi\|_{\mathcal{H}^s}^2) \|\Psi - \Phi\|_{\mathcal{H}^s}.$$ 

**Proof.** The proof relies on the fractional Leibniz rule and fractional integration, see Appendix A. From the Minkowski inequality,

$$\|V[\Psi] \Psi - V[\Phi] \Phi\|_{\mathcal{H}^s} \lesssim (\|V[\Psi] - V[\Phi]\|_{\mathcal{H}^s} + \|\Psi - \Phi\|_{\mathcal{H}^s}).$$  

(2.1)

We begin by estimating the first term on the right.

$$\|V[\Psi] - V[\Phi]\|_{\mathcal{H}^s} \lesssim \sum_{k,l \geq 1} \lambda_k \lambda_l \|w_\gamma * (|\psi_l|^2 - |\phi_l|^2) \psi_k\|_{\mathcal{H}^s}$$

$$\lesssim \sum_{k,l \geq 1} \lambda_k \lambda_l \left(\|w_\gamma * (|\psi_l|^2 - |\phi_l|^2)\|_{L^\infty} \|\psi_k\|_{\mathcal{H}^s}\right.$$  

$$\left.\quad + \|w_\gamma * (|\psi_l|^2 - |\phi_l|^2)\|_{L^\infty} \|\psi_k\|_{\mathcal{H}^s}\right)$$

$$\lesssim \sum_{k,l \geq 1} \lambda_k \lambda_l \left(\|\psi_l - \phi_l\|_{\mathcal{H}^{s/2}} (\|\psi_l\|_{\mathcal{H}^{s/2}} + \|\psi_l\|_{\mathcal{H}^{s/2}}) \|\psi_k\|_{\mathcal{H}^s}\right.$$  

$$\left.\quad + \|\psi_l\|_{L^{2\gamma/\gamma - 1}} \|\psi_k\|_{\mathcal{H}^{s/2}}\right)$$

$$\lesssim (\|\Psi\|_{\mathcal{H}^s}^2 + \|\Phi\|_{\mathcal{H}^s}^2) \|\Psi - \Phi\|_{\mathcal{H}^s}. \quad (2.2)$$
Given \( H \) and \( \gamma > 0 \) such that the unique solution to the Schrödinger-Poisson system of equations is locally well-posed. Using a standard contraction map argument, the generalized semi-relativistic Schrödinger-Poisson system of equations is locally well-posed.

**Proposition 2.2.** Consider the system of equations (1.4)-(1.6), with \( m \geq 0 \), \( 0 < \gamma \leq 1 \) if \( n \geq 2 \), and \( 0 < \gamma < 1 \) if \( n = 1 \). Suppose that \( (\Psi(0), \Lambda) \in \mathbb{S}^s, s \geq \gamma/2 \). Then there exists a positive time \( T \) such that the unique solution \( \Psi \in C([0, T]; \mathcal{H}^s) \). Furthermore, there exists a maximal time \( \tau^* \in (0, \infty) \) such that \( \lim_{t \to \tau^*} \|\Psi(t)\|_{\mathcal{H}^s} = \infty \).

**Proof.** Given \( \rho, T > 0 \), consider the Banach space

\[
\mathcal{B}^s_{\rho} = \{ \Psi \in L^\infty_{\mathcal{F}}(\mathcal{H}^s) : \|\Psi\|_{L^\infty_{\mathcal{F}} \mathcal{H}^s} \leq \rho \}.
\]

Let \( U^{(m)} = e^{-itH_m} \), the unitary operator generated by the semi-relativistic Hamiltonian \( H_m = \sqrt{-\Delta + m^2} - m \). We define the mapping \( \mathcal{N} \) by

\[
\mathcal{N}(\Psi)(t) = U^{(m)}(t)\Psi(0) - i \int_0^t U^{(m)}(t - t')V[\Psi(t')][\Psi(t')]dt',
\]

which is the solution given by the Duhamel formula. First we show that \( \mathcal{N} \) is a mapping from \( \mathcal{B}^s_{\rho} \) into itself.

\[
\|\mathcal{N}(\Psi)\|_{L^\infty_{\mathcal{F}} \mathcal{H}^s} \leq \|\Psi(0)\|_{\mathcal{H}^s} + T \|V[\Psi]\|_{L^\infty_{\mathcal{F}} \mathcal{H}^s},
\]

\[
\leq \|\Psi(0)\|_{\mathcal{H}^s} + T \sum_{k,l \geq 1} \lambda_k \lambda_l \|w_{\gamma} \ast |\psi|^2\|_{L^\infty_{\mathcal{F}} \mathcal{H}^s},
\]

\[
\leq \|\Psi(0)\|_{\mathcal{H}^s} + T \sum_{k,l \geq 1} \lambda_k \lambda_l \{\|w_{\gamma} \ast (|\psi|^2)\|_{L^\infty_{\mathcal{F}} \mathcal{H}^s} + \|w_{\gamma} \ast (|\psi|^2)\|_{L^\infty_{\mathcal{F}} W^s_{-\frac{1}{2}} \mathcal{H}^s}\},
\]

where we have used fractional Leibniz rule (Lemma A.1) in the last inequality. It follows from fractional integration (Lemma A.2) and Sobolev embedding...
H^{\frac{\gamma}{2}} \hookrightarrow L^{2+\gamma}_{\mathbb{R}^n} \) that

\| \mathcal{N}(\Psi) \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}} \leq \| \Psi(0) \|_{H^{\frac{\gamma}{2}}} + T \sum_{k, l \geq 1} \lambda_k \lambda_l \{ \| \psi_l \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}} \| \psi_k \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}} \}

\leq \| \Psi(0) \|_{H^{\frac{\gamma}{2}}} + T \sum_{k, l \geq 1} \lambda_k \lambda_l \left\{ \| \psi_l \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}} \| \psi_k \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}} \right\}^2

\leq \| \Psi(0) \|_{H^{\frac{\gamma}{2}}} + \| \Psi \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}} \| \Psi \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}}.

where we have used the fact that \( \lambda_k \geq 1 \) and \( \sum_{k \geq 1} \lambda_k = 1 \) before the last inequality.

Since \( s \geq \frac{\gamma}{2} \), and since by assumption, \( \Psi \in B^s_{T, \rho} \), we can choose \( T \) and \( \rho \) such that

\| \Psi(0) \|_{H^{\frac{\gamma}{2}}} \leq \frac{\rho}{2}, \quad T \rho^2 < \frac{1}{2},

it follows from the last inequality and the Duhamel formula that

\| \Psi \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}} \leq 2 \| \Psi(0) \|_{H^{\frac{\gamma}{2}}} \leq \rho.

Second, since the nonlinearity is locally Lipschitz (Lemma 2.1), \( \mathcal{N} \) is a contraction map for sufficiently small \( T \).

\| \mathcal{N}(\Psi) - \mathcal{N}(\Phi) \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}} \leq T \| V[\Psi] \Psi - V[\Phi] \Phi \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}}

\leq T \rho^2 \| \Psi - \Phi \|_{L^{2+\gamma}_{\mathbb{R}^n} H^{\frac{\gamma}{2}}}.

Local well-posedness follows from a standard contraction mapping argument, see for example, [6].

It follows from local well-posedness that for every \( k \in \mathbb{N} \), \( \| \psi_k \|_{L^2} \) is conserved.

**Lemma 2.3.** Suppose that the hypotheses of Proposition 2.2 hold. Then

\| \psi_k(t) \|_{L^2} = \| \psi_k(0) \|_{L^2}, \quad t \in [0, \tau^*).

**Proof.** Multiplying (1.4) by \( \bar{\psi}_k \) and integrating over space yields

\[ \frac{i}{2} \partial_t \| \psi_k \|^2 = \langle \psi_1, H_m \psi_k \rangle + \langle \psi_1, V[\Psi] \psi_k \rangle. \]

Taking the imaginary part of both sides of the equation yields \( \partial_t \| \psi_k \|^2 = 0 \).

The energy functional associated with the semi-relativistic Schrödinger-Poisson system is

\[ \mathcal{E}(\Psi) = \frac{1}{2} \langle \Psi, H_m \Psi \rangle_{L^2} + \frac{1}{4} \langle \Psi, V[\Psi] \Psi \rangle_{L^2}. \]

Formally, conservation of energy follows from multiplying (1.4) by \( \lambda_k \partial_k \bar{\psi}_k \), integrating over space, and summing over \( k \geq 1 \). To make the argument precise, we need a regularization procedure.
Lemma 2.4. Suppose that the hypotheses of Proposition 2.2 hold. Then $E(\Psi(t)) = E(\Psi(0))$, $t \in [0, \tau^*)$, is satisfied for solutions $\Psi \in C([0, \tau^*), H^s)$ with $s \geq \frac{1}{2}$.

Proof. Let $J_\varepsilon = (\varepsilon H_m + 1)^{-1}$, $\varepsilon > 0$, act on the sequence of embedding spaces

$$
\cdots \ H^{\frac{s}{2}} \hookrightarrow H^{\frac{1}{2}} \hookrightarrow H^{-\frac{1}{2}} \hookrightarrow H^{-\frac{3}{2}} \cdots
$$

It follows from fractional calculus that

(i) $J_\varepsilon$ is a bounded operator from $H^s$ to $H^{s+1}$,

(ii) $\|J_\varepsilon \Psi\|_{H^s} \leq \|\Psi\|_{H^s}$, and

(iii) $J_\varepsilon \Psi \rightarrow \Psi$ strongly in $H^s$ as $\varepsilon \rightarrow 0$.

Now,

$$
E(J_\varepsilon \Psi(t_2)) - E(J_\varepsilon \Psi(t_1)) = \int_{t_1}^{t_2} \partial_t E(J_\varepsilon \Psi(t)) \, dt
$$

$$
= \operatorname{Re}\left\{ \int_{t_1}^{t_2} -i \langle H_m J_\varepsilon \Psi(t), H_m J_\varepsilon \Psi(t) \rangle_{L^2} + \langle J_\varepsilon V[J_\varepsilon \Psi(t)], J_\varepsilon \Psi(t) \rangle_{L^2} + \langle J_\varepsilon V[J_\varepsilon \Psi(t)], J_\varepsilon V[J_\varepsilon \Psi(t)] \Psi(t) \rangle_{L^2} \right\}.
$$

The first term is trivially zero, since $H_m J_\varepsilon = J_\varepsilon H_m$. Let

$$
g_\varepsilon(t) = \operatorname{Re}\left\{ \langle H_m J_\varepsilon \Psi(t), J_\varepsilon V[J_\varepsilon \Psi(t)] \Psi(t) \rangle_{L^2} + \langle J_\varepsilon V[J_\varepsilon \Psi(t)], J_\varepsilon \Psi(t) \rangle_{L^2} + \langle J_\varepsilon V[J_\varepsilon \Psi(t)], J_\varepsilon V[J_\varepsilon \Psi(t)] \Psi(t) \rangle_{L^2} \right\}.
$$

Then

$$
E(J_\varepsilon \Psi(t_2)) - E(J_\varepsilon \Psi(t_1)) = \int_{t_1}^{t_2} g_\varepsilon(t) \, dt.
$$

It follows from the above properties (i)-(iii) of $J_\varepsilon$ that $\lim_{\varepsilon \rightarrow 0} g_\varepsilon(t) = 0$. Furthermore,

$$
g_\varepsilon(t) \leq \|V[\Psi(t)]\Psi(t)\|_{L^2} \|H_m \Psi(t)\|_{L^2} + \|V[\Psi(t)]\Psi(t)\|^2_{L^2}. \quad (2.4)
$$

Using Lemma A.3, we have

$$
\|V[\Psi]\|_{L^2} \lesssim \sum_{k,l \geq 1} \lambda_k \lambda_l \|\psi_l\|^2_{H^\frac{1}{2}} \|\psi_k\|_{L^2}.
$$
The Gagliardo-Nirenberg inequality,
\[ \| \psi_l \|_{H^{2\gamma}} \lesssim \| \psi_l \|_{H^2} \| \psi_l \|_{L^2}^{1-\gamma}, \]
together with conservation of charge (Lemma 2.3), yields
\[ \| V[\Psi] \|_{L^2} \lesssim \sum_{l \geq 1} \lambda_l \| \psi_l \|_{H^{2\gamma}}^2 \]
\[ \lesssim \left( \sum_{l \geq 1} \lambda_l \| \psi_l \|_{H^{2\gamma}}^2 \right)^\gamma \]
\[ \lesssim \| \Psi \|_{H^2}^\gamma, \]
where we have used in the second inequality the fact that \[ \sum_{l \geq 1} \lambda_l = 1, \quad \lambda_l \geq 0, \]
and \( f(x) = x^\gamma, \quad 0 < \gamma < 1, \) is concave (equality when \( \gamma = 1 \) is trivially satisfied). Substituting back in (2.4) yields
\[ g_\varepsilon(t) \lesssim \| \Psi \|_{H^{2\gamma}}^{\gamma+1} + \| \Psi \|_{H^{2\gamma}} \]
which is finite for \( t < \tau^* \). By the Dominated Convergence Theorem,
\[ \mathcal{E}(\Psi(t_2)) - \mathcal{E}(\Psi(t_1)) = \int_{t_1}^{t_2} \lim_{\varepsilon \to 0} g_\varepsilon(t) \, dt = 0, \]
as claimed.

Global well-posedness in \( H^s \), for \( s \geq \frac{1}{2} \), follows from conservation of charge and energy.

Proposition 2.5. Suppose that the hypotheses of Proposition 2.2 hold. Then, if \( g > 0 \) or \( g < 0 \) with \( \| \Psi(0) \|_{L^2} \) small enough, and for \( s \geq \frac{1}{2} \),
\[ \| \Psi(t) \|_{H^s} \leq C \| \Psi(0) \|_{H^s} e^{C(t E(\Psi(0)))^{\alpha} + \| \Psi(0) \|_{L^2}^\delta t}, \]
where \( C, \alpha \) and \( \delta \) are positive constants that are independent of \( m \geq 0 \).

Proof. We start by bounding \( \| \Psi(t) \|_{H^s} \) from above, uniformly in time.
\[ \langle \Psi, V[\Psi] \Psi \rangle_{L^2} = \sum_{l \geq 1} \lambda_l \langle \psi_l, V[\Psi] \psi_l \rangle \]
\[ \leq \| V[\Psi] \|_{L^\infty} \| \Psi \|_{L^2}^2 \]
\[ \lesssim \left( \sum_{k \geq 1} \lambda_k \| \psi_k \|_{H^{2\gamma}}^2 \right) \| \Psi \|_{L^2}^2 \]
\[ \lesssim \left( \sum_{k \geq 1} \lambda_k \| \psi_k \|_{H^{2\gamma}}^{2\gamma} \right) \| \Psi \|_{L^2}^2 \]
\[ \leq \left( \sum_{k \geq 1} \lambda_k \| \psi_k \|_{H^{2\gamma}}^{2\gamma} \right)^\gamma \| \Psi \|_{L^2}^2 \]
\[ \lesssim \| \Psi \|_{H^2}^\gamma \| \Psi \|_{L^2}^2. \]
Here, we used Hölder’s inequality in the second line, Lemma A.3 in the third line, the Gagliardo-Nirenberg inequality and conservation of charge in the fourth line, and $\sum_{k \geq 1} \lambda_k = 1$, $\lambda_k \geq 0$, the fact that $x^\gamma$, $0 < \gamma < 1$, is concave in the fifth line (equality when $\gamma = 1$ is trivially satisfied). Together with conservation of energy (Lemma 2.4), this implies that for $g > 0$ or $g < 0$ with $\|\Psi(0)\|_{L^2}$ small enough,

$$\|\Psi\|_{H^s} \leq \alpha (\mathcal{E}(\Psi(t)) + \|\Psi(0)\|_{L^2}^2), \tag{2.5}$$

where $\alpha$ and $\delta$ are constants independent of the mass $m \geq 0$. Now, it follows from the Duhamel formula that

$$\|\Psi(t)\|_{H^s} \leq \|\Psi(0)\|_{H^s} + \int_0^t \|\Psi(t')\|_{H^s}^2 \|\Psi(t')\|_{H^s} dt' \leq \|\Psi(0)\|_{H^s} + \alpha (\mathcal{E}(\Psi(t)) + \|\Psi(0)\|_{L^2}^4) \int_0^t \|\Psi(t')\|_{H^s} dt',$n

where we used Hölder’s and Minkowski inequalities in the first line, and (2.5) in the second line. By Gronwall’s lemma,

$$\|\Psi(t)\|_{H^s} \leq \|\Psi(0)\|_{H^s} e^{\alpha (\mathcal{E}(\Psi(0)) + \|\Psi(0)\|_{L^2}^4)t}$$

follows.

**Proof of Theorem 1.1.** It follows from Propositions 2.2 and 2.5 that $\tau^* = \infty$, i.e., the generalized semi-relativistic Schrödinger-Poisson system of equations is globally well-posed.

We now prove the claim of Remark 1.5 about the asymptotic behaviour of the system as the mass tends to zero.

**PROPOSITION 2.6.** Consider the system of equations (1.4)-(1.6) with initial condition $(\Delta \Psi(0))$. Let $\Psi^{(0)}$ denote the solution of the initial value problem with mass $m = 0$, and fix $T > 0$. Under the hypotheses of Proposition 2.5, $\Psi \to \Psi^{(0)}$ strongly in $L^\infty_T (H^s)$, $s \geq 1/2$, as $m \to 0$.

**Proof.** Proposition 2.5 implies that, given $T > 0$, there exists finite $\rho > 0$ such that

$$\sup_{m \in [0,1]} \|\Psi\|_{L^\infty_T H^s} < \rho. \tag{2.6}$$

We now compare the norm of the difference of $\Psi(t)$ and $\Psi^{(0)}(t)$, $t \in [0, T]$. It
follows from the Duhamel formula that
\[
\|\Psi(t) - \Psi^{(0)}(t)\|_{H^s} \lesssim \| \left( U^{(m)}(t) - U^{(0)}(t) \right) \Psi(0) \|_{H^s} + \int_0^t \| V[\Psi(t')]\Psi(t') - V[\Psi^{(0)}(t')]\Psi^{(0)}(t') \|_{H^s} \ dt' + \| \left( U^{(m)}(t') - U^{(0)}(t') \right) V[\Psi^{(0)}(t')]\Psi^{(0)}(t') \|_{H^s} \ dt' \\
\lesssim mT\|\Psi(0)\|_{H^s} + \int_0^t \| V[\Psi(t')]\Psi(t') - V[\Psi^{(0)}(t')]\Psi^{(0)}(t') \|_{H^s} \ dt' + \frac{mT^2}{2}\| V[\Psi^{(0)}]\Psi^{(0)} \|_{L^\infty H^s},
\]
where we used Minkowski inequality in the first inequality and Hölder’s inequality in the second. We also used \( \sqrt{-\Delta + m^2 - m} \leq m \).

It follows from the fact that the nonlinearity is locally Lipschitz (Lemma 2.1) and (2.6) that
\[
\| V[\Psi(t')]\Psi(t') - V[\Psi^{(0)}(t')]\Psi^{(0)}(t') \|_{H^s} \lesssim \rho^2\|\Psi(t') - \Psi^{(0)}(t')\|_{H^s},
\]
\[
\| V[\Psi^{(0)}]\Psi^{(0)} \|_{L^\infty H^s} \lesssim \rho^3.
\]
Hence
\[
\|\Psi(t) - \Psi^{(0)}(t)\|_{H^s} \lesssim m\rho T + m\rho^3 T + \rho^2 \int_0^t \|\Psi(t') - \Psi^{(0)}(t')\|_{H^s} \ dt'.
\]
By Gronwall’s lemma, \( \Psi \to \Psi^{(0)} \) strongly in \( L^\infty(\mathcal{H}^s) \) as \( m \to 0 \).*

3 Asymptotic behaviour of solutions as mass tends to infinity

In this section, we discuss the asymptotics of the solution as the mass \( m \) tends to infinity.

Proof of Theorem 1.6. Recall that from the proof of local well-posedness in Section 2.1, there exists \( T > 0 \) independent of \( m \) such that \( \|\Psi\|_{L^\infty_T H^s} \leq C\|\Psi(0)\|_{H^s}, \ s \geq \gamma/2, \) where \( C \) is independent of \( m \). Similarly, one can show that there exists \( T' > 0 \) independent of \( m \) such that \( \|\Gamma\|_{L^\infty_T H^r} \leq C\|\Psi(0)\|_{H^r}, \) where \( C \) is independent of \( m \). Let \( \tau = \min( T, T' ) \). Let \( \tilde{\Gamma} = \{ \tilde{\gamma}_k \}_{k \in \mathbb{N}} \) satisfy the system of equations
\[
\begin{cases}
 i\partial_t \tilde{\Gamma} = V[\tilde{\Gamma}]\tilde{\Gamma}, \\
 V[\tilde{\Gamma}] = w_\gamma \ast n[\tilde{\Gamma}], \\
 n[\tilde{\Gamma}] = \sum_{k=1}^\infty \lambda_k |\tilde{\gamma}_k|^2,
\end{cases}
\]
with initial condition \( \tilde{\Gamma}(0) = \Psi(0) \). Alternatively, \( \tilde{\Gamma} \) satisfies the integral equation
\[
\tilde{\Gamma}(t) = \Psi(0) - i \int_0^t V[\tilde{\Gamma}(t')]\tilde{\Gamma}(t') \ dt'.
\]
Uniqueness of the solution follows from the fact that the nonlinearity is locally Lipschitz (Lemma 2.1). We are going to compare Ψ to $\tilde{\Gamma}$, and then $\tilde{\Gamma}$ to $\Gamma$.

\[
\|\Psi(t) - \tilde{\Gamma}(t)\|_{\mathcal{H}^s} \leq \| \left( U^{(m)}(t) - 1 \right) \Psi(0) \|_{\mathcal{H}^s} + \int_0^t \| \left( U^{(m)}(t-t') - 1 \right) V[\tilde{\Gamma}(t')]\tilde{\Gamma}(t') \|_{\mathcal{H}^s} dt' + \int_0^t \| V[\Psi(t')]\Psi(t') - V[\tilde{\Gamma}(t')]\tilde{\Gamma}(t') \|_{\mathcal{H}^s} dt'.
\]

(3.1)

To estimate the first term on the right-hand-side, we apply the Fourier transform and use Parseval’s Theorem,

\[
\| \left( U^{(m)}(t) - 1 \right) \Psi(0) \|_{\mathcal{H}^s}^2 = \sum_{l \geq 1} \lambda_l \int_{\mathbb{R}^d} |e^{-it(\sqrt{m^2 + |k|^2} - m)} - 1|^2 (1 + |k|^2) |\hat{\psi}(0, k)|^2 dk \leq \sum_{l \geq 1} \lambda_l \int_{|k| \leq m^+} |e^{-it(\sqrt{m^2 + |k|^2} - m)} - 1|^2 (1 + |k|^2) |\hat{\psi}(0, k)|^2 dk + \int_{|k| > m^+} |e^{-it(\sqrt{m^2 + |k|^2} - m)} - 1|^2 (1 + |k|^2) |\hat{\psi}(0, k)|^2 dk \leq \sum_{l \geq 1} \lambda_l \int_{|k| \leq m^+} \frac{l^2 |k|^4}{(\sqrt{m^2 + |k|^2} + m)^2} (1 + |k|^2) |\hat{\psi}(0, k)|^2 dk + 4 \int_{|k| > m^+} (1 + |k|^2) |\hat{\psi}(0, k)|^2 dk \leq \frac{\nu^2}{4m} \| \Psi(0) \|_{\mathcal{H}^s}^2 + 4 \sum_{l \geq 1} \int_{|k| > m^+} (1 + |k|^2) |\hat{\psi}(0, k)|^2 dk \to 0 \text{ as } m \to \infty.
\]

Since $V[\tilde{\Gamma}]\tilde{\Gamma} \in \mathcal{H}^s$, it follows from the Dominated Convergence Theorem that

\[
\lim_{m \to \infty} \int_0^t \| \left( U^{(m)}(t-t') - 1 \right) V[\tilde{\Gamma}(t')]\tilde{\Gamma}(t') \|_{\mathcal{H}^s} dt' = 0.
\]

To estimate the third term, let $\rho > 0$ be a constant such that

\[
\sup_{m \geq 1} (\| \Psi \|_{L^\infty_{\mathcal{H}^s}} + \| \Gamma \|_{L^\infty_{\mathcal{H}^s}}) + \| \tilde{\Gamma} \|_{L^\infty_{\mathcal{H}^s}} \leq \rho.
\]

It follows from the fact that the nonlinearity is locally Lipschitz that

\[
\| V[\Psi(t')]\Psi(t') - V[\tilde{\Gamma}(t')]\tilde{\Gamma}(t') \|_{\mathcal{H}^s} \leq C \rho^2 \| \Psi(t') - \tilde{\Gamma}(t') \|_{\mathcal{H}^s},
\]

where $C$ is a positive constant independent of $m$. 

Documenta Mathematica 18 (2013) 343–357
Therefore,

$$\|\Psi(t) - \tilde{\Gamma}(t)\|_{\mathcal{H}} \leq f_m + C\rho^2 \int_0^t \|\Psi(t') - \tilde{\Gamma}(t')\|_{\mathcal{H}} dt',$$

where $f_m$ bounds the first two terms on the r.h.s. of (3.1). As shown above, $\lim_{m \to \infty} f_m = 0$ and $C$ is independent of $m$, so that application of Gronwall’s lemma yields

$$\lim_{m \to \infty} \|\Psi - \tilde{\Gamma}\|_{L^\infty \mathcal{H}} = 0.$$

Similarly, one can show that

$$\|\Gamma(t) - \tilde{\Gamma}(t)\|_{\mathcal{H}} \leq g_m + C\rho^2 \int_0^t \|\Psi(t') - \tilde{\Gamma}(t')\|_{\mathcal{H}} dt',$$

where $\lim_{m \to \infty} g_m = 0$ and $C$ is independent of $m$, and it follows that

$$\lim_{m \to \infty} \|\Gamma - \tilde{\Gamma}\|_{L^\infty \mathcal{H}} = 0.$$

Since

$$\|\Psi - \Gamma\|_{L^\infty \mathcal{H}} \leq \|\Psi - \tilde{\Gamma}\|_{L^\infty \mathcal{H}} + \|\Gamma - \tilde{\Gamma}\|_{L^\infty \mathcal{H}},$$

it follows that

$$\lim_{m \to \infty} \|\Psi - \Gamma\|_{L^\infty \mathcal{H}} = 0,$$

as desired. \(\square\)

**Acknowledgements**

WAS acknowledges the financial support of a Discovery grant from the Natural Sciences and Engineering Research Council of Canada. T.C. was supported by NSF grants DMS-1009448 and DMS-1151414 (CAREER).

**A Appendix**

The following result about the fractional Leibniz rule can be found in [9].

**Lemma A.1.**

$$\|\mathcal{D}^s(uv)\|_{L^p} \lesssim \|\mathcal{D}^s u\|_{L^{q_i}} \|v\|_{L^{r_1}} + \|u\|_{L^{q_i}} \|\mathcal{D}^s v\|_{L^{r_2}},$$

where $\frac{1}{p} = \frac{1}{q_i} + \frac{1}{r_i}$, $i = 1, 2$.

The second result is about inequality involving fractional integral operators, which can be found, for example, in [12].
Lemma A.2. Let $I_\alpha$, for $0 < \alpha < n$, be the fractional integral operator

$$I_\alpha(u) = \int_{\mathbb{R}^n} |x - y|^{\alpha - n} u(y) \, dy.$$ 

Then

$$\|I_\alpha(u)\|_{L^p} \lesssim \|u\|_{L^q}, \quad \frac{1}{p} = \frac{1}{q} - \frac{\alpha}{n}.$$ 

We also recall the following useful Hardy-type inequality.

Lemma A.3. Let $0 < \gamma < n$. Then,

$$\sup_{x \in \mathbb{R}^n} \left| \int_{\mathbb{R}^n} \frac{1}{|x - y|^{\gamma}} |u(y)|^2 \, dy \right| \lesssim \|u\|^2_{H^\gamma}.$$ 

References


W. Abou Salem
Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon S7N 5E6
Canada
walid.abousalem@usask.ca

Thomas Chen
Department of Mathematics
University of Texas at Austin
Saskatoon S7N 5E6
Canada
walid.abousalem@usask.ca

documenTa mathematiCa 18 (2013) 343–357

V. Vougalter
University of Cape Town
Department of Mathematics and Applied Mathematics
Private Bag
Rondebosch 7701
South Africa
Vitali.Vougalter@uct.ac.za
REALIZABILITY AND ADMISSIBILITY
UNDER EXTENSION OF $p$-ADIC AND NUMBER FIELDS

DANNY NEFTIN AND UZI VISHNE

Received: July 12, 2012
Communicated by Max Karoubi

Abstract. A finite group $G$ is $K$-admissible if there is a $G$-crossed product $K$-division algebra. In this manuscript we study the behavior of admissibility under extensions of number fields $M/K$. We show that in many cases, including Sylow metacyclic and nilpotent groups whose order is prime to the number of roots of unity in $M$, a $K$-admissible group $G$ is $M$-admissible if and only if $G$ satisfies the easily verifiable Liedahl condition over $M$.

2010 Mathematics Subject Classification: 16K20, 12F12
Keywords and Phrases: admissible group; adequate field; tame admissibility; Liedahl’s condition.

1. Introduction

Let $K$ be a field. A field $L \supseteq K$ is called $K$-adequate if it is contained as a maximal subfield in a finite dimensional central $K$-division algebra. A group $G$ is $K$-admissible if there is a $G$-extension $L/K$, i.e. $L/K$ is a Galois extension with $\text{Gal}(L/K) \cong G$, so that $L$ is $K$-adequate. Equivalently, $G$ is $K$-admissible if there is a $G$-crossed product $K$-division algebra. Ever since adequacy and admissibility were introduced in [19], they were studied extensively over various types of fields, especially over number fields.

As oppose to realizability of groups as Galois groups, there are known restrictions on the number fields $K$ over which a given group is $K$-admissible. Liedahl’s condition (which was shown by Schacher [19] over $\mathbb{Q}$, and generalized by Liedahl [9, Theorem 28]) describes such a restriction. We say that $G$ satisfies Liedahl’s condition over $K$, if for every prime $p$ dividing $|G|$, one of the following holds:

(i) $p$ decomposes in $K$ (has at least two prime divisors),
(ii) $p$ does not decompose in $K$, and a $p$-Sylow subgroup $G(p)$ of $G$ is metacyclic and admits a Liedahl presentation over $K$ (for details see Definition 2.4 which is based on [9]).

1The second named author is partially supported by a BSF US-Israel grant 2010/149.
In [19, Theorem 9.1], Schacher showed that any finite group $G$ is admissible over some number field $K$. However, for many groups $G$ it is an open problem to determine the number fields over which they are admissible. In fact, searching for an explicit description for all groups seems hopeless.

In this paper we fix a field $K$ over which $G$ is admissible and ask over which finite extensions of $K$, $G$ is still admissible. We assume our group $G$ is realizable over $M$ and furthermore can be realized over $M$ with prescribed local conditions, i.e. it admits the Grunwald-Neukirch (GN) property. This question then leads (see Section 5.1) via Schacher’s criterion (Theorem 2.1) to the following local realization problem:

**Problem 1.1.** Let $m/k$ be an extension of $p$-adic fields and $G$ a group that is realizable over $k$. Is there a subgroup $H$ of $G$ which is realizable over $m$ and contains a $p$-Sylow subgroup of $G$?

At first we consider the case of $p$-groups, where the problem is whether a $p$-group that is realizable over $k$, is realizable over an extension $m$ of $k$. For $p$ odd, we notice that the maximal pro-$p$ quotient $G_k(p)$ of the absolute Galois group $G_k$ is covered by $G_m(p)$, providing a positive answer:

**Proposition 1.2.** Let $m/k$ be a finite extension of $p$-adic fields where $p$ is an odd prime. Then any $p$-group that is realizable over $k$ is also realizable over $m$.

The simplest behavior one can hope for in terms of admissibility, is that a $K$-admissible group $G$ would be $M$-admissible if and only if it satisfies Liedahl’s condition over $M$. This is indeed the case for various classes of groups:

1. When all Sylow subgroups of $G$ are cyclic [19, Theorem 2.8];
2. When $G$ is abelian and does not fall into a special case over $M$ [2];
3. For metacyclic groups [9],[10];
4. For $G = \text{SL}_2(5)$ [5];
5. For $G = A_6$ or $G = A_7$ [20];
6. For $G = \text{PGL}_2(7)$ [1], and
7. For the Symmetric groups $G = S_n$, $1 \leq n \leq 17$, $n \neq 12, 13$ (by [4], [9] and [19]).

Using Proposition 1.2, we are able to add all the odd-order $p$-groups having the GN-property over $M$ to this list:

**Proposition 1.3.** Let $M/K$ be an extension of number fields and $p$ an odd prime. Let $G$ be a $p$-group that is $K$-admissible and has the GN-property over $M$. Then $G$ is $M$-admissible if and only if $G$ satisfies Liedahl’s condition over $M$.

Propositions 1.2 and 1.3 are proved in Subsection 3.1. We note (in Section 3) that Proposition 1.3 extends to nilpotent groups of odd order and (by Remark 5.1) to Sylow metacyclic groups (having metacyclic Sylow subgroups). However, the following example shows that Problem 1.1 can have a negative answer for some 2-groups:

**Documenta Mathematica 18 (2013) 359–382**
Example 1.4. There is a group $G$, of order $2^6$, which is realizable over $\mathbb{Q}_2$ but not over $\mathbb{Q}_2(\sqrt{-1})$.

For a proof, see Corollary 3.6. In Proposition 3.8 we interpret this example globally:

Example 1.5. There is a rational prime $q$ for which the group $G$ of Example 1.4 is $\mathbb{Q}(\sqrt{q})$-admissible, satisfies Liedahl’s condition over $\mathbb{Q}(\sqrt{-1}, \sqrt{q})$ but is not $\mathbb{Q}(\sqrt{-1}, \sqrt{q})$-admissible.

Liedahl showed that a similar phenomena happens for the groups $S_n$, $n = 12, 13$ and the local extension $\mathbb{Q}_2(\sqrt{-3})/\mathbb{Q}_2$ (see [4]). We shall restrict our discussion to groups which are either of odd order, or with metacyclic 2-Sylow subgroups. For some $p$-adic extensions $m/k$, for $p$ odd, including extensions in which the inertia degree $f(m/k)$ is a $p$-power and $[m:k] > 5$ we show that $G_m$ covers the maximal quotient of $G_k$ with a normal $p$-Sylow subgroup.

We use this method to answer Problem 1.1 positively for odd primes, under the following assumptions. The list of ‘sensitive’ extensions of $p$-adic fields, (16 with $p = 3$ and one for $p = 5$) is described in Subsection 4.2.

Theorem 1.6. Let $p$ be an odd prime. Let $m/k$ be a non-sensitive extension of $p$-adic fields and $G$ a group with a normal $p$-Sylow subgroup $P$. Assume $G$ is realizable over $k$. Then there is a subgroup $H \leq G$ that contains $P$ and is realizable over $m$.

The question as to whether the non-sensitivity assumption can be removed remains open. However, the assumption that the Sylow subgroup is normal is essential:

Example 1.7. Let $G = C_7 \rtimes D$ where
\[ D = \langle a, b \mid a^7 = b^9 = 1, a^{-1}ba = b^7 \rangle; \]
thus the 7-Sylow subgroups of $G$ are neither normal nor metacyclic.

In Example 4.11 we show there exists an extension $m/k$ of 7-adic fields such that $G$ is realizable over $k$, although no subgroup of $G$ that contains a 7-Sylow subgroup is realizable over $m$.

We say that an extension of number fields $M/K$ is sensitive if it has a sensitive completion. The main theorem follows from Theorem 1.6 by combining the local data (see Subsection 5.1):

Theorem 1.8. Let $M/K$ be a non-sensitive extension of number fields. Let $G$ be a group for which every Sylow subgroup is either normal or metacyclic, and the 2-Sylow subgroups are metacyclic.

Assume $G$ is $K$-admissible and has the GN-property over $M$. Then $G$ is $M$-admissible if and only if $G$ satisfies Liedahl’s condition over $M$.

Let $\mu_n$ denote the set of $n$-th roots of unity. As a consequence of Theorem 1.8 and [13, Corollary 2] we have:
Corollary 1.9. Let $G$ be an odd order group for which every Sylow subgroup is either normal or metacyclic. Let $M/K$ be a non-sensitive extension of number fields so that $G$ is $K$-admissible and $\mu_{|G|} \cap M = \{1\}$. Then $G$ is $M$-admissible if and only if $G$ satisfies Liedahl’s condition over $M$.

In particular, if every prime dividing $|G|$ decomposes in $M$ or if $M \cap \mathbb{Q}(\mu_{|G|}) = K \cap \mathbb{Q}(\mu_{|G|})$, then $G$ is $M$-admissible.

We also show that the assumption that every Sylow subgroup is either normal or metacyclic is essential in Theorem 1.8:

Example 1.10. Let $G$ be the group defined in Example 1.7. In Example 5.7, we show furthermore that there is an extension of number fields $M/K$ so that $G$ is $K$-admissible, satisfies Liedahl’s condition over $M$, has the GN-property over $M$, but is not $M$-admissible.

As an example we use Theorem 1.8 to understand the behavior of admissibility for a specific group (see Example 5.4):

Example 1.11. Let $K = \mathbb{Q}(\sqrt{14})$ and $G = C_{13} \wr M_{3^3}$, where $M_{3^3}$ is the modular group
$$
\langle x, y \mid x^{-1}yx = y^4, x^3 = y^9 = 1 \rangle,
$$
and $\wr$ is the standard wreath product. In Example 5.4, we show $G$ is $K$-admissible and deduce from Theorem 1.8 that for a number field $M \supseteq K$, $G$ is $M$-admissible if and only if $G$ satisfies Liedahl’s condition. We therefore deduce the admissibility behavior in Figure 1 by checking Liedahl’s condition.

Similar examples (also given in Example 5.4) show that the rank (the minimal number of generators) of the $p$-Sylow subgroups of $K$-admissible groups is not bounded (as opposed to the case of admissible $p$-groups discussed in [19, Section 10]).
The basic facts about admissibility of groups over number fields are reviewed in Section 2. We also discuss the behavior of wild and tame admissibility under extension of number fields and the connection between these types of admissibility to parts (i) and (ii) (respectively) in Liedahl’s condition.

The authors are grateful to Jack Sonn for his valuable advice and useful comments.

2. Preliminaries

2.1. Admissibility and Preadmissibility. For a prime \( v \) of a field \( K \), we denote by \( K_v \) the completion of \( K \) with respect to \( v \). If \( L/K \) is a finite Galois extension, \( L_v \) denotes the completion of \( L \) with respect to some prime divisor of \( v \) in \( L \).

The basic criterion for admissibility over global fields is due to Schacher:

**Theorem 2.1 (Schacher, [19]).** Let \( L/K \) be a finite Galois extension of global fields. Then \( L \) is \( K \)-adequate if for every rational prime \( p \) dividing \(|G|\), where \( G = \text{Gal}(L/K) \), there is a pair of primes \( v_1, v_2 \) of \( K \) such that each of \( \text{Gal}(L_v/K_{v_i}) \) contains a \( p \)-Sylow subgroup of \( G \).

Extracting the necessary local conditions for \( K \)-admissibility from Theorem 2.1, we arrive at the following definition. For a group \( G \), \( G(p) \) denotes a \( p \)-Sylow subgroup.

**Definition 2.2.** Let \( K \) be a number field. The group \( G \) is \( K \)-preadmissible if \( G \) is realizable over \( K \), and there exists a finite set \( S = \{v_i(p) : p \mid |G|, i = 1, 2\} \) of primes of \( K \), and, for each \( v \in S \), a subgroup \( G^v \subseteq G \), such that

1. \( v_1(p) \neq v_2(p) \) for every \( p \) dividing \(|G|\),
2. \( G^{v_i(p)} \supseteq G(p) \) for every \( p \) and \( i = 1, 2 \), and
3. \( G^v \) is realizable over \( K_v \) for every \( v \in S \).

(Notice that a \( p \)-group \( G \) is \( K \)-preadmissible if and only if there is a pair of primes \( v_1 \) and \( v_2 \) of \( K \), such that \( G \) is realizable over \( K_{v_1} \) and over \( K_{v_2} \)).

Clearly, every \( K \)-admissible group is also \( K \)-preadmissible. However the opposite does not always hold (see [11, Example 2.14]).

For an extension of fields \( L/K \), \( \text{Br}(L/K) \) denotes the kernel of the restriction map \( \text{res}: \text{Br}(K) \to \text{Br}(L) \). For number fields we have the following isomorphism of groups, where \( \Pi_K \) is the set of places of \( K \):

\[
\text{Br}(L/K) \cong \bigoplus_{\pi \in \Pi_K} \frac{1}{\gcd_{\pi'|\pi} [L_{\pi'}:K_{\pi}]} \mathbb{Z}/\mathbb{Z},
\]

where \( (\cdot)_0 \) denotes that the sum of invariants is zero.

Over a number field \( K \), the exponent of a division algebra is equal to its degree, and so \( L \) is \( K \)-adequate if and only if there is an element of order \([L:K]\) in \( \text{Br}(L/K) \) [19, Proposition 2.1].
2.2. Tame and wild admissibility. We denote by $k_{un}$ the maximal unramified extension of a local field $k$, and by $k_{tr}$ the maximal tamely ramified extension.

The tamely ramified subgroup $\text{Br}(L/K)_{tr}$ of $\text{Br}(L/K)$ is the subgroup of algebras which are split by the tamely ramified part of every completion of $L$; namely the subgroup corresponding under the above isomorphism to

$$
\left( \bigoplus_{\pi \in \Pi_K} \frac{1}{\gcd(x|\pi)} [L_{\pi'} \cap (K_{\pi})_{tr} : K_{\pi}] \mathbb{Z}/\mathbb{Z} \right)_0
$$

Following the above local description of adequacy we define:

**Definition 2.3.** We say that a finite extension $L$ of $K$ is *tamely $K$-adequate* if there is an element of order $[L : K]$ in $\text{Br}(L/K)_{tr}$.

Likewise, a finite group $G$ is *tamely $K$-admissible* if there is a tamely $K$-adequate Galois $G$-extension $L/K$.

The structure of tamely admissible groups is related to Liedahl presentations:

For $t$ prime to $n$, let $\sigma_{t,n}$ be the automorphism of $\mathbb{Q}(\mu_n)/\mathbb{Q}$ defined by $\sigma_{t,n}(\zeta) = \zeta^t$ for $\zeta \in \mu_n$.

**Definition 2.4 ([9]).** We say that a metacyclic $p$-group has a *Liedahl presentation* over $K$, if it has a presentation

$$
\mathcal{M}(m, n, i, t) := \langle x, y \mid x^m = y^i, y^n = 1, x^{-1}yx = y^i \rangle
$$

such that $\sigma_{t,n}$ fixes $K \cap \mathbb{Q}(\mu_n)$.

**Example 2.5.** The dihedral group $D_4$ has a Liedahl presentation over $\mathbb{Q}$, but not over $\mathbb{Q}(\sqrt{-1})$. Thus $D_4$ satisfies the Liedahl condition over $\mathbb{Q}$, but not over $\mathbb{Q}(\sqrt{-1})$.

The existence of Liedahl’s presentation for a $p$-group $G$ over $K$ implies $G$ is $K$-tame-preadmissible (namely, Definition 2.2 holds with realizability within $(K_v)_{tr}$ in 2.2.(3)).

**Remark 2.6 (Liedahl, follows directly from [9, Proofs of theorems 28 and 29]).** Let $G$ be a finite group. If $G$ is realizable over infinitely many completions of $K$ (at infinitely many primes), then $G$ has a presentation as above. If $G$ is a $p$-group then the converse also holds. In addition a $p$-group is realizable over infinitely many completions of $K$ if and only if it is realizable over a completion $K_v$ at one prime $v$ that does not divide $p$.

This allows us to simplify the definition of preadmissibility by noticing that the primes $v_i(p)$, $i = 1, 2, p \mid |G|$, in Definition 2.2 can be chosen to be distinct:

**Lemma 2.7.** Let $K$ be a number field. A group $G$ is $K$-preadmissible if and only if it is realizable over $K$, and there are distinct primes $v_i(p)$, $p$ runs over the primes dividing $|G|$ and $i = 1, 2$, such that for every $p$ and $i = 1, 2$, there is a subgroup $H \leq G$ that contains a $p$-Sylow subgroup of $G$ and is realizable over $K_{v_i(p)}$. 
Proof. The if part holds by definition. To prove the only if part let
\[ T = \{ v_i(p) \mid p \mid |G|, i = 1, 2 \} \]
be a set of primes of \( K \) and for every prime \( v \in T \) a corresponding subgroup \( G^v \) so that
\begin{enumerate}
  \item \( v_1(p) \neq v_2(p) \),
  \item \( G^v \) is realizable over \( K_v \),
  \item \( G^{v_i(p)} \) contains a \( p \)-Sylow subgroup of \( G \),
\end{enumerate}
for every \( p \) dividing \( |G| \) and \( i = 1, 2 \). We shall define primes \( w_i(p) \), \( p \mid |G| \) and \( i = 1, 2 \), such that all primes are distinct and for every \( w_i(p) \) there is a subgroup of \( G \) that contains a \( p \)-Sylow subgroup of \( G \) and is realizable over \( K_{w_i(p)} \).

If \( v_i(p) \) divides \( p \) define \( w_i(p) = v_i(p) \) for any \( p \mid |G| \) and \( i = 1, 2 \). If \( v_i(p) \) does not divide \( p \) then \( G(p) \) is metacyclic and has a Liedahl presentation over \( K \) (by Remark 2.6). Thus, there are infinitely many primes \( w \) of \( K \) for which \( G(p) \) is realizable over \( K_w \). For all primes \( v_i(p) \) that do not divide \( p \) (running over both \( i \) and \( p \)) choose distinct primes \( w_i(p) \) which are not in \( T \) and for which \( G(p) \) is realizable over \( K_{w_i(p)} \) (such a choice is possible since there are infinitely many such \( w \)'s). We have chosen distinct primes \( w_i(p) \), \( p \mid |G| \) and \( i = 1, 2 \), as required. \( \Box \)

Remark 2.8. If a \( p \)-group \( G \) has a Liedahl presentation over \( M \), then \( G \) also has a Liedahl presentation over any subfield \( K \) of \( M \).

Theorem 2.9 (Liedahl [9], see also [11]). If \( G \) is tamely \( K \)-admissible, then \( G(p) \) has a Liedahl presentation over \( K \) for every prime \( p \) dividing \( |G| \).

There are no known counterexamples to the opposite implication. However, the following two results are proved for \( p \)-groups in [9, Theorem 30] and in general in [11]:

Theorem 2.10. Let \( K \) be a number field and let \( G \) be a solvable group with metacyclic Sylow subgroups. Then \( G \) is tamely \( K \)-admissible if and only if its Sylow subgroups have Liedahl presentations.

Theorem 2.11. Let \( K \) be a number field. Let \( G \) be a solvable group such that the rational primes dividing \( |G| \) do not decompose in \( K \). Then \( G \) is \( K \)-admissible if and only if its Sylow subgroups are metacyclic and have Liedahl presentations.

In particular if a solvable group is tamely \( M \)-admissible then it is also tamely \( K \)-admissible for every \( K \subseteq M \), i.e. tame admissibility goes down for solvable groups. Also, if \( G \) is solvable and any prime \( p \mid |G| \) satisfies Item (i) in Liedahl’s condition over \( M \), i.e. does not decompose in \( M \), then \( G \) is \( M \)-admissible if and only if \( G \) is tamely \( M \)-admissible.

In particular for \( M = \mathbb{Q} \) one has that any solvable group \( G \) that is \( \mathbb{Q} \)-admissible is tamely \( \mathbb{Q} \)-admissible. However over larger number fields this is no longer the case. Let us define wild \( K \)-admissibility:
**Definition 2.12.** A $G$-extension $L/K$ is wildly $K$-adequate if $L/K$ is $K$-adequate and there is a prime $p$ dividing $|G|$ such that every prime $v$ of $K$ for which

$$\text{Gal}(L_v/K_v) \supseteq G(p),$$

divides $p$. A $K$-admissible group $G$ is called wildly $K$-admissible if every $K$-adequate $G$-extension is wildly $K$-adequate.

Clearly a tamely $K$-admissible group is not wildly $K$-admissible. Theorems 2.10 and 2.11 guarantee that a solvable group which is $K$-admissible but not wildly, is tamely $K$-admissible. In particular:

**Remark 2.13.** Every $K$-admissible $p$-group which is not tamely $K$-admissible is wildly $K$-admissible. So, every non-metacyclic $K$-admissible $p$-group is wildly $K$-admissible.

### 2.3. The Grunwald-Neukirch (GN) property.

A group $G$ has the GN-property (named after Grunwald and Neukirch) over a number field $K$ if for every finite set $S$ of primes of $K$ and corresponding subgroups $G_v \leq G$ for $v \in S$, there is a Galois $G$-extension $L/K$ for which $\text{Gal}(L_v/K_v) \cong G_v$ for every $v \in S$.

The Grunwald-Wang Theorem shows that except for special cases (see [25]), abelian groups $A$ have the GN-property over $K$. A large set of examples comes from a Theorem of Neukirch [13, Corollary 2]. Let $m(K)$ denote the number of roots of unity in a number field $K$.

**Theorem 2.14 (Neukirch, [13]).** Let $K$ be a number field and $G$ a group for which $|G|$ is prime to $m(K)$. Then $G$ has the GN-property over $K$.

Another important source of examples is having a generic extension ([18, Theorem 5.9]):

**Theorem 2.15 (Saltman).** If $G$ has a generic extension over a number field $K$ then $G$ has the GN-property over $K$.

By [17], if $\mu_p \subseteq K$ then any group of order $p^3$ which is not the cyclic group of order 8 has a generic extension over $K$. In [16], many groups are proved to have a generic extension over number fields, in particular, any abelian group that does not have an element of order 8. In [16] it is also proved that the class of groups with a generic extension is closed under wreath products. In particular we have:

**Corollary 2.16 (Saltman).** Let $q$ be an odd prime and let $K$ be a number field that contains the $q$-th roots of unity. Then any iterated wreath product of odd order cyclic groups and groups of order $q^3$ has the GN-property over $K$.

For more examples see [11]. Under the assumption of the GN-property one has the following characterization of wild admissibility:

**Lemma 2.17.** Let $K$ be a number field and $G$ a $K$-admissible group that has the GN-property over $K$. Then $G$ is wildly $K$-admissible if and only if there is a prime $p \mid |G|$ for which $G(p^\alpha)$ does not have a Liedahl presentation over $K$. 


Theorem 2.19 (Jannsen, Wingberg, [7], see also [15, Theorem 7.5.10])

Moreover, $\sigma_p$ such that for every $p \mid |G|$ there is a prime $v$ of $K$ that does not divide $p$, with $\text{Gal}(L_v/K_v) \supseteq G(p)$. Then $G(p)$ has a Liedahl presentation over $L^{G(p)}$ and by Remark 2.8 $G(p_0)$ has a Liedahl presentation over $K$, contradiction.

On the other hand if all Sylow subgroups have Liedahl presentation s then by Remark 2.6 every Sylow subgroup is realizable over infinitely many completions. One can therefore choose distinct primes $\{v_i(p) \mid i = 1, 2, p \mid |G|\}$ of $K$ such that $G(p)$ is realizable over $K_{v_i(p)}$ and $v_i(p)$ $fp$ for every $p \mid |G|, i = 1, 2$. Since $G$ has the GN-property it follows that $G$ is namely $K$-admissible. \hfill \Box

2.4. Galois groups of local fields. Let $k$ be a $p$-adic field of degree $n$ over $\mathbb{Q}_p$. Let $q$ be the size of the residue field $\tilde{k}$, and let $p^\sigma$ be the size of the group of $p$-power roots of unity inside $k$. Then

1. $\text{Gal}(k_{un}/k)$ is (topologically) generated by an automorphism $\sigma$, and isomorphic to $\mathbb{Z}$;
2. $\text{Gal}(k_{ur}/k_{un})$ is (topologically) generated by an automorphism $\tau$, isomorphic to $\mathbb{Z}(p^\sigma)$ (which is the complement of $\mathbb{Z}_p$ in $\mathbb{Z}$);
3. The group $\text{Gal}(k_{ur}/k)$ is a pro-$\mathbb{Z}_p$ group generated by $\sigma$ (lifting the above mentioned automorphism) and $\tau$, subject to the single relation $\sigma^{-1} \tau \sigma = \tau^q$.

Moreover, $\sigma$ and $\tau$ act on $\mu_{p^n}$ by exponentiation by some $g \in \mathbb{Z}_p$ and $h \in \mathbb{Z}_p$, respectively (Note that $g$ and $h$ are well defined modulo $p^n$).

Let $\overline{G_k(p)}$ denote the Galois group $\text{Gal}(m/k)$, where $m$ is the maximal $p$-extension of $k$ inside a separable closure $k$. Let $p^{\sigma_0}$ be the number of roots of unity of $p$-power order in $k$. Note that if $p^{\sigma_0} \neq 2$ then $n$ must be even. The following Theorem summarizes results of Shafarevich [23], Demushkin [3], Serre [21] and Labute [8]:

Theorem 2.18 ([22, Section II.5.6]). When $p^{\sigma_0} \neq 2$, $\overline{G_k(p)}$ has the following presentation of pro-$p$ groups:

$$\overline{G_k(p)} \cong \left\{ \langle x_1, \ldots, x_{n+2} \mid x_1^{2^{s_0}} [x_1, x_2] \cdots [x_{n+1}, x_{n+2}] = 1 \rangle, \quad \text{if } s_0 > 0 \right\} \cup \left\{ \langle x_1, \ldots, x_{n+2} \rangle, \quad \text{if } s_0 = 0 \right\}$$

When $p^{\sigma_0} = 2$ and $n$ is odd,

$$\overline{G_k(p)} \cong \langle x_1, \ldots, x_{n+2} \mid x_1^2 [x_1, x_2] x_2^2 [x_2, x_3] \cdots [x_{n+1}, x_{n+2}] = 1 \rangle,$$

otherwise there is an $f \geq 2$ for which $\overline{G_k(p)}$ has one of the pro-$p$ presentations:

$$\langle x_1, \ldots, x_{n+2} \mid x_1^2 [x_1, x_2] x_2^2 [x_3, x_4] \cdots [x_{n+1}, x_{n+2}] = 1 \rangle,$$

or

$$\langle x_1, \ldots, x_{n+2} \mid x_1^2 [x_1, x_2] \cdots [x_{n+1}, x_{n+2}] = 1 \rangle.$$
if $n$ is even, and
\[ G_k = \langle \sigma, (\tau)p, (x_0, \ldots, x_n) \mid \tau^a = \tau, \]
\[ x_0^a = \langle x_0, \tau \rangle^a x_1^a [x_1, y_1] [x_2, x_3] \cdots [x_{n-1}, x_n] \]
if $n$ is odd, where $(\tau)p$ denotes that $\tau$ is a pro-$p'$ element (has order prime to $p$ in every finite quotient), and $(x_0, \ldots, x_n)^N_p$ denotes the condition that the closed normal subgroup generated by $x_0, \ldots, x_n$ is required to be a pro-$p$ group. Here, the closed subgroup generated by $\sigma$ and $\tau$ is isomorphic to $\text{Gal}(k_n/k)$.

The notation $\langle x_0, \tau \rangle$ stands for $(x_0, \tau x_0 \tau^{-1}, \ldots, x_0 \tau^{n-1})$, where $\pi_p \in \hat{\mathbb{Z}}$ is an element such that $\pi_p \mathbb{Z} = \mathbb{Z}_p$. Also, $y_1$ is a multiple of $x_1^{\tau_2^{p+1}}$ by an element in the maximal pro-$p$ quotient of the pro-finite group generated by $x_1, \sigma^{\tau_2}$ and $\tau^{\tau_2}$. In particular, in every pro-odd quotient of $G_k$, $[x_1, y_1]$ is trivial.

Remark 2.20. Notice that $G_k$ is a semidirect product of a pro-$p$ group $P_k$ and a profinite metacyclic group $D_k$, where $P_k$ is the closed normal subgroup generated by $x_0, \ldots, x_n$ and $D_k$ is the closed subgroup generated by $\sigma$ and $\tau$.

The $p$-Sylow subgroup of $G_k$ is therefore the pro-$p$ closure of $\langle \sigma^{\tau_2} \rangle \cdot P_k$.

Remark 2.21. If $G$ is admissible over a number field $K$, then for every $p$ there is a subgroup $H \supseteq G(p)$ which is realizable over a completion of $K$. In particular, $H$ is a product of a metacyclic group and a normal $p$-subgroup.

The following result on realizability of metacyclic $p$-groups will be used in Section 5.

**Lemma 2.22.** Let $k$ be a $p$-adic field. Then any metacyclic $p$-group $G$ is realizable over $k$.

**Proof.** Let $G = \mathcal{M}(m, n, i, t)$ (see (2.2)). The proof for $k \neq \mathbb{Q}_p$ is in [12]. For $k = \mathbb{Q}_p$ we cover 2-groups, so $m$ and $n$ are 2-powers and $t$ is odd. In this case $G_k(2)$ has the pro-2 presentation $\langle a, b, c \mid a^2 b^4 [b, c] = 1 \rangle$ (by Theorem 2.18), i.e. $G_k(2)$ is isomorphic to the free pro-2 group on three generator modulo the normal closure of the single relation. So the map $\phi : G_k(2) \to G$ defined by:

\[ a \mapsto x^{-2} y^s, \quad b \mapsto x, \quad c \mapsto y, \]

is well defined (and surjective) whenever $(x^{-2} y^s)^2 x^4 [x, y] = 1$.

As $t$ is odd, $t^2 + 1 \equiv 2$ (mod 4), $\frac{t^2 + 1}{2}$ is odd and we can choose an $s \equiv \frac{1}{t^2 + 1} \frac{1}{2} \frac{k - 1}{2}$ (mod n) so that $s(t^2 + 1) \equiv \frac{1}{2}t^4$ (mod $n$). For such $s$ one has:

\[ (x^{-2} y^s)^2 x^4 [x, y] = x^{-4} y^{s(t^2 - 1)} x^4 y^{t-1} = y^{s(t^2 + 1)t^4 + t - 1} = 1. \]

Thus $\phi$ is well defined. \(\square\)

### 3. $p$-Groups

A nilpotent group $G$ is $K$-admissible if and only if all Sylow subgroups of $G$ are $K$-admissible. In particular studying the behavior of the admissibility of $G$ under extension of number fields is reduced to understanding the behavior of its Sylow subgroups.
3.1. The case $p$ odd. We begin by proving the observation on realizability over extensions of $p$-adic fields, $p$ odd.

**Proof of Proposition 1.2.** Let $n$ denote the rank $[k: \mathbb{Q}_p]$ and let $t = [m:k]$. If $t = 1$ there is nothing to prove. For $t = 2$, $([m:k], p) = 1$ and hence from a $G$-extension $I/k$ we can form a $G$-extension $lm/m$. Now let $t > 2$. It suffice to show that $G_k(p)$ is a quotient of $G_m(p)$.

By Theorem 2.18, $G_k(p)$ and $G_m(p)$ have the following presentations of pro-$p$ groups:

$$G_k(p) \cong \left\{ \left( x_1, \ldots, x_{n+2} \mid x_1^{p^{s_0}} [x_1, x_2] \cdots [x_{n+1}, x_{n+2}] = 1 \right), \quad \text{if } s_0 > 0 \right\},$$

and

$$G_m(p) \cong \left\{ \left( x_1, \ldots, x_{nt+2} \mid x_1^{p^{s_0}} [x_1, x_2] \cdots [x_{nt+1}, x_{nt+2}] = 1 \right), \quad \text{if } s_0' > 0 \right\},$$

where $p^{s_0}$ and $p^{s_0'}$ are the numbers of $p$-power roots of unity in $k$ and $m$, respectively. Clearly $s_0 \leq s_0'$. Let $F_p(y_1, \ldots, y_k)$ denote the free pro-$p$ group of rank $k$ with generators $y_1, \ldots, y_k$. If $s_0 = 0$ then we are done since $F_p(x_1, \ldots, x_{n+1})$ is a quotient of $F_p(x_1, \ldots, x_{nt+1})$.

Suppose $s_0 > 0$. Let $\phi : G_m(p) \to F_p(y_1, \ldots, y_{\frac{nt+2}{2}})$ be the epimorphism defined by $\phi(x_{2i-1}) = 1$ and $\phi(x_{2i}) = y_i$, $i = 1, \ldots, \frac{nt+2}{2}$. Now as $t > 2$ we have:

$$nt + 2 = n \frac{t+2}{2} + 1 \geq n + 2$$

and hence there is a projection $\pi$:

$$\pi : F_p(y_1, \ldots, y_{\frac{nt+2}{2}}) \to G_k(p).$$

Thus $\pi \circ \phi : G_m(p) \to G_k(p)$ is an epimorphism. We deduce that every epimorphic image of $G_k(p)$ is also an epimorphic image of $G_m(p)$. \hfill \Box

We can now prove Proposition 1.3. It suffices to prove:

**Proposition 3.1.** Let $M/K$ be an extension of number fields. Let $p$ be an odd prime and $G$ a $p$-group that is $K$-admissible and has the GN-property over $M$. If $G$ satisfies Liedahl's condition over $M$, then $G$ is $M$-admissible.

**Proof.** As $G$ is $K$-admissible, $G$ is realizable over $K_{v_1}, K_{v_2}$ for two primes $v_1, v_2$ of $K$. We claim there are two primes $w_1, w_2$ of $M$ for which $G$ is realizable over $M_{v_1}, M_{v_2}$. Since $G$ has the GN-property over $M$ proving the claim shows $G$ is $M$-admissible. There are two cases:

Case I: $p$ decomposes in $M$. If one of the primes $v_1, v_2$ does not divide $p$, then $G$ is metacyclic and hence by Lemma 2.22, $G$ is realizable over any $M_{w_1}, M_{w_2}$ for any two primes $w_1, w_2$ of $M$ that divide $p$. If on the other hand both $v_1, v_2$ divide $p$ then by Proposition 1.2, $G$ is realizable over $M_{w_i}$ for $w_i|v_i$, $i = 1, 2$. 

\hfill Documenta Mathematica 18 (2013) 359–382
Case II: $p$ does not decompose in $M$. Since $G$ satisfies Liedahl’s condition over $M$, $G$ has a Liedahl presentation over $M$. In particular by Theorem 2.11, $G$ is $M$-admissible.

As a corollary we deduce that wild admissibility goes up for $p$-groups:

**Corollary 3.2.** Let $p$ be an odd prime. Let $M/K$ be an extension of number fields and $G$ a wildly $K$-admissible $p$-group that has the GN-property over $M$. Then $G$ is also wildly $M$-admissible.

**Proof.** Since $G$ is wildly $K$-admissible, $p$ decomposes in $K$ and hence in $M$. Thus $G$ satisfies Liedahl’s condition over $M$ and by Proposition 3.1 $G$ is $M$-admissible. By remarks 2.6 and 2.13, the wild $K$-admissibility of $G$ implies that $G$ is not realizable over $K_v$ for any prime $v$ which does not divide $p$. By Remark 2.8, $G$ is also not realizable over $M_w$ for any prime $w$ of $M$ which does not divide $p$. Therefore an $M$-adequate $G$-extension must also be wildly $M$-admissible. □

Apply Theorem 2.14, we have:

**Corollary 3.3.** Let $p$ be an odd prime. Let $M/K$ be an extension of number fields so that $M$ does not contain the $p$-th roots of unity. Let $G$ be a $p$-group that is wildly $K$-admissible. Then $G$ is wildly $M$-admissible.

3.2. The case $p = 2$. As in Proposition 1.2 we have:

**Lemma 3.4.** Let $m/k$ is a finite extension of $2$-adic fields, which is either of degree greater than $2$, or such that $m$ and $k$ contain $\sqrt{-1}$ and have the same $2$-power roots of unity. Then any $2$-group realizable over $k$ is also realizable over $m$.

**Proof.** If $[m:k] > 2$, the same proof as of Proposition 1.2 holds in all cases of Theorem 2.18. If $\sqrt{-1} \in k$, and $k$ and $m$ have the same number of $2$-power roots of unity then $G_k(p)$ and $G_m(p)$ have the same type of presentations in Theorem 2.18 and one can obtain an epimorphism: $G_m(p) \rightarrow G_k(p)$ simply by dividing by the redundant generators of $G_m(p)$. □

However, we show that Proposition 1.2 may fail for $p = 2$. We begin with some group-theoretic preparations.

**Lemma 3.5.** The group

$$G = \langle a_1, a_2, a_3 \mid G' \text{ is central of exponent } 2, \quad a_1^2 = [a_2, a_3], \quad a_2^2 = a_3^2 = 1 \rangle$$

is not a quotient of the pro-$2$ group

$$\Gamma = \langle x_1, \ldots, x_4 \mid x_1^2[x_1,x_2][x_3,x_4] = 1 \rangle.$$

**Proof.** For $j,k = 1,2,3$, write $\alpha_{j,k} = [a_j, a_k] \in G$. Suppose $x_i \mapsto a_1^{h_i} a_2^{h_2} a_3^{h_3} z_i$ ($i = 1, \ldots, 4$) is an epimorphism $\Gamma \rightarrow G$, where $z_i \in G'$. Then $[x_2i-1, x_2i] \mapsto$
\[
\prod_{j,k=1}^{3}[a_{j_1k_1}, a_{j_2k_2}, a_{j_3k_3}] = \prod_{1 \leq k < j \leq 3} a_{j,k}^{t_{21-1,k}t_{21,k}t_{22-1,j}t_{22,j}}.
\]
Since \( \exp(G) = 4 \), the defining relation of \( \Gamma \) translates to
\[
\prod_{1 \leq k < j \leq 3} a_{j,k}^{t_{21-1,k}t_{21,k}t_{22-1,j}t_{22,j}} = 1,
\]
from which it follows that \( t_{1,j}t_{2,k} - t_{1,k}t_{2,j} + t_{3,j}t_{4,k} - t_{3,k}t_{4,j} \equiv 0 \pmod{2} \) for every \( 1 \leq k < j \leq 3 \).

Let \( V \) denote the vector space \( \mathbb{F}_2^4 \), endowed with the bilinear form \( b: V \times V \to \mathbb{F}_2 \) defined by \( b((v_1)_{i=1}^4, (v'_i)_{i=1}^4) = v_1v'_1 + v_2v'_2 + v_3v'_3 - v_4v'_4 \). This is an alternating non-degenerate form (in fact, hyperbolic), and letting \( \vartheta \in V \) be the vectors \( \vartheta_1 = t_{i,j} \), we have that \( b(\vartheta, \vartheta) = 0 \) for every \( j, k = 1, 2, 3 \). It follows that \( T = \text{span} \{ \vartheta_1, \vartheta_2, \vartheta_3 \} \subset V \) is orthogonal to itself. But then \( \dim T \leq \frac{1}{2} \dim V = 2 \), contradicting the assumption that the induced map \( \Gamma \to G/G' = C_2^3 \) is surjective.

**Corollary 3.6.** There is a group of order \( 2^6 \) which is realizable over \( k = \mathbb{Q}_2 \) but not over \( m = \mathbb{Q}_2(\sqrt{-1}) \).

**Proof.** As before, we construct a quotient of \( \overline{G_k(2)} \) which is not a quotient of \( \overline{G_m(2)} \). Let \( G \) and \( \Gamma \) be as in Lemma 3.5. By Theorem 2.18, \( \overline{G_m(2)} \cong \Gamma \) and
\[
\overline{G_k(2)} = \langle x_1, x_2, x_3 \mid x_1^2x_2^2x_3^2 = 1 \rangle.
\]

Mapping \( x_i \mapsto a_i \) projects \( \overline{G_k(2)} \) onto \( G \), which is not a quotient of \( \Gamma \). \( \square \)

It seems that \( 2^6 \) is the minimal possible order for such a 2-group.

**Remark 3.7.** Let \( m/k \) be an extension of local fields. If there is one 2-group which is realizable over \( k \) but not over \( m \), then there are infinitely many such groups. Indeed, let \( G \) be such a group, and let \( k' \) be a \( G \)-Galois extension of \( k \); the Galois group of any 2-extension of \( k' \) which is Galois over \( k \) has \( G \) as a quotient, and so is not realizable over \( m \).

Let us apply this example to construct an extension of number fields \( M/K \) for which the group \( G \) of Lemma 3.5 is wildly \( K \)-admissible but not \( M \)-admissible and not even \( M \)-preadmissible.

Let \( p \) and \( q \) be two primes for which:
1) \( p \equiv 5 \pmod{8} \)
2) \( q \equiv 1 \pmod{8} \)
3) \( q \) is not a square mod \( p \).

**Proposition 3.8.** Let \( K = \mathbb{Q}(\sqrt{-7}), M = K(i) \) and \( G \) be the group from Lemma 3.5. Then \( G \) is wildly \( K \)-admissible but not \( M \)-preadmissible.

**Proof.** Since \( G \) is a 2-group that is not metacyclic it is realizable only over completions at primes dividing 2. In particular if \( G \) is \( K \)-admissible then \( G \) is wildly \( K \)-admissible. As 2 splits in \( K \), any (of the two) prime divisor \( v \) of 2 in \( M \) has a completion \( M_v \cong \mathbb{Q}_2(\sqrt{-7}) \). By Corollary 3.6, \( G \) is not realizable...
Danny Neftin and Uzi Vishne

over \( \mathbb{Q}_2(i) \) and hence not \( M\)-preadmissible. It therefore remains to show \( G \) is \( K\)-admissible.

The rational prime \( p \) is inert in \( K \). Let \( \mathfrak{p} \) be the unique prime of \( K \) that divides \( p \). We have \( N(p) := [K_{\mathfrak{p}}] = p^2 \equiv 1 \pmod{8} \). Thus, \( K_{\mathfrak{p}} \) has a totally ramified \( C_8 \)-extension. Let \( q_1, q_2 \) be the two primes of \( K \) dividing 2.

Consider the field extension \( L_0 = K(\mu_8, \sqrt{p})/K \). It has a Galois group

\[
\text{Gal}(L_0/K) \cong C_2^2 \cong G/Z(G).
\]

This extension is ramified only at \( q_1, q_2 \) and \( \mathfrak{p} \). As \( K_{q_i} \cong \mathbb{Q}_2 \) and \( \text{Gal}(L_0_{q_i}/K_{q_i}) \cong C_2^2 \), \( (L_0)_{q_i}/K_{q_i} \) is the maximal abelian extension of \( K_{q_i} \) of exponent 2, for \( i = 1, 2 \). Note that \( N(p) \equiv 1 \pmod{8} \) and hence \( K_{\mathfrak{p}} \) contains \( \mu_8 \).

Notice that \( Z(G) = G' \cong C_2^2 \) and \( G/Z(G) \cong C_2^3 \). Let us show that the central embedding problem

\[
(3.1)
\]

has a solution. Let \( \pi \) denote the epimorphism \( G \to G/Z(G) \). By theorems 2.2 and 4.7 in [14], there is a global solution to Problem 3.1 if and only if there is a local solution at every prime of \( K \). There is always a solution at primes of \( K \) which are unramified in \( L_0 \) so it suffices to find solutions at \( \mathfrak{p}, q_1, q_2 \). Any \( G \)-extension of \( K_{q_i} \) contains \( (L_0)_{q_i} \) (as it is the unique \( C_2^3 \) extensions of \( \mathbb{Q}_2 \), \( i = 1, 2 \). Since \( G \) is realizable over \( \mathbb{Q}_2 \) we deduce that the induced local embedding problem

\[
(3.2)
\]

has a solution for \( i = 1, 2 \). Since \( (L_0)_{\mathfrak{p}} \) is the ramified \( C_2 \)-extension of \( K_{\mathfrak{p}} \), it can be embedded into the totally ramified \( C_4 \)-extension and hence the local embedding problem at \( \mathfrak{p} \) has a solution.

Therefore, Embedding problem 3.1 has a solution. Let \( L \) be the corresponding solution field. As Problem 3.1 is a Frattini embedding problem such a solution must be surjective globally and at \( \{q_1, q_2\} \). Thus, \( L_0/K \) can be embedded in a Galois \( G \)-extension \( L/K \) for which \( \text{Gal}(L_{q_i}/K_{q_i}) \cong G \), for \( i = 1, 2 \). The field \( L \) is clearly \( K \)-adequate and hence \( G \) is \( K \)-admissible.
4. Realizability under extension of local fields

Realizability of a group $G$ as a Galois group over a field $k$ is clearly a necessary condition for $k$-admissibility. When $k$ is a local field, the conditions are equivalent since a division algebra of index $n$ is split by every extension of degree $n$.

In this section we study realizability of groups under field extensions, assuming the fields are local.

4.1. Totally ramified extensions. We first note what happens under prime to $p$ local extensions:

**Lemma 4.1.** Let $G_1$ be a subgroup of $G$ that contains a $p$-Sylow subgroup of $G$ and is realizable over the $p$-adic field $k$. Let $m/k$ be a finite extension for which $([m:k], p) = 1$. Then there is a subgroup $G_2 \leq G_1$ that contains a $p$-Sylow subgroup of $G$ and is realizable over $m$.

**Proof.** Let $l/k$ be a $G_1$-extension. Then $lm/m$ is a Galois extension with Galois group $G_2$ which is a subgroup of $G_1$ and for which $[l : l \cap m] = |G_2|$. Since $([l \cap m:k], p) = 1$, any $p^r | [l:k] = |G_1|$ also divides $p^r | [l : l \cap m] = |G_2|$. Thus $G_2$ must also contain a $p$-Sylow subgroup of $G$. □

The case where $p$ divides the degree $[m:k]$ is more difficult. Let us consider next totally ramified extensions:

**Lemma 4.2.** Let $p \neq 2$. Let $G$ be a group, $k$ a $p$-adic field with $n = [k : \mathbb{Q}_p]$ and $m/k$ a totally ramified finite extension. Assume furthermore that $m/k$ is not the extension $\mathbb{Q}_3(\zeta_9 + \zeta_9)/\mathbb{Q}_3$. If $G$ is realizable over $k$ then $G$ is also realizable over $m$.

**Remark 4.3.** This shows that if $G$ has a subgroup $G_1$ that contains a $p$-Sylow of $G$ and is realizable over $k$ then we can pick $G_2 := G_1$ as a subgroup which contains a $p$-Sylow subgroup of $G$ and is realizable over $m$.

**Proof.** Let $m/k$ be a totally ramified extension of degree $r = [m:k]$. We shall construct an epimorphism $G_m \to G_k$. For this we shall consider the presentations given in Theorem 2.19. Denote the parameters of $k$ by $n, q, s, g$ and $h$. Then the degree of $m$ over $\mathbb{Q}_p$ is $nr$ and its residue degree remains $q$. Denote the rest of the parameters over $m$ by $s', g'$ and $h'$ (the parameters that correspond to $s, g$ and $h$ in Theorem 2.19). Then by Theorem 2.19, $G_m$ has the following presentation (as a profinite group):

$G_m = \langle \sigma, (\tau)^{p^r}, (x_0, \ldots, x_{nr})\rangle_p^n |$

$\tau^s = \tau^q, x_0^a = (x_0, \tau)^g x_1^{p^r} [x_1, x_2] \cdots [x_{nr-1}, x_{nr}],$

if $nr$ is even and

$G_m = \langle \sigma, (\tau)^{p^r}, (x_0, \ldots, x_{nr})\rangle_p^n |$

$\tau^s = \tau^q, x_0^a = (x_0, \tau)^g x_1^{p^r} [x_1, y] [x_2, x_3] \cdots [x_{nr-1}, x_{nr}].$
if \( nr \) is odd. Let \( P_k \) be the closed normal subgroup of \( G_k \) generated by \( x_0,\ldots,x_n \) and let \( D_k \) (resp. \( D_m \)) be the closed subgroup generated by \( \sigma \) and \( \tau \). By assumption, \( P_k \) is a pro-\( p \) group. Note that as \( k \) and \( m \) have the same residue degree (same \( q \)), \( D_k \cong D_m \).

Let us construct the epimorphism from \( G_m \) to \( G_k \). First send \( x_0 \) and every \( x_k \) with \( k \) odd in the presentation of \( G_m \) to 1. We get an epimorphism

\[
G_m \twoheadrightarrow (\sigma, (\tau)^{p, (z_1, \ldots, z_d)^N} | \sigma \tau \sigma^{-1} = \tau^q)
\]

where \( d = \lceil \frac{nr-1}{2} \rceil \) and \( \lceil \gamma \rceil \) denotes the smallest integer \( \geq \gamma \). Let us continue under the assumption \( d \geq n + 1 \). Then there is an epimorphism \( F_p(d) \twoheadrightarrow F_p(n+1) \). We therefore obtain epimorphisms:

\[
G_m \twoheadrightarrow (\sigma, (\tau)^{p, (z_1, \ldots, z_n+1)^N} | \sigma \tau \sigma^{-1} = \tau^q) \twoheadrightarrow D_k \cong P_k \twoheadrightarrow G_k.
\]

The numerical condition \( \lceil \frac{nr-1}{2} \rceil \geq n + 1 \) fails if and only if:

1. \( r = 1 \), or
2. \( r = 2 \), or
3. \( r = 3 \) and \( n = 1 \).

The case \( r = 1 \) is trivial. Since \( p \) is an odd prime, the cases \( r = 2 \) and \( r = 3 \) are done by Lemma 4.1, unless \( p = 3 \) and \( \lceil \frac{m}{k} \rceil = r = 3 \), in which case \( n = 1 \), so \( k = \mathbb{Q}_3 \). If \( m \neq \mathbb{Q}_3(\zeta_9 + \overline{\zeta}_9) \) then \( m \cap k_{tr} = k \) and the parameters \( g, h, s \) in the presentation of \( G_k \) remain the same in the presentation of \( G_m \). In such case there is an epimorphism from \( G_m \) onto \( G_k \) whose kernel is generated by \( \langle x_2, x_3 \rangle \).

So the case \( k = \mathbb{Q}_3 \) and \( m = \mathbb{Q}_3(\zeta_9 + \overline{\zeta}_9) \) remains open. This will be one of several sensitive cases.

4.2. The sensitive cases.

**Definition 4.4.** We call the extension \( m/k \) sensitive if it is one of the following:

1. \( k = \mathbb{Q}_3 \) and \( m \) is the totally ramified 3-extension \( \mathbb{Q}_3(\zeta_9 + \overline{\zeta}_9) \),
2. \( k = \mathbb{Q}_3 \) and \( m = \mathbb{Q}_5(\zeta_{11}) \) is the unramified 5-extension,
3. \( [k: \mathbb{Q}_3] = 1, 2, 3 \) and \( m/k \) is the unramified 3-extension,
4. \( k = \mathbb{Q}_3 \) and \( m = \mathbb{Q}_3(\zeta_7) \) is the unramified 6-extension.

**Remark 4.5.** There are 17 sensitive field extensions, up to isomorphism: one over \( \mathbb{Q}_5 \) and 16 over \( \mathbb{Q}_3 \). Fixing the algebraic closures of the respective \( p \)-adic fields, there is one sensitive 5-adic extension and 27 3-adic ones. This can be verified using the automated tools in [6]. We provide details in the Appendix.

Let us formulate the problem in case (1) for odd order groups:

**Remark 4.6.** Given a field \( F \) denote by \( G_F^{\text{odd}} \) the Galois group that corresponds to the maximal pro-odd Galois extension of \( F \). In the \( p \)-adic case, for odd \( p \), this is obtained from the presentation of \( G_k \) (see Theorem 2.19) simply by dividing by the 2-part of \( \sigma \) and \( \tau \). In such case we get a presentation of \( G_F^{\text{odd}} \) by identifying \( \sigma_2 = \tau_2 = 1 \). We get that \( y_1 \) is a power of \( x_1 \) and hence \( [x_1, y_1] = 1 \).
Question 4.7. Let $m/k$ be the sensitive extension (1). Then $q = 3$; also $p^s = 3$ so we can choose $h = -1$. For $m$ we have $p^m = 9$ and $\tau(\zeta_9 + \zeta_9^{-1}) = \zeta_9 + \zeta_9^{-1}$, so $h_m = -1$ as well. Theorem 2.19 gives us the presentations:

$$G_{k}^{\text{odd}} = \langle \sigma, (\tau)p, (x_0, x_1) \rangle | \text{ } \tau^p = \tau^3, \sigma = (x_0, \tau)x_1^2 \rangle,$$

while:

$$G_{m}^{\text{odd}} = \langle \sigma, (\tau)p, (x_0, x_1, x_2, x_3)^N \rangle | \text{ } \tau^p = \tau^3, \sigma = (x_0, \tau)x_1^2[x_2, x_3] \rangle,$$

where $\sigma, \tau$ are of order prime to 2 and $(x_0, \tau) = (x_0\tau x_0^{-1}\tau)^{\mathbb{F}_p}$, which has order a power of 3 in every finite quotient. Does the following hold: Let $G$ be an epimorphic image of $G_{k}^{\text{odd}}$, is there necessarily a subgroup $G(p) \leq G_0 \leq G$ so that $G_0$ is an epimorphic image of $G_{m}^{\text{odd}}$?

Note that for a 3-group $G$ the claim was proved in Proposition 1.2.

Remark 4.8. In fact quotients of $G_{k}^{\text{odd}}$ with $\tau = 1$ can be covered: the group $\langle \sigma, (x_0, x_1)^N \rangle | \text{ } x_0^p = x_0, \sigma = (x_0, \tau)x_1^2 \rangle$ is covered by $\sigma \mapsto \sigma, \tau \mapsto 1, x_0 \mapsto 1, x_1 \mapsto 1, x_2 \mapsto x_1$ and $x_3 \mapsto 1$. This corresponds to realization of $G$ over $k$ whose ramification index is a 3-power.

4.3. Extensions of local fields. We can now approach the general case. Recall the presentation of $G_k$ from Theorem 2.19. Let $P_k$ denote the closed normal subgroup generated by $x_0, \ldots, x_n$ and $D_k$ the closed subgroup generated by $\sigma$ and $\tau$, as in Remark 2.20.

Remark 4.9. (1) Decompose $\langle \sigma \rangle$ into its $p$-primary part generated by $\sigma_p$ and its complement generated by $\sigma_{p'}$ so that $\sigma = \sigma_p\sigma_{p'}$, where $[\sigma_p, \sigma_{p'}] = 1$. Then the pro-$p$ closure of $\langle \sigma_p \rangle \cdot P_k$ is a $p$-Sylow subgroup of $G_k$.

(2) In every finite quotient $\sigma_p$ is a power of $\sigma$, and so normalizes $\tau$. It follows that $[\tau, \sigma_p]$ is a power of $\tau$, and so a pro-$p'$ element.

(3) The image of the closure of $\langle \sigma_p \rangle P_k$ is normal in a quotient of $G_k$ if and only if $\tau$ conjugates $\sigma_p$ into the closure of $\langle \sigma_p \rangle P_k$; but then the image of $[\tau, \sigma_p]$ is a pro-$p$ element, so by (2) this is the case if and only if the image of $[\tau, \sigma_p]$ is trivial.

(4) Let $\tilde{G}_k$ denote the maximal quotient of $G_k$ with a normal $p$-Sylow subgroup. It follows from (3) that $\tilde{G}_k$ is defined by the relation $[\tau, \sigma_p] = 1$.

Lemma 4.10. Let $p$ be an odd prime. Let $m/k$ be an extension of $p$-adic fields with $f = [m:k]$ a $p$-power, and $\left\lceil \frac{n-1}{2} \right\rceil \geq n+2$ where $n = [k:Q_p]$ and $r = [m:k]$. Then $\tilde{G}_k$ is a quotient of $G_m$.

Proof. We construct an epimorphism from $G_m$ to $\tilde{G}_k$. Let $s_m, g_m, h_m$ be the invariants $s, g, h$ in Theorem 2.19 that correspond to $m$, and let $n = [k:Q_p]$. Theorem 2.19 gives the following presentation of $G_m$:

$$G_m = \langle \sigma, (\tau)p, (x_0, \ldots, x_n)^N \rangle | \text{ } \tau^p = \tau^q, x_0^p = (x_0, \tau)x_1^2[x_1, x_2] \cdots [x_{nr-1}, x_{nr}] \rangle,$$
if \( nr \) is even and
\[
G_m = \langle \sigma, (\tau)_{p'}, (x_0, \ldots, x_{nr})^{N}_{p} \mid \tau' = \tau'^f, x_0^\sigma = (x_0, \tau)^{m_{p'}} x_1^{m_{p'}} [x_1, y_1] [x_2, x_3] \cdots [x_{nr-1}, x_{nr}] \rangle,
\]
if \( nr \) is odd.
Let \( P_k \) (resp. \( P_m \)) be the closed normal subgroup generated by \( x_0, \ldots, x_n \) (resp. \( x_0, \ldots, x_{nr} \)) in \( G_k \) (resp. \( G_m \)) and let \( D_k \leq G_k \) (resp. \( D_m \leq G_m \)) be the closed subgroup generated by \( \sigma, \tau \) in \( G_k \) (resp. in \( G_m \)).
Set \( d = \left\lceil \frac{nr}{2} \right\rceil \), so by assumption \( d \geq n + 2 \). Similarly to Lemma 4.2 (noting that this time \( D_m \) can be viewed as a subgroup of index \( f \) in \( D_k \)), we have an epimorphism
\[
G_m \rightarrow \langle \sigma, (\tau)_{p'}, (z_1, \ldots, z_{m})^{N}_{p} \mid \tau' = \tau'^f \rangle \\
\cong \langle (\sigma)_{p'}, (\sigma')_{p'}, (\tau)_{p'}, (z_1, \ldots, z_{m})^{N}_{p} \mid \tau^\sigma \sigma' = \tau'^f, [\sigma, \sigma'] = 1 \rangle.
\]
Let us divide by the relations \( z_0^f = \sigma \) and \( \tau^m = \tau^\sigma \sigma^{-1} \), where \( \sigma^{-1} \) is well defined since \( f \) is a \( p \)-power. We then obtain an epimorphism to
\[
\langle (\sigma')_{p'}, (\tau)_{p'}, (z_1, \ldots, z_{m})^{N}_{p} \mid \tau^m = \tau^\sigma \sigma'^{-1} f, [z_m, \sigma'] = 1 \rangle.
\]
Adding the relation \([z_m, \sigma']\) and sending \( z_m \mapsto \sigma \) maps this group onto
\[
\langle (\sigma')_{p'}, (\tau)_{p'}, (z_1, \ldots, z_{m-1}, \sigma)_{p'}^{N}_{p} \mid \tau^\sigma \sigma'^{-1} f = \tau^q, [\sigma, \sigma'] = 1 \rangle.
\]
Mapping \( \sigma' \mapsto \sigma'^f \), this groups maps onto
\[
\langle (\sigma)_{p'}, (\sigma')_{p'}, (\tau)_{p'}, (x_0, \ldots, x_{n})^{N}_{p} \mid \tau^\sigma \sigma' = \tau^q, [\sigma, \sigma'] = [\sigma, \tau] = 1 \rangle,
\]
since by Remark 4.9 the assumption that the normal subgroup generated by \( \sigma \) is a \( p \)-group is equivalent to \( [\sigma, \tau] = 1 \).
But \( G_k \) is a quotient of this group by Theorem 2.19 and Remark 4.9.(4).

Using Lemma 4.10, we can prove:

**Proof of Theorem 1.6.** Let \( n, q \) be as defined above for \( k \), and let \( r = [m:k] \).
Let \( f = [m:k] = f_p f_{p'} \) where \( f_p \) is a \( p \)-power and \( f_{p'} \) is prime to \( p \). There is an unramified \( C_f \)-extension \( m'/k \) which lies in \( m \), and then \( m/m' \) is totally ramified. Denote by \( m_p \) the subfield of \( m' \) which is fixed by \( C_{f_p} \). Let \( r' = \frac{m'}{m} = [m:m_p] \). By Lemma 4.1, there is a subgroup \( G_0 \leq G \) that contains a \( p \)-Sylow subgroup of \( G \) and an epimorphism \( \phi : G_{m_p} \rightarrow G_0 \). The list of sensitive cases satisfies that if \( m/k \) is non-sensitive and \( m_p/k \) is unramified and prime to \( p \), then \( m/m_p \) is also non-sensitive and therefore we can assume without loss of generality that \( m_p = k \), \( G = G_0 \), i.e \( f_{p'} = 1, f = f_p \), and \( r' = r \).
If \( \left\lceil \frac{m-1}{2} \right\rceil \geq n + 2 \) then \( G \) is a quotient of \( G_m \) by Lemma 4.10. This numerical condition fails if and only if
- (1) \( r = 4, 5 \) and \( n = 1 \);
- (2) \( r = 3 \) and \( n = 1, 2, 3 \);
- (3) \( r = 1, 2 \).

---

*Documenta Mathematica 18 (2013) 359–382*
The cases \( r = 1, 2, 4 \) are covered by Lemma 4.1. We are left with cases \( r = 3, 5 \). For \( r = 5 \), \( n = 1 \) so \( k = \mathbb{Q}_3 \) and by Lemma 4.2 we may assume \( m/k \) is not totally ramified, so \( m/k \) is the unramified 5-extension which is sensitive.

Let \( r = 3 \). Lemma 4.1 covers the case \( p \neq 3 \), so we may assume \( p = 3 \).

Note that \( f | r \). If \( f = 1 \), Lemma 4.2 applies, except for \( m = \mathbb{Q}_3(\zeta_9 + \zeta_9) \) and \( k = \mathbb{Q}_3 \), which is sensitive. If \( f = 3 \), then \( m/k \) is the unramified 3-extension and \( n = [k: \mathbb{Q}_3] = 1, 2, 3 \) which are all sensitive. \( \square \)

The following example shows that the assumption in Theorem 1.6 that the normal \( p \)-Sylow subgroup of \( G \) is normal, is essential.

**Example 4.11.** Let \( p < q \) be odd primes such that \( p^q \equiv 1 \pmod{q} \) and \( p \neq 1 \pmod{q} \) (for example \( p = 7 \) and \( q = 29 \)). Let \( G = C_p \ast D \) where

\[
D = \langle a, b \mid a^p = b^q = 1, a^{-1}ba = b^p \rangle.
\]

Let \( k = \mathbb{Q}_p \) and \( m/k \) the unramified extension of degree \( p \). Then \( G \) is realizable over \( k \) but there is no subgroup of \( G \) that contains a \( p \)-Sylow subgroup of \( G \) and is realizable over \( m \).

**Proof.** Let \( P = C_p^p \), so that \( G = P \times D \). Then one has the projections

\[
G_k \to G_k^{\text{odd}} = \langle (\sigma, (\tau)_p', (x_0, x_1)_p^N \mid \tau^\sigma = \tau^p, x_0^\sigma = (x_0, \tau)x_1^p \rangle \to \langle (\sigma, (\tau)_p', (x_1)_p^N \mid \tau^\sigma = \tau^p, x_1^p = 1 \rangle,
\]

where the latter two group are pro-odd and the second epimorphism is obtained by dividing by \( x_0 \). The latter group maps onto \( G \) by \( \sigma \mapsto a \) and \( \tau \mapsto b \). It is therefore left to prove that for any homomorphism \( \phi : G_m \to G \), \( \text{Im}(\phi) \) does not contain a \( p \)-Sylow subgroup of \( G \). Assume on the contrary that \( H = \text{Im}(\phi) \) does. Recall:

\[
G_m = \langle (\sigma, (\tau)_p', (x_0, \ldots, x_p)_p^N \mid \tau^\sigma = \tau^p, x_0^\sigma = (x_0, \tau)x_1^p[x_1, y_1][x_2, x_3] \cdots [x_{p-1}, x_p] \rangle.
\]

Since \( q \) is the only prime dividing \( |G| \) other than \( p \), and \( \tau \) is pro-\( p' \), any map into \( G \) must split through:

\[
G_m \to \langle (\sigma, (\tau)_p', (x_0, \ldots, x_p)_p^N \mid \sigma, \tau = 1, \tau^q = 1, x_0^\sigma = (x_0, \tau)x_1^p[x_1, y_1][x_2, x_3] \cdots [x_{p-1}, x_p] \rangle.
\]

However the latter group has a normal \( p \)-Sylow subgroup which is the product of the closed normal subgroup generated by the \( x_i \)'s and the pro-\( p \) group generated by \( \sigma^q \). In particular, letting \( \pi : G \to G/P = D \) be the projection, the image of \( \pi\phi \) has a normal \( p \)-Sylow subgroup. This implies \( \pi\phi \) is not surjective. But \( H \) contains a \( p \)-Sylow subgroup of \( G \), so we must have \( \text{Im}(\pi\phi) = C_p \). Again since \( H \) contains a \( p \)-Sylow subgroup, and in particular \( P \), we must have \( H = P \times C \) where \( C = C_p \) is a subgroup of \( D \) and the action of \( C \) on \( P \) is induced from the action of \( D \). Thus:

\[
\text{rank}(H) = \text{rank}(H/[H, H]) = \text{rank}((P/[P, C]) \times C) = q + 1.
\]
Since $H$ is a $p$-group any epimorphism to it must split through $G_m(p)$. However
rank$(G_m(p)) = [m:k] + 1 = p + 1$, leading to a contradiction. \hfill $\Box$

5. Extensions of number fields

We shall now apply Theorem 1.6 to study admissibility and wild admissibility.

5.1. Main Theorem.

Proof of Theorem 1.8. As mentioned in the introduction, Liedahl’s condition is
necessary. Let us show that if $G$ satisfies this condition then $G$ is $M$-admissible.
We claim that one can choose distinct primes $w_i(p)$, $i = 1, 2, p | |G|$, of $M$ and
and corresponding subgroups $H_i(p) \leq G$ so that $H_i(p)$ contains a $p$-Sylow subgroup
of $G$ and is realizable over $M_{w_i(p)}$, $i = 1, 2$.
As $G$ is $K$-admissible, for every $p | |G|$ there are two options:

1) there are two primes $v_1, v_2$ of $K$ dividing $p$ and two subgroups $G(p) \leq
G_i \leq G$ so that $G_i$ is realizable over $K_v$, $i = 1, 2$.

2) $G(p)$ is realizable over $K_v$ for $v$ which does not divide $p$.

In case (1) with $p$ odd and $G(p)$ normal in $G$, by Theorem 1.6, for any prime
$p$ dividing $v_1$ or $v_2$ there is a subgroup $G(p) \leq H_w \leq G$ that is realizable
over $M_w$. Choose two such primes $w_1(p), w_2(p)$ and set $H_i(p) := H_w(w_i(p))$ (the
subgroups Theorem 1.6 constructs). In case $G(p)$ is not normal or $p = 2$, we
assumed $G(p)$ is metacyclic and by Lemma 2.22, $G(p)$ is realizable over any
$M_w$ for any prime $w$ dividing $v_1$ or $v_2$. In such case similarly choose two such
primes $w_1(p), w_2(p)$ and set $H_i(p) = G(p)$.
In case (2), $G(p)$ is metacyclic. If $p$ has more than one prime divisor in $M$ then
there are two primes $w_1(p), w_2(p)$ so that $w_i(p)$ divides $p$ and by Lemma 2.22
$H_i(p) := G(p)$ is realizable over $M_{w_i(p)}$, $i = 1, 2$.
If $p$ has a unique prime divisor in $M$ then $G(p)$ is assumed to have a Liedahl
presentation. Liedahl’s condition implies that there are infinitely many primes
$w(p)$ for which $G(p)$ is realizable over $M_{w(p)}$ (see Theorem 28 and Theorem 30
in [9]). Thus, we can choose two primes $w_1(p), w_2(p)$ for every prime $p | |G|
that has only one prime divisor in $M$, so that the primes $w_i(p)$, $i = 1, 2$, are not
divisors of any prime $q | |G|$ and are all distinct. For such $p$, we also choose
$H_i(p) := G(p)$.
We have covered all cases of behavior of divisors of rational primes in $M$ and
hence proved the claim. It follows that $G$ is $M$-preadmissible and as $G$ has the
GN-property over $M$, $G$ is $M$-admissible. \hfill $\Box$

Remark 5.1. If $G$ has metacyclic Sylow subgroups, the proof of Theorem 1.8
does not use Theorem 1.6 and holds for sensitive extensions as well.

5.2. Wild admissibility. As to wild admissibility Theorem 1.8 and
Lemma 2.17 give:

Corollary 5.2. Let $M/K$ be a non-sensitive extension. Let $G$ be a $K$-
admissible group for which every Sylow subgroup is either normal or metacyclic
and the 2-Sylow subgroups are metacyclic. Assume $G$ has the GN-property over

\textbf{Documenta Mathematica 18 (2013) 359–382}
M, satisfies Liedahl’s condition over M but there is a prime p for which G(p) does not have a Liedahl presentation over M. Then G is wildly M-admissible.

We deduce that for groups as in Theorem 1.8, wild admissibility goes up in the following sense that generalizes Corollary 3.2:

**Corollary 5.3.** Let M/K be a non-sensitive extension. Let G be a wildly K-admissible group for which every Sylow subgroup is either normal or metacyclic and the 2-Sylow subgroups are metacyclic. Assume G has the GN-property over K and M satisfies Liedahl’s condition over M. Then G is wildly M-admissible.

**Proof.** By Theorem 1.8, G is M-admissible. The assertion now follows from Lemma 2.17, applied to both K and M, and the fact that if G(p) does not have a Liedahl presentation over K then G(p) does not have a Liedahl presentation over M (see Remark 2.8). □

5.3. **Examples.** The following is an example in which Theorem 1.8 is used to understand how admissibility behaves under extensions of a given number field:

**Example 5.4.** Let p, q be odd primes and m an integer so that m is not square mod q but is a square mod p and p ≡ q + 1 (mod q^2). For example p = 13, q = 3, m = 14. Let K = Q(√m) and G = C_p ⋊ H, where H is one of the following groups:

1. H = M_q^3 is the modular group of order q^3, i.e.
   \[ H = \langle x, y|x^{-1}yx = y^{q+1}, x^q = y^{q^2} = 1 \rangle. \]

2. H = C_{pq} × C_q.

3. H = C_t where t ∈ N_{odd} is prime to p.

We shall show in each of the cases G is K-admissible. Let M be any non-sensitive extension of K.

By Theorem 2.14, in case (1) G satisfies the GN-property over any number field that does not have any p-th and q-th roots of unity, in particular over K. By Corollary 2.16, in cases (2),(3), G satisfies the GN-property over any M and in case (1) if M contains the q-th roots of unity.

In cases (2),(3), G satisfies Liedahl’s condition over any M and in case (1), G satisfies Liedahl’s condition over M if and only if q decomposes in M or M ∩ Q(μ_{q^2}) ⊆ Q(μ_q).

It follows from Theorem 1.8, that in cases (2) and (3) G is M-admissible. In case (1), if one assumes M does not contain any p-th and q-th roots of unity or that M contains the q-th roots of unity then G is M-admissible if and only if G satisfies Liedahl’s condition.

**Proof.** The prime p splits (completely) in K. Denote it’s prime divisors in K by v_1, v_2. Then K_{v_i} ∼= Q_p for i = 1, 2. Using the presentation of G_{odd} given in Question 4.7 and dividing by x_0 = 1 one obtains an epimorphism:

\[ G_{Q_p} \to \langle \sigma, \tau^{p^i}, (x_1)^{p^i} \mid \tau^p = \sigma, x_1^p = 1 \rangle. \]
Since $p \equiv q + 1 \pmod{q^2}$ there is an epimorphism
\[ \langle \sigma, (\tau)^p \mid \tau^c = \tau^p \rangle \to M_{q^3} \]
which together with Epimorphism 5.1 shows that $C_p \wr M_{q^3}$ is an epimorphic image of $G_{Q_p}$. The group $G$ in case 3 can obtained as an epimorphic image of $G_{Q_p}$ after dividing 5.1 by $\tau = 1$. In case 2, since $q|p - 1$, there is an epimorphism
\[ \langle \sigma, (\tau)^p \mid \tau^c = \tau^p \rangle \to C_{pq} \times C_q, \]
which together with 5.1 can be used to construct an epimorphism onto $G$.
In particular $C_p \wr H$ is realizable over $K_{v_1}, K_{v_2}$ in all cases. Since $M_{q^3}, C_q \times C_q$ and $C_r$ have Liedahl presentations over $K$, they are realizable over completions at infinitely many primes of $K$. As $G$ has the GN-property over $K$, it follows that $G$ is $K$-admissible in all cases.

Remark 5.5. As Case 3 of Example 5.4 shows, the rank of $p$-Sylow subgroups of $K$-admissible groups is not bounded as apposed to the case of admissible $p$-group in which the rank of the group is bounded (see [19, Section 10]).

Remark 5.6. Case 2 in Example 5.4 is an example of a group for which proving $M$-admissibility requires the use of all steps in the proof of Theorem 1.6.

The following example shows that the assumption that every Sylow subgroup is either normal or metacyclic is essential for Theorem 1.8 even for odd order groups and non-sensitive extensions:

Example 5.7. As in Example 4.11, let $p < q$ be odd primes such that $p^2 \equiv 1 \pmod{q}$ and $p \not\equiv 1 \pmod{q}$. Let $G = C_p \wr D$ where
\[ D = \langle a, b \mid a^p = b^q = 1, a^{-1}ba = b^p \rangle. \]
Let $d$ be a non-square integer that is a square mod $pq$ and $K = \mathbb{Q}(\sqrt{d})$. Let $v_1, v_2$ be the primes of $K$ dividing $p$. Let $M/K$ be a $C_p$-extension in which both $v_1$ and $v_2$ are inert and $M$ does not contain any $p$-th and $q$-th roots of unity. Since both $p$ and $q$ have more than one prime divisor in $M$, $G$ satisfies Liedahl’s condition over $M$. As $M$ does not contain any $p$-th and $q$-th roots of unity, by Theorem 2.14, $G$ has the GN-property over $M$ and $K$. We shall now show $G$ is $K$-admissible but not $M$-admissible. Note that the only condition of Theorem 1.8 that fails is that either $G$ has a normal $p$-Sylow subgroup or a metacyclic one.

Proof. By Example 4.11, $G$ is realizable over $\mathbb{Q}_p$ and hence over $K_{v_1}, K_{v_2}$. As $G$ has the GN-property over $K$, $G$ is $K$-admissible. On the other hand, since $G(p)$ is not metacyclic, a subgroup of $G$ that contains $G(p)$ is realizable only over completions of $M$ at prime divisors of $p$. Let $w_i$ be the prime dividing $v_i$ in $M$, $i = 1, 2$. Then $w_1, w_2$ are the only primes dividing $p$ in $M$ but by Example 4.11 $G$ is not realizable over $M_{w_i}$, for $i = 1, 2$. In particular $G$ is not $M$-preadmissible and not $M$-admissible. □
We use the Jannsen-Wingberg presentation of $G_{Q_3}$ to count the sensitive extensions, as defined in Subsection 4.2, up to isomorphism. There is a single extension in each of cases (1), (2) and (4). In case (3) the extension is unramified, so it suffices to count the ground field $k$, which we do by degrees over $Q_3$. In degree 1 there is one case. In degree 2 there are $\phi$ epimorphisms all together. Dividing by the number of automorphisms, we give $[Q_3^x/Q_3^{x^2}] - 1 = 3$ quadratic extensions. Since the abelianization of $G_{Q_3}$ is $C_3^2$, there are $2^3 = 4$ Galois cubic extensions of $Q_3$. For every non-Galois cubic extension $k$ there is a unique $S_3$-Galois extension of $Q_3$ (generated over $k$ by the square root of the discriminant). The $S_3$-Galois extensions of $Q_3$ are in one-to-one correspondence to the normal subgroups of $G_{Q_3}$ with quotient $S_3$. The number of such subgroups is the number of epimorphisms from $G_{Q_3}$ to $S_3$, divided by $|\text{Aut}(S_3)| = 6$. When counting epimorphisms $\phi : G_{Q_3} \to S_3$, we may assume the generators are in $S_3$, which simplifies the presentation a great deal. Since $p^* = q = 3$ and we may assume $g = 1$ and $h = -1$, the presentation is

$$G_{Q_3} = \langle \sigma, (\tau) \rangle, (x_0, x_1) \rangle \mid \tau^3 = x_0^3 = (x_0 \tau x_0^{-1})^2 x_1^3[x_1, y_1] \rangle.$$ 

However, since $x_1$ is a pro-3 element, we may assume $x_1^3 = 1$. Since all elements of order 3 in $S_3$ commute with each other, we may assume $[x_1, y_1] = 1$ (see Theorem 2.19 for more details on $y_1$). Since $\tau$ is a pro-$3'$ element, $\varphi(\tau)$ has order at most 2, so $\varphi(x_0 \tau x_0^{-1})$ is a commutator, whose order must divide 3. Exponentiation by $\frac{3}{2}$ squares such elements. So every epimorphism to $S_3$ splits through

$$\langle \sigma, (\tau) \varphi, (x_0, x_1) \rangle \mid \tau^2 = x_0^3 = x_1^3 = 1, \tau^3 = x_0^3 = (x_0 \tau x_0^{-1})^2,$$

and we count epimorphisms from this group. If $\varphi(\tau) = 1$ then $\varphi(x_0) = 1$, so $\varphi(x_1)$ is a non-trivial element of order 3 and $\varphi(\sigma) \not\in \{\varphi(x_1)\}$. There are 6 such epimorphisms. So assume $\varphi(\tau)$ is a non-trivial involution. The relations give $[\varphi(\sigma), \varphi(\tau)] = 1$, so $\varphi(\sigma)$ is either 1 or $\varphi(\tau)$. Moreover, it turns out that $\varphi(x_0 \tau x_0^{-1} \tau^2 = \varphi(x_0)$ whenever $\varphi(x_0)$ has order dividing 3. For each possible value of $\varphi(\tau)$ we get 8 epimorphisms with $\varphi(\sigma) = 1$ and 2 more with $\varphi(\sigma) = \varphi(\tau)$. There are 3 involutions, providing us with $6 + 3 \cdot 10 = 36$ epimorphisms all together. Dividing by the number of automorphisms, we have 6 Galois extensions of $Q_3$ with Galois group $S_3$. Each Galois extension of this type contains 3 non-Galois cubic extensions of $Q_3$. The $S_3$-extension is determined by the cubic extension, being its Galois closure. So we have $3 \cdot 6 = 18$ non-Galois cubic subfields of a fixed algebraic closure $Q_3$, consisting of 6 isomorphism classes. Summing up, there are $1 + 1 + (1 + 3 + (4 + 6)) + 1 = 17$ sensitive field extensions, up to isomorphism.

References

Danny Neftin and Uzi Vishne

382

Documenta Mathematica 18 (2013) 359–382
Hasse Principle for $G$-Quadratic Forms

Eva Bayer–Fluckiger, Nivedita Bhaskhar, and Raman Parimala

Received: August 19, 2012
Revised: February 14, 2013
Communicated by Ulf Rehmann

Abstract.

Introduction.

Let $k$ be a global field of characteristic $\neq 2$. The classical Hasse–Minkowski theorem states that if two quadratic forms become isomorphic over all the completions of $k$, then they are isomorphic over $k$ as well. It is natural to ask whether this is true for $G$–quadratic forms, where $G$ is a finite group. In the case of number fields the Hasse principle for $G$–quadratic forms does not hold in general, as shown by J. Morales [M 86]. The aim of the present paper is to study this question when $k$ is a global field of positive characteristic. We give a sufficient criterion for the Hasse principle to hold (see th. 2.1.), and also give counter–examples. These counter–examples are of a different nature than those for number fields: indeed, if $k$ is a global field of positive characteristic, then the Hasse principle does hold for $G$–quadratic forms on projective $k[G]$–modules (see cor. 2.3), and in particular if $k[G]$ is semi–simple, then the Hasse principle is true for $G$–quadratic forms, contrarily to what happens in the case of number fields. On the other hand, there are counter–examples in the non semi–simple case, as shown in §3. Note that the Hasse principle holds in all generality for $G$–trace forms (cf. [BPS 13]).

The third named author is partially supported by National Science Foundation grant DMS-1001872.

§1. Definitions, notation and basic facts

Let $k$ be a field of characteristic $\neq 2$. All modules are supposed to be left modules.
Let $G$ be a finite group, and let $k[G]$ be the associated group ring. A $G$–quadratic space is a pair $(V, q)$, where $V$ is a $k[G]$–module that is a finite dimensional $k$–vector space, and $q : V \times V \to k$ is a non–degenerate symmetric bilinear form such that
\[ q(gx, gy) = q(x, y) \]
for all $x, y \in V$ and all $g \in G$.

Two $G$–quadratic spaces $(V, q)$ and $(V', q')$ are isomorphic if there exists an isomorphism of $k[G]$–modules $f : V \to V'$ such that $q'(f(x), f(y)) = q(x, y)$ for all $x, y \in V$. If this is the case, we write $(V, q) \cong_G (V', q')$, or simply $q \cong q'$.

**Hermitian forms**

Let $R$ be a ring endowed with an involution $r \mapsto \sigma(r)$. For any $R$–module $M$, we denote by $M^*$ its dual $\text{Hom}_R(M, R)$. Then $M^*$ has an $R$–module structure given by $(rf)(x) = f(x)\sigma(r)$ for all $r \in R$, $x \in M$ and $f \in M^*$. If $M$ and $N$ are two $R$–modules and if $f : M \to N$ is a homomorphism of $R$–modules, then $f$ induces a homomorphism $f^* : N^* \to M^*$ defined by $f^*(g) = gf$ for all $g \in N^*$, called the adjoint of $f$.

A hermitian form is a pair $(M, h)$ where $M$ is an $R$–module and $h : M \times M \to R$ is biadditive, satisfying the following two conditions:

1. $h(rx, sy) = rh(x, y)\sigma(r)$ and $h(x, y) = h(y, x)$ for all $x, y \in M$ and all $r, s \in R$.
2. The homomorphism $h : M \to M^*$ given by $y \mapsto h(\cdot, y)$ is an isomorphism.

Note that the existence of $h$ implies that $M$ is self-dual, i.e. isomorphic to its dual.

If $G$ is a finite group, then the group algebra $R = k[G]$ has a natural $k$–linear involution, characterized by the formula $\overline{g} = g^{-1}$ for every $g \in G$. We have the following dictionary (see for instance [BPS 13, 2.1, Example])

a) $R$–module $M \iff k$–module $M$ with a $k$–linear action of $G$;

b) $R$–dual $M^* \iff k$–dual of $M$, with the contragredient (i.e. dual) action of $G$.

c) hermitian space $(M, h) \iff$ symmetric bilinear form on $M$, which is $G$–invariant and defines an isomorphism of $M$ onto its $k$–dual.

Therefore a hermitian space over $k[G]$ corresponds to a $G$–quadratic space, as defined above.

**Hermitian elements**

Let $E$ be a ring with an involution $\sigma : E \to E$ and put
\[ E^0 = \{ z \in E^\times \mid \sigma(z) = z \}. \]
If \( z \in E^0 \), the map \( h_z : E \times E \to E \) defined by \( h_z(x,y) = x.z.\sigma(y) \) is a hermitian space over \( E \); conversely, every hermitian space over \( E \) with underlying module \( E \) is isomorphic to \( h_z \) for some \( z \in E^0 \).

Define an equivalence relation on \( E^0 \) by setting \( z \equiv z' \) if there exists \( e \in E^\times \) with \( z' = \sigma(e)ze \); this is equivalent to \((E,h_z) \simeq (E,h_{z'})\). Let \( H(E,\sigma) \) be the quotient of \( E^0 \) by this equivalence relation. If \( z \in E^0 \), we denote by \([z]\) its class in \( H(E,\sigma)\).

**Classifying hermitian spaces via hermitian elements**

Let \((M,h_0)\) be a hermitian space over \( R \). Set \( E_M = \text{End}(M) \). Let \( \tau : E_M \to E_M \) be the involution of \( E_M \) induced by \( h_0 \), i.e.

\[
\tau(e) = h_0^{-1} e^* h_0, \quad \text{for } e \in E_M,
\]

where \( e^* \) is the adjoint of \( e \). If \((M,h)\) is a hermitian space (with the same underlying module \( M \)), we have \( \tau(h_0^{-1}h) = h_0^{-1}(h_0^{-1}h)^* h_0 = h_0^{-1}h^* (h_0^{-1})^* h_0 = h_0^{-1}h \). Hence \( h_0^{-1}h \) is a hermitian element of \((E_M,\tau)\); let \([h_0^{-1}h]\) be its class in \( H(E_M,\tau)\).

**Lemma 1.1.** (see for instance [BPS 13, lemma 3.8.1]) Sending a hermitian space \((M,h)\) to the element \([h_0^{-1}h]\) of \( H(E_M,\tau) \) induces a bijection between the set of isomorphism classes of hermitian spaces \((M,h)\) and the set \( H(E_M,\tau) \).

**Components of algebras with involution**

Let \( A \) be a finite dimensional \( k \)-algebra, and let \( \iota : A \to A \) be a \( k \)-linear involution. Let \( R_A \) be the radical of \( A \). Then \( A/R_A \) is a semi-simple \( k \)-algebra, hence we have a decomposition \( A/R_A = \prod_{i=1}^e M_{n_i}(D_i) \), where \( D_1,\ldots,D_e \) are division algebras. Let us denote by \( K_i \) the center of \( D_i \), and let \( D_i^{op} \) be the opposite algebra of \( D_i \).

Note that \( \iota(R_A) = R_A \), hence \( \iota \) induces an involution \( \iota : A/R_A \to A/R_A \). Therefore \( A/R_A \) decomposes into a product of involution invariant factors. These can be of two types : either an involution invariant matrix algebra \( M_{n_i}(D_i) \), or a product \( M_{n_i}(D_i) \times M_{n_i}(D_i^{op}) \), with \( M_{n_i}(D_i) \) and \( M_{n_i}(D_i^{op}) \) exchanged by the involution. We say that a factor is unitary if the restriction of the involution to its center is the not the identity : in other words, either an involution invariant \( M_{n_i}(D_i) \) with \( \iota|K_i \neq \text{the identity} \), or a product \( M_{n_i}(D_i) \times M_{n_i}(D_i^{op}) \). Otherwise, the factor is said to be of the first kind. In this case, the component is of the form \( M_{n_i}(D_i) \) and the restriction of \( \iota \) to \( K_i \) is the identity. We say that the component is orthogonal if after base change to a separable closure \( \iota \) is given by the transposition, and symplectic otherwise. A component \( M_{n_i}(D_i) \) is said to be split if \( D_i \) is a commutative field.

**Completions**
If $k$ is a global field and if $v$ is a place of $k$, we denote by $k_v$ the completion of $k$ at $v$. For any $k$–algebra $E$, set $E_v = E \otimes_k k_v$. If $K/k$ is a field extension of finite degree and if $w$ is a place of $K$ above $v$, then we use the notation $w/v$.

§2. Hasse principle

In this section, $k$ will be a global field of characteristic $\neq 2$. Let us denote by $\Sigma_k$ the set of all places of $k$. The aim of this section is to give a sufficient criterion for the Hasse principle for $G$–quadratic forms to hold. All modules are left modules, and finite dimensional $k$–vector spaces.

Theorem 2.1. Let $V$ be a $k[G]$–module, and let $E = \text{End}(V)$. Let $R_E$ be the radical of $E$, and set $\overline{E} = E/R_E$. Suppose that all the orthogonal components of $\overline{E}$ are split, and let $(V, q), (V, q')$ be two $G$–forms. Then $q \simeq_G q'$ over $k$ if and only if $q \simeq_G q'$ over all the completions of $k$.

This is announced in [BP 13], and replaces th. 3.5 of [BP 11]. The proof of th. 2.1 relies on the following proposition

Proposition 2.2. Let $E$ be a finite dimensional $k$–algebra endowed with a $k$–linear involution $\sigma : E \to E$. Let $R_E$ be the radical of $E$, and set $\overline{E} = E/R_E$. Suppose that all the orthogonal components of $\overline{E}$ are split. Then the canonical map $H(E, \sigma) \to \prod_{v \in \Sigma_k} H(E_v, \sigma_v)$ is injective.

Proof. The case of a simple algebra. Suppose first that $E$ is a simple $k$–algebra. Let $K$ be the center of $E$, and let $F$ be the fixed field of $\sigma$ in $K$. Let $\Sigma_F$ denote the set of all places of $F$. For all $v \in \Sigma_k$, set $E_v = E \otimes_k k_v$, and note that $E_v = \prod_{w|v} E_w$, therefore $\prod_{v \in \Sigma_k} H(E_v, \sigma_v) = \prod_{w \in \Sigma_F} H(E_w, \sigma_w)$. By definition, $H(E, \sigma)$ is the set of isomorphism classes of one dimensional hermitian forms over $E$. Moreover, if $\sigma$ is orthogonal, then the hypothesis implies that $E$ is split, in other words we have $E \simeq M_n(F)$. Therefore the conditions of [R 11, th. 3.3.1] are fulfilled, hence the Hasse principle holds for hermitian forms over $E$ with respect to $\sigma$. This implies that the canonical map $H(E, \sigma) \to \prod_{v \in \Sigma_k} H(E_v, \sigma_v)$ is injective.

The case of a semi–simple algebra. Suppose now that $E$ is semi–simple. Then

$$E \simeq E_1 \times \ldots \times E_r \times A \times A^{\text{opp}},$$

where $E_1, \ldots, E_r$ are simple algebras which are stable under the involution $\sigma$, and where the restriction of $\sigma$ to $A \times A^{\text{opp}}$ exchanges the two factors. Applying [BPS 13, lemmas 3.7.1 and 3.7.2] we are reduced to the case where $E$ is a simple algebra, and we already know that the result is true in this case.

Documenta Mathematica 18 (2013) 383–392
General case. We have $\overline{E} = E/R_E$. Then $\overline{E}$ is semi–simple, and $\sigma$ induces a $k$–linear involution $\overline{\sigma}: \overline{E} \to \overline{E}$. We have the following commutative diagram

\[
\begin{array}{ccc}
H(E,\sigma) & \xrightarrow{f} & \prod_{v \in \Sigma_k} H(E_v,\sigma) \\
\downarrow & & \downarrow \\
H(E,\overline{\sigma}) & \xrightarrow{\overline{f}} & \prod_{v \in \Sigma_k} H(E_v,\overline{\sigma}),
\end{array}
\]

where the vertical maps are induced by the projection $E \to \overline{E}$. By [BPS 13, lemma 3.7.3], these maps are bijective. As $\overline{E}$ is semi–simple, the map $\overline{f}$ is injective, hence $f$ is also injective. This concludes the proof.

**Proof of th. 2.1.** It is clear that if $q \simeq_G q'$ over $k$, then $q \simeq_G q'$ over all the completions of $k$. Let us prove the converse. Let $(V, h)$ be the $k[G]$–hermitian space corresponding to $(V, q)$, and let $\sigma : E \to E$ be the involution induced by $(V, h)$ as in §1. Let $(V, h')$ be the $k[G]$–hermitian space corresponding to $(V, q')$, and set $u = h^{-1}h'$. Then $u \in E^p$, and by lemma 1.1, the element $[u] \in H(E, \sigma)$ determines the isomorphism class of $(V, q')$; in other words, we have $q \simeq_G q'$ if and only if $[u] = [1]$ in $H(E, \sigma)$. Hence the theorem is a consequence of proposition 2.2.

**Corollary 2.3** Suppose that $\text{char}(k) = p > 0$, and let $V$ be a projective $k[G]$–module. Let $(V, q)$, $(V, q')$ be two $G$–forms. Then $q \simeq_G q'$ over $k$ if and only if $q \simeq_G q'$ over all the completions of $k$.

**Proof.** Since $V$ is projective, there exists a $k[G]$–module $W$ and $n \in N$ such that $V \oplus W \simeq k[G]^n$. The endomorphism ring of $k[G]^n$ is $M_n(k[G])$, and as $\text{char}(k) = p > 0$, we have $k[G] = F_p[G] \otimes_{F_p} k$. Hence $M_n(k[G])$ is isomorphic to $M_n(F_p[G]) \otimes_{F_p} k$. Let $E = \text{End}(V)$, let $R_E$ be the radical of $E$, and let $\overline{E} = E/R_E$. Let us show that all the components of $\overline{E}$ are split. Let $e$ be the idempotent endomorphism of $V \oplus W$ which is the identity of $V$. Set $\Lambda = \text{End}(V \oplus W)$ and let $R_\Lambda$ be the radical of $\Lambda$. Then $eR_\Lambda e = E$ and $eR_\Lambda e = R_E$. Set $\overline{\Lambda} = \Lambda/R_\Lambda$, and and let $\overline{\sigma}$ be the image of $e$ in $\overline{\Lambda}$. Set $k[\overline{\Gamma}] = k[G]/(\text{rad}(k[G])$. Then we have $\overline{E} \cong \overline{\sigma}\overline{\Lambda}\overline{\sigma} \cong \overline{\Lambda} M_n(k[\overline{\Gamma}]) \overline{\sigma}$. This implies that $\overline{E}$ is a component of the semi–simple algebra $M_n(k[\overline{\Gamma}])$. Let us show that all the components of $M_n(k[\overline{\Gamma}])$ are split. As $F_p$ is a finite field, $F_p[G]/(\text{rad}(F_p[G]))$ is a product of matrix algebras over finite fields. Moreover, for any finite field $F$ of characteristic $p$, the tensor product $F \otimes_{F_p} k$ is a product of fields. This shows that $(F_p[G]/(\text{rad}(F_p[G]))) \otimes_{F_p} k$ is a product of matrix algebras over finite extensions of $k$; in particular, it is semi–simple. The natural isomorphism $F_p[G] \otimes_{F_p} k \to k[G]$ induces an isomorphism $[F_p[G]/(\text{rad}(F_p[G]))] \otimes_{F_p} k \to k[G]/(\text{rad}(F_p[G]), k[G])$. Therefore $\text{rad}(F_p[G], k[G])$ is the radical of $k[G]$, and we have an isomorphism $[F_p[G]/(\text{rad}(F_p[G]))] \otimes_{F_p} k \to k[G]/(\text{rad}(k[G]))$. Hence all the components of $k[G]/(\text{rad}(k[G]))$ are split. This implies that all the components of $\overline{E}$ are split as well. Therefore the corollary follows from th. 2.1.

The following corollary is well–known (see for instance [R 11, 3.3.1 (b)]).
Corollary 2.4 Suppose that $\text{char}(k) = p > 0$, and that the order of $G$ is prime to $p$. Then two $G$–quadratic forms are isomorphic over $k$ if and only if they become isomorphic over all the completions of $k$.

Proof. This follows immediately from cor. 2.3.

§3. Counter–examples to the Hasse principle

Let $k$ be a field of characteristic $p > 0$, let $C_p$ be the cyclic group of order $p$, and let $G = C_p \times C_p \times C_p$. In this section we give counter–examples to the Hasse principle for $G \times G$–quadratic forms over $k$ in the case where $k$ is a global field. We start with some constructions that are valid for any field of positive characteristic.

3.1 A construction

Let $D$ be a division algebra over $k$. It is well–known that there exist indecomposable $k[G]$–modules such that their endomorphism ring modulo the radical is isomorphic to $D$. We recall here such a construction, brought to our attention by R. Guralnick, in order to use it in 3.2 in the case of quaternion algebras.

The algebra $D$ can be generated by two elements (see for instance [J 64, Chapter VII, §12, th. 3, p. 182]). Let us choose $i, j \in D$ be two such elements. Let us denote by $D^{op}$ the opposite algebra of $D$, and let $d$ be the degree of $D$. Then we have $D \otimes_k D^{op} \simeq M_d(k)$. Let us choose an isomorphism $f : D \otimes_k D^{op} \simeq M_d(k)$, and set $a_1 = f(1 \otimes 1) = 1$, $a_2 = f(i \otimes 1)$ and $a_3 = f(j \otimes 1)$.

Let $g_1, g_2, g_3 \in G$ be three elements of order $p$ such that the set $\{g_1, g_2, g_3\}$ generates $G$, and let us define a representation $G \to \text{GL}_{2d^2}(k)$ by sending $g_m$ to the matrix

$$
\begin{pmatrix}
I & a_m \\
0 & I
\end{pmatrix}
$$

for all $m = 1, 2, 3$. Note that this is well–defined because $\text{char}(k) = p$. This endows $k^{2d^2}$ with a structure of $k[G]$–module. Let us denote by $N$ this $k[G]$–module, and let $E_N$ be its endomorphism ring. Then

$$
E_N = \left\{ \begin{pmatrix} x & y \\ 0 & x \end{pmatrix} \mid x \in D^{op} \subset M_{d^2}(k), \; y \in M_{d^2}(k) \right\},
$$

and its radical is

$$
R_N = \left\{ \begin{pmatrix} 0 & y \\ 0 & 0 \end{pmatrix} \mid y \in M_{d^2}(k) \right\},
$$

hence $E_N/R_N \simeq D^{op}$. 

Documenta Mathematica 18 (2013) 383–392
3.2. The case of a quaternion algebra

Let $H$ be a quaternion algebra over $k$. Then by 3.1, we get a $k[G]$-module $N = NH$ with endomorphism ring $E_N$ such that $E_N/R_N \simeq H^{op}$, where $R_N$ is the radical of $E_N$. We now construct a $G$-quadratic form $q$ over $N$ in such a way that the involution it induces on $E_N/R_N \simeq H^{op}$ is the canonical involution.

Let $i, j \in H$ such that $i^2, j^2 \in k^\times$ and that $ij = -ji$. Let $\tau : H \to H$ be the orthogonal involution of $H$ obtained by composing the canonical involution of $H$ with $\text{Int}(ij).$ Let $\sigma : H^{op} \to H^{op}$ be the canonical involution of $H^{op}$. Let us consider the tensor product of algebras with involution

$$(H, \tau) \otimes (H^{op}, \sigma) = (M_4(k), \rho).$$

Then $\rho$ is a symplectic involution of $M_4(k)$ satisfying $\rho(a_m) = a_m$ for all $m = 1, 2, 3$, since $\tau(i) = (ij)(-i)(ij)^{-1} = i$, $\tau(j) = (ij)(-j)(ij)^{-1} = j$. Let $\alpha \in M_4(k)$ be a skew-symmetric matrix such that for all $x \in M_4(k)$, we have

$$\rho(x) = \alpha^{-1}x^T \alpha,$$

where $x^T$ denotes the transpose of $x$. Set $A = \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix}$.

Then $A^T = A$. Let $q : N \times N \to k$ be the symmetric bilinear form defined by $A$:

$$q(v, w) = v^T Aw$$

for all $v, w \in N$. Let $\gamma : M_8(k) \to M_8(k)$ be the involution adjoint to $q$, that is

$$\gamma(X) = A^{-1}X^T A$$

for all $X \in M_8(k)$, i.e. $q(fv, w) = q(v, \gamma(f)w)$ for all $f \in M_8(k)$ and all $v, w \in N$. The involution $\gamma$ restricts to an involution of $E_N$, as for all $x, y \in M_4(k)$, we have

$$\gamma \begin{pmatrix} x & y \\ 0 & x \end{pmatrix} = \begin{pmatrix} \alpha^{-1}x^T \alpha & -\alpha^{-1}y^T \alpha \\ 0 & \alpha^{-1}x^T \alpha \end{pmatrix}.$$

It also sends $R_N$ to itself, and induces an involution $\overline{\gamma}$ on $H^{op} \simeq E_N/R_N$ that coincides with the canonical involution of $H^{op}$.

We claim that $q : N \times N \to k$ is a $G$-quadratic form. To check this, it suffices to show that $q(g_m v, g_m w) = q(v, w)$ for all $v, w \in N$ and for all $m = 1, 2, 3$. Since $\rho(a_m) = a_m$ for all $m = 1, 2, 3$, we have

$$\gamma \begin{pmatrix} I & a_m \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & a_m \\ 0 & I \end{pmatrix}^{-1}$$

and hence

$$q(g_m v, g_m w) = q(v, \gamma(g_m)g_m w) = q(v, w)$$

for all $m = 1, 2, 3$ and all $v, w \in N$. Thus $q$ is a $G$-quadratic form, and by construction, the involution of $E_N$ induced by $q$ is the restriction of $\gamma$ to $E_N$. 

Documenta Mathematica 18 (2013) 383–392
3.3. Two quaternion algebras

Let \( H_1 \) and \( H_2 \) be two quaternion algebras over \( k \). By the construction of 3.2, we obtain two indecomposable \( k[G] \)-modules \( N_1 \) and \( N_2 \). Set \( E_1 = E_{N_1} \) and \( E_2 = E_{N_2} \). Let \( R_i \) be the radical of \( E_i \) for \( i = 1, 2 \), and set \( \overline{E}_i = E_i/R_i \). We also obtain \( G \)-quadratic spaces \( q_i : N_i \times N_i \to k \) inducing involutions \( \gamma_i : E_i \to E_i \) such that the involutions \( \overline{\gamma}_i : \overline{E}_i \to \overline{E}_i \) coincide with the canonical involution of \( H_{i, op} \), for all \( i = 1, 2 \).

Let us consider the tensor product \( (N, q) = (N_1, q_1) \otimes_k (N_2, q_2) \). Then \( (N, q) \) is a \( G \times G \)-quadratic space. Set \( E = \text{End}_{k[G \times G]}(N_1 \otimes N_2) \). Then \( E \simeq E_1 \otimes E_2 \). Let \( I \) be the ideal of \( E \) generated by \( R_1 \) and \( R_2 \). Then there is a natural isomorphism \( f : E_1 \otimes E_2 \to E \) with \( f(I) = R_E \), where \( R_E \) is the radical of \( E \). Set \( \overline{E} = E/R_E \). Then \( \overline{E} \simeq \overline{E}_1 \otimes \overline{E}_2 \simeq H_{1, op} \otimes H_{2, op} \).

Set \( \gamma = \gamma_1 \otimes \gamma_2 \). Then \( \gamma : E \to E \) is the involution induced by the \( G \times G \)-quadratic space \((N, q)\). We obtain an involution \( \overline{\gamma} : \overline{E} \to \overline{E} \), and \( \overline{\gamma} = \overline{\gamma}_1 \otimes \overline{\gamma}_2 \).

Let us recall that \( \overline{E}_i = H_{i, op} \) for \( i = 1, 2 \), and that \( \overline{\gamma}_i \) is the canonical involution of \( H_{i, op} \). Hence \( \overline{\gamma} : \overline{E} \to \overline{E} \) is an orthogonal involution.

3.4. A counter-example to the Hasse principle

Suppose now that \( k \) is a global field of characteristic \( p \), with \( p > 2 \), and suppose that \( H_i \) is ramified at exactly two places \( v_1, v'_1 \) of \( k \), such that \( v_1, v'_1, v_2, v'_2 \) are all distinct. We have \( H_{1, op} \otimes H_{2, op} \simeq M_2(Q) \) where \( Q \) is a quaternion division algebra over \( k \), and \( Q \) is ramified exactly at the places \( v_1, v'_1, v_2, v'_2 \) of \( k \). Recall that the involution \( \overline{\gamma} : M_2(Q) \to M_2(Q) \) is the tensor product of the canonical involutions of \( H_{i, op} \). In particular, \( \overline{\gamma} \) is of orthogonal type. Note that at all \( v \in \Sigma_k \), one of the algebras \( H_{1, op} \) or \( H_{2, op} \) is split. This implies that at all \( v \in \Sigma_k \), the involution \( \overline{\gamma} \) is hyperbolic.

Let \( \delta : Q \to Q \) be an orthogonal involution of the division algebra \( Q \). Then \( \overline{\gamma} \) is induced by some hermitian space \( h : Q^2 \times Q^2 \to Q \) with respect to the involution \( \delta \). As for all \( v \in \Sigma_k \), the involution \( \overline{\gamma} \) is hyperbolic at \( v \), the hermitian form \( h \) is also hyperbolic at \( v \). By Lemma 1.1 the set of isomorphism classes of hermitian spaces on \( Q^2 \) is in bijection with the set \( H(\overline{E}, \overline{\gamma}) \), the hermitian space \((Q^2, h)\) corresponding to the element \([1] \in H(\overline{E}, \overline{\gamma})\).

Let \((Q^2, h')\) be a hermitian space which becomes isomorphic to \((Q^2, h)\) over \( Q_v \), for all \( v \in \Sigma_k \), but is not isomorphic to \((Q^2, h)\) over \( Q \) (this is possible by [Sch 85, 10.4.6]). Let \( u \in \overline{E} \), such that \([u] \in H(\overline{E}, \overline{\gamma})\) corresponds to \((Q^2, h')\) by the bijection of lemma 1.1. Then \([u] \neq [1] \in H(\overline{E}, \overline{\gamma})\), and the images of \([u] \) and \([1] \) coincide in \( \prod_{v \in \Sigma_k} H(\overline{E}_v, \overline{\gamma}) \).

Recall that \( H(E, \gamma) \) is in bijection with the isomorphism classes of \((G \times G)\)-quadratic forms over \( N \), the element \([1] \in H(E, \gamma) \) corresponding to the isomorphism class of \((N, q)\). Let \( \pi : E \to \overline{E} \) be the projection, and let \( \tilde{u} \in E^0 \) be...
such that $\pi(\tilde{u}) = u$ (cf. lemma 1.1). Let $(N, q')$ be a $(G \times G)$–quadratic form corresponding to $\tilde{u}$. The diagram

$$
\begin{align*}
H(E, \gamma) & \xrightarrow{f} \prod_{v \in \Sigma_k} H(E_v, \gamma) \\
\downarrow & \downarrow \\
H(E', \tau) & \xrightarrow{f'} \prod_{v \in \Sigma_k} H(E'_v, \tau),
\end{align*}
$$

is commutative, and the vertical maps are bijective by [BPS 13, lemma 3.7.3]. Hence $(N, q)$ and $(N, q')$ are become isomorphic over all the completions of $k$, but are not isomorphic over $k$.

Bibliography


Documenta Mathematica 18 (2013) 383–392
ERRATUM FOR
“GALOIS ALGEBRAS, HASSE PRINCIPLE, AND INDUCTION–RESTRICTION METHODS”
CP. DOCUMENTA MATH. 16 (2011), 677–707

E. BAYER–FLUCKIGER AND R. PARIMALA

Revised: February 17, 2013
Communicated by Ulf Rehmann


Theorem 3.5 of [BP 11] is not correct as stated, and should be replaced by

THEOREM. Let $V$ be a $k[G]$–module that is a finite dimensional $k$–vector space, and let $E = \text{End}(V)$. Let $R_E$ be the radical of $E$, and set $\overline{E} = E/R_E$. Suppose that all the orthogonal components of $\overline{E}$ are split, and let $(V, q), (V, q')$ be two $G$–forms. Then $q \simeq_G q'$ over $k$ if and only if $q \simeq_G q'$ over all the completions of $k$.

This is proved in [BPN 13], Theorem 2.1. Note however that very few changes are needed in [BP 11]. Indeed, Theorem 3.5 and its proof are correct when $k[G]$ is semi–simple, and this is the only case that is used in the sequel of [BP 11].

Bibliography


Documenta Mathematica 18 (2013) 393–394
On Zagier’s Conjecture for Base Changes of Elliptic Curves

François Brunault

Received: June 15, 2012
Revised: December 21, 2012

Communicated by Stephen Lichtenbaum

Abstract. Let $E$ be an elliptic curve over $\mathbb{Q}$, and let $F$ be a finite abelian extension of $\mathbb{Q}$. Using Beilinson’s theorem on a suitable modular curve, we prove a weak version of Zagier’s conjecture for $L(E_F, 2)$, where $E_F$ is the base change of $E$ to $F$.

2010 Mathematics Subject Classification: 11G40, 11G55, 19F27

Keywords and Phrases: Elliptic curves, $L$-functions, elliptic dilogarithm, Zagier’s conjecture, regulators, Beilinson’s conjecture

Introduction

Zagier conjectured in [19] very deep relations between special values of zeta functions at integers, special values of polylogarithms at algebraic arguments and $K$-theory. While the original conjectures concerned the Dedekind zeta function of a number field and Artin $L$-functions, theoretical and numerical results by many authors suggested an extension of these conjectures to elliptic curves (see [20] for an historical account). A precise formulation for elliptic curves over number fields was given by Wildeshaus in [17]. The conjecture on $L(E, 2)$, where $E$ is an elliptic curve over $\mathbb{Q}$, was proved by Goncharov and Levin in [11]. In this article, we prove an analogue of Goncharov and Levin’s result for the base change of $E$ to an arbitrary abelian number field.

Let $E$ be an elliptic curve defined over $\mathbb{Q}$. Let $F \subset \overline{\mathbb{Q}}$ be a finite abelian extension of $\mathbb{Q}$, and let $E_F$ be the base change of $E$ to $F$. The $L$-function $L(E_F, s)$ admits a factorization $\prod_{\chi \in \hat{G}} L(E \otimes \chi, s)$, where $\hat{G}$ is the group of $\overline{\mathbb{Q}}^*$-valued characters of $G = \text{Gal}(F/\mathbb{Q})$. Each factor $L(E \otimes \chi, s)$ has an analytic continuation to $\mathbb{C}$ with a simple zero at $s = 0$. The functional equation relates $L(E_F, 2)$ with the leading term of $L(E_F, s)$ at $s = 0$. 
Let $D_E$ (resp. $J_E$) be the Bloch elliptic dilogarithm (resp. its “imaginary” cousin) on $E(C) \cong C/(Z + \tau Z)$ (see [28] for the definitions). These functions induce linear maps on the free abelian group $\mathbb{Z}[E(C)]$, and thus on $\mathbb{Z}[E(\overline{Q})]$ after fixing an embedding $\iota : \overline{Q} \to C$. Let $\mathbb{Z}[E(\overline{Q})]^{G_F}$ denote the subgroup of divisors fixed by $G_F := \text{Gal}(\overline{Q}/F)$. It carries a natural action of $G$. The main theorem of this article can be stated as follows.

**Theorem 1.** There exists a divisor $\ell \in \mathbb{Z}[E(\overline{Q})]^{G_F}$ such that for every character $\chi \in \hat{G}$, the following identity holds

$$L'(E \otimes \chi, 0) \sim \chi \left\{ \begin{array}{ll} \frac{1}{\pi} \sum_{\sigma \in G} \chi(\sigma) D_E(\ell^\sigma) & \text{if $\chi$ is even}, \\ \frac{1}{\pi(\tau)} \sum_{\sigma \in G} \chi(\sigma) J_E(\ell^\sigma) & \text{if $\chi$ is odd}. \end{array} \right. \quad (1)$$

Using the Dedekind-Frobenius formula for group determinants, we deduce from Theorem 1 the following result. Let us denote the elements of $G$ by $\sigma_1, \ldots, \sigma_d$ (resp. $\sigma_1, \overline{\sigma_1}, \ldots, \sigma_d, \overline{\sigma_d}$) if $F$ is real (resp. complex), with $d = [F : \mathbb{Q}]$.

**Corollary (Weak version of Zagier’s conjecture for $L(E_F, 2)$).**

Let $\ell \in \mathbb{Z}[E(\overline{Q})]^{G_F}$ be a divisor satisfying the identities (1) of Theorem 1. For any $i$, define $\ell_i = \ell^{\sigma_i}$. If $F$ is real, then we have

$$L(E_F, 2) \sim \chi \pi^d \cdot \det(D_E(\ell_i^{\sigma_i}))_{1 \leq i, j \leq d}. \quad (2)$$

If $F$ is complex, then we have

$$L(E_F, 2) \sim \chi \pi^d \left\{ \begin{array}{ll} \frac{1}{\pi(\tau)} \sum_{\sigma \in G} \chi(\sigma) D_E(\ell^\sigma) & \text{if $\chi$ is even}, \\ \frac{1}{\pi(\tau)} \sum_{\sigma \in G} \chi(\sigma) J_E(\ell^\sigma) & \text{if $\chi$ is odd}. \end{array} \right. \quad (3)$$

**Remarks.**

1. Wildeshaus’s formulation of the conjecture [17] Conjecture, Part 2, p. 366] uses Kronecker double series instead of $D_E$ and $J_E$. The link between these objects is classical (see the proof of Prop. 6).

We have chosen to formulate our results in terms of $D_E$ and $J_E$ because these functions are easier to compute numerically and make apparent the distinction according to the parity of $\chi$.

2. Because of the definition of $\ell_i$, the determinant appearing in (2) is a group determinant, indexed by $G$. In fact, the eigenvalues of the matrix $(D_E(\ell_i^{\sigma_i}))$ are precisely the sums $\sum_{\sigma \in G} \chi(\sigma) D_E(\ell^\sigma)$ appearing in Theorem 1. This is an algebraic counterpart of the factorization of the $L$-value of $E_F$ as a product of twisted $L$-values.

3. The divisor $\ell$ produced by Theorem 1 satisfies Goncharov and Levin’s conditions [11] (2)-(4). Following [28], let $A_{E/F} \subset \mathbb{Z}[E(\overline{Q})]^{G_F}$ be the group of divisors satisfying these conditions. The strong version of Zagier’s conjecture predicts that if $F$ is real (resp. complex), then for any divisors $\ell_1, \ldots, \ell_d \in A_{E/F}$ (resp. $\ell_1, \ldots, \ell_d \in A_{E/F}$), the right-hand side of (2) (resp. (3)) is a rational multiple of $L(E_F, 2)$ (possibly equal to zero). As in the case $F = \mathbb{Q}$, this strong conjecture is beyond the reach of current technology.
In order to prove Theorem 1, we prove a weak version of Beilinson’s conjecture for the special value $L^{(d)}(E_F, 0)$ (see [3] for the definition of the objects involved in the following theorem).

**Theorem 2.** There exists a subspace $\mathcal{P}_{E/F} \subset H^2_{\text{mot}}(E_F, \mathbb{Q}(2))$ such that $R_{E/F} := \text{reg}_{E/F}(\mathcal{P}_{E/F})$ is a $\mathbb{Q}$-structure of $H^1(E_F(\mathbb{C}), \mathbb{R})^-$ and

$$\det(R_{E/F}) = L^{(d)}(E_F, 0) \cdot \det(H^1(E_F(\mathbb{C}), \mathbb{Q})^-).$$ (4)

We prove Theorem 2 by using Beilinson’s theorem on a suitable modular curve. More precisely, we make use of a result of Schappacher and Scholl [15] on the (non geometrically connected) modular curve $X_1(N)_F$, where $N$ is the conductor of $E$. We therefore need to work in the adelic setting. We establish a divisibility statement in the Hecke algebra of $X_1(N)_F$ in order to get the desired result for $E_F$.

The methods used in this article are of inexplicit nature and do not give rise, in general, to explicit divisors. However, Theorem 1 and its corollary can be made explicit in the particular case of the elliptic curve $E = X_1(11)$ and the maximal real subfield $F = \mathbb{Q}(\zeta_{11})^+$ inside the cyclotomic field $\mathbb{Q}(\zeta_{11})$. In fact, we may take for $\ell$ a divisor on the cuspidal subgroup of $E$. The tools for proving this are Kato’s explicit version of Beilinson’s theorem for the modular curve $X_1(N)_{\mathbb{Q}(\zeta_m)}$, the work of the author [3], as well as a technique used by Mellit [13] to get new relations between values of the elliptic dilogarithm. We hope to give soon an expanded account of this example.

The organization of the article is as follows. In §1, we recall well-known facts about $L(E_F, s)$. In §2 and §3, we recall the definition of the regulator map and we compute it for $E_F$ (Prop. 9). In §4, we explain the adelic setting for modular curves. In §5, we prove the divisibility we need in the Hecke algebra (Prop. 16). Finally, we give in §6 the proofs of the main results. We conclude with some remarks and a conjecture in the case $F/\mathbb{Q}$ is not abelian.

**Acknowledgements.** I would like to thank Anton Mellit for the very inspiring discussions which led to the discovery of the example alluded to above, which in turn motivated all the results presented here.

1 The $L$-function of the base change

By the Kronecker-Weber theorem, we have $F \subset \mathbb{Q}(\zeta_m)$ for some $m \geq 1$, so that $G$ is a quotient of $(\mathbb{Z}/m\mathbb{Z})^*$ and $\overline{G}$ can be identified with a subgroup of the Dirichlet characters modulo $m$.

Let $f = \sum_{n \geq 1} a_n q^n \in S_2(\Gamma_0(N))$ be the newform associated to $E$. For any $\chi \in \overline{G}$, define $L(E \otimes \chi, s) := L(f \otimes \chi, s)$, where $f \otimes \chi$ is the unique newform of weight 2 whose $p$-th Fourier coefficient is $a_p \chi(p)$ for every prime $p \nmid Nm$. The $L$-function of $E_F$ has the following description.

\[ \text{Documenta Mathematica 18 (2013) 395–412} \]
Proposition 3. The following identity holds:

\[ L(E_F, s) = \prod_{\chi \in \hat{G}} L(f \otimes \chi, s). \]  

(5)

Proof. Let \( \rho = (\rho_\ell) \) be the compatible system of 2-dimensional \( \ell \)-adic representations of \( G_\mathbb{Q} \) attached to \( f \) by Deligne [5]. By modularity \( L(E_F, s) = L(\rho|_{G_F}, s) \). Using Artin’s formalism for \( L \)-functions, we have

\[ L(\rho|_{G_F}, s) = L(\text{Ind}_{G_\mathbb{Q}}^{G_F}(\rho|_{G_F}), s). \]  

(6)

If \( 1_{G_F} \) denotes the trivial representation of \( G_F \), we have

\[ \text{Ind}_{G_\mathbb{Q}}^{G_F}(\rho|_{G_F}) \cong \text{Ind}_{G_\mathbb{Q}}^{G_F}(1_{G_F}) \otimes \rho \]  

(7)

\[ \cong \bigoplus_{\chi \in \hat{G}} \rho \otimes \chi. \]  

(Here we chose embeddings \( \overline{\mathbb{Q}} \to \overline{\mathbb{Q}_\ell} \).) Finally, since an irreducible \( \ell \)-adic representation of \( G_\mathbb{Q} \) is determined by the traces of all but finitely many Frobenius elements, the compatible system associated to \( f \otimes \chi \) is \( \rho \otimes \chi \), so that

\[ L(\rho \otimes \chi, s) = L(f \otimes \chi, s) \]  

for any \( \chi \in \hat{G} \).

Proposition 4. We have

\[ L(E_F, 2) \sim_{\mathbb{Q}^*} \pi^{2d} L^{(d)}(E_F, 0), \]  

where \( L^{(d)}(E_F, 0) \) denotes the \( d \)-th derivative at \( s = 0 \).

Proof. Since each \( L(f \otimes \chi, s) \) has a simple zero at \( s = 0 \), we get

\[ \frac{L^{(d)}(E_F, 0)}{d!} = \prod_{\chi \in \hat{G}} L'(f \otimes \chi, 0), \]  

(8)

Let \( N_{f \otimes \chi} \) be the level of the newform \( f \otimes \chi \). Putting \( \Lambda(f \otimes \chi, s) = N_{f \otimes \chi}^{s/2}(2\pi)^{-s} \Gamma(s)L(f \otimes \chi, s) \), we have [7, §5]

\[ \Lambda(f \otimes \chi, s) = -w_{f \otimes \chi} \Lambda(f \otimes \overline{\chi}, 2 - s) \quad (s \in \mathbb{C}) \]  

(9)

where \( w_{f \otimes \chi} \) is the pseudo-eigenvalue of \( f \otimes \chi \) with respect to the Atkin-Lehner involution of level \( N_{f \otimes \chi} \). Note that [10] implies \( w_{f \otimes \chi} w_{f \otimes \overline{\chi}} = 1 \). Letting \( w = \prod_{\chi \in \hat{G}} w_{f \otimes \chi} \), we have

\[ w^2 = \prod_{\chi \in \hat{G}} w_{f \otimes \chi} w_{f \otimes \overline{\chi}} = 1 \]  

(10)

so that \( w = \pm 1 \). Moreover \( \Lambda(f \otimes \chi, 0) = L'(f \otimes \chi, 0) \) and \( \Lambda(f \otimes \overline{\chi}, 2) = (N_{f \otimes \chi}/4\pi^2)L(f \otimes \overline{\chi}, 2) \). Taking the product over \( \chi \) yields the result. \( \square \)
2. The regulator map on Riemann surfaces

In this section, we recall the definition of the regulator map on compact Riemann surfaces [5, §1], and its computation in the case of elliptic curves. Let $X$ be a compact connected Riemann surface, and $\mathcal{M}(X)$ be its field of meromorphic functions. For any $f, g \in \mathcal{M}(X)^*$, consider the 1-form

$$\eta(f, g) := \log |f| \cdot \text{darg}(g) - \log |g| \cdot \text{darg}(f).$$

(11)

For any $f \in \mathcal{M}(X) \setminus \{0, 1\}$, the differential form $\eta(f, 1 - f)$ is exact on $X \setminus f^{-1}(\{0, 1, \infty\})$. More precisely $\eta(f, 1 - f) = d(D \circ f)$, where $D$ is the Bloch-Wigner dilogarithm function [13]. Let $K_2(\mathcal{M}(X))$ be the Milnor $K_2$-group associated to $\mathcal{M}(X)$. The regulator map on $X$ is the unique linear map

$$\text{reg}_X : K_2(\mathcal{M}(X)) \to H^1(X, \mathbb{R})$$

(12)

such that for any $f, g \in \mathcal{M}(X)^*$ and any holomorphic 1-form $\omega$ on $X$, we have

$$\int_X \text{reg}_X \{f, g\} \wedge \omega = \frac{1}{2\pi} \int_X \eta(f, g) \wedge \omega.$$

(13)

The map $\text{reg}_X$ is well-defined by exactness of $\eta(f, 1 - f)$ and Stokes’ theorem. The construction of $\text{reg}_X$ easily extends to the case where $X$ is compact but not connected. Indeed, put $\mathcal{M}(X) := \prod_{i=1}^r \mathcal{M}(X_i)$ where $X_1, \ldots, X_r$ are the connected components of $X$. Then $K_2(\mathcal{M}(X)) \cong \bigoplus_{i=1}^r K_2(\mathcal{M}(X_i))$ as well as $H^1(X, \mathbb{R}) \cong \bigoplus_i H^1(X_i, \mathbb{R})$, and we define $\text{reg}_X$ to be the direct sum of the maps $\text{reg}_{X_i}$ for $1 \leq i \leq r$.

Let us recall the classical computation of the regulator map on a complex torus [11, §4]. Let $E_\tau := \mathbb{C}/(\mathbb{Z} + \tau \mathbb{Z})$ with $\tau \in \mathbb{C}$, $\text{Im}(\tau) > 0$. The map $z \mapsto \exp(2\pi i z)$ induces an isomorphism $E_\tau \cong \mathbb{C}^*/q^\mathbb{Z}$, where $q := \exp(2\pi i \tau)$. Let $D_q : E_\tau \to \mathbb{R}$ be the Bloch elliptic dilogarithm, defined by $D_q([x]) = \sum_{n=-\infty}^{\infty} D(q^n x)$ for any $x \in \mathbb{C}^\times$. We will also use the function $J_q : E_\tau \to \mathbb{R}$, which is defined as follows. Let $J : \mathbb{C}^* \to \mathbb{R}$ be the function defined by $J(x) = \log |x| \cdot \log |1 - x|$ if $x \neq 1$, and $J(1) = 0$. Following [13], we put

$$J_q([x]) = \sum_{n=0}^{\infty} J(q^n x^n) - \sum_{n=1}^{\infty} J(x^{-1} q^n) + \frac{1}{3} \log^2 |q| \cdot B_3 \left( \frac{|x|}{\log |q|} \right) \quad (x \in \mathbb{C}^*)$$

(14)

where $B_3 = x^3 - \frac{1}{2} x^2 + \frac{1}{12}$ is the third Bernoulli polynomial. The function $J_q$ is well-defined since $J(x) + J(\frac{1}{x}) = \log |x|$ and $B_3(x + 1) - B_3(x) = 3x^2$. Both functions $D_q$ and $J_q$ extend to linear maps $\mathbb{Z}[E_\tau] \to \mathbb{R}$, by setting $D_q(\sum_i n_i[P_i]) := \sum_i n_i D_q(P_i)$ and similarly for $J_q$.

**Definition 5.** For any $f, g \in \mathcal{M}(E_\tau)^*$ with divisors $\text{div}(f) = \sum_i m_i[P_i]$ and $\text{div}(g) = \sum_i n_j[Q_j]$, the divisor $\beta(f, g) \in \mathbb{Z}[E_\tau]$ is defined by

$$\beta(f, g) = \sum_{i,j} m_i n_j [P_i - Q_j].$$

(15)
The following classical result expresses the regulator map on $E_\tau$ in terms of $D_q$ and $J_q$.

**Proposition 6.** For any $f, g \in \mathcal{M}(E_\tau)^\times$, we have

\[
\int_{E_\tau} \eta(f, g) \wedge dz = (D_q - iJ_q)(\beta(f, g)).
\] (16)

**Proof.** We have $\int_{E_\tau} \eta(f, g) \wedge dz = -\frac{2(\tau)^2}{\pi}K_{2,1,\tau}(\beta(f, g))$ by [12 §4.3] and [6 (6.2)], where $K_{2,1,\tau}$ is the linear extension of the following Eisenstein-Kronecker series on $E_\tau$:

\[
K_{2,1,\tau}(z) := \sum_{\lambda \in \mathbb{Z}^+ \tau \mathbb{Z}} \exp\left(\frac{2i\pi}{\lambda}(z^\tau - \bar{z^\tau})\right) / \lambda^2 (z \in \mathbb{C}/(\mathbb{Z}^+ \tau \mathbb{Z})).
\] (17)

The result now follows from the formula $-\frac{2(\tau)^2}{\pi}K_{2,1,\tau} = D_q - iJ_q$, for which we refer to [2, Thm 10.2.1] and [15 §2, p. 616].

3. The Regulator Map on the Base Change

Let $X$ be a connected (but not necessarily geometrically connected) smooth projective curve over $\mathbb{Q}$. Its function field $\mathbb{Q}(X)$ embeds into $\mathcal{M}(X(\mathbb{C}))$, so we get a natural map $K_2(\mathbb{Q}(X)) \to K_2(\mathcal{M}(X(\mathbb{C})))$. Let $c$ denote the complex conjugation on $X(\mathbb{C})$. For any $f, g \in \mathbb{Q}(X)^\times$, we have $c^* \eta(f, g) = -\eta(f, g)$, so that [12] induces a map

\[
K_2(\mathbb{Q}(X)) \to H^3(X(\mathbb{C}), \mathbb{R})^-, \quad (18)
\]

where $(\cdot)^-$ denotes the $(-1)$-eigenspace of $c^*$.

Let $K_2(X)$ be the Quillen algebraic $K_2$-group associated to $X$. Recall that the motivic cohomology group $H^2_M(X, \mathbb{Q}(2)) := K_2^{(2)}(X)$ is defined as the second Adams eigenspace of $K_2(X) \otimes \mathbb{Q}$. The exact localization sequence in $K$-theory yields a canonical injective map $K_2(X) \otimes \mathbb{Q} \to K_2(\mathbb{Q}(X)) \otimes \mathbb{Q}$ which is compatible with the Adams operations, so that in fact $K_2^{(2)}(X) = K_2(X) \otimes \mathbb{Q}$.

The integral subspace $H^2_M(X, \mathbb{Q}(2)) \subset H^2_M(X, \mathbb{Q}(2))$ is the image of the map $K_2(X) \otimes \mathbb{Q} \to K_2(X) \otimes \mathbb{Q}$ for any proper regular model $X/\mathbb{Z}$ of $X$ (see [10] for a definition in a more general setting). Tensoring [15] with $\mathbb{Q}$ and restricting to the integral subspace gives the Beilinson regulator map on $X$:

\[
\text{reg}_X : H^2_M(Z(X, \mathbb{Q}(2))) \to H^1(X(\mathbb{C}), \mathbb{R})^-.
\] (19)

Note that the real vector space $H^1(X(\mathbb{C}), \mathbb{R})^-$ admits the natural $\mathbb{Q}$-structure $H_X := H^1(X(\mathbb{C}), \mathbb{Q})^-$.

Any finite morphism $\varphi : X \to Y$ between smooth projective curves over $\mathbb{Q}$ induces maps $\varphi^* : K_2(Y) \to K_2(X)$ and $\varphi_* : K_2(X) \to K_2(Y)$, the latter being
Let us return to our elliptic curve $E$. Fix an isomorphism $E(\mathbb{C}) \cong E_\tau$ which is compatible with complex conjugation, and let $q = \exp(2i\pi\tau)$. Let $D_E$ and $J_E$ be the real-valued functions on $E(\mathbb{C})$ induced by $D_q$ and $J_q$, respectively. The space $H^1(E(\mathbb{C}), \mathbb{Q})^\ast$ is generated by the 1-form $\eta^\ast$, with

$$\eta^\ast = dz + d\bar{z} \quad \text{and} \quad \eta^- = \frac{dz - d\bar{z}}{\tau - \bar{\tau}}$$

**Lemma 7.** Let $f, g \in \mathbb{C}(E)^\times$ and $\ell = \beta(f, g)$. We have

$$\text{reg}_{E(\mathbb{C})}(f, g) = \frac{1}{2i\pi} (D_E(\ell) \cdot \eta^\ast + J_E(\ell) \cdot \eta^-).$$

**Proof.** Using (13) with Prop. 6 and identifying the real and imaginary parts gives the lemma.

Let $\Sigma$ be the set of embedding of $F$ into $\mathbb{C}$. We consider $E_F = E \times_{\text{Spec} \mathbb{Q}} \text{Spec} F$ as a scheme over $\text{Spec} \mathbb{Q}$, so that $E_F(\mathbb{C})$ is the disjoint union of $d$ copies of $E(\mathbb{C})$. In particular

$$H^1(E_F(\mathbb{C}), \mathbb{R}) \cong \bigoplus_{\psi \in \Sigma} H^1(E(\mathbb{C}), \mathbb{R})$$

and $H^1(E_F(\mathbb{C}), \mathbb{Q})$ decomposes accordingly. The group $G$ acts from the right on $E_F$. This induces a left action of $G$ on $H^1(E_F(\mathbb{C}), \mathbb{Q})$. For any character $\chi \in \widehat{G}$, consider the idempotent

$$e_\chi := \frac{1}{|G|} \sum_{\sigma \in \widehat{G}} \chi(\sigma) \cdot [\sigma] \in \overline{\mathbb{Q}}[G].$$

It acts on $H^1(E_F(\mathbb{C}), \overline{\mathbb{Q}} \otimes \mathbb{R})$. For any $\psi \in \Sigma$, let $\eta^\ast(\psi)$ be the 1-form $\eta^\ast$ sitting in the $\psi$-component of (23). Note that the embedding $\iota : \overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$ induces a distinguished element $\iota \in \Sigma$. Define

$$\eta_\chi = \begin{cases} e_\chi(\eta^-(\iota)) & \text{if } \chi \text{ is even}, \\ e_\chi(\eta^+(\iota)) & \text{if } \chi \text{ is odd}. \end{cases}$$

The lattice $\mathbb{Z} + i\mathbb{Z}$ is uniquely determined by $E$, and $q$ is a well-defined real number such that $0 < |q| < 1$. But the pair $(D_E, J_E)$ is defined only up to sign (choosing an isomorphism $E(\mathbb{C}) \cong E_\tau$ amounts to specifying an orientation of $E(\mathbb{R})$).

---

*Documenta Mathematica* 18 (2013) 395–412
LEMMA 8. If \( c : E_F(C) \to E_F(C) \) is the map induced by complex conjugation on Spec \( C \), then \( c^* \eta_\chi = -\eta_\chi \).

Proof. For any \( \psi \in \Sigma \), we have \( c^* \eta^\chi(\psi) = \pm \eta^\chi(\overline{\psi}) \). It follows that

\[
c^* \eta_\chi = \frac{1}{|G|} \sum_{\sigma \in G} \chi(\sigma) c^* (\sigma \cdot \eta^\chi(\psi)) = -\chi(1) \frac{1}{|G|} \sum_{\sigma \in G} \chi(\sigma) (\overline{\sigma} \cdot \eta^\chi(\psi)).
\]

Since \( \chi(1) \chi(\sigma) = \chi(\overline{\sigma}) \), we get the result. \( \square \)

The map \( \beta \) induces a linear map \( F(E)^\chi \otimes F(E)^\chi \to \mathbb{Z}[E(\overline{\mathbb{Q}})]^{G_\ell} \), which we still denote by \( \beta \). The following proposition computes explicitly the regulator map associated to \( E_\ell \).

PROPOSITION 9. Let \( \gamma \in F(E)^\chi \otimes F(E)^\chi \) and \( \ell = \beta(\gamma) \). For any \( \chi \in \widehat{G} \), we have \( c_\chi \text{reg}_{E/F}(\langle \gamma \rangle) = \mu_\chi(\ell) \cdot \eta_\chi \), where \( \mu_\chi(\ell) \in \mathbb{Q} \otimes \mathbb{R} \) is given by

\[
\mu_\chi(\ell) = \begin{cases} 
-\frac{1}{2\pi} \sum_{\sigma \in G} \chi(\sigma) \otimes D_E(\ell^a) & \text{if } \chi \text{ is even}, \\
\frac{1}{2\pi} \sum_{\sigma \in G} \chi(\sigma) \otimes J_E(\ell^a) & \text{if } \chi \text{ is odd}.
\end{cases}
\]

Proof. Put \( r = \text{reg}_{E/F}(\langle \gamma \rangle) \). By Lemma \( \square \), the \( \psi \)-component of \( r \) is

\[
r_\psi = -\frac{1}{2\pi} \left( D_E(\psi(\ell)) \cdot \eta^+(\psi) + J_E(\psi(\ell)) \cdot \eta^-(\psi) \right).
\]

Since \( c_\chi(r) \) and \( \eta_\chi \) belong to the same \( G \)-eigenspace, it suffices to compare their \( t \)-components. By definition, we have \( (\eta_\chi)_t = \frac{1}{|G|} \eta^\chi(1) \). Moreover

\[
e_\chi(r) = \frac{1}{|G|} \sum_{\sigma \in G} \chi(\sigma) \otimes (\sigma \cdot t), \quad e_\chi(t) = \frac{1}{|G|} \sum_{\sigma \in G} \chi(\sigma) \otimes r_{t, o a} \quad (28)
\]

\[
= -\frac{1}{2\pi |G|} \sum_{\sigma \in G} \chi(\sigma) \otimes (D_E(\ell^a) \cdot \eta^+ + J_E(\ell^a) \cdot \eta^-). \quad (29)
\]

But \( D_E(\overline{\mathbb{P}}) = D_E(P) \) and \( J_E(\overline{\mathbb{P}}) = -J_E(P) \) for any \( P \in E(C) \), so that the terms involving \( J_E \) (resp. \( D_E \)) cancel out if \( \chi \) is even (resp. odd). \( \square \)

4 Modular curves in the adelic setting

Let \( A_f \) be the ring of finite adèles of \( \mathbb{Q} \). For any compact open subgroup \( K \subset GL_2(A_f) \), there is an associated smooth projective modular curve \( \overline{M}_K \) over \( \mathbb{Q} \). For example \( X(N) = \overline{M}_{K(N)} \) and \( X_1(N) = \overline{M}_{K_1(N)} \), where

\[
K(N) = \ker(GL_2(\overline{\mathbb{Z}}) \to GL_2(\mathbb{Z}/N\mathbb{Z})) \quad (30)
\]

\[
K_1(N) = \{ g \in GL_2(\overline{\mathbb{Z}}); g \equiv \begin{pmatrix} * & 1 \\ 0 & 1 \end{pmatrix} \pmod{N} \}. \quad (31)
\]

Documenta Mathematica 18 (2013) 395–412
The Riemann surface $\overline{M}_K(C)$ can be identified with the compactification of $GL_2(\mathbb{Q})/(\mathbb{A}_f^+ \times GL_2(\mathbb{A}_f))/K$. The set of connected components of $\overline{M}_K(C)$ is in bijection with $\hat{Z}/\det(K)$. For any $g \in GL_2(\mathbb{A}_f)$, we have an isomorphism $g:\overline{M}_K \cong \overline{M}_{g^{-1}Kg}$ over $\mathbb{Q}$, which is given on the complex points by $(\tau, h) \mapsto (\tau, hg)$. For any compact open subgroups $K' \subset K$ of $GL_2(\mathbb{A}_f)$, we have a finite morphism $\pi_{K'K} : \overline{M}_{K'} \to \overline{M}_K$.

The Hecke algebra $\mathcal{H}_K$ is the space of functions $K \backslash GL_2(\mathbb{A}_f)/K \to \overline{Q}$ with finite support, equipped with the convolution product $[\mathbb{H}]$. It acts on $H^1(\mathbb{M}_K(C), \overline{Q})$ and $\Omega^1(\mathbb{M}_K) \otimes \overline{Q}$. Let $T_K = T_{\mathcal{M}_K}$ be the image of $\mathcal{H}_K$ in $\text{End}_{\overline{Q}}(\Omega^1(\mathbb{M}_K) \otimes \overline{Q})$. Let

$$\langle \cdot, \cdot \rangle : H^1(\mathbb{M}_K(C), \mathbb{R})^* \times (\Omega^1(\mathbb{M}_K) \otimes \mathbb{R}) \to \mathbb{R}$$

be the perfect pairing induced by Poincaré duality. For any $T \in \mathcal{H}_K$, we have $\langle T \eta, \omega \rangle = \langle \eta, T' \omega \rangle$, where $T' \in \mathcal{H}_K$ is defined by $T'(g) = T(g^{-1})$, so that the action of $\mathcal{H}_K$ on $H^1(\mathbb{M}_K(C), \overline{Q} \otimes \mathbb{R})^*$ factors through $T_K$.

Following [15] 1.1.1, let $Q_K \subset K_2(\mathbb{M}_K) \otimes \mathbb{Q}$ be the subspace of Beilinson elements, and let

$$\mathcal{P}_K = \bigcup_{K' \subset K} \langle \pi_{K'K} \rangle_* Q_{K'} \subset K_2(\mathbb{M}_K) \otimes \mathbb{Q}. \tag{33}$$

Schappacher and Scholl [15] 1.1.2 proved that $\mathcal{P}_K \subset H^2_{\text{ét}}(\mathbb{M}_K, \mathbb{Q}(2))$ and that $\text{res}_{\mathbb{M}_K} \mathcal{P}_K$ is a $\mathbb{Q}$-structure of $H^1(\mathbb{M}_K(C), \mathbb{R})^*$ whose determinant with respect to the natural $\mathbb{Q}$-structure of $\mathcal{H}_{\mathcal{M}_K}$ is given by the leading term of $L(h^1(\mathbb{M}_K), s)$ at $s = 0$.

In the following, we assume $K = \prod_p K_p$, where $K_p$ a compact open subgroup of $GL_2(\mathbb{Q}_p)$. The Hecke algebra then decomposes as a restricted tensor product $\mathcal{H}_K = \bigotimes_p \mathcal{H}_{K_p}$. For any prime $p$, let $\overline{T}(p) \in \mathcal{H}_K$ (resp. $\widehat{T}(p, p) \in \mathcal{H}_K$) be the characteristic function of $K \left( \begin{smallmatrix} \varpi_p & 0 \\ 0 & 1 \end{smallmatrix} \right) K$ (resp. $K \left( \begin{smallmatrix} \varpi_p & 0 \\ 0 & \varpi_p \end{smallmatrix} \right)$), where $\varpi_p \in \mathbb{A}_f^+$ has component $p$ at the place $p$, and 1 elsewhere. Let $T(p)$ (resp. $\overline{T}(p, p)$) be the image of $\overline{T}(p)$ (resp. $\widehat{T}(p, p)$) in $T_K$. When $K$ needs to be specified, we write $T(p)_K$ or $T(p)_{\mathcal{M}_K}$.

For any integer $M \geq 1$, we let $\mathcal{H}_K^{(M)} \subset \mathcal{H}_K$ be the subalgebra generated by the $\mathcal{H}_{K_p}$ for $p \mid M$. We use the notation $T_K^{(M)}$ for the corresponding subalgebra of $T_K$.

**Lemma 10.** If $K(M) \subset K$ then $T_K^{(M)}$ is in the center of $T_K$.

**Proof.** For any prime $p \mid M$, we have $\mathcal{H}_{K_p} \cong GL_2(\mathbb{Z}_p)$ and by Satake the map $\overline{Q}[T, S, S^{-1}] \to \mathcal{H}_{K_p}$ given by $T \mapsto \overline{T}(p)$ and $S \mapsto \widehat{T}(p, p)$ is an isomorphism. In particular $\mathcal{H}_{K_p}$ is contained in the center of $\mathcal{H}_K$, whence the result.

Let $U_F \subset \hat{Z}$ denote the preimage of $\text{Gal}(\mathbb{Q}(\zeta_m)/F) \subset (\mathbb{Z}/m\mathbb{Z})^\times$ under the natural map $\hat{Z}^\times \to (\mathbb{Z}/m\mathbb{Z})^\times$ (note that $U_F$ does not depend on $m$). For any
compact open subgroup $K \subset \text{GL}_2(A_f)$ with $\det(K) = \hat{\mathbb{Z}}^\times$, let

$$K_F := \{ k \in K; \det(k) \in U_F \}. \quad (34)$$

Let $pr : A_f^\times \to \hat{\mathbb{Z}}^\times$ be the projection associated to the decomposition $A_f^\times \cong Q_{>0} \times \hat{\mathbb{Z}}^\times$.

**Definition 11.** Let $\gamma : \text{GL}_2(A_f) \to G$ be the composite morphism

$$\text{GL}_2(A_f) \xrightarrow{\det} A_f^\times \xrightarrow{pr} \hat{\mathbb{Z}}^\times \to G. \quad (35)$$

Note that there is an exact sequence

$$1 \to K_F \to K \xrightarrow{\gamma |_K} G \to 1. \quad (36)$$

The sequence $(36)$ induces a right action of $G$ on $\overline{M}_{K_F}$, and thus a left action of $G$ on $\Omega^1(\overline{M}_{K_F})$. Moreover, the curve $\overline{M}_K$ can be identified with $\overline{M}_{K_F}$ as a curve over $\mathbb{Q}$, and we have a bijection

$$\overline{M}_{K_F}(C) \xrightarrow{\cong} G \times \overline{M}_K(C) \quad (37)$$

$$[\tau, g] \mapsto (\gamma(g), [\tau, g]).$$

The action of $G$ on $\overline{M}_{K_F}(C)$ corresponds via $(37)$ to the action by translation on the first factor of $G \times \overline{M}_K(C)$.

Now let us consider the case $K = K_1(N)$, so that $\overline{M}_{K_F} \cong X_1(N)_F$. By the previous discussion, the image of $G$ in $\text{End} \Omega^1(X_1(N)_F) \otimes \mathbb{Q}$ is contained in $T_{X_1(N)_F}$, in order to ease notations, let $T = T_{X_1(N)_F} \subset \text{End} \Omega^1(X_1(N)_F) \otimes \mathbb{Q}$. Let $TG$ be the subalgebra of $T_{X_1(N)_F}$ generated by $T$ and $G$.

**Lemma 12.** The algebra $TG$ is commutative.

**Proof.** Note that $K(Nm) \subset K_1(N)_F$, so $T$ is commutative and commutes with $G$ by Lemma 10. Since $G$ is abelian, the result follows. \qed

Since $\Omega^1(X_1(N)_F) \cong \Omega^1(X_1(N)) \otimes F$, we can define the base change morphism $\nu_F : \text{End} \Omega^1(X_1(N)) \to \text{End} \Omega^1(X_1(N)_F)$ by $\nu_F(T) = T \otimes \text{id}_F$. For any $\alpha \in (\mathbb{Z}/m\mathbb{Z})^\times$, let $\sigma_\alpha$ be its image in $G$.

**Lemma 13.** For any prime $p \nmid Nm$, we have

$$\nu_F(T(p)_{X_1(N)}) = T(p)_{X_1(N)_F} \cdot \sigma_p \in TG \quad (38)$$

$$\nu_F(T(p,p)_{X_1(N)}) = T(p,p)_{X_1(N)_F} \cdot \sigma_p^2 \in TG. \quad (39)$$
Proof. Let \( g = \begin{pmatrix} \omega_p & 0 \\ 0 & 1 \end{pmatrix} \) and \( K := K_1(N) \cap g^{-1}K_1(N)g = K_1(N) \cap K_0(p) \). Note that \( \det K = \tilde{Z}^* \). Consider the following correspondence

\[
\begin{array}{c}
\mathcal{M}_K \\
\downarrow \alpha \\
X_1(N) \\
\downarrow \beta \\
\tilde{T}(p)X_1(N) \rightarrow X_1(N)
\end{array}
\]

(40)

where \( \alpha = \pi_{K,K_1(N)} \) and \( \beta = g^{-1} \circ \pi_{K,g^{-1}K_1(N)g} = \pi_{K,g^{-1}K_1(N)g} \circ g^{-1} \). Then \( T(p)X_1(N) = \beta \circ \alpha \) on \( \Omega^1(X_1(N)) \). Similarly \( T(p)X_1(N)_{\phi F} \) is defined by

\[
\begin{array}{c}
\mathcal{M}_{K_F} \\
\downarrow \alpha_F \\
X_1(N)_F \\
\downarrow \beta_F \\
\tilde{T}(p)X_1(N)_{\phi F} \rightarrow X_1(N)_F
\end{array}
\]

(41)

where \( \alpha_F \) is the natural projection and \( \beta_F \) is induced by \( g^{-1} \). Using the identification \( \mathcal{M}_{K_F} \approx \mathcal{M}_K \otimes F \) and the description (37) of the complex points, we obtain \( \alpha_F = \alpha \otimes \text{id}_F \) and \( \beta_F = \beta \otimes \gamma(g^{-1}) \). Since \( \gamma(g) = \sigma_{p^{-1}} \), we get \( T(p)X_1(N)_{\phi F} = \nu_F(T(p)X_1(N)) \circ (\sigma_p)_* \) and thus (38). The proof of (39) is similar.

5 A divisibility in the Hecke algebra

In this section we define and study a projection associated to \( E_F \) using the Hecke algebra of \( X_1(N)_F \).

Let \( \phi : X_1(N) \rightarrow E \) be a modular parametrization of the elliptic curve \( E \), and let \( \phi_F : X_1(N)_F \rightarrow E_F \) be the base change of \( \phi \) to \( F \). Consider the map

\[
e_F = \frac{1}{\deg \phi}(\phi_F)^*(\phi_F) \circ (\sigma_p)_*, \quad \Omega^1(X_1(N)_F).
\]

Lemma 14. We have \( e_F^2 = e_F \) and \( e_F \in T(G) \).

Proof. The first equality follows from \( (\phi_F)^*(\phi_F) = \deg \phi_F \).

We have \( e_F = \nu_F(e) \) where \( e = \frac{1}{\deg \phi} \phi^* \phi \in \text{End}_Q \Omega^1(X_1(N)) \). The image of \( e \) is the \( Q \)-vector space generated by \( \omega_f = 2i\pi f(z)dz \). Since \( f \) is a newform of level \( N \), the Atkin-Lehner-Li theory implies that \( e \in T_{X_1(N)} \). The result now follows from Lemma 13.

The space \( \Omega = \varprojlim_K \Omega^1(\mathcal{M}_K) \otimes \mathcal{Q} \) has a natural \( \text{GL}_2(A_f) \)-action and decomposes as a direct sum of irreducible admissible representations \( \Omega_\pi \) of \( \text{GL}_2(A_f) \). For any \( K \) we have \( \Omega^K = \Omega^1(\mathcal{M}_K) \otimes \mathcal{Q} \). Let \( \Pi(K) \) be the set of such \( \pi \) satisfying \( \Omega^K \neq \{0\} \). By (12) p. 393, we have

\[
\Omega^1(\mathcal{M}_K) \otimes \mathcal{Q} = \bigoplus_{\pi \in \Pi(K)} \Omega^K_\pi
\]

(42)
where each $\Omega^K_\pi$ is a simple $T_K$-module. In particular $T_K$ is a semisimple algebra. By Lemma [10] the algebra $T$ is contained in the center of $T_{K_1(N)_F}$. Using [12 Prop 2.11], we deduce that $T$ acts by scalar multiplication on each $\Omega^K_{\pi_1(N)_F}$, so there exists a morphism $\theta_\pi : T \to \mathbb{Q}$ such that $T$ acts as $\theta_\pi(T)$ on $\Omega^\pi_{K_1(N)_F}$. The multiplicity one and strong multiplicity one theorems [14] ensure that the characters $(\theta_\pi)_{\pi \in \Pi(K_1(N)_F)}$ are pairwise distinct.

For any $\chi \in \hat{G}$, let $\pi(f \otimes \chi)$ be the automorphic representation of $\text{GL}_2(A_f)$ corresponding to the modular form $f \otimes \chi$. We have $\pi(f \otimes \chi) \cong \pi(f) \otimes (\chi \circ \det)$, where $\chi : A_f^1/\mathbb{Q}_{\text{am}} \to \mathbb{C}^*$ denotes the ad\'elization of $\chi$, sending $\varpi_p$ to $\chi(p)$ for every $p \nmid m$. Since $\pi(f) \in \Pi(K_1(N)_F)$, it follows that $\pi(f \otimes \chi) \in \Pi(K_1(N)_F)$.

**Lemma 15.** For any prime $p \nmid Nm$, we have

$$\theta_\pi(f \otimes \chi)(T(p)) = a_p \chi(p)$$  \hspace{1cm} (43)

$$\theta_\pi(f \otimes \chi)(T(p,p)) = \chi(p)^2.$$  \hspace{1cm} (44)

**Proof.** We know that $\theta_\pi(f)(T(p)) = a_p$ and $\theta_\pi(f)(T(p,p)) = 1$. The equalities [43] and [44] follow formally from the fact that $\chi \circ \det$ is equal to $\chi(p)$ on the double coset $K_1(N)_F \left( \varpi_p \begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix} \right) K_1(N)_F$. \hfill $\Box$

Let $e_{f \otimes \chi} : \Omega^1(X_1(N)_F) \otimes \overline{\mathbb{Q}} \to \Omega^K_{\pi(f \otimes \chi)}$ be the projection induced by [22]. The multiplicity one theorems imply that $e_{f \otimes \chi} \in T$.

**Proposition 16.** The element $e_\chi e_F$ is divisible by $e_{f \otimes \chi}$ in $TG$.

**Proof.** Since $e_\chi$, $e_F$ and $e_{f \otimes \chi}$ are commuting projections, it suffices to prove that the image of $e_\chi e_F$ is contained in the image of $e_{f \otimes \chi}$. We know that the image of $\varphi^* : \Omega^1(E) \to \Omega^1(X_1(N))$ lies in the kernel of $T(p) - a_p \in T_{X_1(N)}$. Therefore the image of $\varphi^*_F$ lies in the kernel of $\nu_F(T(p)) - a_p$. Using Lemma [13] it follows that in $TG$ we have

$$T(p)\sigma_p e_F = a_p e_F.$$  \hspace{1cm} (45)

Applying $e_\chi$ to both sides and using the identity $e_\chi \sigma_p = \overline{\chi}(p)e_\chi$ yields

$$T(p) e_\chi e_F = a_p \chi(p) e_\chi e_F.$$  \hspace{1cm} (46)

The same argument shows that $T(p,p) e_\chi e_F = \chi(p)^2 e_\chi e_F$. The proposition now follows from Lemma [15] and the multiplicity one theorems. \hfill $\Box$
6 Proof of the main results

Recall that \( \varphi : X_1(N) \to E \) is a modular parametrization, and that \( \varphi_F \) is the base change of \( \varphi \) to \( F \). We have a commutative diagram

\[
\begin{array}{ccc}
K_2(X_1(N)_F) \otimes \mathbb{Q} & \xrightarrow{} & H^1(X_1(N)_F(\mathbb{C}), \mathbb{R})^* \\
\xrightarrow{\langle \varphi_F \rangle} & & \xrightarrow{\langle \varphi_F \rangle}
\end{array}
\]

where the horizontal maps are the regulator maps on \( X_1(N)_F \) and \( E_F \).

The strategy of the proof is to use Beilinson’s theorem on \( X_1(N)_F \) and then to get back to \( E_F \) using the Hecke algebra.

Let \( \mathcal{P}_{E/F} = (\varphi_F)_*\mathcal{P}_{X_1(N)_F} \subset K_2(E_F) \otimes \mathbb{Q} \). By [15, 1.1.2(iii)], we have \( \mathcal{P}_{E/F} \subset H^2_{\text{et}}(E_F, \mathbb{Q}(2)) \). We want to prove that \( R_{E/F} := \text{reg}_{E/F}(\mathcal{P}_{E/F}) \) is a \( \mathbb{Q} \)-structure satisfying \( \mathcal{F} \). Since \( \mathcal{P}_{X_1(N)_F} \) is stable by the Hecke algebra, the spaces \( \mathcal{P}_{E/F} \) and \( R_{E/F} \) are stable by \( G \).

For any \( \chi \in \hat{G} \), let \( R_\chi = e_\chi (R_{E/F} \otimes \overline{\mathbb{Q}}) \) and \( H_\chi = e_\chi (H_{E/F} \otimes \overline{\mathbb{Q}}) \). We want to compare \( R_\chi \) and \( H_\chi \).

We have

\[
\varphi_F^* R_\chi = e_\chi \varphi_F^* (R_{E/F} \otimes \overline{\mathbb{Q}})
\]

Similarly, we have

\[
\varphi_F^* H_\chi = e_\chi \varphi_F^* (H_{X_1(N)_F} \otimes \overline{\mathbb{Q}}).
\]

We will build on the following theorem of Schappacher and Scholl. Let \( \lambda_\chi \) be the unique element of \( (\overline{\mathbb{Q}} \otimes \mathbb{R})^* \) such that for every \( \psi : \overline{\mathbb{Q}} \to \mathbb{C}^* \), we have

\[
\psi(\lambda_\chi) = L'(f \otimes \chi, 0) \in \mathbb{C}^*.
\]

By [15, 1.2.4 and 1.2.6], we have

\[
e_{f \otimes \chi} (\text{reg}_{X_1(N)_F}(\mathcal{P}_{X_1(N)_F} \otimes \overline{\mathbb{Q}})) = \lambda_\chi \cdot e_{f \otimes \chi} (H_{X_1(N)_F} \otimes \overline{\mathbb{Q}}).
\]

By Prop. [16] the equality [50] remains true when \( e_{f \otimes \chi} \) is replaced by \( e_\chi e_F \), so that \( \varphi_F^* R_\chi = \lambda_\chi \cdot \varphi_F^* H_\chi \) by [48] and [49]. Since \( \varphi_F^* \) is injective, we get \( R_\chi = \lambda_\chi \cdot H_\chi \). Put \( V = H^1(E_F, \mathbb{R})^* \) and \( V_\chi = e_\chi (V \otimes \overline{\mathbb{Q}}) \) for any \( \chi \in \hat{G} \).

Lemma 17. The \( \mathbb{R}[G] \)-module \( V \) is free of rank 1.

Proof. By Poincaré duality \( V \cong \text{Hom}_\mathbb{Q}(\Omega^1(E_F), \mathbb{R}) \), and \( \Omega^1(E_F) \cong \Omega^1(E) \otimes F \) is free of rank 1 over \( \mathbb{Q}[G] \) by the normal basis theorem.

We will use the following lemma from linear algebra. Recall that if \( B \) is an \( A \)-algebra and \( N \) is a \( B \)-module, an \( A \)-structure of \( N \) is an \( A \)-submodule \( M \subset N \) such that \( M \otimes_A B \cong N \).

Lemma 18. Let \( M \) be a \( \mathbb{Q}[G] \)-submodule of \( V \). The following conditions are equivalent:

\[
\text{(i) } M \text{ is free of rank } 1.
\text{(ii) } M \text{ is an } A \text{-structure of } N.
\text{(iii) } M \otimes_A B \cong N.
\]
(i) $M$ is a $\mathbb{Q}$-structure of the real vector space $V$.

(ii) For any $\chi \in \hat{G}$, the space $M_\chi := e_\chi(M \otimes \overline{\mathbb{Q}})$ is a $\overline{\mathbb{Q}}$-structure of the $\overline{\mathbb{Q}} \otimes \mathbb{R}$-module $V_\chi$.

Moreover, if these conditions hold, then $M$ is free of rank 1 over $\mathbb{Q}[G]$.

Proof. The implication (i) $\Rightarrow$ (ii) follows from the isomorphisms $M_\chi \otimes \overline{\mathbb{Q}} (\overline{\mathbb{Q}} \otimes \mathbb{R}) \cong e_\chi (M \otimes \overline{\mathbb{Q}} \otimes \mathbb{R}) \cong V_\chi$. Let us assume (ii). By Lemma 18 the $\overline{\mathbb{Q}} \otimes \mathbb{R}$-module $V_\chi$ is free of rank 1, so that $\dim_{\overline{\mathbb{Q}}} M_\chi = 1$. Since $M \otimes \overline{\mathbb{Q}} \cong \oplus_{\chi \in \hat{G}} M_\chi$, we get $\dim_{\overline{\mathbb{Q}}} M = d$. Moreover $M \otimes \overline{\mathbb{Q}} \otimes \mathbb{R}$ generates $V \otimes \overline{\mathbb{Q}} \otimes \mathbb{R}$, so that any $\mathbb{Q}$-basis of $M$ is actually free over $\mathbb{R}$.

Finally, if (i) holds, then $M$ is isomorphic to the regular representation of $G$ by Lemma 17, so that $M$ is free of rank 1 over $\mathbb{Q}[G]$. $\square$

Using Lemma 18 with the $\mathbb{Q}$-structure $H_{EF}$, we see that $H_\chi$ is a $\overline{\mathbb{Q}}$-structure of $V_\chi$. By Lemma 8 the 1-form $\eta_\chi$ is a $\overline{\mathbb{Q}}$-basis of $H_\chi$.

Proof of Theorem 2. Since $R_\chi = \lambda_\chi \cdot H_\chi$ is a $\overline{\mathbb{Q}}$-structure of $V_\chi$, Lemma 18 implies that $R_{EF}$ is a $\mathbb{Q}$-structure of $V$. Moreover, the determinant of $R_{EF} \otimes \mathbb{Q}$ with respect to $H_{EF} \otimes \overline{\mathbb{Q}}$ is represented by $\delta := \prod_{\chi \in \hat{G}} \lambda_\chi \in (\overline{\mathbb{Q}} \otimes \mathbb{R})^\times$. Note that $\sigma(\lambda_\chi) = \lambda_{\sigma \chi}$ for any $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, so that $\delta$ lies in the image of $\mathbb{R}^\times$ in $(\overline{\mathbb{Q}} \otimes \mathbb{R})^\times$. Using the natural evaluation map $(\overline{\mathbb{Q}} \otimes \mathbb{R})^\times \hookrightarrow C^\times$, we get in fact $\delta = \prod_{\chi \in \hat{G}} L(f \otimes \chi, 0)$. Since the natural map $\mathbb{R}^\times/\mathbb{Q}^\times \to (\overline{\mathbb{Q}} \otimes \mathbb{R})^\times/\overline{\mathbb{Q}}^\times$ is injective, we conclude that $\det(R_{EF}) = L(d)(E_F, 0) \cdot \det(H_{EF})$ by 8. $\square$

Proof of Theorem 1. We know from Theorem 2 that $R_{EF}$ is a $\mathbb{Q}$-structure of $V$. Since $R_{EF}$ is stable by $G$, it is free of rank 1 over $\mathbb{Q}[G]$ by Lemma 18. Let $\gamma \in \mathcal{P}_{EF}$ such that $R_{EF} = Q[G] \cdot \text{reg}_{EF}(\gamma)$. Replacing $\gamma$ by a suitable integer multiple, we may assume that $\gamma$ has a representative $\overline{\gamma} \in F(E) \otimes F(E)$. Let $\ell = \beta(\overline{\gamma})$. For any $\chi \in \hat{G}$, we have $R_\chi = \mu_\chi(\ell) H_\chi$ by Prop. 9 where $\mu_\chi(\ell)$ is given by 20. It follows that $\mu_\chi(\ell)/\lambda_\chi \in \overline{\mathbb{Q}}^\times$. Since $\lambda_\chi$ and $\mu_\chi(\ell)$ belong to $\mathbb{Q}(\chi) \otimes \mathbb{R}$, we have in fact $\mu_\chi(\ell)/\lambda_\chi \in \mathbb{Q}(\chi)^\times$. Moreover, the definitions of $\lambda_\chi$ and $\mu_\chi(\ell)$ show that $\tau(\lambda_\chi) = \lambda_{\chi^\tau}$ and $\tau(\mu_\chi(\ell)) = \mu_{\chi^\tau}(\ell)$ $(\tau \in \text{Gal}(\mathbb{Q}(\chi)/\mathbb{Q}))$. (51)

Lemma 19. Let $(a_\chi)_{\chi \in \hat{G}}$ be a family of algebraic numbers, with $a_\chi \in \mathbb{Q}(\chi)^\times$, such that $\tau(a_\chi) = a_{\chi^\tau}$ for any $\chi$ and any $\tau \in \text{Gal}(\mathbb{Q}(\chi)/\mathbb{Q})$. Then there exists a unique $a \in \mathbb{Q}[G]^\times$ such that for every $\chi \in \hat{G}$, we have $\chi(a) = a_\chi$.

Proof. The canonical morphism of $\mathbb{Q}$-algebras $\Psi : \mathbb{Q}[G] \to \prod_{\chi \in \hat{G}} \mathbb{Q}(\chi)$ is injective and its image is contained in the subalgebra $W$ of families $(b_\chi)_\chi$ satisfying $\tau(b_\chi) = b_{\chi^\tau}$ for any $\chi$ and $\tau$. Writing $\hat{G}$ as a disjoint union of Galois orbits, we have $\dim_{\mathbb{Q}} W = \# \hat{G} = d$, so that $\Psi$ is an isomorphism. $\square$
Let $\ell(\chi)$, we get $a \in \mathbb{Q}[G]^{*}$ such that $\mu_{\chi}(\ell) = \chi(a)\lambda_{\chi}$ for any $\chi$. Since $\mu_{\chi}(a\ell) = \chi(a)\mu_{\chi}(\ell)$, replacing $\ell$ with a suitable integer multiple of $a\ell$ results in $\mu_{\chi}(\ell) \sim_{\mathbb{Q}} \lambda_{\chi}$ for any $\chi$. Evaluating everything in $\mathbb{C}$ yields $\chi$. $\hfill \Box$

**Proof of Corollary.** Let us first recall the Dedekind-Frobenius formula for group determinants. If $a : G \to \mathbb{C}$ is an arbitrary function, let $A$ be the matrix $(a(gh^{-1}))_{g,h \in G}$. Then

$$\det(A) = \prod_{\chi \in \hat{G}} \sum_{g \in G} \chi(g) a(g). \quad (52)$$

Let $\ell \in \mathbb{Z}[E(Q)]^{G_{r}}$ be a divisor satisfying the identities (1) of Theorem 1. Assume first $F$ is real. Put $\ell_{i} := \ell^{\sigma_{i}}$ for $1 \leq i \leq d$. Using (52) with $a(\sigma) = D_{E}(\ell^{\sigma})$ yields

$$\det(D_{E}(\ell_{i}^{\sigma}))_{1 \leq i,j \leq d} \sim_{\mathbb{Q}^{*}} \prod_{\chi \in \hat{G}} \pi L(E \otimes \chi, 0) \sim_{\mathbb{Q}^{*}} \pi^{-d} L(E_{F}, 2) \quad (53)$$

where the last relation follows from (8) and Prop. 4.

Assume now $F$ is complex. Put $\ell_{i} := \ell^{\sigma_{i}^{-1}}$ for $1 \leq i \leq d/2$. We use (52) with the function $a(\sigma) = D_{E}(\ell^{\sigma}) + J_{E}(\ell^{\sigma})$. Indexing the lines and columns of $A$ by \( \sigma_{1}, \sigma_{1}, \ldots, \sigma_{d/2}, \sigma_{d/2} \), we see that $A$ consists of blocks of the form

$$\begin{pmatrix} x + y & x - y \\ x - y & x + y \end{pmatrix},$$

where $x = D_{E}(\ell^{\sigma_{i}^{-1}})$ and $y = J_{E}(\ell^{\sigma_{i}^{-1}})$. Elementary operations on the lines and columns of $A$ thus gives

$$\det A = 2^{d} \det(D_{E}(\ell_{i}^{\sigma}))_{1 \leq i,j \leq d/2} \cdot \det(J_{E}(\ell_{i}^{\sigma}))_{1 \leq i,j \leq d/2}. \quad (54)$$

On the other hand, we have

$$\sum_{\sigma \in G} \chi(\sigma) a(\sigma) = \begin{cases} \sum_{\sigma \in G} \chi(\sigma) D_{E}(\ell^{\sigma}) & \text{if } \chi \text{ is even}, \\ \sum_{\sigma \in G} \chi(\sigma) J_{E}(\ell^{\sigma}) & \text{if } \chi \text{ is odd}, \end{cases} \quad (55)$$

so that we conclude as in the first case. $\hfill \Box$

**Further remarks and a conjecture**

The proof of Theorem 2 relies crucially on the hypothesis that $F/\mathbb{Q}$ is abelian. Since the field of constants of a modular curve is always an abelian extension of $\mathbb{Q}$, it is not possible to cover a non-abelian base change of $E$ by a usual modular curve. In fact, in the case $F/\mathbb{Q}$ is not abelian, we have no example of a (non CM) elliptic curve $E$ over $\mathbb{Q}$ for which we can prove Zagier’s conjecture for $L(E_{F}, 2)$. However, Theorem 1 suggests the following conjecture for Artin-twisted $L$-values. For simplicity, we restrict to the case $F$ is totally real.
Conjecture 20. Let $E$ be an elliptic curve defined over $\mathbb{Q}$, and let $F$ be a finite Galois totally real extension of $\mathbb{Q}$. There exists a divisor $\ell \in \mathbb{Z}[E(\mathbb{Q})]^\text{Gal}(\mathbb{Q}/F)$ satisfying Goncharov and Levin’s conditions such that for every Artin representation $\rho : \text{Gal}(F/\mathbb{Q}) \to \text{GL}_d(\mathbb{C})$, we have

$$L^{(d)}(E \otimes \rho, 0) \sim_{\mathbb{Q}} \pi^{-d} \det\left( \sum_{\sigma \in G} \rho(\sigma) D_E(\ell^\sigma) \right).$$

(56)

Conversely, for every $\ell \in \mathbb{Z}[E(\mathbb{Q})]^\text{Gal}(\mathbb{Q}/F)$ satisfying Goncharov and Levin’s conditions and for every $\rho : \text{Gal}(F/\mathbb{Q}) \to \text{GL}_d(\mathbb{C})$, we have

$$\pi^{-d} \det\left( \sum_{\sigma \in G} \rho(\sigma) D_E(\ell^\sigma) \right) \in L^{(d)}(E \otimes \rho, 0) \cdot \mathbb{Q}(\text{tr } \rho),$$

(57)

where $\mathbb{Q}(\text{tr } \rho)$ is the field generated by the traces of $\rho$.

Noting that the identities (56) and (57) are compatible with taking direct sums of Artin representations. In fact, Conjecture 20 is a refinement of Zagier’s conjecture for $L(E_F, 2)$, in the sense that taking the product over irreducibles $\rho$ with multiplicities $\text{dim}(\rho)$ gives the conjecture for $E_F$. Note that the analytic continuation and the functional equation of $L(E \otimes \rho, s)$ are only conjectural in general.

It would be interesting to investigate the rational factors arising in Theorem 1. As a matter of fact, even for $F = \mathbb{Q}$, we don’t know how to predict the rational factor appearing in Zagier’s conjecture. The Bloch-Kato conjecture predicts the exact value of $L(E_F, 2)$ (at least up to a unit in the ring of integers of $F$), but the link between both conjectures remains to be worked out. In fact, in this setting it may be more natural to investigate the equivariant Tamagawa number conjecture of Burns and Flach [9, Part 2, Conjecture 3], which predicts the equivariant $L$-value $L(F_E, 2) \in \mathbb{R}[G]^+$ up to a unit in an order of $\mathbb{Q}[G]$. Taking norms down to $\mathbb{Q}$, this predicts $L(E_F, 2)$ up to sign. The deep work of Gealy [10] on the Bloch-Kato conjecture for modular forms, which uses Kato’s Euler system, could be used to tackle this equivariant conjecture. Note also that if $F$ is abelian and real, then Theorem 1 gives a link between $L(F_E, 2)$ and the vector-valued elliptic dilogarithm $D_E(\ell) = \sum_{\sigma \in G} D_E(\ell^\sigma)[\sigma]$.

Finally, although the divisor $\ell$ produced by Theorem 1 is inexplicit in general, it would be interesting to try to bound the number field generated by the support of $\ell$, as well as the heights of the points involved.

References


François Brunault
ÉNS Lyon, UMPA
46 allée d’Italie
69007 Lyon
France
francois.brunault@ens-lyon.fr
Algebraic Groups of Type $D_4$, Triality, and Composition Algebras

Vladimir Chernousov\textsuperscript{1}, Alberto Elduque\textsuperscript{2}, Max-Albert Knus, and Jean-Pierre Tignol\textsuperscript{3}

Received: February 20, 2013
Revised: May 6, 2013
Communicated by Ulf Rehmann

Abstract. Conjugacy classes of outer automorphisms of order 3 of simple algebraic groups of classical type $D_4$ are classified over arbitrary fields. There are two main types of conjugacy classes. For one type the fixed algebraic groups are simple of type $G_2$; for the other type they are simple of type $A_2$ when the characteristic is different from 3 and are not smooth when the characteristic is 3. A large part of the paper is dedicated to the exceptional case of characteristic 3. A key ingredient of the classification of conjugacy classes of trialitarian automorphisms is the fact that the fixed groups are automorphism groups of certain composition algebras.

2010 Mathematics Subject Classification: 20G15, 11E57, 17A75, 14L10.
Keywords and Phrases: Algebraic group of type $D_4$, triality, outer automorphism of order 3, composition algebra, symmetric composition, octonions, Okubo algebra.

\textsuperscript{1}Partially supported by the Canada Research Chairs Program and an NSERC research grant.

\textsuperscript{2}Supported by the Spanish Ministerio de Economía y Competitividad and FEDER (MTM2010-18370-C04-02) and by the Diputación General de Aragón—Fondo Social Europeo (Grupo de Investigación de Álgebra).

\textsuperscript{3}Partially supported by the F.R.S.–FNRS (Belgium). J.-P. Tignol gratefully acknowledges the hospitality of the Zukunftskolleg of the Universität Konstanz, where he was in residence as a Senior Fellow while this work was developing.
1. Introduction

The projective linear algebraic group $\text{PGL}_n$ admits two types of conjugacy classes of outer automorphisms of order two. For one type the fixed subgroups are symplectic groups; for the other type the fixed groups are orthogonal groups, which are not smooth when the base field has characteristic 2.

The picture is similar for trialitarian automorphisms (i.e., outer automorphisms of order three) of the algebraic groups $G = \text{PGO}_8^+$ or $G = \text{Spin}_8$. There are two types of conjugacy classes. For one type the fixed groups are simple of type $G_2$; for the other type they are simple of type $A_2$ when the characteristic is different from 3, and not smooth when the characteristic is 3. The first case is well known: there is a split exact sequence

$$1 \to \text{Int}(G) \to \text{Aut}(G) \to S_3 \to 1$$

where the permutation group of three elements $S_3$ is viewed as the group of automorphisms of the Dynkin diagram of type $D_4$.

\[ \begin{array}{c}
\alpha_1 \\
\alpha_2 \\
\hdashline \\
\alpha_3 \\
\alpha_4
\end{array} \]

(The simple roots $\alpha_i$ are numbered after [Bou81]). Viewing $G$ as a Chevalley group, there is a canonical section of the exact sequence (1.1) which leads to a distinguished trialitarian automorphism of $G$, known as a graph automorphism.

The permutation $\rho: \alpha_1 \mapsto \alpha_3 \mapsto \alpha_4 \mapsto \alpha_1, \alpha_2 \mapsto \alpha_2$ of the Dynkin diagram above induces a permutation of all simple roots $\alpha$ of $G$, also denoted by $\rho$.

The graph automorphism is such that $\rho(x_\alpha(u)) = x_{\rho(\alpha)}(u)$ for each simple root group $x_\alpha(u)$ of $G$ (see [Ste68b] or Section 2). The subgroup of $G$ fixed by the graph automorphism is of type $G_2$.

The aim of this paper is to describe all conjugacy classes of trialitarian automorphisms of simple algebraic groups of classical inner type $D_4$ over arbitrary fields, from the (functorial) point of view of algebraic groups. In particular the case of characteristic 3 has not been considered before in this setting, and a large part of this paper is dedicated to it. In a previous paper [CKT12], three of the authors gave a description based on the (known) classification of symmetric composition algebras of dimension 8 and a correspondence between symmetric compositions and trialitarian automorphisms. Here we take a different approach: we classify trialitarian automorphisms directly, and then use the correspondence to derive a new proof of the classification of symmetric composition algebras of dimension 8.

In the first part of the paper, consisting of §§2–7, we consider algebraic groups over algebraically closed fields. The first step, achieved in §§2–3, is to prove

\[ \text{Trialitarian automorphisms (or more generally automorphisms of finite order) of simple Lie algebras have been extensively studied, see i.a., [Hel78], [Kac69], [Knu09], [Rec10] and [WG68].} \]
that every trialitarian automorphism admits an invariant maximal torus. This is clear in characteristic different from 3 since trialitarian automorphisms are semisimple and semisimple automorphisms admit invariant tori (see [BM55], [Ste68a] or [Pia85]). However we did not find references for the result over fields of characteristic 3. In §§4–5, we recall the correspondence between symmetric compositions and trialitarian automorphisms set up in [CKT12], and use it to define two standard trialitarian automorphisms $\rho_0$ and $\rho_0$ corresponding respectively to the para-Zorn composition $\diamond$ and the split Okubo composition $\triangle$ on the 8-dimensional quadratic space $(C,n)$ of Zorn matrices. We then define in §6 a split maximal torus $T$ of $PGO^+(n)$ invariant under $\rho_0$ and $\rho_0$, and use it in §7 to show that over an algebraically closed field, every trialitarian automorphism of $PGO^+(n)$ is conjugate to $\rho_0$ or to $\rho_0$. In view of the results of §3, and since over a separably closed field all the maximal tori are conjugate, it suffices to consider trialitarian automorphisms that preserve the given torus $T$.

In the second part of the paper, we describe the subgroups of $PGO^+_8$ fixed under trialitarian automorphisms, over an arbitrary field $F$. These fixed subgroups are the automorphism groups of the corresponding symmetric composition algebras. They are divided into two classes according to their isomorphism class over an algebraic closure. In §8 we show that for one of the classes the symmetric composition algebras are para-octonion algebras and the automorphism groups are of type $G_2$; for the other the symmetric composition algebras are Okubo algebras and the automorphism groups are of type $A_2$ when the characteristic is different from 3. The following §§9–11 deal with the case of Okubo algebras in characteristic 3. A particular type of idempotents, which we call quaternionic idempotents, is singled out in §9, where we show that the existence of a quaternionic idempotent characterizes split Okubo algebras, and that split Okubo algebras over a field of characteristic 3 contain a unique quaternionic idempotent. This idempotent is used to describe in §§10–11 the group scheme $\text{Aut}(\triangle)$ of automorphisms of the split Okubo algebra in characteristic 3. (The corresponding groups of rational points already occur in various settings, see [Tit59], [GL83] and [Eld99].) Using the description of $\text{Aut}(\triangle)$, we recover in §12 the classification of Okubo algebras over arbitrary fields of characteristic 3 by a cohomological approach. Finally, in §13 we give a cohomological version of the correspondence between symmetric compositions and trialitarian automorphisms, and show that the only groups of classical type $^{1,2}D_4$ that admit trialitarian automorphisms are $PGO^+(n)$ and $\text{Spin}(n)$ for $n$ a 3-Pfister quadratic form. Note that the cohomology we use is faithfully flat finitely presented cohomology, or fppf-cohomology, since we need to deal with inseparable base field extensions. (Galois cohomology would be sufficient for fields of characteristic different from 3.) We refer to [DG70], [Wat79] and [KO74] for details on fppf-cohomology and descent theory.

The fact that the composition algebras of octonions could be used to define trialitarian automorphisms was already known to Élie Cartan (see [Car25]). We refer to [KMRT98] and [SV00] for historical comments on triality.
Chernousov, Elduque, Knus, Tignol

If not explicitly mentioned $F$ denotes throughout the paper an arbitrary field and the algebras considered are defined over $F$.

We gratefully thank Philippe Gille for his help during the preparation of this paper.

2. Generators and relations in Chevalley groups

We recall in this section results about Chevalley groups which will be used later. Let $G$ be a split simple simply connected group over $F$. Since $G$ is a Chevalley group over $F$, its $F$-structure is well known. For details and proofs of all standard facts about $G(F)$ used in this section we refer to [Ste68b]. Let $G$ be the Lie algebra of $G$. Choose a split maximal torus $T \subset G$ and a Borel subgroup $T \subset B \subset G$. Let $\Sigma = \Sigma(G,T)$ be the root system of $G$ relative to $T$. The Borel subgroup $B$ determines an ordering of $\Sigma$, hence the system of simple roots $\Pi = \{\alpha_1, \ldots, \alpha_n\}$. We pick a Chevalley basis [Ste68b]

$$\{H_{\alpha_1}, \ldots, H_{\alpha_n}, X_{\alpha}, \alpha \in \Sigma\}$$

in $G$ corresponding to the pair $(T,B)$. This basis is unique up to signs and automorphisms of $G$ which preserve $B$ and $T$ (see [Ste68b], §1, Remark 1). The group $G(F)$ is generated by the so-called root subgroups $U_\alpha = \langle x_\alpha(u) | u \in F \rangle$, where $\alpha \in \Sigma$ and

$$x_\alpha(u) = \sum_{n=0}^{\infty} u^n X^n_\alpha / n!$$

(we refer to [Ste68b] for the definition of the operators $X^n_\alpha / n!$ in the case $\text{char}(F) = p > 0$, see also [SGA3]).

If $\alpha \in \Sigma$ and $t \in G_m(F)$, the following elements are also of great importance:

$$w_\alpha(t) = x_\alpha(t)x_{-\alpha}(-t^{-1})x_\alpha(t) \quad \text{and} \quad h_\alpha(t) = w_\alpha(t)w_\alpha(1)^{-1}.$$

To ease notation we set $w_\alpha = w_\alpha(1)$. The elements $h_\alpha(t)$ give rise to a cocharacter $h_\alpha : G_m \to T$ whose image is $T_\alpha = T \cap G_\alpha$ where $G_\alpha$ is the subgroup generated by $U_{\pm \alpha}$.

Example 2.2. The group $G_\alpha$ is isomorphic in a natural way to $\text{SL}_2$. This isomorphism is given by

$$x_\alpha(u) \mapsto \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, \quad x_{-\alpha}(u) \mapsto \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix},$$

$$h_\alpha(t) \mapsto \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}, \quad w_\alpha \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The following relations hold in $G$ (cf. [Ste68b], Lemma 19 a), Lemma 28 b), Lemma 20 a)):

$$T = T_{\alpha_1} \times \cdots \times T_{\alpha_n};$$

for any two roots $\alpha, \beta \in \Sigma$ we have

$$w_\alpha h_\beta(t) w_\alpha^{-1} = h_{w_\alpha(\beta)}(t);$$

Documenta Mathematica 18 (2013) 413–468
where \( \varepsilon = \pm 1 \).
We finally recall a commutator formula due to Chevalley, see [Ste68b, Lemma 32']. For arbitrary roots \( \alpha, \beta \), one has
\[
(w_\alpha X_\beta w_\alpha^{-1} = \varepsilon X_{w_\alpha(\beta)}
\]
where \( \varepsilon = \pm 1 \).

3. Invariant Tori

The main result in this section is the existence of invariant tori under trialitarian automorphisms. As already mentioned in the introduction the claim is known over fields of characteristic different from 3. We give here a proof for any field \( F \).

We first reduce to the case of simply connected groups. Let \( G \) be a split simple simply connected group of type \( D_4 \) over \( F \), and let \( G^{\text{ad}} \) be the isogenous adjoint group. We have a canonical exact sequence
\[
1 \to \mu_2 \times \mu_2 \to G \to G^{\text{ad}} \to 1,
\]
where \( \mu_2 \) is the group scheme of square roots of unity. Every automorphism of \( G \) induces an automorphism of \( G^{\text{ad}} \).

Lemma 3.2. Every trialitarian automorphism \( \phi \) of \( G^{\text{ad}} \) lifts to a trialitarian automorphism \( \tilde{\phi} \) of \( G \). If \( T \) is a torus in \( G \) invariant under \( \tilde{\phi} \), the image of \( T \) in \( G^{\text{ad}} \) is invariant under \( \phi \).

Proof. Viewing \( G \) as a Chevalley group, we have a group of graph automorphisms of \( G \) isomorphic to \( S_3 \) defined over \( F \), and we may consider the semidirect products \( G \rtimes S_3 \) and \( G^{\text{ad}} \rtimes S_3 \). Note that
\[
\text{Aut}(G^{\text{ad}}) = \text{Int}(G^{\text{ad}}) \rtimes S_3 = G^{\text{ad}} \rtimes S_3,
\]
and the exact sequence (3.1) yields an exact sequence
\[
1 \to \mu_2 \times \mu_2 \to G \rtimes S_3 \to G^{\text{ad}} \rtimes S_3 \to 1.
\]
Consider the associated exact sequence in faithfully flat finitely presented cohomology (see [DG70, p. 272–273]):
\[
(\mu_2 \times \mu_2)(F) \to (G \rtimes S_3)(F) \to (G^{\text{ad}} \rtimes S_3)(F) \to H^1_{\text{fppf}}(F, \mu_2 \times \mu_2).
\]
Viewing \( \phi \) as an element in \( (G^{\text{ad}} \rtimes S_3)(F) \), we may consider its image \( \phi' \) in \( H^1_{\text{fppf}}(F, \mu_2 \times \mu_2) \). We have \( \phi'^2 = 1 \) because \( H^1_{\text{fppf}}(F, \mu_2 \times \mu_2) \) has exponent 2, and \( \phi'^3 = 1 \) because \( \phi'^3 = 1 \), hence \( \phi' = 1 \). It follows that \( \phi \) has a preimage \( \phi'' \) in \( (G \rtimes S_3)(F) \). We have \( \phi''' \in (\mu_2 \times \mu_2)(F) \) since \( \phi' = 1 \), hence conjugation by \( \phi''' \) in \( G \rtimes S_3 \) restricts to an automorphism \( \tilde{\phi} \) of order 3 of \( G \). The induced automorphism on \( G^{\text{ad}} \) is \( \phi \), so \( \tilde{\phi} \) is a lift of \( \phi \) and is an outer automorphism.
A torus $T$ in $G$ is invariant under $\tilde{\phi}$ if and only if it centralizes $\phi''$. It is then clear that the image of $T$ in $G^{ad}$ is invariant under $\phi$. \qed

The crucial tool to prove the existence of invariant tori is the following result of G. Harder:

**Proposition 3.3** ([Har75, Lemma 3.2.4]). Let $F$ be an infinite field and let $G$ be a split simple simply connected algebraic group or a split simple adjoint algebraic group of type $\mathbb{D}_4$ over $F$. Let $\phi \in \text{Aut}(G)(F)$ be an outer automorphism of order 3 of $G$ over $F$. There exists a parabolic subgroup $P \subset G$ such that $H = P \cap P^{\phi} \cap P^{\phi^2}$ is a reductive subgroup of $G$ of dimension 10 defined over $F$ whose semi-simple part $H' = [H,H]$ is a simple group of type $A_2$.

We first make some comments about properties of $H$. Assume that $F$ is separably closed. Being a reductive group, $H$ is an almost direct product of its central 2-dimensional torus $S$ and its derived subgroup $H'$. Let $T$ be an arbitrary maximal torus in $H$. Since $\dim T = 4$, $T$ is also maximal in $G$. It then follows that $H'$ is generated by some root subgroups $U_{\pm \gamma}$ and $U_{\pm \delta}$ where $\gamma, \delta$ are roots in $G$ with respect to $T$. Note that every such subgroup in $G$ is simply connected, hence $H' \cong \text{SL}_3$. Let $\alpha_1, \ldots, \alpha_4$ be the basis of the root system $\Sigma$ of type $\mathbb{D}_4$ as defined in [Bou81], see also (1), and let $\beta = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \in \Sigma$. For an arbitrary pair of roots $\gamma, \delta$ with $(\gamma, \delta) = -1$ in $\Sigma$ there exists an element $w$ in the corresponding Weyl group such that $w(\gamma) = \alpha_2$, $w(\delta) = \beta$. In view of (2.5) we may then assume without loss of generality that $H'$ is the subgroup $G_{(\alpha_2,\beta)}$ generated by the root subgroups $U_{\pm \alpha_2}$ and $U_{\pm \beta}$.

Another ingredient in the proof of existence of invariant tori is the following lemma. Let $G$ be a split simple simply connected algebraic group of type $\mathbb{D}_4$ over a field $F$. Let $T \subset B$ be a maximal split torus and a Borel subgroup in $G$. As usual, the subgroup $B$ determines the ordering in the root system $\Sigma = \Sigma(G,T)$ of $G$ with respect to $T$ and hence a basis of $\Sigma$. As above we set $\beta = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$.

**Lemma 3.4.** Assume that $F$ is a field of characteristic 3. Let $\rho$ be the graph automorphism of $G$. If $x = x_{\alpha_2}(u)$, where $u \in F^\times$, there exists $g \in B(F) \times \langle \rho \rangle$ such that $g(\rho x)g^{-1} = \rho x_{\alpha_2}(u_1)x_\beta(u_2)x_{\alpha_2 + \beta}(u_3) \in B(F) \times \langle \rho \rangle$, where $u_1, u_2 \in F^\times$ and $u_3 \in F$.

**Proof.** The proof is based on the commutator formula (2.6). Note that for a group of type $G_2$ in the commutator formula applied to two simple roots $\gamma_1, \gamma_2$ of $G_2$ (numbered as in [Bou81]), all integers $c_{ij}$ are non-zero modulo 3.

Consider the subgroup $\mathcal{G}$ of $G$ of elements invariant under $\rho$, which is of type $G_2$. It contains the subgroup of $G$ generated by $G_{\alpha_2}, G_{\beta}$ of type $A_2$. Also, the root subgroups in $\mathcal{G}$ corresponding to its two simple roots $\gamma_1$ and $\gamma_2$ are generated by elements $x_{\alpha_2}(u)$ and $x_{\alpha_1}(v)x_{\alpha_3}(v)x_{\alpha_4}(v)$. Taking into consideration (2.6) and the fact that $\rho$ commutes with $x_{\alpha_1}(v)x_{\alpha_3}(v)x_{\alpha_4}(v)$ we can write the element

$$X = (x_{\alpha_1}(1)x_{\alpha_3}(1)x_{\alpha_4}(1)) (\rho x_{\alpha_2}(u)) (x_{\alpha_1}(1)x_{\alpha_3}(1)x_{\alpha_4}(1))^{-1}$$

\[2\text{We are indebted to Philippe Gille for pointing out this result to us.}\]
in the form
\begin{equation}
X = \rho(x_{\alpha_2}(v_1)x_{\beta}(v_2)x_{\alpha_2+\beta}(v_3))YZ
\end{equation}
where $v_1, v_2 \neq 0$ (see the above remark about the integers $c_{ij}$) and
\begin{align*}
Y &= x_{\alpha_1+\alpha_2}(v_4)x_{\alpha_3+\alpha_2}(v_5)x_{\alpha_4+\alpha_2}(v_6) \\
Z &= x_{\alpha_1+\alpha_3+\alpha_2}(v_7)x_{\alpha_1+\alpha_2+\alpha_2}(v_8)x_{\alpha_3+\alpha_4+\alpha_2}(v_9)
\end{align*}
where $v_i \in F$. We now note that in the expression (3.5) for $X$ the root elements $x_{\alpha_1}(\cdot), x_{\alpha_3}(\cdot)$ and $x_{\alpha_4}(\cdot)$ are missing and the commutator of the other root subgroups which appear in (3.5) is either 1 or of the form $x_{\alpha_2+\beta}(\cdot)$ which is in the center of $B(F) \rtimes \langle \rho \rangle$. So there is no harm if we ignore factors $x_{\alpha_2+\beta}(\cdot)$ in the computations below and thus we shall assume that all root subgroups $x_\gamma$ which appear commute with each other. Since $\rho x$ has order 3 (because $\rho$ and $x$ commute and $x^3 = x_{\alpha_2}(3\gamma) = 1$) it follows almost immediately that $Z$ satisfies the “cocycle condition” $Z\rho(Z)\rho^2(Z) = 1$. Since $\rho$ acts freely on the subgroup
\begin{equation*}
A = \langle x_{\alpha_1+\alpha_2+\alpha_1}(\cdot) \rangle \times \langle x_{\alpha_1+\alpha_2+\alpha_2}(\cdot) \rangle \times \langle x_{\alpha_3+\alpha_2+\alpha_4}(\cdot) \rangle
\end{equation*}
one can see that $Z$ can be written in the form $Z = \rho^{-1}Z_1\rho Z_1^{-1}$ where $Z_1 \in A$. We then have
\begin{align*}
Z_1^{-1}XZ_1 &= Z_1^{-1}\rho x_{\alpha_2}(v_1)x_{\beta}(v_2)Y(\rho^{-1}Z_1\rho) \\
&= \rho(\rho^{-1}Z_1^{-1}\rho) x_{\alpha_2}(v_1)x_{\beta}(v_2)Y(\rho^{-1}Z_1\rho) \\
&= \rho x_{\alpha_2}(v_1)x_{\beta}(v_2)Y.
\end{align*}
Arguing similarly we can now up to conjugacy eliminate $Y$ and this completes the proof. We emphasize once more that all the above equalities are considered modulo the central subgroup $x_{\alpha_2+\beta}(\cdot)$ of $B(F) \rtimes \langle \rho \rangle$. \hfill \qed

**Theorem 3.6.** Let $F$ be a separably closed field and let $G$ be a simple simply connected algebraic group or a simple adjoint algebraic group of type $D_4$. Let $\phi \in \text{Aut}(G)(F)$ be an outer automorphism of order 3 of $G$. There exists a maximal $F$-torus $T$ which is invariant under $\phi$. Moreover, if $\tau$ is another outer automorphism of order 3, there exists a conjugate of $\tau$ in $\text{Aut}(G)(F)$ which leaves the $F$-torus $T$ invariant.

**Proof.** The last claim follows from the fact that over a separably closed field two maximal tori in $G$ are always conjugate ([Bor56]). To show the existence of an invariant torus we treat the cases $\text{char } F \neq 3$ and $\text{char } F = 3$ separately. We may assume that $G$ is simply connected by Lemma 3.2. We keep the above notation. Let $H$ be the subgroup in $G$ provided by Proposition 3.3. By construction it is $\phi$-stable. Since $H'$ has no outer automorphisms of order 3 the restriction of $\phi$ to $H'$ is an inner automorphism of $H'$ given by an element $x \in H'$.

A) **Characteristic $F \neq 3$.**

Note that $\phi$ is a semi-simple automorphism of $G$ in characteristic different from 3, hence $x$ is also semi-simple and therefore $x$ is contained in a maximal torus, say $S'$, in $H'$. By our construction we have $\dim S' = 2$ and $\dim S = 2$. 

Documenta Mathematica 18 (2013) 413–468
and this implies that the torus $T = S \cdot S'$ is maximal in $G$ and stable with respect to $\phi$ (because so are $S$ and $S'$).

**B) Characteristic $F = 3$.**

We keep the notation used above. Since $x$ has order 3 it is unipotent. Therefore there are two possibilities for its Jordan normal form.

We first assume that

$$
\begin{bmatrix}
1 & 1 & 0 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{bmatrix}
$$

Let $\{e_1, e_2, e_3\}$ be the basis of the vector space over $F$ in which $x$ has form (3.7). Let $S' \subset H'$ be a maximal torus in $H'$ whose three weight subspaces are spanned by $e_1 + e_2$, $2e_1 + e_2$, $2e_1 + 2e_2 + e_3$. One easily checks that $S'$ is stable with respect to conjugation given by $x$. As above we then get that the maximal torus $T = S \cdot S'$ in $G$ is invariant with respect to $\phi$.

We next assume that $x$ is of the form

$$
x = x_{\alpha_2}(1) = \begin{bmatrix}
1 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}.
$$

Let $S'$ be the corresponding diagonal torus in $H' = \text{SL}_3$ and let $T = S \cdot S'$. Choose a graph automorphism $\sigma$ of $G$ such that it acts trivially on $H' = G_{\alpha_2, \beta}$ and the images of $\phi$ and $\sigma$ with respect to $G \times \mathfrak{S}_3 \to \mathfrak{S}_3$ are equal. Let $\psi = \phi \circ (\sigma x)^{-1}$. It is an inner automorphism of $G$ acting trivially on $H'$ and hence on $S'$. It follows that $\psi$ stabilizes $C_G(S') = T$. Note that the last equality follows from the fact that in general case $C_G(S')$ is a reductive group generated by $T$ and root subgroups $U_\gamma$, such that $\gamma$ is orthogonal to $\alpha_2$ and $\beta$ and in case $D_4$ there are no roots orthogonal to $\alpha_2, \beta$. Thus we showed that $\psi$ is an inner conjugation in $G$ given by an element say $y$ which belongs to $N_G(T)(F)$. By our construction $y$ commutes with $S'$ and hence with its generators $h_{\alpha_2}(u)$ and $h_{\beta}(u)$. In view of (2.4) we conclude that the image $\hat{y}$ of $y$ in the Weyl group $W = N_G(T)/T$ acts trivially on the roots $\alpha_2, \beta$. One checks that in the Weyl group of type $D_4$ none of the elements acts trivially on two roots whose scalar product is $-1$. It follows $\hat{y} = 1$ implying $y \in T$. Thus we showed that $\phi = y\sigma x$ where $y \in T(F)$. By (2.3) we can write $y$ in the form $y = h_{\alpha_2}(u_2)h_{\alpha_1}(u_1)h_{\alpha_2}(u_3)h_{\alpha_4}(u_4)$ where $u_i \in F^\times$. Since $x$ commutes with $\sigma$ it is easy to see that the condition that $\phi$ has order 3 implies $h_{\alpha_1}(u_1)h_{\alpha_2}(u_3)h_{\alpha_4}(u_4)\sigma$ also has order 3 and $h_{\alpha_2}(u_2)^3 = 1$. The latter of course implies that $u_2 = 1$. Then arguing as in Lemma 3.4 we may write $h_{\alpha_1}(u_1)h_{\alpha_2}(u_3)h_{\alpha_4}(u_4)\sigma$ in the form $t\sigma t^{-1}$ for some element $t \in T(F)$. It follows $\phi = t\sigma t^{-1} x$ or $t^{-1}\phi t = \sigma(t^{-1}xt)$. Note that $\tilde{x} = t^{-1}xt \in U_{\alpha_2}$. Let $B \subset G$ be the corresponding Borel subgroup in $G$. By Lemma 3.4 the element $\sigma \tilde{x}$ is conjugate in $B(F) \rtimes \mathbb{Z}/3\mathbb{Z}$ to an element of the form $\sigma x'$ where $x' = x_{\alpha_2}(u_1)x_{\beta}(u_2)x_{\alpha_2+\beta}(u_3)$ and $u_1, u_2$ are non-zero elements in $F$. The Jordan
normal form of such unipotent element is of the shape (3.7) and this reduces the proof to the previous case. □

Corollary 3.8. Let $G$ be a simple simply connected algebraic group or a simple adjoint algebraic group of type $\text{D}_4$ over an arbitrary field $F$ and let $\phi$ be a trialitarian automorphism of $G$ over $F$. There exists a finite separable field extension $\tilde{F}$ of $F$ such that over $\tilde{F}$ there is a split maximal torus $T$ which is invariant under $\phi$. Moreover, if $\tau$ is another trialitarian automorphism of $G$ over $F$, there exists a conjugate of $\tau$ in the automorphism group of $G$ over a finite separable field extension of $F$ which leaves the torus $T$ invariant.

Let $G$ be split over $F$ and $T \subset G$ be a split maximal torus over $F$. We denote the group of automorphisms of $G$ which leave the torus $T$ invariant by $\text{Aut}(G \supset T)$. Let $\rho \in \text{Aut}(G \supset T)(F)$ be a fixed trialitarian automorphism of $G$, for example the graph automorphism of $G$ described in the introduction.

The normalizer $N$ of $T$ in $G$ is also invariant under $\rho$ and $\rho$ induces a well-defined action $\rho$ on $N/T$, which is the Weyl group $W$ of $G$. This action is not inner. The group $W$ also admits a group of outer automorphisms isomorphic to $S_3$ and we still call trialitarian the outer automorphisms of $W$ of order 3 (see [Ban69], [Fra01] or [FH03] for automorphism groups of Weyl groups).

Proposition 3.9. (i) Any trialitarian automorphism $\alpha \in \text{Aut}(G)(F)$ leaving $T$ invariant induces a trialitarian automorphism $\alpha$ of $W$.

(ii) If two trialitarian automorphisms $\alpha$ and $\beta$ of $G$ over $F$ leaving the same torus $T$ invariant satisfy $\alpha = \beta$ in $\text{Aut}(W)$, then $\alpha = \text{Int}(t) \circ \beta$ for some element $t \in T(F)$.

Proof. We may assume that $G$ is adjoint and we may view trialitarian automorphisms as elements of $G \rtimes S_3$. In particular trialitarian automorphisms of $G$ which leave $T$ invariant are represented by elements of $N \rtimes S_3$. Let now $\alpha$ be represented by an element $\alpha_0 \in (N \rtimes S_3)(F)$. Let $\alpha_0'$ be the class of $\alpha_0$ in $(W \rtimes S_3)(F) = W \rtimes S_3$ according to the last map in the exact sequence

$$1 \rightarrow T \rightarrow N \rtimes S_3 \rightarrow W \rtimes S_3 \rightarrow 1.$$ (3.10)

Let $Z(W) = \mathbb{Z}/2\mathbb{Z}$ be the center of $W$. The class of the automorphism $\overline{\alpha}$ of $W$ induced by $\alpha$ is the class $\overline{\alpha} = \overline{\alpha_0'}$ of $\alpha_0'$ in $\text{Aut}(W) \subset W/Z(W) \rtimes S_3$, hence Claim (i). For Claim (ii), the image of an element $\alpha \circ \beta^{-1}$ in $W$ induces the trivial automorphism of $W$. Since the center of $W$ has order 2 and $\alpha$, $\beta$ have order 3 this image is trivial and this implies $\alpha = \text{Int}(t) \circ \beta$ for some $t \in T(F)$. □

4. Trialitarian automorphisms and symmetric compositions

In the following sections, we recall a construction of trialitarian automorphisms given in [KMRT98, §35.B], [CKT12, Th. 4.6], and we construct explicitly invariant tori for some of these automorphisms.
Let \((V,q)\) be a quadratic space over \(F\), i.e., \(V\) is a finite-dimensional vector space over \(F\) and \(q: V \to F\) is a quadratic form. We always assume that \(q\) is nonsingular, in the sense that the polar bilinear form \(b_q\) defined by
\[
b_q(x, y) = q(x + y) - q(x) - q(y) \quad \text{for } x, y \in V
\]
has radical \(\{0\}\). We also assume throughout that \(\dim V\) is even.

Let \(\GO(q)\) be the \(F\)-algebraic group of similarities of \((V,q)\), whose group of rational points \(\GO(q)(F)\) consists of linear maps \(f: V \to V\) for which there exists a scalar \(\mu(f) \in F^\times\), called the multiplier of \(f\), such that
\[
q(f(x)) = \mu(f)q(x) \quad \text{for all } x \in V.
\]

Let also \(O(q)\) be the \(F\)-algebraic group of isometries of \((V,q)\), i.e., the kernel of the multiplier map \(\mu: \GO(q) \to G_m\). The center of \(\GO(q)\) is the multiplicative group \(G_m\), whose rational points are viewed as homotheties. For \(f \in \GO(q)(F)\), we let \([f]\) be the image of \(f\) in \(\PGO(q)(F) = \GO(q)(F)/G_m(F)\). For simplicity, we write
\[
\GO(q) = \GO(q)(F) \quad \text{and} \quad \PGO(q) = \PGO(q)(F) = \GO(q)/F^\times.
\]

Let \(C(V,q)\) be the Clifford algebra of the quadratic space \((V,q)\) and let \(C_0(V,q)\) be the even Clifford algebra. We let \(\sigma\) be the canonical involution of \(C(V,q)\), such that \(\sigma(x) = x\) for \(x \in V\), and use the same notation for its restriction to \(C_0(V,q)\). Every similarity \(f \in \GO(q)\) induces an automorphism \(C_0(f)\) of \((C_0(V,q),\sigma)\) such that
\[
C_0(f)(xy) = \mu(f)^{-1}f(x)f(y) \quad \text{for } x, y \in V,
\]
see [KMRT98, (13.1)]. This automorphism depends only on the image \([f] = fF^\times\) of \(f\) in \(\PGO(q)\), and we shall use the notation \(C_0[f]\) for \(C_0(f)\). The similarity \(f\) is proper if \(C_0[f]\) fixes the center of \(C_0(V,q)\) and improper if it induces a nontrivial automorphism of the center of \(C_0(V,q)\) (see [KMRT98, (13.2)]).

Proper similarities define an algebraic subgroup \(\GO^+(q)\) in \(\GO(q)\), and we let \(\PGO^+(q) = \GO^+(q)/G_m\), a subgroup of \(\PGO(q)\). The groups \(\GO^+(q)\) and \(\PGO^+(q)\) are the connected components of the identity in \(\GO(q)\) and \(\PGO(q)\) respectively, see [KMRT98, §23.B]. We write \(O^+(q)\) for the algebraic group of proper isometries of \((V,q)\).

The \(F\)-rational points of the \(F\)-algebraic group \(\Spin(q)\) are given by
\[
\Spin(q)(F) = \Spin(q) = \{c \in C_0(V,q)^\times \mid cVc^{-1} \subset V \quad \text{and} \quad c\sigma(c) = 1\},
\]
see for example [KMRT98, §35.C]. For any element \(c \in \Spin(q)\) the linear map \(x \mapsto cxc^{-1}, x \in V\), is a proper similitude of \((V,q)\). Thus there exists a morphism of algebraic groups \(\pi: \Spin(q) \to \PGO^+(q)\). In the next proposition we collect known results about the algebraic groups \(\Spin(q)\) and \(\PGO^+(q)\), see for example [KMRT98, 26.A].

**Proposition 4.2.** Let \((V,q)\) be a quadratic space and let \(\dim V = 2n\) with \(n \equiv 0 \pmod{2}\).
(i) The algebraic group $\text{Spin}(q)$ is simple, simply connected of type $D_n$ and its center is isomorphic to $\mu_2^2 = \mu_2 \times \mu_2$, with $\mu_2(F) = \{\pm 1\}$.

(ii) The algebraic group $\text{PGO}^+(q)$ is simple, adjoint of type $D_n$.

(iii) The sequence of algebraic groups

$$1 \to \mu_2^2 \to \text{Spin}(q) \xrightarrow{\eta} \text{PGO}^+(q) \to 1$$

is exact.

Following [KMRT98, §35.B] or [CKT12, §4] we next recall how trialitarian automorphisms can be generated with the help of symmetric compositions, which we define next, following [KMRT98, §34] or [CKT12, §3].

**Definition 4.3.** A symmetric composition on a nonsingular quadratic space $(S, n)$ is a bilinear map $*: S \times S \to S$ such that for all $x, y, z \in S$

1. $n(x \ast y) = n(x)n(y)$,
2. $b_n(x \ast y, z) = b_n(x, y \ast z)$,
3. $x \ast (y \ast x) = n(x)y = (x \ast y) \ast x$.

If $*$ is a symmetric composition on $(S, n)$, the triple $(S, *, n)$ is called a symmetric composition algebra. The composition $*$ is its multiplication.

Symmetric composition algebras exist in dimension 2, 4, and 8. Their norms are Pfister forms. From now on, we restrict to dimension 8.

**Theorem 4.4.** Let $(S, *, n)$ be an 8-dimensional symmetric composition algebra. Let $f$ be a similitude of $n$ with multiplier $\mu(f)$.

(i) If $f$ is proper there exist proper similitudes $f_1, f_2$ of $n$ such that for all $x, y \in S$

$$\mu(f)^{-1}f(x \ast y) = f_2(x) \ast f_1(y)$$

$$\mu(f_1)^{-1}f_1(x \ast y) = f(x) \ast f_2(y)$$

$$\mu(f_2)^{-1}f_2(x \ast y) = f_1(x) \ast f(y).$$

(ii) If $f$ is improper there exist improper similitudes $f_1, f_2$ of $n$ such that for all $x, y \in S$

$$\mu(f)^{-1}f(x \ast y) = f_2(y) \ast f_1(x)$$

$$\mu(f_1)^{-1}f_1(x \ast y) = f(y) \ast f_2(x)$$

$$\mu(f_2)^{-1}f_2(x \ast y) = f_1(y) \ast f(x).$$

The pair $(f_1, f_2)$ is determined by $f$ up to a factor $(\mu, \mu^{-1})$, $\mu \in F^\times$, and we have $\mu(f_1)\mu(f_2) = 1$. Furthermore any of the three formulas in (i) (resp. (ii)) implies the two others.

For a proof, see [KMRT98, (35.4)] or [CKT12, Th. 4.5].

---

3The symmetric compositions considered in [CKT12] are slightly more general; the ones we use here are referred to as normalized in [CKT12].
In view of Theorem 4.4, the elements $[f_1], [f_2]$ in $\text{PGO}^+(n)(F)$ are uniquely determined by $[f]$. Thus they give rise to well-defined maps

$$\rho_* : \text{PGO}^+(n)(F) \to \text{PGO}^+(n)(F), \quad [f] \mapsto [f_2],$$

$$\hat{\rho}_* : \text{PGO}^+(n)(F) \to \text{PGO}^+(n)(F), \quad [f] \mapsto [f_1].$$

**Theorem 4.5.** The mappings $\rho_*$ and $\hat{\rho}_*$ are outer automorphisms over $F$ of order 3 of $\text{PGO}^+(n)$, and $\hat{\rho}_* = \rho_2^*.$

**Proof.** See [KMRT98, (35.6)]. □

The algebraic group $\text{Spin}(n)$ can be described with the help of a symmetric composition $\star$ on $(S,n)$:

**Proposition 4.6.** Let $(S,\star,n)$ be a symmetric composition algebra with norm $n$. Then

$$\text{Spin}(n)(F) = \{ (f,f_1,f_2) \mid f_i \in \mathcal{O}(n)(F), f(x \star y) = f_2(x) \star f_1(y), x,y \in S \}.$$  

Moreover any of the three relations

$$f(x \star y) = f_2(x) \star f_1(y)$$

$$f_1(x \star y) = f(x) \star f_2(y)$$

$$f_2(x \star y) = f_1(x) \star f(y)$$

implies the two others.

**Proof.** See [KMRT98, (35.2)] for a proof in characteristic different from 2 and [KT13] for a proof in any characteristic. □

We have an obvious trialitarian automorphism $\rho_*$ of $\text{Spin}(n)$:

$$\rho_* : (f,f_1,f_2) \mapsto (f_2,f,f_1)$$

and properties similar to those for $\text{PGO}^+(n)$ hold.

Viewing $\mu_3^2$ as kernel of the multiplication map

$$p : \mu_3^2 \to \mu_2, \quad p(\varepsilon_1,\varepsilon_2,\varepsilon_3) = \varepsilon_1\varepsilon_2\varepsilon_3,$$

we get a natural $\mathfrak{A}_3$-action on the center $\mu_3^2$ of $\text{Spin}(n)$, compatible with the trialitarian action on $\text{Spin}(n)$ and the sequence

$$1 \to \mu_2^2 \to \text{Spin}(n) \to \text{PGO}^+(n) \to 1$$

is $\mathfrak{A}_3$-equivariant (see [KMRT98, (35.13)]). Here $\mathfrak{A}_3 \subseteq \mathfrak{S}_3$ is the subgroup of even permutations. In particular the trialitarian action on $\text{Spin}(n)$ is a lift of the trialitarian action on $\text{PGO}^+(n)$.

Isomorphisms of symmetric compositions are defined in the obvious way. The trialitarian automorphisms associated to isomorphic symmetric compositions are conjugate; more precisely, we have:

**Lemma 4.7.** Let $\phi : (S,\diamond,n) \to (S,\star,n)$ be an isomorphism of symmetric composition algebras. Then

$$\rho_* = \text{Int}(\phi) \circ \rho_0 \circ \text{Int}(\phi)^{-1}.$$
Proof. See [CKT12, Proposition 6.1]. \[\square\]

The main result of [CKT12] is the following:

**Theorem 4.8.** Let \((S, n)\) be a 3-Pfister quadratic space over a field \(F\). The assignment \(* \mapsto \rho_*\) defines a one-to-one correspondence between isomorphism classes of symmetric compositions on \((S, n)\) and conjugacy classes over \(F\) of trialitarian automorphisms of \(\text{PGO}^+(n)\) defined over \(F\).

**Proof.** This follows from Theorems 5.8 and 6.4 of [CKT12]. \[\square\]

Automorphisms of order 3 of a given symmetric composition lead to new symmetric compositions by “twisting”:

**Proposition 4.9.** Let \((S, \star, n)\) be a symmetric composition algebra and let \(\phi\) be an automorphism of order 3 of \((S, \star, n)\).

(i) The multiplication 
\[
(\xi, \eta) \mapsto \xi \star \phi \eta = \phi(\xi) \star \phi^2(\eta), \quad \text{for } \xi, \eta \in S
\]
defines a symmetric composition on the quadratic space \((S, n)\).

(ii) \(\rho_{\star \phi} = \text{Int}(\phi^{-1}) \circ \rho_\star\).

**Proof.** See [Pet69] for the first claim and [CKT12, Lemma 5.2] for the second. \[\square\]

The symmetric composition \(*_\phi\) is called the (Petersson) twist of \(*\).

5. **Zorn matrices**

In this section we use Zorn’s nice description of the split octonion algebra in [Zor30] to give two examples of 8-dimensional symmetric compositions. The associated trialitarian automorphisms play a fundamental rôle in this work.

The **Zorn matrix algebra** is defined as follows. Let \(\bullet\) be the usual scalar product on the 3-dimensional space \(F^3\) and let \(\times\) the vector product: for vectors \(\vec{a} = (a_1, a_2, a_3)\) and \(\vec{b} = (b_1, b_2, b_3) \in F^3\), we have \(\vec{a} \bullet \vec{b} = a_1b_1 + a_2b_2 + a_3b_3\) and \(\vec{a} \times \vec{b} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)\). The Zorn algebra is the set of matrices
\[
C = \left\{ \begin{pmatrix} \alpha & \vec{a} \\ \vec{b} & \beta \end{pmatrix} \big| \alpha, \beta \in F, \, \vec{a}, \vec{b} \in F^3 \right\}
\]
with the product
\[
(5.1) \quad \begin{pmatrix} \alpha & \vec{a} \\ \vec{b} & \beta \end{pmatrix} \cdot \begin{pmatrix} \gamma & \vec{c} \\ \vec{d} & \delta \end{pmatrix} = \begin{pmatrix} \alpha\gamma + \vec{a} \bullet \vec{d} & \alpha \vec{c} + \delta \vec{a} - \vec{b} \times \vec{d} \\ \gamma \vec{b} + \beta \vec{d} + \vec{a} \times \vec{c} & \beta \delta + \vec{b} \bullet \vec{c} \end{pmatrix},
\]
the norm
\[
n(\begin{pmatrix} \alpha & \vec{a} \\ \vec{b} & \beta \end{pmatrix}) = \alpha \beta - \vec{a} \bullet \vec{b},
\]
and the conjugation
\[
\overline{(\begin{pmatrix} \alpha & \vec{a} \\ \vec{b} & \beta \end{pmatrix})} = \begin{pmatrix} \beta & -\vec{a} \\ -\vec{b} & \alpha \end{pmatrix}.
\]

**Documenta Mathematica** 18 (2013) 413–468
which is such that $\xi \cdot \overline{\xi} = \xi \cdot \xi = n(\xi)$ for all $\xi \in C$ (see [Zor30, p. 144]). The element $1 = \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right)$ is an identity element for the product. One checks that the norm is multiplicative:

(5.2) $n(\xi \cdot \eta) = n(\xi)n(\eta)$ for all $\xi, \eta$ in $C$.

Thus $(C, \cdot, n)$ is an 8-dimensional composition algebra with identity. Since the norm $n$ is obviously a hyperbolic form, $(C, \cdot, n)$ is a split octonion algebra.

Let $\{\vec{a}_1, \vec{a}_2, \vec{a}_3\}$ be the standard basis of $F^3$. The elements

(5.3) $e_1 = \left( \begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array} \right)$, $f_1 = \left( \begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right)$, $u_i = \left( \begin{array}{cc} 0 & -\vec{a}_i \\ \vec{a}_i & 0 \end{array} \right)$, $v_i = \left( \begin{array}{cc} 0 & \vec{a}_i \\ \vec{a}_i & 0 \end{array} \right)$, $i = 1, 2, 3$, form a hyperbolic basis, called the canonical basis of the Zorn algebra $(C, \cdot, n)$.

We have $e_1 = f_1$, and $u_i = -u_i$, $v_i = -v_i$ for $i = 1, 2, 3$.

Next we associate to $(C, \cdot, n)$ a symmetric composition. More generally, let $(H, \cdot, n)$ be a Hurwitz algebra (i.e., a quadratic algebra, a quaternion algebra or an octonion algebra) with conjugation $x \mapsto \overline{x}$, $x \in H$.

**Lemma 5.4.** Setting $x \diamond y = x \cdot y$ we get a symmetric composition $(H, \diamond, n)$.

**Proof.** See for example [KMRT98, §34A].

We call $(H, \diamond, n)$ the para-Hurwitz algebra (resp. para-quadratic, para-quaternion or para-octonion algebra) associated with $H$. Applying the same construction to the Zorn algebra, we obtain the para-Zorn algebra with multiplication

(5.5) $\left( \begin{array}{cc} \alpha & \vec{a} \\ \vec{b} & \beta \end{array} \right) \circ \left( \begin{array}{cc} \gamma & \vec{c} \\ \vec{d} & \delta \end{array} \right) = \left( \begin{array}{cc} \beta \vec{d} + \vec{a} \bullet \vec{d} & -\beta \vec{c} - \gamma \vec{a} - \vec{b} \times \vec{d} \\ -\delta \vec{d} - \alpha \vec{d} + \vec{a} \times \vec{c} & \alpha \gamma + \vec{b} \bullet \vec{c} \end{array} \right)$.

As we shall see in Remark 8.11 the trialitarian automorphism associated to the para-Zorn algebra is conjugate to the graph automorphism described in the introduction.

The group $\text{GL}_3$ acts on the vector space $C$ of Zorn matrices by

(5.6) $s_g \left( \begin{array}{cc} \alpha & \vec{a} \\ \vec{b} & \beta \end{array} \right) = \left( \begin{array}{cc} \alpha g^2(\vec{b}) & g(\vec{a}) \\ g^2(\vec{b}) & \beta \end{array} \right)$,

where $g \in \text{GL}_3(F)$ and $g^4 = (g^{-1})^4$.

**Proposition 5.7.** (i) The mappings $s_g$ for $g \in \text{SL}_3(F)$ are automorphisms of the para-Zorn algebra $(C, \diamond, n)$.

(ii) The mappings $s_g$ for $g \in \text{GL}_3(F)$ are isometries of $(C, n)$.

---

4Zorn’s definition is different, but isomorphic under the map $(\alpha \vec{a} \vec{b} \beta) \mapsto (\alpha \vec{a} \vec{b} \beta)$. 

---

**Documenta Mathematica 18 (2013) 413–468**
Proof. Let \( g \in \text{GL}_3(F) \). For \( \vec{x}, \vec{y} \in F^3 \) we have \( g(\vec{x}) \cdot \vec{y} = \vec{x} \cdot g^t(\vec{y}) \), hence
\[
(5.8) \quad g(\vec{x}) \cdot g^t(\vec{y}) = \vec{x} \cdot \vec{y}.
\]

This readily implies (ii). To prove (i), observe that \( \vec{x} \times \vec{y} \) is characterized by the condition that for all \( \vec{z} \in F^3 \)
\[
[(\vec{x} \times \vec{y}) \cdot \vec{z}] \vec{a}_1 \wedge \vec{a}_2 \wedge \vec{a}_3 = \vec{x} \wedge \vec{y} \wedge \vec{z} \quad \text{in } \bigwedge^3 F^3.
\]

For all \( \vec{x}, \vec{y}, \vec{z} \in F^3 \) we have
\[
(\det g) \vec{x} \wedge \vec{y} \wedge \vec{z} = g(\vec{x}) \wedge g(\vec{y}) \wedge g(\vec{z}),
\]

hence
\[
(\det g) (\vec{x} \times \vec{y}) \cdot \vec{z} = (g(\vec{x}) \times g(\vec{y})) \cdot g(\vec{z}) = g^t (g(\vec{x}) \times g(\vec{y})) \cdot \vec{z}.
\]

Since these equations hold for all \( \vec{z} \in F^3 \) it follows that
\[
(\det g) (\vec{x} \times \vec{y} = g^t (g(\vec{x}) \times g(\vec{y})) \quad \text{for all } \vec{x}, \vec{y} \in F^3,
\]

hence
\[
(5.9) \quad (\det g) g^t(\vec{x} \times \vec{y}) = g(\vec{x}) \times g(\vec{y}) \quad \text{for all } \vec{x}, \vec{y} \in F^3.
\]

A straightforward computation then yields (i). \( \square \)

Let \( \pi \) be the cyclic permutation of the standard basis vectors \( \pi: \vec{a}_1 \mapsto \vec{a}_2 \mapsto \vec{a}_3 \mapsto \vec{a}_1 \), extended by linearity to \( F^3 \), i.e.,
\[
(5.10) \quad \pi = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \in \text{SL}_3(F).
\]

The symmetric composition algebra \((C, \circ_\pi, n)\) obtained by twisting the parazorn algebra \((C, \circ, n)\) with the automorphism \( s_\pi \) is called the split Okubo algebra. We let \( \triangle \) denote its multiplication. Thus
\[
\begin{pmatrix} \alpha & \vec{a} \\ \beta & \vec{b} \end{pmatrix} \triangle \begin{pmatrix} \gamma & \vec{c} \\ \delta & \vec{d} \end{pmatrix} = \begin{pmatrix} \beta \delta + \pi(\vec{a}) \cdot \pi^2(\vec{d}) & -\beta \pi^2(\vec{c}) - \gamma \pi(\vec{a}) - \pi(\vec{b}) \times \pi^2(\vec{d}) \\ -\delta \pi(\vec{b}) - \alpha \pi^2(\vec{d}) + \pi(\vec{a}) \times \pi^2(\vec{c}) & \alpha \gamma + \pi(\vec{b}) \cdot \pi^2(\vec{c}) \end{pmatrix}.
\]

The multiplication table of \( \triangle \) with respect to the canonical basis (5.3) is

---

\(^5\)Okubo ([Oku78]) gave a different description of this algebra, valid over fields of characteristic not 3, see also Proposition 8.5 in this paper.
If the field $F$ contains a primitive cube root $\omega$ of 1 (so that in particular $F$ has characteristic different from 3), there is another way to twist the multiplication of the para-Zorn algebra to get the split Okubo algebra: let $\omega = \text{Diag}(1, \omega, \omega^2) \in \text{SL}_3(F)$ and let $s_\omega$ be the corresponding automorphism of $(C, \circ, n)$, see Proposition 5.7. We may then consider the Petersson twist $\triangledown = \circ s_\omega$. The next lemma shows that the symmetric composition $\triangledown$ is isomorphic to $\triangle$.

**Lemma 5.12.** Let

$$g = \frac{\omega - \omega^2}{3} \begin{pmatrix} 1 & 1 & 1 \\ -1 & -\omega & -\omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix} \in \text{SL}_3(F).$$

The isometry $s_g$ of $(C, n)$ is an isomorphism of symmetric compositions $\triangle \sim \triangledown$.

**Proof.** The matrix $g$ satisfies $\omega g = g, \pi$, hence also $\omega^2 g^2 = g^2 \pi^2 = g^2 \pi$. The lemma is then verified by straightforward computations, using the identities (5.8), (5.9).

By contrast, the para-Zorn multiplication $\circ$ and the split Okubo multiplication $\triangle$ are not isomorphic, even if the base field $F$ is algebraically closed. This can be seen directly because the para-Zorn algebra contains a para-unit (see Definition 9.1 below), whereas the split Okubo algebra does not.

### 6. A Maximal Torus for $\text{PGO}_8^+$ and $\text{Spin}_8$

Let $G$ denote one of the algebraic groups $\text{PGO}_8^+$ or $\text{Spin}_8$. In view of Theorem 3.6, to classify triallitarian automorphisms over a separably closed field, we may fix a maximal torus $T$ and consider only triallitarian automorphisms of $G$ which leave $T$ invariant. We now describe a special maximal split torus which is invariant under the triallitarian automorphisms associated to the para-Zorn algebra $(C, \circ, n)$ and the split Okubo algebra $(C, \triangle, n)$.

Consider the following torus

$$T_0 = G_m^5 = G_m^2 \times \text{Diag}_3$$

**References:**

Chernousov, Elduque, Knus, Tignol

**Documenta Mathematica** 18 (2013) 413–468
where $\text{Diag}_3$ is the torus of diagonal matrices in $\text{GL}_3$. The torus $T_0$ acts by similitudes on the quadratic space $(C,n)$ of Zorn matrices as follows:

\[(6.2) \quad (\lambda, \mu, t) \oplus \left( \begin{array}{cc} \alpha & \tilde{a} \\ \beta & \tilde{b} \end{array} \right) = \left( \begin{array}{cc} \lambda \alpha & \lambda \mu t^{-1} \tilde{a} \\ \mu \beta & \mu \end{array} \right).\]

The group $T_0$ is a maximal torus of $\text{GO}^+_4 (= \text{GO}^+(n))$, and $G_m$ diagonally embedded into $T_0$ is the center of $\text{GO}^+_4$. Thus $T_0/G_m$ is a maximal torus of $\text{PGO}^+_4$, which we denote by $T$. For $(\lambda, \mu, t) \in T_0(F)$, we write $[\lambda, \mu, t]$ for $[(\lambda, \mu, t)] \in T(F)$. We compute the action on $T$ of the trialitarian automorphism $\rho_0$ associated to the para-Zorn algebra $(C, \diamond, n)$:

**Lemma 6.3.** Let $(\lambda, \mu, t) \in T_0(F)$. If $\lambda', \mu', \lambda'', \mu'' \in F^\times$ are such that

\[(6.4) \quad \mu' \mu'' = \lambda, \quad \lambda' \lambda'' = \mu, \quad \text{and} \quad \lambda' \lambda'' = (\det t)(\lambda \mu)^{-1},\]

then for $\xi, \eta \in C$, we have

\[(\lambda, \mu, t) \circ (\xi \circ \eta) = ((\lambda', \mu', \mu''^{-1} t) \circ \xi) \circ ((\lambda'', \mu'', \lambda' t) \circ \eta);\]

hence

\[\rho_0[\lambda, \mu, t] = [\lambda', \mu', \mu''^{-1} t] \quad \text{and} \quad \rho_0^2[\lambda, \mu, t] = [\lambda'', \mu'', \lambda^{-1} t].\]

The lemma is verified by a straightforward computation. It shows that the torus $T$ is invariant under $\rho_0$, provided a solution to (6.4) exists for all $(\lambda, \mu, t) \in T_0(F)$. This is indeed the case:

**Proposition 6.5.** For $(\lambda, \mu, t) \in T_0(F)$,

\[\rho_0[\lambda, \mu, t] = [\det t, \lambda^2, \lambda \mu t] = [(\lambda \mu)^{-1} \det t, \lambda, t]\]

and

\[\rho_0^2[\lambda, \mu, t] = [\mu(\det t)^{-1}, (\lambda \mu)^{-1}, (\det t)^{-1} t] = [\mu, (\lambda \mu)^{-1} \det t, t].\]

**Proof.** Check that the equations (6.4) have the following solution:

\[\lambda' = \det t, \quad \mu' = \lambda^2, \quad \lambda'' = \mu(\det t)^{-1}, \quad \mu'' = (\lambda \mu)^{-1} \cdot \]

\[\square\]

We next show that the torus $T$ is also stable under the trialitarian automorphism $\rho_0$ associated to the split Okubo algebra $(C, \triangle)$. We recall that the multiplicity $\triangle$ is the Petersson twist $\otimes_{s_K}$ of the multiplication $\diamond$ by the automorphism $s_K$ of the para-Zorn algebra.

**Proposition 6.6.** For $(\lambda, \mu, t) \in T_0(F)$ we have

\[s_K^{-1} \circ (\lambda, \mu, t) \circ s_K = (\lambda, \mu, \pi^{-1} t \pi),\]

and

\[\rho_0[\lambda, \mu, t] = [(\lambda \mu)^{-1} \det t, \lambda, \pi^{-1} t \pi], \quad \rho_0^2[\lambda, \mu, t] = [\mu, (\lambda \mu)^{-1} \det t, \pi^{-2} t \pi^2].\]

**Proof.** The first equation is easily checked by direct computation. The others follow from Proposition 4.9 and Proposition 6.5. \[
\square\]
For the algebraic group $\text{Spin}_8$ we consider the torus
\[ T'_0 = G^4_m = G_m \times \text{Diag}_3 \]
as a subtorus of the torus $T_0$ defined in (6.1) through the embedding $(\lambda, t) \mapsto (\lambda, \lambda^{-1}, t)$. The torus $T'_0$ is a maximal torus of $O^+_{8}$ through the action $\otimes$ defined in (6.2). The spinor norm of the element $(\lambda, t)$ is the class of $\lambda \det t$ in $F^\times / F^\times 2$. Thus $(\lambda, t)$ lifts to an element in $\text{Spin}_8(F)$ if and only if the product $\lambda \det t$ is a square in $F$. By Lemma 6.3, the equation
\[ (\lambda, t) \otimes (\xi \odot \eta) = ((\lambda', t') \otimes \xi) \odot ((\lambda'', t'') \otimes \eta) \]
holds for all $\xi, \eta \in C$ if
\[ \lambda' \lambda'' = 1, \quad t' = \lambda'' t, \quad t'' = \lambda'^{-1} t, \quad \text{and} \quad \lambda' \lambda'^{-1} = \det t. \]
These equations imply
\[ \lambda'' = \lambda^{-1} \det t \quad \text{and} \quad \lambda'^{-1} = (\lambda \det t)^{-1}. \]
Thus the set of triples $((\lambda, t), (\lambda', t'), (\lambda'', t''))$ satisfying the above conditions is a maximal torus $\tilde{T}$ in $\text{Spin}_8$. For the twisted composition $\triangle$ we have
\[ \lambda \lambda'' = 1, \quad t' = \lambda'' t, \quad t'' = \lambda'^{-1} t, \quad \text{and} \quad \lambda' \lambda'^{-1} = \det t \]
instead of (6.7).

The natural coverings $\tilde{T} \to T'_0$ and $T'_0 \to T$ give natural embeddings of the character groups $X(T) \hookrightarrow X(T'_0) \hookrightarrow X(\tilde{T})$. Letting $t = \text{Diag}(t_1, t_2, t_3)$, we define characters $\varepsilon_1, \ldots, \varepsilon_4$ of $T'_0$ generating $X(T'_0)$ by
\[ \varepsilon_1 : (\lambda, t) \mapsto t_1, \quad \varepsilon_2 : (\lambda, t) \mapsto t_2^{-1}, \quad \varepsilon_3 : (\lambda, t) \mapsto t_3^{-1}, \quad \text{and} \quad \varepsilon_4 : (\lambda, t) \mapsto \lambda. \]
View each $\varepsilon_i$ as a character of $\tilde{T}$ through the embedding $X(T'_0) \hookrightarrow X(\tilde{T})$. Using additive notation, it follows from (6.8) that
\[ \rho_0(\varepsilon_1) = \frac{1}{2}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 - \varepsilon_4) \]
\[ \rho_0(\varepsilon_2) = \frac{1}{2}(\varepsilon_1 + \varepsilon_2 - \varepsilon_3 + \varepsilon_4) \]
\[ \rho_0(\varepsilon_3) = \frac{1}{2}(\varepsilon_1 - \varepsilon_2 + \varepsilon_3 + \varepsilon_4) \]
\[ \rho_0(\varepsilon_4) = \frac{1}{2}(\varepsilon_1 - \varepsilon_2 - \varepsilon_3 + \varepsilon_4). \]
Formulas (6.10) imply that the action of $\rho_0$ on the vector space $\mathbb{R}^4 = X(\tilde{T}) \otimes_{\mathbb{Z}} \mathbb{R} = X(T'_0) \otimes_{\mathbb{Z}} \mathbb{R} = X(T) \otimes_{\mathbb{Z}} \mathbb{R}$ is given by the matrix
\[ R_\phi = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{pmatrix}. \]
In view of (6.9) the action of $\rho_\delta$ is given by

$$R_\delta = \frac{1}{2} \begin{pmatrix}
-1 & -1 & 1 & 1 \\
1 & -1 & 1 & -1 \\
-1 & -1 & 1 & -1 \\
1 & 1 & 1 & -1
\end{pmatrix}. $$

Recall from [Bou81] that the simple roots of $\text{Spin}_8$ are

$$\alpha_1 = \varepsilon_1 - \varepsilon_2, \quad \alpha_2 = \varepsilon_2 - \varepsilon_3, \quad \alpha_3 = \varepsilon_3 - \varepsilon_4, \quad \text{and} \quad \alpha_4 = \varepsilon_3 + \varepsilon_4.$$ 

Let also $\tilde{\alpha} = \alpha_1 + 2\alpha_2 + \alpha_3 + \alpha_4$ be the highest root.

**Corollary 6.11.** The action of $\rho_\delta$ on $X(\tilde{T})$ permutes the simple roots as follows:

$$\rho_\delta : \alpha_1 \mapsto \alpha_3, \quad \alpha_3 \mapsto \alpha_4, \quad \alpha_4 \mapsto \alpha_1, \quad \alpha_2 \mapsto \alpha_2.$$ 

The action of $\rho_\delta$ is given by

$$\alpha_1 \mapsto \alpha_3 + \alpha_2, \quad \alpha_3 \mapsto \alpha_4 + \alpha_2, \quad \alpha_4 \mapsto \alpha_1 + \alpha_2, \quad \alpha_2 \mapsto -\tilde{\alpha}.$$

**Proof.** This is readily verified by computation using the matrices $R_\xi$ and $R_\delta$. 

**Remark 6.12.** Viewing the Weyl group $W$ of $\text{Spin}_8$ as the subgroup of the real orthogonal group $O_4(\mathbb{R})$ generated by the reflections with respect to the roots $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, we conclude from the above discussion that the trialitarian automorphism $\overrightarrow{\rho}$ (resp. $\overleftarrow{\rho}$) of $W$ induced by $\rho_\delta$ (resp. $\rho_\Delta$) is given by conjugation in $O_4(\mathbb{R})$ by the matrix $R_\xi$ (resp. $R_\Delta$). In particular the action $\overrightarrow{\rho}$ induced by the graph automorphism $\rho$ (see the introduction) is the same as $\overrightarrow{\rho}$.

7. Trialitarian Automorphisms of $\text{PGO}_8^+$

In this section we show that every trialitarian automorphism of $\text{PGO}_8^+$ over an algebraically closed field is conjugate to either $\rho_\delta$ or $\rho_\Delta$, which we call the standard trialitarian automorphisms.

We denote by $\text{Aut}(\text{PGO}_8^+ \supset T)$ the subgroup of $\text{Aut}(\text{PGO}_8^+)$ consisting of automorphisms which map the maximal split torus $T$ introduced in Section 6 to itself. It follows from Theorem 3.6 that every trialitarian automorphism of $\text{PGO}_8^+$ is conjugate over a separable closure to an automorphism which belongs to $\text{Aut}(\text{PGO}_8^+ \supset T)$. Let $N_0 \subset \text{GO}_8^+$ be the normalizer of the torus $T_0$ (see (6.1)) and let $N = N_0 / G_m \subset \text{PGO}_8^+$. Let further

$$W = N_0 / T_0 = N / T$$

be the Weyl group of $\text{PGO}_8^+$. For $\alpha \in \text{Aut}(\text{PGO}_8^+ \supset T)(F)$, we have obviously $\alpha(N) = N$, hence $\alpha$ induces an automorphism $\overrightarrow{\alpha} \in \text{Aut}(W)$ defined by

$$\overrightarrow{\alpha}([g] \cdot T) = \alpha([g]) \cdot T \quad \text{for} \ g \in N_0.$$

In particular, for $g \in N_0(F)$, we have $\text{Int} \ (\overrightarrow{[g]}) \in \text{Aut}(\text{PGO}_8^+ \supset T)(F)$, and

$$\text{Int} \ (\overrightarrow{[g]}) = \text{Int}(gT_0).$$

\[\text{We could as well have discussed the case of Spin}_n.\]
By Proposition 3.9, $\overline{\rho}_c$ and $\overline{\rho}_d$ are outer automorphisms of order 3 of $W$.

**Proposition 7.1.** Every outer automorphism of order 3 of $W$ is conjugate in $\text{Aut}(W)$ to exactly one of $\overline{\rho}_c$ or $\overline{\rho}_d \in \text{Aut}(W)$.

**Proof.** This is well-known; see for instance [KT10, Proposition 5.8] or [KT12, Theorem 5.11].

**Proposition 7.2.** Let $F$ be an arbitrary field and $\phi \in \text{Aut}(\text{PGO}^+_8(F))$ be a trialitarian automorphism which admits an invariant maximal split torus over $F$. Then $\phi$ is conjugate in $\text{Aut}(\text{PGO}^+_8)(F)$ to an automorphism $\beta \in \text{Aut}(\text{PGO}^+_8 \supset T)(F)$ such that $\overline{\beta} = \overline{\rho}_c$ or $\overline{\rho}_d$.

**Proof.** Two maximal split tori in $\text{PGO}^+_8$ are conjugate over $F$. Thus $\phi$ is conjugate to a trialitarian automorphism $\beta \in \text{Aut}(\text{PGO}^+_8 \supset T)(F)$. Its image $\overline{\beta} \in \text{Aut}(W)$ is well defined and is also trialitarian. Note that $N(F) \to W$ is surjective (because $T$ is split) and therefore so is $N(F) \times \mathfrak{S}_3 \to \text{Aut}(W)$. By Proposition 7.1, one can find $g \in N(F) \times \mathfrak{S}_3 \subset \text{Aut}(\text{PGO}^+_8)(F)$ such that

$$\overline{\beta} = \text{Int}(gT) \circ \overline{\rho}_c \circ \text{Int}(gT)^{-1} \quad \text{or} \quad \overline{\beta} = \text{Int}(gT) \circ \overline{\rho}_d \circ \text{Int}(gT)^{-1}.$$  

Let $\beta_0 = \text{Int}([g])^{-1} \circ \beta \circ \text{Int}([g])$. Then $\beta_0$ is conjugate to $\beta$, and $\overline{\beta_0} = \overline{\rho}_c$ or $\overline{\rho}_d$.

From now on we only consider automorphisms $\beta \in \text{Aut}(\text{PGO}^+_8 \supset T)(F)$ such that $\overline{\beta} = \overline{\rho}_c$ or $\overline{\rho}_d$.

**Proposition 7.3.** Let $\beta \in \text{Aut}(\text{PGO}^+_8 \supset T)(F)$ be a trialitarian automorphism such that $\overline{\beta} = \overline{\rho}_d$. There exists an extension $\tilde{F} = F(\xi, \eta)$ of $F$, with $\xi^3 \in F^x$ and $\eta^3 \in F^x$, such that $\beta$ and $\rho_d$ are conjugate in $\text{Aut}(\text{PGO}^+_8)(\tilde{F})$.

**Proof.** Since $\overline{\beta} = \overline{\rho}_d$, we have $\beta \rho_d^{-1} = 1_W$, thus by Proposition 3.9 we have $\beta = \text{Int}([\lambda, \mu, t]) \circ \rho_d$ for some $(\lambda, \mu, t) \in T_0(F)$. Let $t = \text{Diag}(t_1, t_2, t_3)$. In an algebraic closure of $F$, choose $\xi$ such that $\xi^3 = \det t$ and $\eta$ such that $\eta^3 = \lambda^{-1} \mu t_1 t_2$. Then $\beta F = F(\xi, \eta)$. Let also

$$u = \text{Diag}(t_3, \xi^t t_2^{-1}, \xi) \in \text{Diag}_3(\tilde{F})$$

Consider $[\eta, \mu^{-1} \xi, u] \in T(\tilde{F})$. By Proposition 6.6 we have

$$\rho_d [\eta, \mu^{-1} \xi, u] = [(\mu^{-1} \xi)^{-1} \det u, \eta, \pi^{-1} u \bar{x}]$$

On the other hand,

$$[\eta, \mu^{-1} \xi, u] \cdot [\lambda, \mu, t] = [\lambda \eta, \xi, ut]$$

and computation shows that

$$\xi (\mu^{-1} \xi \eta)^{-1} \det u = \lambda \eta \quad \text{and} \quad \xi (\pi^{-1} u \bar{x}) = ut.$$  

Therefore, we have

$$[\eta, \mu^{-1} \xi, u] \cdot [\lambda, \mu, t] \cdot \rho_d [\eta, \mu^{-1} \xi, u]^{-1} = 1 \quad \text{in} \quad \text{PGO}^+_8(\tilde{F}),$$

**References:**

hence, letting \( \tau = \text{Int} ([\eta, \mu^{-1} \xi, u]) \),
\[
\tau \circ (\text{Int} ([\lambda, \mu, t]) \circ \rho_\delta) \circ \tau^{-1} = \rho_\delta.
\]
Since \( \beta = \text{Int} ([\lambda, \mu, t]) \circ \rho_\delta \), this shows that \( \beta \) and \( \rho_\delta \) are conjugate. \( \Box \)

We next consider the case where \( \overline{\beta} = \overline{\rho_\delta} \).

**Proposition 7.4.** Let \( \beta \in \text{Aut}(\PGO_4^+ \times T)(F) \) be such that \( \overline{\beta} = \overline{\rho_\delta} \) and \( \beta^3 = 1 \). If \( F \) does not contain any \( \omega \neq 1 \) such that \( \omega^3 = 1 \) (in particular if \( F \) has characteristic 3), then \( \beta \) is conjugate in \( \text{Aut}(\PGO_4^+)(F) \) to \( \rho_\delta \). If \( F \) contains an element \( \omega \neq 1 \) such that \( \omega^3 = 1 \), then \( \beta \) is conjugate to \( \rho_\sigma \) or to \( \rho_\phi \).

**Proof.** As in Proposition 7.3, we may assume that \( \beta = \text{Int} ([\lambda, \mu, t]) \circ \rho_\sigma \) for some \( (\lambda, \mu, t) \in T(F) \). Since \( \beta^3 = 1 \), we must have
\[
\text{Int} ([\lambda, \mu, t]) \circ \rho_\sigma [\lambda, \mu, t] \cdot \rho_\sigma^2 [\lambda, \mu, t] = 1,
\]
By Proposition 6.5, we have
\[
[\lambda, \mu, t] \cdot \rho_\sigma [\lambda, \mu, t] \cdot \rho_\sigma^2 [\lambda, \mu, t] = [\det t, \det t, t^3].
\]
Therefore, letting \( t = \text{Diag}(t_1, t_2, t_3) \), we see that the condition \( \beta^3 = 1 \) implies
\[
t_1^3 = t_2^3 = t_3^3 = t_1 t_2 t_3.
\]
Let \( \omega = t^3 t_1^{-1} \). We then have \( \omega^3 = 1 \) and \( t_3 = \omega^2 t_1 \), hence \( t = t_1 w \) where \( w = \text{Diag}(1, \omega, \omega^2) \). For \( u = \text{Diag}(\lambda, \mu, \mu) \), computations using Proposition 6.5 show that
\[
[\mu, t_1, u] \cdot [\lambda, \mu, t_1 w] \cdot \rho_\sigma [\mu, t_1, u]^{-1} = [t_1, t_1, t_1 w] = [1, 1, w].
\]
Therefore, letting \( \sigma = \text{Int} ([\mu, t_1, u]) \) we have
\[
(7.5) \quad \sigma \circ \text{Int} ([\lambda, \mu, t_1 w]) \circ \rho_\sigma \circ \sigma^{-1} = \text{Int} ([1, 1, w]) \circ \rho_\sigma.
\]
If \( \omega = 1 \), then \( w = 1 \), and (7.5) shows that \( \beta \) is conjugate to \( \rho_\sigma \). If \( \omega \neq 1 \), observe that \( (1, 1, w) \in T_0(F) \) acts on the quadratic space of Zorn matrices as the automorphism \( s_\omega^{-1} \) of (5.6): we have
\[
[1, 1, w] \circ \xi = s_\omega^{-1} (\xi) \quad \text{for all } \xi \in C.
\]
Therefore, for the Petersson twist \( \nabla = o_s \), we have \( \text{Int} ([1, 1, w]) \circ \rho_\sigma = \rho_\nabla \) by Proposition 4.9(ii), and (7.5) shows that \( \beta \) is conjugate to \( \rho_\nabla \). But we saw in Lemma 5.12 that \( \nabla \) is isomorphic to \( \triangle \), hence by Lemma 4.7 the trialitarian automorphisms \( \rho_\nabla \) and \( \rho_\sigma \) are conjugate. Therefore, \( \beta \) is conjugate to \( \rho_\delta \). \( \Box \)

We get to the main result:

**Theorem 7.6.** Let \( n \) be a 3-Pfister form over a field \( F \), let \( G \) be either \( \PGO_\pm(n) \) or \( \text{Spin}(n) \) and let \( \phi \in \text{Aut}(G)(F) \) be any trialitarian automorphism. There is a finite field extension \( \overline{F} \) of \( F \) splitting \( G \) and such that \( \phi \) is conjugate over \( \overline{F} \) to one of the two standard trialitarian automorphisms of \( \PGO_4^+ \) or \( \text{Spin}_4 \).
Proof. After enlarging $F$ to split $G$, the claim follows from Propositions 7.2, 7.3, 7.4 and Theorem 3.6.

Thus there are two types of trialitarian automorphisms of $\text{PGO}^+(n)$ or $\text{Spin}(n)$. We call those conjugate over a field extension of $F$ to $\rho_\diamond$ of octonion type and the others of Okubo type. Using the correspondence between trialitarian automorphisms and symmetric compositions in Theorem 4.8, we readily derive the classification of 8-dimensional symmetric composition algebras, first established by Petersson [Pet69, Satz 2.7] over fields of characteristic different from 2 and 3, and by Elduque-Pérez [EP96] over arbitrary fields:

**Theorem 7.7.**

(i) Let $F$ be an algebraically closed field. The para-Zorn algebra $(C, \diamond, n)$ and the split Okubo algebra $(C, \triangle, n)$ are non-isomorphic symmetric composition algebras.

(ii) For every 8-dimensional symmetric composition algebra $(S, \star, n)$ over an arbitrary field $F$ there is a finite field extension $\tilde{F}$ of $F$ such that $(S, \star, n) \otimes F \tilde{F}$ is either isomorphic to the para-Zorn algebra or to the split Okubo algebra over $\tilde{F}$.

It follows from Theorem 7.7 that symmetric composition algebras of dimension 8 over an arbitrary base field can be divided into two types, according to their isomorphism class over an algebraic closure. Those that after scalar extension are isomorphic to the split Okubo algebra are called Okubo algebras. As we shall see in Theorem 8.2, those that after scalar extension are isomorphic to the para-Zorn algebra are the para-octonion algebras defined after Lemma 5.4. Thus:

**Corollary 7.8.** Symmetric composition algebras of dimension 8 are either para-octonion algebras or Okubo algebras.

8. AUTOMORPHISMS OF SYMMETRIC COMPOSITIONS

Let $(S, \star, n)$ be a symmetric composition algebra of dimension 8 with associated trialitarian automorphism $\rho_\star$. We let $\text{Aut}(\star)$ or $\text{Aut}(S, \star, n)$ denote the $F$-algebraic group of automorphisms of $\star$, whose group of $F$-rational points $\text{Aut}(\star)$ consists of the automorphisms $f: \star \to \star$. This group is related to $\rho_\star$ as follows:

**Proposition 8.1.** The canonical map $\text{GO}^+(n) \to \text{PGO}^+(n)$ induces an isomorphism from $\text{Aut}(\star)$ to the subgroup $\text{PGO}^+(n)^{\rho_\star}$ of $\text{PGO}^+(n)$ fixed under the trialitarian automorphism $\rho_\star$.

**Proof.** See [CKT12, Theorem 6.6].

In view of Theorem 7.6 we have two types of subgroups fixed under trialitarian automorphisms, those which are automorphism groups of para-octonion algebras and those which are automorphism groups of Okubo algebras.

We next determine the type of the algebraic group $\text{Aut}(\star)$, and show that this group determines the composition $\star$ up to isomorphism or anti-isomorphism.
Type I: Para-octonion algebras. Let $(C, \cdot, n)$ be an octonion algebra over an arbitrary field $F$, and let $\circ$ be the corresponding para-octonion multiplication on $(C, n)$ (see Lemma 5.4).

Theorem 8.2. (i) The octonion algebra $(C, \cdot, n)$ and the para-octonion algebra $(C, \circ, n)$ have the same groups of automorphisms.

(ii) Forms of para-octonion algebras are para-octonion algebras.

(iii) The group of automorphisms of any para-octonion algebra is an algebraic group of type $G_2$.

Proof. The equality $\text{Aut}(\circ) = \text{Aut}(\cdot)$ is proved in [KMRT98, (34.4)]. Claim (i) and (ii) now follow from the fact that algebraic groups of type $G_2$ are the automorphism groups of octonion algebras (see i.a. [SV00, §1.7]). □

We collect more related known results in the next proposition. Sketches of proofs are given for completeness.

Proposition 8.3. Let $(C, \cdot, n)$ and $(C', \cdot', n')$ be octonion algebras, and let $\circ$ resp. $\circ'$ be the corresponding para-octonion multiplications. The following properties are equivalent:

(i) $(C, \cdot, n)$ and $(C', \cdot', n')$ are isomorphic.

(ii) $(C, \circ, n)$ and $(C', \circ', n')$ are isomorphic.

(iii) The norms $n$ and $n'$ are isometric.

(iv) The algebraic groups $\text{Aut}(\cdot)$ and $\text{Aut}(\cdot')$ are isomorphic.

(v) The algebraic groups $\text{Aut}(\circ)$ and $\text{Aut}(\circ')$ are isomorphic.

(vi) The algebraic groups $\text{PGO}^+(n)$ and $\text{PGO}^+(n')$ are isomorphic.

Proof. The equivalence of (i) and (ii) is proved in [KMRT98, (34.4)] and the equivalence of (i) and (iii) is for example in [KMRT98, (33.19)] or [SV00, §4.3]. Claims (i), (ii), and (iii) imply (iv), (v), and (vi) are clear, and the equivalence of (iv) and (v) is Theorem 8.2. If $\text{Aut}(\cdot) \xrightarrow{\sim} \text{Aut}(\cdot')$, then $\text{Aut}(\cdot)(F) \xrightarrow{\sim} \text{Aut}(\cdot')(F)$ and (i) follows by [vdBS59, (2.3)]. The implication (iv) $\implies$ (i) can also be obtained by a cohomological argument. Let $G = \text{Aut}(\cdot)$ and let $G' = \text{Aut}(\cdot')$. Forms of $(C, \cdot, n)$ are classified by $H^1(F, G)$ and forms of $G$ by $H^1(F, \text{Aut}(G))$, for étale or fppf-cohomology. The algebra $(C', \cdot', n')$, being a form of $(C, \cdot, n)$, gives a cohomology class $[\xi] \in H^1(F, G)$ and $G'$ is the twisted group $^G\xi$. If $G' \xrightarrow{\sim} G$, then the image of $\xi$ in $H^1(F, \text{Aut}(G))$ under the map $H^1(F, G) \to H^1(F, \text{Aut}(G))$ induced by inner conjugation $\iota: G \to \text{Aut}(G)$ is trivial. However the map $\iota$ is an isomorphism for groups of type $G_2$. Thus $[\xi]$ is trivial and $(C, \cdot, n) \xrightarrow{\sim} (C', \cdot', n')$. The implication (vi) $\implies$ (iii) can also be obtained by a similar cohomological argument. Namely, let $[\xi] \in H^1(F, \text{Aut}(\cdot))$ correspond to $(C', \cdot', n')$ and assume that its image in $H^1(F, \text{PGO}^+(n))$ is trivial. First we recall that $\text{Aut}(\cdot)$ is a subgroup of $\text{GO}^+(n)$ and that similar Pfister forms are isomorphic. Thus it suffices to consider the norm forms $n$.
and \( n' \) up to similarity and we are reduced to show that the image of \([\xi]\) in \(H^1(F, \text{GO}^+(n))\) is trivial. We next remark that the map \(H^1(F, \text{GO}^+(n)) \to H^1(F, \text{PGO}^+(n))\) induced by the projection has trivial kernel. This follows from the fact that in the exact sequence
\[
H^1(F, \mathbb{G}_m) \to H^1(F, \text{GO}^+(n)) \to H^1(F, \text{PGO}^+(n))
\]
induced by
\[
1 \to \mathbb{G}_m \to \text{GO}^+(n) \to \text{PGO}^+(n) \to 1,
\]
the set \(H^1(F, \mathbb{G}_m)\) is trivial by Hilbert 90. Let \( H = \text{PGO}^+(n) \). It remains to verify that \( H^1(F, H) \to H^1(F, \text{Aut}(H)) \) has trivial kernel. Since \( \text{Aut}(H)(F) = (H \rtimes \mathbb{S}_3)(F) \to \mathbb{S}_3(F) = \mathbb{S}_3 \) is surjective this follows from the cohomology sequence
\[
(H \rtimes \mathbb{S}_3)(F) \to \mathbb{S}_3(F) \to H^1(F, H) \to H^1(F, \text{Aut}(H))
\]
induced by the exact sequence
\[
1 \to H \to H \rtimes \mathbb{S}_3 \to \mathbb{S}_3 \to 1.
\]
Thus, the composition
\[
H^1(F, \text{Aut}(\cdot)) \to H^1(F, \text{GO}^+(n)) \to H^1(F, H) \to H^1(F, \text{Aut}(H))
\]
has trivial kernel and we are done. \(\square\)

**Corollary 8.4.** Forms of para-Zorn algebras are para-octonion algebras. In particular there is up to isomorphism a unique para-octonion algebra with given 3-Pfister form \( n \).

**Proof.** In view of Proposition 8.3 the classification of para-octonion algebras is equivalent to the classification of octonion algebras. Thus the claim follows from the known fact that forms of octonion algebras are octonion algebras. \(\square\)

For Okubo algebras we treat separately fields of characteristic not 3 and fields of characteristic 3.

**Type II: Okubo algebras in characteristic different from 3.** We distinguish two subtypes IIa and IIb, depending on whether the base field contains a primitive cube root of unity or not. Suppose first \( F \) contains a primitive cube root of unity \( \omega \). Let \( A \) be a central simple \( F \)-algebra of degree 3. For the reduced characteristic polynomial of \( a \in A \), we use the notation
\[
X^3 - \text{Trd}(a)X^2 + \text{Srd}(a)X - \text{Nrd}(a)1,
\]
so Trd is the reduced trace map on \( A \), Srd is the reduced quadratic trace map, and Nrd is the reduced norm. Let \( A^0 \subset A \) be the kernel of Trd.

**Proposition 8.5.** Assume that \( F \) contains a primitive cubic root of unity.
(i) The multiplication $\star$ on $A^0$ given by

$$x \star y = \frac{yx - \omega xy}{1 - \omega} - \frac{1}{3} \text{Trd}(xy)$$

(8.6)

together with the quadratic form

$$n(x) = -\frac{1}{7} \text{Srd}(x).$$

(8.7)

define an Okubo algebra on the quadratic space $(A^0, n)$, which is hyperbolic.

(ii) The symmetric composition $(A^0, \star, n)$ and the algebra $A$ have the same groups of automorphisms.

Proof. See for instance [KMRT98, (34.19), (34.25)] and [KMRT98, (34.35)]. □

If $F$ does not contain a primitive cube root of unity $\omega$, we denote $K = F(\omega)$ the separable quadratic extension generated by $\omega$. Let $B$ be a central simple $K$-algebra of degree 3 with a unitary involution $\tau$ leaving $F$ fixed and let $\text{Sym}(B, \tau)^0$ be the $F$-vector space of $\tau$-symmetric elements of reduced trace zero.

**Proposition 8.8.** Assume that $F$ does not contain a primitive cube root of unity $\omega$.

(i) The quadratic space $(\text{Sym}(B, \tau)^0, n)$ is a 3-fold Pfister quadratic space that becomes hyperbolic over $K$, and the restriction of the product $\star$ and of the norm $n$ of $B$ to $\text{Sym}(B, \tau)^0$ define a symmetric composition on this space.

(ii) The symmetric composition $((\text{Sym}(B, \tau)^0, \star, n)$ and the algebra with involution $(B, \tau)$ have the same groups of automorphisms.

Proof. See [KMRT98, (34.35)]. □

**Corollary 8.9.**

(i) If $F$ contains a primitive cube root of unity $\omega$, Okubo algebras over $F$ are in functorial bijective correspondence with central simple algebras over $F$ of degree 3.

(ii) Let $K = F(\omega)$ if $F$ does not contain $\omega$. Okubo algebras over $F$ are in functorial bijective correspondence with central simple algebras of degree 3 over $K$, with unitary $K/F$-involution.

(iii) The group of automorphisms of an Okubo algebra over a field $F$ of characteristic not 3 is an algebraic group of inner type $^1A_2$ if $F$ contains a primitive cube root of unity $\omega$ and of outer type $^2A_2$ if $F$ does not contain $\omega$. In the latter case it becomes a group of inner type over $F(\omega)$.

Proof. It follows from Proposition 8.5 and Proposition 8.8 that the classification of Okubo algebras over $F$ is equivalent to the classification of central simple algebras of degree 3 over $F$, resp. of central simple algebras of degree 3 over $F(\omega)$. In particular they have isomorphic groups of automorphisms. □
Proposition 8.10. Two Okubo algebras \((S,\star,n)\) and \((S',\star',n')\) over a field \(F\) of characteristic not 3 are isomorphic or anti-isomorphic if and only if their groups of automorphisms \(\text{Aut}(\star)\) and \(\text{Aut}(\star')\) are isomorphic.

Proof. In view of Corollary 8.9, the claim follows from the corresponding fact for central simple algebras, which is certainly known. However, since the only reference we could find was at the level of Lie algebras ([Jac79, Chap. X, §4]), we give here a cohomological proof in the spirit of the proof of Proposition 8.3. Assume that \(F\) contains a cube root of unity. Let \(A\) and \(A'\) be central simple \(F\)-algebras of degree 3 and let \(G = \text{Aut}(A)\), \(G' = \text{Aut}(A')\). First assume that \(A\) is split, so that \(G = \text{PGL}_3\) and \(\text{Aut}(G) = G \rtimes \mathfrak{S}_2\). Let \([\xi] \in H^1(F,G)\) be a cocycle defining \((S',\star',n')\) or equivalently \(A'\). If \(G' \to G\), the image of \([\xi]\) in \(H^1(F,\text{Aut}(G))\) is trivial. From the exact sequence

\[
G(F) \to \text{Aut}(G)(F) \xrightarrow{\lambda} \mathfrak{S}_2 \xrightarrow{\nu} H^1(F,G) \to H^1(F,\text{Aut}(G))
\]

we conclude that \([\xi]\) lies in the image of \(\nu\). But \(\lambda\) is surjective because we are in the split case, hence \([\xi] = 1\) and \(A' \to A\). If \(A\) is not split, \(\lambda\) is not surjective and the kernel of \(H^1(F,G) \to H^1(F,\text{Aut}(G))\) consists of two elements. Thus we have exactly two choices for \([\xi]\), resp. for the algebra \(A'\). On the other hand we have two candidates, the algebra \(A\) and the opposite algebra \(A^{\text{op}}\). Obviously \(A\) and \(A^{\text{op}}\) have the same group of automorphisms and they are not isomorphic: if \(A \to A^{\text{op}}\), the class of \(A\) in the Brauer group of \(F\) would have order smaller than or equal to 2. But since \(A\) is of degree 3 and is not split, its Brauer class has order 3. A similar argument can be given if \(F\) does not contain a cube root of unity. \(\square\)

Type III: Okubo algebras in characteristic 3. The algebraic group of automorphisms of the split Okubo algebra in characteristic 3 is more intricate than in characteristic different from 3. In particular the group is not smooth. Its set of \(F\)-points was computed in [Eld99, Theorem 7], see also [Tit59, §10] and [GL83, (9.1)] for the computation of fixed \(F\)-points under triality. The next sections are devoted to the computation of this algebraic group of automorphisms.

Remark 8.11. We conclude this section by verifying that the graph automorphism \(\rho\) and the trialitarian automorphism \(\rho_\circ\) defined through the para-Zorn algebra \((C,\circ,n)\) are conjugate in \(\text{Aut}(\text{PGO}_4^+)(F)\). We know already that the induced actions on the Weyl group are conjugate (see Remark 6.12). Thus, by Proposition 7.4 \(\rho\) is conjugate to \(\rho_\circ\) or to \(\rho_\Delta\), the automorphism associated to the split Okubo algebra. Since the subgroup fixed by \(\rho\) is of type \(G_2\), while the fixed subgroup of \(\rho_\Delta\) is a simple group of type \(A_2\) if the characteristic is different from 3 (Proposition 8.10), or not smooth if the characteristic is 3 (Section 10), we obtain that \(\rho\) is conjugate to \(\rho_\circ\).
9. Idempotents of Composition Algebras in Characteristic 3

For simplicity we assume throughout this section that the base field $F$ has characteristic 3, even if some of the results hold for arbitrary fields. Let $(H, \cdot, n)$ be a Hurwitz algebra and let $(H, \circ, n)$ be the associated para-Hurwitz symmetric composition (see Lemma 5.4). The identity element of $(H, \cdot)$ also plays a special role for the associated para-Hurwitz algebra. It is an idempotent (i.e., $1 \circ 1 = 1$) and satisfies $1 \circ x = x \circ 1 = -x$ for $x \in H$ such that $b_n(1, x) = 0$.

**Definition 9.1.** An idempotent $e$ of a symmetric composition $(S, \star, n)$ is called a para-unit if $e \star x = x \star e = -x$ for $x \in S$ such that $b_n(e, x) = 0$.

We recall that a symmetric composition is para-Hurwitz if and only if it admits a para-unit (see [KMRT98, (34.8)]).

Let now $(S, \star, n)$ be an Okubo algebra.

**Definition 9.2.** We say that an idempotent $e$ of $(S, \star, n)$ is quaternionic if $e$ is the para-unit of a para-quaternion subalgebra of $(S, \star, n)$. We say that an idempotent $e$ of $(S, \star, n)$ is quadratic if $e$ is not quaternionic, and $e$ is the para-unit of a para-quadratic subalgebra of $(S, \star, n)$.

We give examples of such elements in the split Okubo algebra $(C, \triangle, n)$, as described in (5.5) using Zorn matrices:

**Lemma 9.3.**

(i) The Zorn matrix $e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is a quadratic idempotent.

(ii) The Zorn matrix $\hat{e} = \begin{pmatrix} 1 & -(1,1,1) \\ (1,1,1) & 1 \end{pmatrix}$ is a quaternionic idempotent.

**Proof.** (i) The element $e$ is the para-unit of the para-quadratic subalgebra

$$Q = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \mid \alpha, \beta \in F \right\}.$$

Moreover $e$ cannot be the para-unit of any para-quaternion subalgebra, as this would imply that this subalgebra (with nondegenerate norm!) would be contained in the centralizer of $e$. But the centralizer $\text{Cent}_{(C, \Delta, n)}(e) = \{ x \in C \mid x \Delta e = e \Delta x \}$ of $e$ is the set of Zorn matrices

$$Z = \left\{ \begin{pmatrix} \alpha \\ (y,y,y) \end{pmatrix} \begin{pmatrix} \alpha & x \\ x, x \end{pmatrix} \mid \alpha,\beta,x,y \in F \right\},$$

and the restriction of the norm to this centralizer is degenerate.

(ii) In the same vein, $\hat{e}$ is an idempotent and the set of Zorn matrices

$$H = \left\{ \begin{pmatrix} \alpha \\ (x,y,z) \end{pmatrix} \begin{pmatrix} \alpha & -(x,y,z) \\ (x,y,z) \end{pmatrix} \mid \alpha,\beta,x,y,z \in F \right\}$$

is a para-quaternion subalgebra with para-unit $\hat{e}$. \qed

Documenta Mathematica 18 (2013) 413–468
We next give another construction of an Okubo algebra with a quaternionic idempotent:

**Example 9.4.** Let \( Q = \text{Mat}_2(F) \) be the split quaternion algebra (multiplication denoted by juxtaposition) with conjugation:

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.
\]

The split octonion algebra can be constructed through the Cayley-Dickson doubling process as \( C = Q \oplus Qv \), with multiplication, conjugation and norm given by

\[
(a + bv) \cdot (c + dv) = (ac + \overline{d}b) + (da + b\overline{c})v,
\]

(9.5)

\[
a + bv = \overline{a} - bv,
\]

\[
n(a + bv) = \det(a) - \det(b).
\]

The element \( w = (\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}) \) satisfies \((w - 1)^2 = 0, w\overline{w} = 1, \overline{w} = w^2 = -1 - w\), and the map

\[
\vartheta: a + bv \mapsto a + (wb)v, \quad a, b \in Q
\]

(9.6)

is an order 3 automorphism of \((C, \cdot)\). Let \( \star = \cdot \vartheta \) be the associated Petersson twist, i.e.,

\[
x \star y = \vartheta(\overline{x}) \cdot \vartheta^2(\overline{y}) \quad \text{for } x, y \in C.
\]

(9.7)

Obviously, \( e = 1 \) is an idempotent of the symmetric composition \((C, \star, n)\), and \((Q, \star, n)\) is a para-quaternion subalgebra, since the restriction of \( \vartheta \) to \( Q \) is the identity, with para-unit \( e \). Thus \( e \) is also quaternionic.

**Lemma 9.8.** The symmetric composition algebra \((C, \star, n)\) is an Okubo algebra.

**Proof.** In view of Corollary 7.8 it suffices to show that \((C, \star, n)\) is not a para-octonion algebra. The decomposition \( C = Q \oplus Qv \) is a \( \mathbb{Z}_2 \)-grading of both \((C, \cdot)\) and \((C, \star)\). Hence the commutative center

\[
K(C, \star) = \{ x \in C \mid x \star y = y \star x, \ \forall y \in C \}
\]

is graded: \( K(C, \star) = (K(C, \star) \cap Q) \oplus (K(C, \star) \cap Qv) \). If \((C, \star, n)\) were a para-octonion algebra, its commutative center would be one-dimensional and spanned by a nonzero idempotent. Hence we would have \( K(C, \star) = K(C, \star) \cap Q \), and this is contained in the commutative center of the para-quaternion algebra \((Q, \star)\), which is \( Fe \). But \( e \star v = -w^2v \) while \( v \star e = -uv \). \( \Box \)

**Proposition 9.9.** If an Okubo algebra \((S, \star, n_S)\) admits a quaternionic idempotent \( f \), there is an isomorphism \( \varphi: (S, \star, n_S) \rightarrow (C, \star, n) \) such that \( \varphi(f) = e \). In particular \((C, \star, n)\) is isomorphic to the split Okubo algebra, and the unique Okubo algebra that contains a quaternionic idempotent is, up to isomorphism, the split Okubo algebra.
Proof. For the new multiplication
\[ x \cdot y = (f \ast x) \ast (y \ast f), \]
\((S, \cdot, n_S)\) is a Cayley algebra with unity \(f\), the map
\[ \vartheta_S(x) = (f \ast (f \ast x)) = n_S(f, x)f - x \ast f \]
is an automorphism of \((S, \cdot)\) and of \((S, \ast)\), of order 3, such that \(\ast = \cdot \vartheta_S\) is the twist with respect to \(\vartheta_S\) (see [KMRT98, (34.9)]). Let \(P\) be a para-quaternion subalgebra of \((S, \ast)\) with para-unit \(f\). Then the restriction of \(\vartheta_S\) to \(P\) is the identity. Let \(z \in P^\perp\) with \(n_S(z) \neq 0\), so \(P^\perp = P \cdot z\). Then \(S = P \oplus P \cdot z\) and \(\vartheta_S(z) = u \cdot z\) for some \(u \in P\). But then \(u^3 = 1\), so \((u - 1)^3 = 0 \neq u - 1\).

We conclude that \((P, \cdot)\) is the split quaternion algebra. Then \(z\) can be taken with \(n_S(z) = -1\). Moreover, there is an isomorphism \((P, \cdot) \to \text{Mat}_2(F)\) that takes \(u\) to the element \(w\) in Example 9.4, and this isomorphism extends to a unique isomorphism \(\varphi: (S, \cdot) \to (C, \cdot)\) such that \(\varphi(z) = w\). Obviously \(\varphi(f) = e\) and \(\varphi \vartheta_S = \vartheta \varphi\). This last condition implies that \(\varphi\) is also an isomorphism \((S, \ast, n_S) \to (C, \ast, n)\). Moreover, since the split Okubo algebra contains a quaternionic idempotent (see Lemma 9.3), there is a unique Okubo algebra up to isomorphism containing quaternionic idempotents.

Our aim now is to prove that the quaternionic idempotent in the split Okubo algebra is unique. The Okubo algebra \((C, \ast, n)\) in Example 9.4 is isomorphic to the split Okubo algebra, since it contains a quaternionic idempotent. Thus we may use it as a model of the split Okubo algebra. We keep the notation used in Example 9.4. The subalgebra of \((C, \ast)\) of elements fixed under the order 3 automorphism \(\vartheta\) is \(Q \oplus IV\), where \(I = \{a \in Q \mid wa = a\} = \{a \in Q \mid (w - 1) a = 0\} = (w - 1)Q\), because \((w - 1)^2 = 0\). (Note that it is clear that \((w - 1)Q\) is contained in the subspace of elements of \(Q\) annihilated by \(w - 1\), the equality holds since both subspaces are proper right ideals of \(Q\), and hence both of them have dimension 2.) Therefore, the centralizer of \(e\) in \((C, \ast, n)\) is
\[
\text{Cent}_{(C, \ast, n)}(e) = \{x \in C \mid e \ast x = x \ast e\} = \{x \in C \mid \vartheta^2(x) = \vartheta(x)\} = \{x \in C \mid \vartheta(x) = x\} = Q \oplus IV,
\]
and it is a six-dimensional subalgebra of both \((C, \cdot)\) and \((C, \ast)\). Note also that 
\[
IV = \{x \in \text{Cent}_{(C, \ast, n)}(e) \mid b_n(x, \text{Cent}_{(C, \ast, n)}(e)) = 0\}
\]
is precisely the radical of \(\text{Cent}_{(C, \ast, n)}(e)\) relative to the polar form of the norm, and \((IV) \ast (IV) = TIV = Q(w - 1)^2Q = 0\), so \(IV\) is a nilpotent ideal of \((\text{Cent}_{(C, \ast, n)}(e), \ast)\). Since \((Q, \ast)\) is simple, \(IV\) is the largest nilpotent ideal of \((\text{Cent}_{(C, \ast, n)}(e), \ast)\).

Let \(Z = \text{Cent}_{(C, \ast, n)}(e)\) and let \(K(Z, \ast) = \{x \in Z \mid x \ast y = y \ast x, \forall y \in Z\}\).

Lemma 9.11. \(K(Z, \ast) = Fe\).

(i) \(K(Z, \ast) = Fe\).

(ii) Any subalgebra of \((Z, \ast)\) of dimension \(\geq 4\) contains \(e\).
Proof. Clearly $e$ belongs to $K(Z, *)$. Also $(Z, *)$ is $\mathbb{Z}_2$-graded, so
\[ K(Z, *) = (K(Z, *) \cap Q) \oplus (K(Z, *) \cap I_v). \]

Since $(Q, *)$ is a para-quaternion algebra, its commutative center is spanned by $e$, which is its para-unit. On the other hand, for any $a \in Q$ and $b \in I$, $a \star (bv) = -\overline{a} \cdot (bv) = -(\overline{a}v)\overline{b}$, while $(bv) \star a = -(bv) \cdot \overline{a} = -(ba)v$. Hence, if $bv$ is in $K(Z, *)$, then $ba = b\overline{a}$ for any $a \in Q$, so $b(a - \overline{a}) = 0$ for any $a \in Q$, and $b = 0$ since we may find elements $a \in Q$ with $n(a - \overline{a}) \neq 0$.

For the second part, let $Z_0 = \{x \in Z \mid b_n(x, x) = 0\}$, so $Z = F\alpha \oplus Z_0$. If $T$ is a subalgebra of $(Z, *)$ and $e \notin T$, then the projection $\pi: T \to Z_0$ is one-to-one, and hence for any $x \in \pi(T)$ there is a unique $\alpha \in F$ such that $\alpha e + x \in T$. But $(\alpha e + x)^2 = (\alpha e - x)^2 = \alpha^2 e - 2\alpha x + x^2 = \alpha(\alpha e - x) - n(x)e$, so $n(x)e \in T$ for any $x \in \pi(T)$. Since we are assuming $e \notin T$, $\pi(T)$ consists of isotropic vectors and hence $\dim T = \dim \pi(T) \leq 3$, because $\dim Z_0 = 5$ and the rank of the restriction of $n$ to $Z_0$ is $3$. \qed

**Corollary 9.12.** If $f$ is a quaternionic idempotent of the split Okubo algebra $(C, *, n)$, then $f \in Z$.

**Proof.** We know by Proposition 9.9 and by formulas (9.10) that $Z = Cent(C, *, n)(e)$ and $Cent(C, *, n)(f)$ are six-dimensional subalgebras. Hence $Cent(C, (*, n))(f) \cap Z$ has dimension at least 4, so it contains $e$ by Lemma 9.11. Hence $e$ lies in $Cent(C, (*, n))(f)$, $e \star f = f \star e$, and this is equivalent to $f \in Z$. \qed

**Theorem 9.13.** The split Okubo algebra contains a unique quaternionic idempotent.

**Proof.** By Corollary 9.12 it suffices to study the nonzero idempotents in the subalgebra $Z$ of the split Okubo algebra $(C, *, n)$. For $a \in Q$ and $\beta \in I = (w - 1)Q$, we have
\[
(a + bv)^2 = \overline{(a - bv)} = b_n(0, 1)(a - bv) - n(a - bv)1 = b_n(a, 1)(a - bv) - n(a)1,
\]
and this is an idempotent if and only if
\[
\begin{align*}
(a + bv)^2 &= n(a)1, \\
b &= -b_n(a, 1)b.
\end{align*}
\]
Hence $a = b_n(a, 1)(b_n(a, 1)1 - a) - n(a)1$, or
\[
(1 + b_n(a, 1))a = (b_n(a, 1)^2 - n(a))1.
\]
Now, if $b_n(a, 1) \neq -1$ we conclude from (9.14) that $a \in F1$ and $b = 0$, so $a = 1 = e$.

On the other hand, if $b_n(a, 1) = -1$, then we conclude from (9.15) that $n(a) = 1$ and hence $(a - 1)^2 = 0$. Then we assume from now on that $(a - 1)^2 = 0$.

If $b = 0$, then either $a = 1 = e$, or $a$ is conjugate in $Q$ to $w = (\frac{1}{0} \frac{1}{1})$. In this last case, for $x, y \in Q$, $x + yv$ is in the centralizer of $a$ if and only if so are $x$ and $yv$, but $x \star a = a \star x$ if and only if $\overline{a} = \overline{a}$, if and only if $ax = xa$, if and only if $x \in F1 + Fa$. And $(yw) \cdot a = a \cdot (yw)$ if and only if $-((wy)v) = -(\overline{w}((w^2)y)v)$, if and only if $wya = w^2y\overline{a}$. Since $\overline{a} = a^2 = a^{-1}$, this is equivalent to $ya^2 = wy$v.
But this equation does not hold for all $y \in Q$, and hence the dimension of $\text{Cent}_{(C,\ast,n)}(a)$ is not 6, and $a$ is not a quaternionic idempotent.

If $b \neq 0$, we compute again $\text{Cent}_{(C,\ast,n)}(a + bv)$. For $x, y \in Q$,

$$(x + yv) \ast (a + bv) = (\pi - (wy)v) \cdot (\pi - bv)$$

$$= ((\pi \pi + byw) - (b\pi + wy)v),$$

$$(a + bv) \ast (x + yv) = (\pi - bv) \cdot (\pi - (w^2y)v)$$

$$= (\pi \pi + w^2yb) - (bx + w^2y\pi)v.$$

Hence, $x + yv$ is in $\text{Cent}_{(C,\ast,n)}(a + bv)$ if and only if

$$\pi \pi + byw = \pi \pi + ywb \quad \text{and} \quad b\pi + wyv = bx + w^2y\pi.$$  

But $wb = b$, because $a + bv \in Z$, so $\pi w = w^2b = b$. Thus, this is equivalent to

$$(9.16) \quad \pi \pi + by = \pi \pi + yb \quad \text{and} \quad b\pi + wy = bx + w^2y\pi.$$  

If $a = 1$, this gives $by = yb$, so $\pi y \in F1$. But $n(b) = 0$, so $\pi y = 0$, or $y \in bQ$.

Note that $I$ is a minimal right ideal of $Q$, so $bQ = I$ and $y \in I$. Hence the centralizer of $a + bv$ is contained in $Q + Iv$. But then $wy = y$ and we obtain $b\pi = bx$, or $b(x - \pi) = 0$. We conclude that $Q$ is not contained in the centralizer of $a + bv$, so $\dim \text{Cent}_{(C,\ast,n)}(a + bv)$ is not 6 and $a + bv$ is not a quaternionic idempotent.

Finally, if $a \neq 1$ and we take $y \in bQ = I$, we have as before $\pi y = 0 = yb$, so $\pi \pi = \pi \pi$, and hence $x \in F1 + Fa$. If $a + bv$ were a quaternionic element, the dimension of $\text{Cent}_{(C,\ast,n)}(a + bv)$ would be 6, and hence the dimension of $\text{Cent}_{(Z,\ast,n)}(a + bv)$ would be at least 4. Therefore we would have $\text{Cent}_{(Z,\ast)}(a + bv) = (F1 + Fa) + Iv$. But for $y \in I$, $yv \in \text{Cent}_{(Z,\ast)}((a + bv)$ if and only if $ya = y\pi$ because of $(9.16)$, if and only if $y(a - 1) = 0$. We conclude that $I(a - 1) = 0$, or $(w - 1)Q(a - 1) = 0$, and this is impossible since $Q$ is simple (and hence prime).

For any Okubo algebra $(S,\ast,n)$, the cube form $g(x) = b_n(x, x \ast x)$ plays a key role (see [EP96, §5], [KMT98, (36.11)]).

**Lemma 9.17** (see [Eld97]). Let $(S,\ast,n)$ be an Okubo algebra. Then:

(i) The cubic form $g$ is semilinear, i.e. $g(\alpha x + y) = \alpha^3 g(x) + g(y)$ for any $
\alpha \in F$ and $x, y \in S$.

(ii) For any $x \in S$, we have $g(\alpha^2) = g(x)^2 + n(x)^3$.

(iii) The set $F^3 + g(S)$ is a subfield of $F$ (a purely inseparable field extension of $F^3$ of exponent 1).

(iv) The quadratic form $n$ is isotropic (and hence its Witt index is 4).

**Theorem 9.18.** Let $(C,\triangle,n)$ be the split Okubo algebra represented by Zorn matrices (see §5), over an algebraically closed field $F$. Consider the subset

$$O = \{x \in C \mid n(x) = 0, g(x) = 1\}.$$

Then:
(i) The set \( O \) is an irreducible closed subset of \( C \) in the Zariski topology.

(ii) For any \( x \in O \), \( e = x + x \triangle x \) is an idempotent of \((C, \triangle, n)\) which

is either quadratic or quaternionic.

(iii) The subset \( O_0 = \{ x \in O \mid x + x \triangle x \) is a quadratic idempotent\} coincides

with the orbit of \( e_1 \) under the automorphism group of \((C, \triangle, n)\).

(iv) For any \( x \in O_0 \) and any \( y \in C \) such that \( x \triangle y = 0 \), \( b_n(x, y) = 0 \) and \( g(y) = 1 \), there is an

automorphism \( \varphi \) of \((C, \triangle, n)\) such that \( \varphi(e_1) = x \) and \( \varphi(u_1) = y \).

(v) The subset \( O_0 \) is a nonempty open subset of \( O \), and the subset \( O_1 = \{ x \in O \mid x + x \triangle x \) is a quaternionic idempotent\} is a closed subset of \( O \).

Proof. (i) Let \( e_1 \) and \( f_1 \) be as in (5.3). Note first that \( e_1 \) belongs to \( O \), as

\( e_1 \triangle e_1 = f_1 \) and \( g(e_1) = b_n(e_1, f_1) = 1 \). Since the cubic form \( g \) is semilinear

(Lemma 9.17), the subset \( \{ x \in C \mid g(x) = 1 \} \) coincides with the set \( e_1 + \{ x \in C \mid g(x) = 0 \} \),

which is isomorphic to an affine space of dimension 7. It follows

that \( O \) is a quadric in this affine space, and one easily checks that this quadric

is irreducible.

(ii) For any \( x \in O \), \( x \triangle (x \triangle x) = n(x)x = 0 \), while

\[ (x \triangle x) \triangle (x \triangle x) + ((x \triangle x) \triangle x) \triangle x = b_n(x, x \triangle x)x \]

(see 4.3), so \( (x \triangle x) \triangle (x \triangle x) = x \). Hence \( e = x + x \triangle x \) is an idempotent, the

para-unit of the para-quadratic subalgebra spanned by \( x \) and \( x \triangle x \). Thus \( e \) is
either quadratic or quaternionic.

(iii) Note that \( e_1 \in O_0 \) as \( e = e_1 + e_1 \triangle e_1 = e_1 + f_1 \) is a quadratic idempotent

(Lemma 9.3). Recall that the Hurwitz product \( \cdot \) is recovered as

\[ x \circ y = (e \triangle x) \triangle (y \triangle e). \]

For any \( z \in O_0 \), let \( f = z + z \triangle z \), and consider the new multiplication on \( C \):

\[ x \circ y = (f \triangle x) \triangle (y \triangle f). \]

Then \((C, o, n)\) is a Cayley algebra with identity element \( f \) and the map \( \vartheta(x) =

\[ f \triangle (f \triangle x) = b_n(f, x)f - x \triangle f \] is an automorphism of \((C, o, n)\) of order 3.

Besides, \( x \triangle y = \vartheta(\widehat{x}) \circ \vartheta^2(\widehat{y}) \), where \( \widehat{x} = b_n(f, x) - x \) for any \( x \), so the

Okubo algebra \((C, \triangle, n)\) is the twist \((C, o, n)\) of \((C, o, n)\) by \( \vartheta \) (see [KMRT98,

(34.9)]). The automorphism \( \vartheta \) fixes the elements \( z \) and \( z \triangle z \) of \((C, o, n)\), and

hence the subalgebra \( K = \text{span} \{ z, z \triangle z \} \) of \((C, o, n)\). Moreover, \( \vartheta(x) = x \) if

and only if \( b_n(f, x)f - x \triangle f = x \), and this implies \( f \triangle x = b_n(f, x) - f \triangle

(f \triangle f) = b_n(f, x)f - x = x \triangle f \). Conversely, if \( f \triangle x = x \triangle f \), then \( \vartheta(x) = f \triangle (f \triangle x) = f \triangle (x \triangle f) = n(f)x = x \), so the subalgebra of \((C, o, n)\)
of the elements fixed by \( \vartheta \) coincides with the centralizer of \( f \) in \((C, \triangle, n)\).

Besides, \( \vartheta^3 = \text{Id} \), so \( (\vartheta - \text{Id})^3 = 0 \). If \( (\vartheta - \text{Id})^2 \) were 0, then there would exist
an element \( x \) orthogonal to \( K \) such that \( \vartheta(x) = x \) and \( n(x) \neq 0 \), but then the
subalgebra \( K \oplus K \circ x \) would be a para-quaternion subalgebra with para-unit \( f \),
a contradiction with \( f \) being quadratic. Hence the restriction of \( \vartheta \) to \( K^\perp \) has
minimal polynomial \((X - 1)^3\) and the arguments in [EP96, proof of Proposition 1.1] show that there exists an isomorphism \(\varphi: (C, \cdot, n) \to (C, \circ, n)\) such that \(\varphi(\lambda_1) = z, \varphi(\lambda_1) = z \triangle z\) and such that \(\varphi_{\pi} = \varphi\) (recall that \((C, \triangle, n)\) is the twist \((C, \circ, n)_s\)). Therefore \(\varphi\) is an automorphism of \((C, \triangle, n)\) and \(\varphi(\lambda_1) = z\).

(iv) Note that \(0 = y \triangle (x \triangle y) = n(y)x, \) so \(n(y) = 0\). Also,
\[
b_n(x \triangle x, y) = -b_n(x, x \triangle y) = 0, \quad b_n(x, y \triangle y) = b_n(x \triangle y, y) = 0
\]
and
\[
b_n(x \triangle x, y \triangle y) = b_n(y, y \triangle (x \triangle x)) = b_n(y, -b_n(x, y)x + x \triangle (x \triangle y)) = 0.
\]
Therefore the subalgebras span \(\{x, x \triangle x\}\) and span \(\{y, y \triangle y\}\) are orthogonal and the result follows now from [Eld09, Theorem 3.12].

(v) As \(\text{Cent}_{(C, \triangle, n)}(e_1 + e_1 \triangle e_1) = \text{span} \{e_1, f_1, u_1 + u_2 + u_3, v_1 + v_2 + v_3\}\), we have that \(\text{Cent}_{(C, \triangle, n)}(x + x \triangle x)\) has dimension 4 for any \(x \in O_0\). On the other hand, for any \(x \in O_1, x + x \triangle x\) is the unique quaternionic idempotent, whose centralizer has dimension 6. Since for any \(x, \text{Cent}_{(C, \triangle, n)}(x) = \ker(l_x - r_x), \)
where \(l_x\) and \(r_x\) denote the left and right multiplication by \(x\) in \((C, \triangle, n)\), we obtain:
\[
O_0 = \{x \in O \mid \text{rank}(l_{x+\Delta x} - r_{x+\Delta x}) = 4\}
\]
\[
= \{x \in O \mid \text{rank}(l_{x+\Delta x} - r_{x+\Delta x}) > 2\},
\]
and this is open in \(O\).

**Corollary 9.19.** Let \((C, \triangle, n)\) be the Okubo algebra with the multiplication table (5.11) over an algebraically closed field \(F\), and let \(\lambda \in F\) be nonzero. The subset
\[
O^\lambda = \{x \in C \mid n(x) = 0, g(x) = \lambda\}
\]
is an irreducible closed subset of \(C\), and its subset \(O_0^\lambda = \{x \in O^\lambda \mid \lambda^{1/3}x \in O_0\}\) is a nonempty open subset of \(O^\lambda\).

Given an Okubo algebra \((S, \star, n)\) over a field \(F\) and a field extension \(K/F\), denote by \(O(K)\) the subset \(\{x \in S \otimes_F K \mid n(x) = 0, g(x) = 1\}\) (the extensions of \(n\) and \(g\) to \(S \otimes_F K\) are denoted by the same letters). The same applies to \(O^\lambda(K)\). For any field \(F\), let \(\overline{F}\) denote an algebraic closure of \(F\).

**Corollary 9.20.** Let \((S, \star, n)\) be an Okubo algebra over an infinite field \(F\) and let \(\lambda \in F\) be nonzero. If the set \(O^\lambda(F)\) is not empty, then it is a closed irreducible subset of \(S\) and its subset \(O_0^\lambda(F) := O^\lambda(F) \cap O_0^\lambda(\overline{F})\) is a nonempty open subset.

10. **Automorphisms of the split Okubo algebra**

The rational points of the groups of automorphisms of Okubo algebras over fields of characteristic 3 have been computed in [Eld99], by relating the Okubo algebras with some noncommutative Jordan algebras. In this section the existence of a unique quaternionic idempotent in the split Okubo algebra will be used to compute its group of automorphisms inside the group of automorphisms of the split Cayley algebra. We systematically assume that the base field \(F\) has
characteristic 3. As in Example 9.4, consider the split Cayley algebra $(C, \cdot, n)$, with $C = Q \oplus Q e$, and the order 3 automorphism $\vartheta$ in (9.6) such that its restriction to $Q$ is the identity and $\vartheta(v) = wv$ with $1 \neq w \in Q$ such that $w^3 = 1$. The Petersson twist $\ast = \ast_\vartheta$ is the split Okubo algebra multiplication.

**Proposition 10.1.** $\text{Aut}(C, \ast, n)(F) = \{ \psi \in \text{Aut}(C, \cdot, n)(F) \mid \psi \vartheta = \vartheta \psi \}$.

**Proof.** Obviously, any automorphism of $(C, \cdot, n)$ commuting with $\vartheta$ is an automorphism of $(C, \ast, n)$. Conversely, if $\psi$ is an automorphism of $(C, \ast, n)$, then $\psi(e) = e$, because $e$ is the unique quaternionic idempotent, and since $\vartheta(x) = e \ast (e \ast x)$ for any $x \in C$, we conclude that $\psi$ commutes with $\vartheta$. \hfill \Box

It is no longer true that the quaternionic idempotent $e$ of $(C, \ast, n)$ is fixed by all automorphisms of the extension $(C \otimes_F R, \ast, n)$, for any unital commutative associative algebra $R$. However, the automorphisms that fix $e \otimes 1_R$ coincide with the automorphisms of $(C \otimes_F R, \cdot, n)$ that commute with $\vartheta$ (more precisely with the automorphism $\vartheta \otimes 1_R$). Denote by $H$ the centralizer of the automorphism $\vartheta$ in the group scheme of automorphisms $\text{Aut}(C, \cdot, n)$. Alternatively, $H$ is the stabilizer of $e$ (or of $Fe$). By (9.10) any $\varphi \in H(R)$ leaves $Iv \otimes_F R$ invariant, since $Iv$ is the radical of the restriction of the norm to the subalgebra of fixed elements by $\vartheta$. Therefore there is a morphism

$$(10.2) \quad \Phi : H \rightarrow \text{GL}(I),$$

that takes any automorphism $\varphi \in \text{Aut}((C \otimes F R, \ast, n))$ commuting with $\vartheta \otimes 1_R$ to the linear automorphism $\Phi(\varphi) = f$ of $I \otimes F R$ defined by $\varphi(xv) = f(x)v$, for any $x \in I \otimes F R$. Given any element $a \in \text{SL}_2(R)$ there is a unique automorphism $\psi_a$ of $(C \otimes F R, \cdot, n)$ such that $\psi_a(xa^{-1}) = axa^{-1}$ and $\psi_a(xv) = (xa^{-1})v$ for $x \in Q \otimes F R = \text{Mat}_2(R)$. This automorphism commutes with $\vartheta$. Thus we get a closed embedding of $\text{SL}_2$ into $H$. Denote by $S$ its image. Moreover, $\text{GL}(I)$ will be identified with $\text{GL}_2$, with $a \in \text{GL}_2(R) = (Q \otimes F R, \ast)$ being identified with the map $xv \mapsto (xa^{-1})v$. With these identifications we get the following result:

**Theorem 10.3.**

(i) $\Phi(H) = \text{SL}_2$.

(ii) $H$ is the semidirect product of $\ker \Phi$ and $S$.

**Proof.** As in the proof of Theorem 9.13, we may assume $w = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Hence $I = (w - 1)Q$ is the right ideal of $2 \times 2$ matrices with trivial second row. Let $J$ be the right ideal of $2 \times 2$ matrices with trivial first row, so $Q = I \oplus J$. For simplicity, given a unital commutative associative algebra $R$, we write $Q_R = Q \otimes F R$ and similarly $I_R, J_R$. Take $\varphi \in H(R)$ and let $a \in \text{GL}_2(R)$ be its image under $\Phi$, so $\varphi(xv) = (xa^{-1})v$ for any $x \in I$. For $x \in Q_R$, $\varphi(x) = \varphi_0(x) + \sigma(x)v$ for $R$-linear maps $\varphi_0 : Q_R \rightarrow Q_R$ and $\sigma : Q_R \rightarrow I_R$. Then, for $x \in Q_R$ and $y \in I_R$,

$$\varphi(x \cdot (yv)) = \begin{cases} \varphi((yx)v) = (yxa^{-1})v, \\
\varphi(x) \cdot \varphi(yv) = (\varphi_0(x) + \sigma(x)v) \cdot ((ya^{-1})v) = (ya^{-1})\varphi_0(x)v,
\end{cases}$$

with $[\varphi] = a \otimes 1_R$.
as $(iv)^2 = 0$. Thus $y(xa^{-1} - a^{-1} \varphi_0(x)) = 0$ for any $x \in Q_R$, $y \in I_R$, so
\[ \varphi_0(x) = axa^{-1}. \]
Now, for $x \in J_R$, $\varphi(xv) = \mu(x)v + \nu(x) + \delta(x)v$ for some $R$-linear maps $\mu: J_R \to J_R$, $\nu: J_R \to Q_R$ and $\delta: J_R \to I_R$. Hence, with $\vartheta$ as in (9.6), we have
\[ \varphi \vartheta(xv) = \varphi((wx)v) = \varphi(xv) + \varphi((w-1)x)v \]
\[ = \mu(x)v + \nu(x) + (\delta(x) + (w-1)xax^{-1})v, \]
\[ \vartheta \varphi(xv) = \vartheta(\mu(x)v + \nu(x) + \delta(x)v) \]
\[ = (w\mu(x))v + \nu(x) + \delta(x)v \]
\[ = \mu(x)v + \nu(x) + ((w-1)\mu(x) + \delta(x))v. \]
We get $(w-1)xax^{-1} = (w-1)\mu(x)$, so $\mu(x) - xa^{-1}$ lies in $J_R$ and it is annihilated by $(w-1)$, and hence lies in $I_R$. Thus $\mu(x) = xa^{-1}$ for any $x \in J_R$.
But $\varphi$ is an isometry of the norm too, so for any $x \in J_R$ and $y \in I_R$,
\[ b_n(x,y) = -b_n(xy,yv) = -b_n(\varphi(xy),\varphi(yv)) \]
\[ = -b_n((xa^{-1})v, (ya^{-1})v) = b_n(xa^{-1}, ya^{-1}) = n(a^{-1})b_n(x,y), \]
and we conclude that $\det(a) = n(a) = 1$. This proves (i). Finally, the assignment $a \mapsto \psi_a$ gives a section of the epimorphism $\Phi: H \to \text{SL}_2$. \hfill \Box

We next compute the group $\ker \Phi$.

**Lemma 10.4.** Let $(C, \cdot, n)$ be the Zorn algebra as in §5. Let $R$ be a unital commutative associative algebra over $F$ and let $W$ be a free $R$-submodule of $C \otimes_F R$ of rank $3$ with $n(W) = 0 = b_n(W,1)$ and $1 \in b_n(W, W^{-2})$. Let $\{w_1, w_2, w_3\}$ be a basis of $W$ with $b_n(w_1, w_2, w_3) = 1$. Let $\{e_1, f_1, u_1, u_2, u_3, v_1, v_2, v_3\}$ be the canonical basis of $C$. Then there is a unique automorphism $\varphi \in \text{Aut}(C \otimes_F R, \cdot, n)$ such that $\varphi(e_i) = w_i$ for $i = 1, 2, 3$.

**Proof.** The uniqueness follows from the fact that $u_1, u_2, u_3$ generate $(C, \cdot)$. Consider the trilinear form given by $\Lambda(x, y, z) = b_n(x, y \cdot z)$. For any $x \in W$, $n(x) = 0 = b_n(x, 1)$, so $x^2 = 0$, and hence $x \cdot y = -y \cdot x$ for any $x, y \in W$.
Moreover, for all $x, y \in W$,
\[ \Lambda(x, x, y) = b_n(x, x \cdot y) = n(x)b_n(1, y) = 0, \]
\[ \Lambda(x, y, y) = b_n(x, y^2) = 0, \]
so $\Lambda$ is alternating on $W$. In particular, we have $\Lambda(x, y, z) = \Lambda(y, z, x) = \Lambda(z, x, y)$ for any $x, y, z \in W$.
Let $W$ be the linear span (over $R$) of $w_2 \cdot w_3$, $w_3 \cdot w_1$ and $w_1 \cdot w_2$. Taking indices modulo 3 we get $n(w_1 \cdot w_{i+1}) = 0$ and
\[ b_n(w_1 \cdot w_{i+1}, w_{i+1} \cdot w_{i+2}) = -b_n(w_{i+1} \cdot w_i, w_{i+1} \cdot w_{i+2}) \]
\[ = -n(w_{i+1})b_n(w_i, w_{i+2}) = 0, \]
so $W$ is isotropic. Also, $b_n(w_1 \cdot w_{i+1}, 1) = b_n(w_i, w_{i+1}) = 0$. Moreover
\[ b_n(w_i, w_{i+1} \cdot w_{i+2}) = \Lambda(w_i, w_{i+1}, w_{i+2}) = 1, \]
while
\[ b_n(w_i, w_i \cdot w_{i+1}) = n(w_i)b_n(1, w_{i+1}) = 0 = b_n(w_i, w_{i+2} \cdot w_i). \]

Therefore, \( W \) and \( \hat{W} \) are paired by the polar form of the norm.

Now, since we have \( x \cdot (y \cdot z) + y \cdot (x \cdot z) = (x \cdot y + y \cdot x) \cdot z, \) we get
\[ w_1 \cdot (w_2 \cdot w_3) = w_2 \cdot (w_1 \cdot w_3) = w_3 \cdot (w_1 \cdot w_3), \]
and also
\[ (w_1 \cdot w_2) \cdot w_3 = (w_2 \cdot w_3) \cdot w_1 = (w_3 \cdot w_1) \cdot w_2. \]

But
\[ w_1 \cdot (w_2 \cdot w_3) + (w_2 \cdot w_3) \cdot w_1 = -b_n(w_1, w_2 \cdot w_3)1 = -1. \]

As \( n(w_1 \cdot (w_2 \cdot w_3)) = 0 \) and \( b_n(w_1 \cdot (w_2 \cdot w_3), 1) = -b_n(w_2 \cdot w_3, w_1) = -1, \) it turns out that \( \hat{e}_1 = -w_1 \cdot (w_2 \cdot w_3) \) is an idempotent, and so is \( \hat{f}_1 = 1 - \hat{e}_1 = -(w_2 \cdot w_3) \cdot w_1. \) Then
\[ w_1 \cdot \hat{e}_1 = w_1 \cdot (w_1 \cdot (w_{i+1} \cdot w_i + 1)) = w_1^2 \cdot (w_{i+1} \cdot w_i + 1) = 0, \]
and we get \( \hat{f}_1 \cdot w_i = 0 \) in a similar vein. Also \( \hat{e}_1 \cdot (w_i \cdot w_{i+1}) = 0 = (w_i \cdot w_{i+1}) \cdot \hat{f}_1, \)
and it follows easily that \( \{\hat{e}_1, \hat{f}_1, w_1, w_2, w_3, w_2 \cdot w_3, w_3 \cdot w_1, w_1 \cdot w_2\} \) is a basis of \( C \otimes F R \) with the same multiplication table as for the canonical basis over \( F. \)

Consider now an automorphism \( \varphi \in \text{Aut}(C \otimes F R, \cdot, n) \) that commutes with \( \vartheta \) and which belongs to \( \text{ker } \Phi, \) that is, \( \varphi \) fixes the elements in \( I_{RV}. \) Then for any \( x \in Q \otimes R, \) \( x \) is fixed by \( \vartheta, \) and hence so is \( \varphi(x). \) Thus, as in the proof of Theorem 10.3 (with \( a = 1 \)), we obtain \( \varphi(x) = x + \sigma(x)v \) for any \( x \in Q, \) for an \( R \)-linear map \( \sigma: Q \otimes R \to I_{RV}. \)

We may take a canonical basis \( \{e_1, f_1, u_1, u_2, u_3, v_1, v_2, v_3\} \) of \( C \) such that \( Q = \text{span } \{e_1, f_1, u_1, v_1\}, \) \( w = 1 - u_1 \) and \( v = u_3 - v_3. \) Then \( I = (w - 1)Q = \text{span } \{e_1, u_1\}, \) \( I_v = \text{span } \{u_3, v_3\}. \) Also, \( \vartheta(u_2) = \vartheta(v_2) = (wv_1)v = ((1 - u_1)v_1)v = (v_1 - e_1) \cdot (u_3 - v_3) = u_2 + u_3. \) In particular, \( \varphi(u_3) = u_3, \) since we are assuming that \( \varphi \) leaves the elements in \( I_{RV}, \) and \( \varphi(u_1) = u_1 + \alpha u_3 + \beta v_2 \) for some \( \alpha, \beta \in R, \) since \( \varphi(u_1) - u_1 \in I_{RV}. \) Besides, for any \( x \in I_{RV}, \)
\[ b_n(\varphi(u_2), x) = b_n(\varphi(u_2), \varphi(x)) = b_n(u_2, x), \] so \( \varphi(u_2) - u_2 \in (I_{RV})^2 = Q \otimes I_{RV}. \) As \( b_n(\varphi(u_2), 1) = b_n(u_2, 1) = 0, \) we conclude that \( \varphi(u_2) = u_2 + \gamma(e_1 - f_1) + \delta u_1 + \mu v_1 + \nu u_3 + \rho v_2 \) for some \( \gamma, \delta, \mu, \nu, \rho \in R. \) Hence
\[ \varphi(u_1) = u_1 + \alpha u_3 + \beta v_2, \]
(10.5)
\[ \varphi(u_2) = u_2 + \gamma(e_1 - f_1) + \delta u_1 + \mu v_1 + \nu u_3 + \rho v_2, \]
\[ \varphi(u_3) = u_3, \]
and this determines \( \varphi \) completely because the elements \( u_1, u_2, u_3 \) generate \( C. \) Conversely, Lemma 10.4 shows that for any \( \alpha, \beta, \gamma, \delta, \mu, \nu, \rho \in R, \) there is a unique automorphism \( \varphi \) satisfying (10.5) if and only if the elements \( u_1 + \alpha u_3 + \beta v_2, u_2 + \gamma(e_1 - f_1) + \delta u_1 + \mu v_1 + \nu u_3 + \rho v_2 \) and \( u_3 \) span an isotropic space and
\[ b_n(u_1 + \alpha u_2 + \beta v_2, (u_2 + \gamma(e_1 - f_1) + \delta u_1 + \mu v_1 + \nu u_3 + \rho v_2) \cdot u_3) = 1. \]
But
\[ n(u_1 + \alpha u_3 + \beta v_2) = 0 = n(u_3), \]
\[ n(u_2 + \gamma(e_1 - f_1) + \delta u_1 + \mu v_1 + \nu u_3 + \rho v_2) = -\gamma^2 + \rho + \delta \mu, \]
while
\[ b_n(u_1 + \alpha u_3 + \beta u_2, u_3) = 0 = b_n(u_2 + \gamma(e_1 - f_1) + \delta u_1 + \mu v_1 + \nu u_3 + \rho v_2, u_3) \]
\[ b_n(u_1 + \alpha u_3 + \beta v_2, u_2 + \gamma(e_1 - f_1) + \delta u_1 + \mu v_1 + \nu u_3 + \rho v_2) = \mu + \beta, \]
and
\[ b_n(u_1 + \alpha u_3 + \beta u_2, (u_2 + \gamma(e_1 - f_1) + \delta u_1 + \mu v_1 + \nu u_3 + \rho v_2) \cdot u_3) = b_n(u_1 + \alpha u_3 + \beta v_2, v_1 + \gamma u_3 - \delta v_2) = 1. \]
Hence the only restrictions on the scalars \(\alpha, \ldots, \rho\) so that equation (10.5) determines an automorphism are
\[ \beta = -\mu, \quad \rho = \gamma^2 - \delta \mu. \]
If these conditions are satisfied, then \(\varphi \vartheta(u_1) = \vartheta \varphi(u_1)\) and \(\varphi \vartheta(u_3) = \vartheta \varphi(u_3)\) since \(u_1, u_3, \varphi(u_1), \varphi(u_3)\) are fixed by \(\vartheta\). Also, \(\varphi \vartheta(u_2) = \varphi(u_2 + u_3) = \varphi(u_2) + u_3\), while \(\vartheta(\varphi(u_2) - u_2) = \varphi(u_2) - u_2\), so \(\vartheta \varphi(u_2) = \varphi(u_2) - u_2 + \vartheta(u_2) = \varphi(u_2) + u_3 = \varphi \vartheta(u_2)\). Hence \(\varphi \vartheta = \vartheta \varphi\).
Since \(u_1, u_2, u_3\) generate the Cayley algebra \((C, \cdot)\), we can easily compute the images of any basic element. It turns out that the coordinate matrix in the basis \(\{u_2, v_3, e_1, f_1, u_1, v_1, u_3, v_2\}\) of an arbitrary element in \(\ker \Phi(R)\) is of the form
\[
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\gamma & -\mu & 1 & 0 & 0 & 0 & 0 & 0 \\
-\gamma & \mu & 0 & 1 & 0 & 0 & 0 & 0 \\
\delta & -\gamma & 0 & 0 & 1 & 0 & 0 & 0 \\
\mu & -\alpha & 0 & 0 & 0 & 1 & 0 & 0 \\
\nu & \mu^2 - \alpha \gamma & -\mu & \mu & \alpha & \gamma & 1 & 0 \\
\gamma^2 - \delta \mu & \alpha \delta - \mu \gamma - \nu & -\mu & \mu & \alpha & \gamma & 1 & 0 \\
-\gamma & -\gamma & -\mu & -\delta & 0 & 1 & 0 & 1
\end{pmatrix}
\]
\[
(10.6)
\]
Proposition 10.7. The algebraic group \(\ker \Phi\) is smooth unipotent of dimension 5. It is represented by the Hopf algebra which, as an algebra, is the polynomial ring in 5 variables \(F[\alpha, \gamma, \delta, \mu, \nu]\), with comultiplication given by \(\alpha, \gamma, \delta, \mu\) being primitive and \(\Delta(\nu) = \nu \otimes 1 + 1 \otimes \nu + \mu \otimes \gamma + \gamma \otimes \mu + \alpha \otimes \delta\).

Proof. Everything follows from the coordinate expression in equation (10.6). \(\Box\)

Corollary 10.8. \(H\) is a smooth group scheme.

Proof. Both \(\ker \Phi\) and \(SL_2\) are smooth group schemes, so \(H\) is smooth ([KMRT98, (22.12)]) \(\Box\)

Proposition 10.9. The group \(\ker \Phi\) is split unipotent.

\textbf{Documenta Mathematica} 18 (2013) 413–468
Proof. Consider the ideal in the ring of regular functions $F[\alpha, \gamma, \delta, \mu, \nu]$ of $\ker \Phi$ generated by $\alpha, \gamma, \delta, \mu$. It is a Hopf ideal because $\alpha, \gamma, \delta, \mu$ are primitive elements. Hence it corresponds to a closed group subscheme, say $Z$, isomorphic to $G_a$ and consisting of all matrices of the form

$$
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
\nu & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

It is straightforward to check that $Z$ is in the center of $\ker \Phi$, hence normal. The quotient map $\ker \Phi \to (\ker \Phi)/Z$ corresponds to the embedding $F[\alpha, \gamma, \delta, \mu, \nu] \hookrightarrow F[\alpha, \gamma, \delta, \mu, \nu]$ of Hopf algebras. Since $\alpha, \gamma, \delta, \mu$ are primitive elements it follows that $(\ker \Phi)/Z \cong G_a \times G_a \times G_a \times G_a$ and the claim follows. □

**Proposition 10.10.** The scheme $H$ is a closed subgroup of the group scheme of automorphisms $\text{Aut}(C, \star, n)$ of the split Okubo algebra and $H = \text{Aut}(C, \star, n)_{\text{red}}$.

**Proof.** The scheme $H$ is the stabilizer in $\text{Aut}(C, \star, n)$ of the quaternionic idempotent $e$. Let $A$ be the Hopf algebra representing $\text{Aut}(C, \star, n)$, and let $I$ be the Hopf ideal such that $A/I = \text{H\text{g}(C, \star, n)}$. Since $H$ is smooth, $I$ is a radical ideal. But $H(F) = \text{Aut}(C \otimes F, \star, n) = \text{Aut}(C, \star, n)(F)$ by Proposition 10.1. By Hilbert’s Nullstellensatz $I$ coincides with the radical of $A$. Hence the radical of $A$ is a Hopf ideal, and $H = \text{Aut}(C, \star, n)_{\text{red}}$. □

**11. The Group Scheme of Automorphisms of the Split Okubo Algebra**

We continue assuming that the characteristic of the ground field $F$ is 3.

Given any Okubo algebra $(S, \star, n)$, the arguments in [Eld97] and [Eld99] show that the vector space $A = F1 \oplus S$, with the multiplication given by

$$1 \cdot x = x = x \cdot 1,$$

$$a \cdot b = \frac{1}{2}(a \star b + b \star a + 2b_n(a, b)1) = -(a \star b + b \star a) + b_n(a, b)1,$$

for $x \in A$ and $a, b \in S$, is a 9-dimensional commutative and associative algebra.

We give a proof for completeness. We start with

**Lemma 11.2.** Every Okubo algebra $(S, \star, n)$ is Lie admissible, i.e., $S$ becomes a Lie algebra with bracket $[x, y]^\star = x \star y - y \star x$.

**Proof.** This can be proved using the multiplication table in (5.11), but we can proceed in a different way, which has its own independent interest. The result
is clear for the unique (split) Okubo algebra over the complex numbers, since it can be defined as the space of zero trace $3 \times 3$ matrices with the product
\[
x \star y = \frac{yx - \omega xy}{1 - \omega} - \frac{1}{3} \text{trace}(xy) 1
\]
where $\omega$ is a primitive cube root of unity and $xy$ denotes the usual matrix multiplication (see (8.6)). In this case $[x,y]^* = \frac{\omega^2}{1-\omega}(xy - yx)$, so the Jacobi identity holds, and the Okubo algebra is Lie admissible. It has a basis \{e1, f1, v1, u2, v2, v3, v4\} with the same multiplication table as in (5.11). Consider the subalgebra $E$ over $\mathbb{Z}$ spanned by the elements of this basis. Then $(E, \star)$ is a Lie algebra over $\mathbb{Z}$ with the bracket $[x,y]^\star = x \star y - y \star x$, and so is $(E \otimes_\mathbb{Z} F, \star)$ over $F$. But $(E \otimes_\mathbb{Z} F, \star)$ is the split Okubo algebra over $F$. Hence the split Okubo algebra is Lie-admissible, and the result follows since any Okubo algebra splits after a scalar extension. \hfill \Box

**Proposition 11.3.** The algebra $(A, \cdot)$, with multiplication (11.1), is a commutative and associative algebra. Moreover, if $(S, \ast, n)$ is split, then $(A, \cdot)$ is isomorphic to the algebra of truncated polynomials $F[X, Y]/(X^3, Y^3)$.

**Proof.** By its own definition $(A, \cdot)$ is commutative. To prove the associativity, it is enough to check that any associator $(a, b, c) = (a \cdot b) \cdot c - a \cdot (b \cdot c)$ is trivial. This is clear if one of the elements is 1, and hence we may assume that $a, b, c$ are in $S$. Then we get, using Lemma 11.2,
\[
0 = \left( [a, b]^*, c \right)^* + \left( [b, c]^*, a \right)^* + \left[ [c, a]^*, b \right]^* \\
= (a \cdot (c \cdot (a \cdot b) - b \cdot (a \cdot c - c \cdot a)) + (c \cdot b - b \cdot c) - a \cdot (b \cdot c - c \cdot b) + (b \cdot a - a \cdot b) \cdot c - c \cdot (b \cdot a - a \cdot b) \\
= - (b \cdot (c \cdot a) + b_n(a, b)c + (b \cdot a) \cdot c - b_n(b, c)a + c \cdot (a \cdot b) \\
- b_n(b, c)a - a \cdot (c \cdot b) + b_n(a, b)c + (c \cdot b - b \cdot c) \cdot a \\
- a \cdot (b \cdot (c \cdot b - c \cdot a) + (b \cdot a - a \cdot b) \cdot c - c \cdot (b \cdot a - a \cdot b) \\
= (c \cdot b + b \cdot c) \cdot a + a \cdot (c \cdot b + b \cdot c) + b_n(b, c)a \\
- (b \cdot a + a \cdot b) \cdot c - c \cdot (b \cdot a + a \cdot b) - b_n(a, b)c \\
= a \cdot (b \cdot c) - (a \cdot b) \cdot c = -(a, b, c),
\]
where we have used that $b_n(a, b \cdot c) = b_n(a \cdot b, c)$ and $b_n(a, c \cdot b) = b_n(b \cdot a, c)$. Therefore $(A, \cdot)$ is associative. Take now a canonical basis of the split Okubo algebra $(C, \Delta, n)$ as in (5.11). The elements
\[
x = e_1 - 1 \quad \text{and} \quad y = u_1 - 1
\]
satisfy $x^3 = 0 = y^3$, and the algebra $(A, \cdot)$ is isomorphic to the truncated polynomial algebra $F[X, Y]/(X^3, Y^3)$, with $x$ and $y$ corresponding to the classes of $X$ and $Y$ modulo $(X^3, Y^3)$. \hfill \Box

**Remark 11.5.** For any $a \in S$, the third power $a^3$ in $A$ is
\[
a^3 = (a \cdot a) \cdot a = (a \cdot a - n(a)1) \cdot a \\
= (a \cdot a) \cdot a + b_n(a \cdot a, a)1 - n(a)a = g(a)1,
\]
where $g(x) = b_n(x, x \cdot x)$ is the cubic form considered in Lemma 9.17.
Consider now the split Okubo algebra \((C, \triangle, n)\) and the associated commutative and associative algebra \((A, \cdot)\), \(A = F1 \oplus C\), as above. With the notation used in the previous proof, the ideal \(N\) generated by the elements \(x\) and \(y\) in \((A, \cdot)\) is the radical of \((A, \cdot)\) and \(N^4 = F(x^2 \cdot y^2)\). A straightforward computation gives
\[
(11.6) \quad x^2 \cdot y^2 = 1 + (e_1 + f_1 + u_1 + v_1 + u_2 + v_2 + u_3 + v_3) = 1 + e,
\]
where \(e\) is the quaternionic idempotent of \((C, \triangle, n)\) (Lemma 9.3).

Any automorphism \(\varphi\) of \((C, \triangle, n)\) extends to a unique automorphism of \((A, \cdot)\) by means of \(\varphi(1) = 1\). This is also valid if we extend scalars to a unital commutative associative algebra over \(F\). As usual, given a unital commutative associative algebra \(R\over F\), \(N_R\) will denote \(N \otimes_F R\). Recall that the algebraic group \(H\) is the stabilizer in \(\text{Aut}(C, \triangle, n)\) of the quaternionic idempotent \(e\). (Initially, \(H\) was defined in terms of the model \((C, \ast, n)\) in Example 9.4 of the split Okubo algebra, but this should cause no confusion.)

**Lemma 11.7.** Let \((C, \triangle, n)\) be the split Okubo algebra. Given any unital commutative associative algebra \(R\) over \(F\), an element \(\varphi\) in \(\text{Aut}(C, \triangle, n)(R)\) lies in \(H(R)\) if and only if, when extended to an automorphism of \(A_R, \cdot\), it satisfies \(\varphi((N_R)^4) = (N_R)^4\).

**Proof.** If an automorphism \(\varphi\) of \((C_R, \triangle, n)\) fixes \(e\), then its extension to \((A_R, \cdot)\) fixes \(1 + e\) and hence \(\varphi\) preserves \((N_R)^4\). Conversely, if \(\varphi\) preserves \((N_R)^4\), then there is an element \(\alpha \in R\) such that \(\varphi(1 + e) = \alpha(1 + e)\). But \(\varphi(1) = 1\) and \(\varphi(e) \in C_R\). Hence \(\alpha = 1\) and \(\varphi\) fixes \(e\). \(\square\)

The elements \(e_1\) and \(u_1\) generate the algebra \((C, \triangle)\), and the assignment \(\deg(e_1) = (1, 0), \deg(u_1) = (\overline{1}, \overline{1})\), endows \((C, \triangle)\) with a grading by \(Z_3 \times Z_3\). This gives a morphism of group schemes
\[
\mu_3 \times \mu_3 \to \text{Aut}(C, \triangle, n),
\]
such that for any \(\alpha, \beta \in \mu_3(R)\times \mu_3(R)\), i.e., \(\alpha^3 = \beta^3 = 1\), the image of \((\alpha, \beta)\) is the automorphism \(\psi_{\alpha, \beta}\) of \((C \otimes_F R, \triangle)\) with the property that \(\psi_{\alpha, \beta}(e_1) = \alpha e_1\) and \(\psi_{\alpha, \beta}(u_1) = \beta u_1\). The image \(D\) of this morphism is isomorphic to \(\mu_3 \times \mu_3\). Moreover, since \(b_3(e_1, e_1 \triangle e_1) = 1\) and \(b_3(u_1, u_1 \triangle u_1) = 1\), it turns out that \(D\) is the group scheme of diagonal automorphisms relative to the basis of \((C, \triangle, n)\) in Table (5.11).

**Theorem 11.8.** Let \((C, \triangle, n)\) be the split Okubo algebra. For any unital commutative associative algebra \(R\) over \(F\), \(\text{Aut}(C, \triangle, n)(R) = H(R)D(R)\). Moreover \(H(R) \cap D(R) = 1\).

**Proof.** Since the quaternionic idempotent is \(e = e_1 + f_1 + u_1 + v_1 + u_2 + v_2 + u_3 + v_3\) (the sum of the elements in the basis, see Lemma 9.3), an automorphism fixes \(e\) and is diagonal relative to this basis if and only if it is the identity. Hence we have \(H(R) \cap D(R) = 1\).

Let \(\varphi\) be an element in \(\text{Aut}(C, \triangle, n)(R)\). Extend it to an automorphism of \((A_R, \cdot)\). Then, with \(x\) and \(y\) as in (11.4), there are elements \(\mu, \nu \in R\) such that
\( \varphi(x) - \mu 1 \in N_R \) and \( \varphi(y) - \nu 1 \in N_R \). Since \( x^3 = 0 = y^3 \) and \( N_R \) is an ideal, we have \( \mu^2 = \nu^3 = 0 \). Consider the element \((\alpha, \beta) := (\mu + 1, \nu + 1) \in \mu_3(R) \times \mu_3(R)\) and the automorphism \( \psi_{\alpha, \beta} \in D(R) \). Then we have

\[
\varphi^{-1}_{\alpha, \beta}(x) = \varphi \psi_{\alpha, \beta}(x) \\
= \varphi \psi_{\alpha, \beta}(e - 1) \\
= \varphi(\alpha^2 e - 1) = \varphi(\alpha^2 x + (\alpha^2 - 1)1) \\
\equiv (\alpha^2 - 1)1 + \alpha^2 \mu 1 \pmod{N_R} \\
\equiv (\alpha^2 - 1 + \alpha^2(\alpha - 1))1 \equiv 0 \pmod{N_R}, \text{ as } \alpha^3 = 1,
\]

so \( \varphi^{-1}_{\alpha, \beta}(x) \in N_R \). In a similar vein we get \( \varphi^{-1}_{\alpha, \beta}(y) \in N_R \), and this proves \( \varphi^{-1}_{\alpha, \beta}(N_R) = N_R \), and hence \( \varphi^{-1}_{\alpha, \beta}((N_R)^4) = (N_R)^4 \). By Lemma 11.7 we conclude that \( \varphi^{-1}_{\alpha, \beta} \) lies in \( H(R) \), and hence \( \varphi = (\varphi^{-1}_{\alpha, \beta}) \psi_{\alpha, \beta} \in H(R)D(R) \).

**Proposition 11.9.** Neither of the subgroups \( H \) and \( D \) of \( \text{Aut}(C, \triangle, n) \) is normal.

**Proof.** Consider the automorphism \( \varphi \) of the split Cayley algebra \((C, \cdot, n)\) in (5.1) defined on the canonical basis as

\[
\varphi(e_i) = e_i, \quad \varphi(f_i) = f_i \quad \text{and} \quad \varphi(u_i) = -(u_i + u_{i+1}), \quad \varphi(v_i) = v_i + v_{i+1} - v_{i+2},
\]

for \( i = 1, 2, 3 \) (indices modulo 3). It is straightforward to check that \( \varphi \) is indeed an automorphism, and that it commutes with the automorphism \( s_\tau \) that permutes cyclically the \( u_i \)’s and the \( v_i \)’s. Therefore \( \varphi \) is an automorphism of the split Okubo algebra \((C, \triangle, n)\), where \( \triangle = s_\tau \) is the Petersson twist. And \( \varphi \) preserves the quaternionic idempotent, so \( \varphi \) is in \( H(F) \). Its inverse fixes \( e_1 \) and \( f_1 \) and satisfies

\[
\varphi^{-1}(u_i) = u_i - u_{i+1} + u_{i+2}, \quad \varphi^{-1}(v_i) = v_i - v_{i+2}
\]

for \( i = 1, 2, 3 \). For any unital commutative and associative algebra \( R \) over \( F \) containing an element \( 1 \neq \alpha \in \mu_3(R) \), consider the automorphism \( \psi_{\alpha, 1} \in D(R) \). Then the commutator \( [\varphi, \psi_{\alpha, 1}] = \varphi^{-1}\psi_{\alpha, 1}\varphi \psi_{\alpha, 1} \) fixes \( e_1 \) and takes \( u_1 \) to

\[
\varphi^{-1}\psi_{\alpha, 1}(u_1) = \varphi^{-1}\psi_{\alpha, 1}(u_1) \\
= \varphi^{-1}\psi_{\alpha, 1}(-u_1 - u_2) \\
= \varphi^{-1}(-u_1 - \alpha u_2) \\
= -(u_1 - u_2 + u_3) - \alpha(u_2 - u_3 + u_1) \\
= -(1 + \alpha)u_1 + (1 - \alpha)u_2 - (1 - \alpha)u_3.
\]

In particular \( \Phi = [\varphi, \psi_{\alpha, 1}] \) is not in \( D(R) \), as it does not act diagonally in our basis. But it does not belong to \( H(R) \) either because it does not fix the quaternionic idempotent \( e = e_1 + f_1 + u_1 + e_1 + u_2 + v_2 + u_3 + v_3 \). To check this, note that \( \Phi(u_2) = -\Phi(u_1 \triangle e_1) = -(1 + \alpha)u_2 + (1 - \alpha)u_3 - (1 - \alpha)u_1 \) and

\[
\text{Documenta Mathematica 18 (2013) 413–468}
\]
\[ \Phi(u_3) = -\Phi(u_2 \triangle e_1) = -(1 + \alpha)u_3 + (1 - \alpha)u_1 - (1 - \alpha)u_2. \]

In the same vein \( \Phi(e_i) \) belongs to the linear span of \( v_1, v_2, v_3 \). Then

\[ \Phi(e) = \Phi(e_1 + f_1) + \Phi(u_1 + u_2 + u_3) + \Phi(v_1 + v_2 + v_3) = e_1 + f_1 - (1 + \alpha)(u_1 + u_2 + u_3) + \text{a linear combination of } v_1v_2, v_1 \]

\( \neq e. \)

**Remark 11.10.** The subscheme \( H \) embeds in the simple group scheme \( \text{Aut}(C, \cdot, n) \) of type \( G_2 \) (Proposition 10.1). However, there is no such embedding for \( G = \text{Aut}(C, \triangle, n) \). Actually, an embedding \( G \to \text{Aut}(C, \cdot, n) \) would give a monomorphism of Lie algebras

\[ \iota: \text{Lie}(G) = \text{Der}(C, \triangle, n) \to \text{Der}(C, \cdot, n). \]

But \( \text{Der}(C, \triangle, n) \) contains a simple nonclassical ideal \( i \) of dimension 8 [Eld99], while \( g \) is not simple (the characteristic is 3!) but contains a simple ideal \( j \) of dimension 7 such that the quotient \( g/j \) is again simple and isomorphic, as a Lie algebra, to \( j \) (see [AEMN02]). Then \( \iota \) would induce a Lie algebra homomorphism \( i \to g/j \), which must be 0 by dimension count, and this would give that \( i \) embeds in \( j \), a contradiction, again by dimension count.

**12. The classification of Okubo algebras revisited**

The classification of the Okubo algebras in characteristic 3 was obtained in [Eld97, §5], using previous classification results in [EP96, Theorems B and 5.1]. Here this classification will be revisited in light of the results in the preceding sections.

Still the characteristic of \( F \) is assumed to be 3 in this section.

**Proposition 12.1.** The cohomology set \( H^1_{\text{fppf}}(F, H) \) is trivial.

**Proof.** The group \( H \) is the semidirect product of the split unipotent group \( \ker \Phi \) (see Theorem 10.3) and of \( \text{SL}_2 \). Hence we have an exact sequence

\[ H^1_{\text{fppf}}(F, \ker \Phi) \to H^1_{\text{fppf}}(F, H) \to H^1_{\text{fppf}}(F, \text{SL}_2). \]

The left and right terms are trivial, so is the central term.

**Remark 12.2.** Alternatively, one may proceed as follows. Let \( (C, \triangle, n) \) be the split Okubo algebra and let \( e \) be its (unique) quaternionic idempotent. The group scheme \( H \) is the subscheme of \( \text{Aut}(C, \triangle, n) \) consisting of those elements fixing \( e \):

\[ H = \text{Aut}(C, \triangle, n, e). \]

The twisted forms of \( (C, \triangle, n, e) \) are the Okubo algebras with a quaternionic idempotent. But Proposition 9.9 asserts that the only Okubo algebra containing a quaternionic idempotent is the split Okubo algebra. This shows that \( H^1_{\text{fppf}}(F, H) \) is trivial.
Also, $H$ is smooth and hence we have ([Wat79, (18.5) and (17.8)]):

$$H^1_{\text{fppf}}(F, H) = H^1_{\text{ét}}(F, H) = H^1(\Gamma, H(F_{\text{sep}}))$$

$$= H^1(\Gamma, \text{Aut}(C \otimes F, \Delta, n)) = H^1_{\text{ét}}(F, \text{Aut}(C, \Delta, n)),$$

where $F_{\text{sep}}$ denotes the separable closure of $F$ in $\overline{F}$ and $\Gamma$ the Galois group of the extension $F_{\text{sep}}/F$. (Note that over any field extension $K$ of $F$, $H(K)$ exhausts $\text{Aut}(C \otimes K, \Delta, n)$, since $D(K) = 1$.)

Hence we conclude, without making any appeal to the classification in [Eld97], that the only twisted $F_{\text{sep}}/F$-form of $(C, \Delta, n)$ is, up to isomorphism, $(C, \Delta, n)$ itself:

**Corollary 12.3.** The cohomology set $H^1_{\text{ét}}(F, \text{Aut}(C, \Delta, n))$ is trivial. In particular, if $F$ is perfect, there is up to isomorphism a unique Okubo algebra over $F$.

Take now nonzero elements $\alpha, \beta \in F$, and let $(C_{\alpha, \beta}, \Delta, n)$ be the $F$-subalgebra of $(C \otimes_F \overline{F}, \Delta, n)$ generated by the elements $e_1 \otimes \alpha^\frac{1}{3}$ and $u_1 \otimes \beta^\frac{1}{3}$. This is a twisted form of $(C, \Delta, n)$. Denote by $g_{\alpha, \beta}$ the cubic form on $(C_{\alpha, \beta}, \Delta, n)$ given by $g_{\alpha, \beta}(x) = b_{\alpha}(x, x \Delta x)$. The image $g_{\alpha, \beta}(C_{\alpha, \beta})$ is a $F^3$-subspace of $F$ spanned by the elements $\alpha, \alpha^3, \beta, \beta^3, \alpha \beta, \alpha^2 \beta, \alpha^2 \beta^2$. Moreover, under the bijection

$$H^1_{\text{fppf}}(F, \text{Aut}(C, \Delta, n)) \cong \{\text{Isomorphism classes of Okubo algebras}\},$$

the isomorphism class of $(C_{\alpha, \beta}, \Delta, n)$ corresponds to the class of the cocycle

$$\Psi^{\alpha, \beta} \in \text{Aut}(C, \Delta, n)(\overline{F} \otimes_F \overline{F}) = \text{Aut}(C \otimes_F \overline{F}, \overline{F}, \overline{F})$$

determined by

$$\Psi^{\alpha, \beta}(e_1 \otimes 1 \otimes 1) = e_1 \otimes \alpha^{-\frac{1}{3}} \otimes \alpha^{\frac{1}{3}}, \quad \Psi^{\alpha, \beta}(u_1 \otimes 1 \otimes 1) = u_1 \otimes \beta^{-\frac{1}{3}} \otimes \beta^{\frac{1}{3}}.$$ 

Note that $\alpha^{-\frac{1}{3}} \otimes \alpha^{\frac{1}{3}}$ and $\beta^{-\frac{1}{3}} \otimes \beta^{\frac{1}{3}}$ are cube roots of unity, so $\Psi^{\alpha, \beta}$ belongs to the subscheme $D$.

**Remark 12.5.** The scheme $D$ is isomorphic to $\mu_3 \times \mu_3$, so

$$H^1_{\text{fppf}}(F, D) = \{[\Psi^{\alpha, \beta}] \mid \alpha, \beta \in F^\times\},$$

where $[\Psi^{\alpha, \beta}] = [\Psi^{\alpha', \beta'}]$ if and only if $\alpha' \alpha^{-1}, \beta' \beta^{-1} \in F^3$, see [Wat79, 18.2(a)]. In particular, over perfect fields, $H^1_{\text{fppf}}(F, D)$ is trivial.

**Theorem 12.6.** The mapping $H^1_{\text{fppf}}(F, \mu_3 \times \mu_3) \to H^1_{\text{fppf}}(F, \text{Aut}(C, \Delta, n))$, induced by the inclusion $D \hookrightarrow \text{Aut}(C, \Delta, n)$, is surjective.

**Proof.** This is trivial if $F$ is perfect, in particular if $F$ is finite. Hence we will assume that $F$ is infinite. Let $\xi \in Z^1_{\text{fppf}}(F, \text{Aut}(C, \Delta, n))$ be a cocycle and let $(C^S, \Delta, n)$ be the Okubo algebra $C^S = \{x \in C \otimes F \overline{F} \mid \zeta(x \otimes 1) = \theta(x \otimes 1)\}$, where $\theta : C \otimes_F \overline{F} \otimes_F \overline{F} \to C \otimes_F \overline{F} \otimes_F \overline{F}$, $x \otimes \alpha \otimes \beta \mapsto x \otimes \beta \otimes \alpha$. 

Documenta Mathematica 18 (2013) 413–468
Since the norm of an Okubo algebra is isotropic (Lemma 9.17), $C^\xi$ has a basis of isotropic elements and hence there are elements $x \in C^\xi$ with $n(x) = 0$ and $g(x) = \alpha \neq 0$. Then, by Corollary 9.20, we may take

$$x \in O^\partial_0(F) \subset C^\xi \subset C^\xi \otimes_F \mathbb{F} = C \otimes_F \mathbb{F}.$$ 

Then Theorem 9.18 shows that there is an automorphism $\varphi \in \text{Aut}(C \otimes_F \mathbb{F}, \vartriangle, n)$ such that $\varphi(e_1 \otimes 1) = \alpha^{-\frac{1}{2}} x$. Replacing $\zeta$ by an equivalent cocycle we may assume $x = e_1 \otimes \alpha^\frac{1}{2}$. The subspace $\{ z \in C \mid e_1 \vartriangle z = 0, b_\alpha(e_1, z) = 0 \}$ is spanned by $u_1, u_2, u_3$, thus the subspace $\{ y \in C^\xi \mid x \vartriangle y = 0, b_\alpha(x, y) = 0 \}$ has dimension 3 and it is not contained in the kernel of the semilinear map $g$. Take an element $y \in C^\xi$ with $x \vartriangle y = 0$, $b_\alpha(x, y) = 0$ and $g(y) = \beta \neq 0$. Theorem 9.18 shows that there is an automorphism $\psi \in \text{Aut}(C \otimes_F \mathbb{F}, \vartriangle, n)$ such that $\psi(e_1 \otimes 1) = e_1 \otimes 1, \psi(u_1 \otimes 1) = \beta^{-\frac{1}{2}} y$. Replacing $\zeta$ by an equivalent cocycle we may assume $x = e_1 \otimes \alpha^\frac{1}{2}$ and $y = u_1 \otimes \beta^\frac{1}{2}$ belong to $C_3$. Therefore $\zeta(e_1 \otimes \alpha^\frac{1}{2} \vartriangle 1) = e_1 \otimes 1 \otimes \alpha^\frac{1}{2}$, so $\zeta(e_1 \otimes 1 \otimes 1) = e_1 \otimes \alpha^{-\frac{1}{2}} \otimes \alpha^\frac{1}{2}$, and similarly $\zeta(u_1 \otimes 1 \otimes 1) = u_1 \otimes \beta^{-\frac{1}{2}} \otimes \beta^\frac{1}{2}$. Hence we get $\zeta = \psi \zeta_0$ (see (12.4)).

This shows that any Okubo algebra is isomorphic to $(C_{\alpha, \beta}, \vartriangle, n)$ for some nonzero $\alpha, \beta \in F$. Moreover, the set of isomorphism classes of Okubo algebras is a quotient set of $H^1_{\text{fppf}}(F, \mathfrak{d}_3 \times \mathfrak{d}_3) \cong F^\times/(F^\times)^3 \times F^\times/(F^\times)^3$. It remains to consider the isomorphism conditions:

**Theorem 12.7** (see [Eld97]). For $\alpha, \beta, \alpha', \beta' \in F^\times$, the algebras $(C_{\alpha, \beta}, \vartriangle, n)$ and $(C_{\alpha', \beta'}, \vartriangle, n)$ are isomorphic if and only if

$$g_{\alpha, \beta}(C_{\alpha, \beta}) = g_{\alpha', \beta'}(C_{\alpha', \beta'})$$

and if this common $F^3$-subspace has dimension 8 (i.e., $g_{\alpha, \beta}$ and $g_{\alpha', \beta'}$ are bijective), then the following restriction is required too:

$$g_{\alpha, \beta, n}^{-1}(\alpha') \otimes g_{\alpha, \beta, n}^{-1}(\beta') = 0$$

(this is a product of two elements in $(C_{\alpha, \beta}, \vartriangle, n)$).

**Proof.** The result is trivial for perfect fields, and hence for finite fields. Hence we will assume that the ground field $F$ is infinite. If the compositions $(C_{\alpha, \beta}, \vartriangle, n)$ and $(C_{\alpha', \beta'}, \vartriangle, n)$ are isomorphic, any isomorphism $\varphi$ satisfies $n(\varphi(x)) = n(x)$ and $g_{\alpha, \beta}(\varphi(x)) = g_{\alpha', \beta'}(x)$ for any $x$, so $g_{\alpha, \beta}(C_{\alpha, \beta}) = g_{\alpha', \beta'}(C_{\alpha', \beta'})$. Moreover, Lemma 9.17 shows that $F^3 + g_{\alpha, \beta}(C_{\alpha, \beta})$ is a subfield of $F$. If $g_{\alpha, \beta}(C_{\alpha, \beta}) = F^3$, then the elements $x$ and $y$ in the proof of Theorem 12.6 can be taken with $g(x) = 1 = g(y)$, and hence $C_{\alpha, \beta}$ is isomorphic to $C_{1,1}$, which is the split Okubo algebra. If the semilinear map $g_{\alpha, \beta}$ is not bijective, then $\dim_{F^3}(F^3 + g_{\alpha, \beta}(C_{\alpha, \beta}))$ equals 1 or 3, and the dimension of $\ker g_{\alpha, \beta}$ is at least 5. Hence the restriction of the norm to $\ker g_{\alpha, \beta}$ is isotropic, so there exists a hyperbolic pair $a, b$ in $\ker g_{\alpha, \beta}$: $n(a) = n(b) = 0$, $b_\alpha(a, b) = 1$. Take $x$ in $\ker g_{\alpha, \beta}$ with $n(x) \neq 0$. Then $g_{\alpha, \beta}(x \vartriangle x) = n(x)^3$ (Lemma 9.17), so $F^3 \subseteq g_{\alpha, \beta}(C_{\alpha, \beta})$. If $\gamma \in g_{\alpha, \beta}(C_{\alpha, \beta}) \setminus F^3$ and $x \in C_{\alpha, \beta}$ satisfies $g(x) = \gamma$, then we may find $\mu, \nu \in F$.
such that the element $x' = x + \mu a + \nu b$ is isotropic (and satisfies $g(x') = \gamma$). The three-dimensional subspace $\{ z \in C_{\alpha,\beta} \mid x' \triangleleft z = 0, b_n(x', z) = 0 \}$ intersects nontrivially the six-dimensional subspace $g_{\alpha,\beta}^{-1}(F^3)$, so we may take an element

$y'$ satisfying $x' \triangleleft y' = 0, b_n(x', y') = 0$ and $g_{\alpha,\beta}(y') = 1$. The proof of Theorem 12.6 shows that $C_{\alpha,\beta}$ is isomorphic to $C_{\gamma,1}$ for any $0 \neq \gamma \in g_{\alpha,\beta}(C_{\alpha,\beta}) \setminus F^3$.

We are left with the case in which $g_{\alpha,\beta}$ is one-to-one, so $F^3 \cap g_{\alpha,\beta}(C_{\alpha,\beta}) = 0$ and $F^3 + g_{\alpha,\beta}(C_{\alpha,\beta})$ is a field extension of degree 9 of $F^3$. Then, if $g_{\alpha,\beta}(C_{\alpha,\beta}) = g_{\alpha',\beta'}(C_{\alpha',\beta'})$, then both $g_{\alpha,\beta}$ and $g_{\alpha',\beta'}$ are one-to-one, so there is a unique linear map $\phi: C_{\alpha,\beta} \to C_{\alpha,\beta}$ such that $g_{\alpha',\beta'}(x) = g_{\alpha,\beta}(\phi(x))$ for any $x \in C_{\alpha',\beta'}$ (so $\phi = g_{\alpha,\beta}^{-1} g_{\alpha',\beta'}$). Then, by the uniqueness of $\phi$, $C_{\alpha,\beta}$ and $C_{\alpha',\beta'}$ are isomorphic if and only if $\phi$ is an isomorphism. But Lemma 9.17 shows that

$$g_{\alpha,\beta}(\phi(x)^{\Delta 2}) = g_{\alpha,\beta}(\phi(x))^2 + n(\phi(x))^3$$

$$= g_{\alpha',\beta'}(x)^2 + n(\phi(x))^3$$

$$= g_{\alpha',\beta'}(x^{\Delta 2}) - n(x)^3 + n(\phi(x))^3,$$

thus $g_{\alpha,\beta}(\phi(x)^{\Delta 2}) - g_{\alpha',\beta'}(x^{\Delta 2}) \in g_{\alpha,\beta}(C_{\alpha,\beta}) \cap F^3 = 0$ and $\phi(x \triangle x) = \phi(x) \triangle \phi(x)$ for any $x \in C_{\alpha',\beta'}$. This implies, using $(x \triangle x) \triangle x = x \triangle (x \triangle x) = n(x)x$, that $n(\phi(x)) = n(x)$ for any $x \in C_{\alpha',\beta'}$. If $\phi$ is an isomorphism, then $\phi(e_1 \otimes (\alpha')^{\frac{1}{2}}) \triangle \phi(u_1 \otimes (\beta')^{\frac{1}{2}}) = 0$, which is equivalent to $g_{\alpha,\beta}^{-1}(\alpha') \triangle g_{\alpha,\beta}^{-1}(\beta') = 0$.

Conversely, if $\phi(e_1 \otimes (\alpha')^{\frac{1}{2}}) \triangle \phi(u_1 \otimes (\beta')^{\frac{1}{2}}) = 0$, then since the subspaces

$$\{\phi(e_1 \otimes (\alpha')^{\frac{1}{2}}), \phi(e_1 \otimes (\alpha')^{\frac{1}{2}})^{\Delta 2}\}$$

and

$$\{\phi(u_1 \otimes (\beta')^{\frac{1}{2}}), \phi(u_1 \otimes (\beta')^{\frac{1}{2}})^{\Delta 2}\}$$

are orthogonal subspaces of $C_{\alpha,\beta}$, [Eld09, Theorem 3.12] shows that $C_{\alpha,\beta}$ and $C_{\alpha',\beta'}$ are isomorphic. $\square$

In the last part of this section we show that Okubo algebras with isomorphic automorphisms groups (as algebraic groups) are isomorphic or anti-isomorphic (compare with Proposition 8.10). We start with a definition:

**Definition 12.8.** Let $(S, \ast, n_{\mathbf{S}})$ be an Okubo algebra and let $K$ be a 2-dimensional composition subalgebra (i.e., the restriction of $n_{\mathbf{S}}$ to the subalgebra $K$ is nondegenerate). Then $K$ is said to be a regular subalgebra if the para unit of $K \otimes F \mathbf{T}$ is a quadratic idempotent.

**Lemma 12.9.** Let $(C, \Delta, n)$ be the (split) Okubo algebra over an algebraically closed field $F$ (as in Table (5.11)), and let $K$ be a two-dimensional composition subalgebra. Then the following conditions are equivalent:

(i) $K$ is regular.

(ii) There is an automorphism $\varphi$ of $(C, \Delta, n)$ such that $\varphi(K) = Fe_1 + Ff_1$.

(iii) There is another two-dimensional composition subalgebra $K'$ of $(C, \Delta, n)$ orthogonal to $K$: $b_n(K, K') = 0$.

**Proof.** (i)⇒(ii) $K$ is, up to isomorphism, the unique para-quadratic algebra over the algebraically closed field $F$, so there is a basis $\{a, b\}$ of $K$ such that

$\text{DOCUMENTA MATHEMATICA 18 (2013) 413–468}$
\[ a \triangle b = a, b \triangle a = b, a \triangle b = 0 = b \triangle a, \text{ and } n(a) = 0 = n(b), \quad b_\alpha(a, b) = 1. \]

Since \( K \) is regular \( a + a \triangle a = a + b \) is a quadratic idempotent, and by Theorem 9.18 there is an automorphism \( \varphi \) of \( (C, \triangle, n) \) such that \( \varphi(a) = e_1; \)
then \( \varphi(K) = Fe_1 + Ff_1. \)

\[(ii) \Rightarrow (iii) \] This is clear since \( Fu_1 + Fv_1 \) is a two-dimensional composition subalgebra of \( (C, \triangle, n) \) orthogonal to \( Fe_1 + Ff_1. \)

\[(iii) \Rightarrow (i) \] Both \( K \) and \( K' \) are para-quadratic algebras. Take bases \( \{x, x \triangle x\} \)
and \( \{y, y \triangle y\} \) of \( K \) and \( K' \) respectively, with \( n(x) = 0 = n(y) \) and \( b_\alpha(x, x \triangle x) = 1 = b_\alpha(y, y \triangle y). \)
Then [Ekl09, Theorem 3.12] shows that either \( x \triangle y = 0 \)
or \( y \triangle x = 0. \) But \( y \triangle x = 0 \) implies \( x \triangle (y \triangle y) = -y \triangle (y \triangle x) = 0, \)
so replacing \( y \) by \( y \triangle y \) we may assume \( x \triangle y = 0. \) Then by Theorem 9.18
there is an automorphism of \( (C, \triangle, n) \) such that \( \varphi(x) = e_1 \) and \( \varphi(y) = u_1. \)
The para-unit \( x + x \triangle x \) of \( K \) corresponds to \( e_1 + f_1, \) which is a quadratic idempotent.

**Lemma 12.10.** Let \( K \) be a regular two-dimensional composition subalgebra of an Okubo algebra \((S, \ast, n_S).\) Let \( x, y \in K \) be two nonzero elements such that the endomorphism \( \text{ad}_x \ast \text{ad}_y \) is diagonalizable (where \( \text{ad}_z \ast: z \mapsto [x, z] = x \ast z - z \ast x). \)
Then \( n(x) = 0, y \) is a scalar multiple of \( x \ast x, \) and the eigenvalues of \( (\text{ad}_x \ast)^3 \)
are 0, with multiplicity 2, and \( \pm g_S(x), \) each with multiplicity 3. (Recall that \( g_S(x) = b_{\alpha\beta}(x, x \ast x).) \)

**Proof.** Extending scalars to \( \overline{F} \) we may assume by Lemma 12.9 that \((S, \ast, n_S)\)
is the algebra \((C, \triangle, n)\) in Table (5.11) and that \( K = Fe_1 + Ff_1. \) Hence \( x = \alpha e_1 + \beta f_1 \)
and \( y = \alpha' e_1 + \beta' f_1. \) Without loss of generality we may assume \( \alpha \neq 0. \) The subspace \( U = Fu_1 + Fv_2 + Fv_3 \)
is invariant under \( \text{ad}_x \ast \text{ad}_y \ast, \) and the coordinate matrix of the restriction of \( \text{ad}_x \ast \text{ad}_y \ast \)
to \( U \) in the basis \( \{u_1, u_2, u_3\} \) are \( \alpha C - \beta C^2 \) and \( \alpha' C - \beta' C^2 \) where \( C =\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \)
Hence the coordinate matrix of the restriction of \( \text{ad}_x \ast \text{ad}_y \ast \) to \( U \) is
\[
(\alpha C - \beta C^2)(\alpha' C - \beta' C^2) = -((\alpha \beta' + \beta \alpha') I + \beta \beta' C + \alpha \alpha' C^2,
\]
because \( C^3 = I \) (the identity matrix). Therefore we have
\[
(\text{ad}_x \ast \text{ad}_y \ast |_U)^3 = (-\alpha \beta' - \beta \alpha' + \beta \beta' + \alpha \alpha') I = ((\alpha - \beta)(\alpha' - \beta'))^3 I,
\]
and hence \( \text{ad}_x \ast \text{ad}_y \ast |_U - (\alpha - \beta)(\alpha' - \beta') I \) is simultaneously diagonalizable and nilpotent. It follows that \( \text{ad}_x \ast \text{ad}_y \ast |_U \) is a scalar multiple of the identity, so \( \beta \beta' = \alpha \alpha' = 0. \) But \( \alpha \neq 0, \) so \( \alpha' = 0 \) and hence (as \( y \neq 0 \)) \( \beta' \neq 0 \) and \( \beta = 0. \)
Thus \( x = \alpha e_1, y = \beta' f_1, \) and the result follows, since \( (\text{ad}_x \ast)^3 \) acts trivially on \( K, \) by multiplication by \( \alpha \beta' \) on \( U, \) and by multiplication by \( -\alpha^3 \) on \( V = Fu_1 + Fv_2 + Fv_3. \)

Even if the group \( \text{Aut}(S, \ast, n_S) \) is not smooth for an Okubo algebra over a field of characteristic 3, a result corresponding to Proposition 8.10 holds:

**Theorem 12.11.** Let \((S, \ast, n_S)\) and \((S', \ast, n_{S'})\) be two Okubo algebras. The following conditions are equivalent:
(i) The algebraic groups \( \text{Aut}(S, *, n) \) and \( \text{Aut}(S', *, n_{S'}) \) are isomorphic.

(ii) The Lie algebras (see 11.2) \((S, *)^−\) and \((S', *)^−\) are isomorphic. (Here \((S, *)^−\) denotes the Lie algebra defined over \( S \) with bracket \([x, y]^* = x \ast y - y \ast x\).)

(iii) The Okubo algebras \((S, *, n_S)\) and \((S', *, n_{S'})\) are either isomorphic or anti-isomorphic.

Proof. (i)⇒(ii) If the algebraic groups \( \text{Aut}(S, *, n) \) and \( \text{Aut}(S', *, n_{S'}) \) are isomorphic, so are their Lie algebras \( \text{Der}(S, *) \) and \( \text{Der}(S', *) \). But the simple Lie algebra \((S, *)^−\) is isomorphic to the only minimal ideal of \( \text{Der}(S, *) \) [Eld99, Theorem 4], and the same holds for \((S', *)^−\), whence the result.

(iii)⇒(i) This is straightforward.

(ii)⇒(iii) There are nonzero elements \( \alpha, \beta \in F \) such that the Okubo algebra \((S, *, n_S)\) is, up to isomorphism, the Okubo algebra \((C_{\alpha, \beta}, \delta, n)\) (see Theorem 12.6), which is the \( F \)-subalgebra of \((C \otimes F, \delta, n)\) generated by \( x = e_1 \otimes \alpha^+ \) and \( y = u_1 \otimes \beta^+ \). Note that \( x \ast y = 0 \). Then, \((S, *, n_S)\) is graded by \( \mathbb{Z}^3_2 \), with \( \deg(x) = (1,0,0) \) and \( \deg(y) = (0,0,1) \). Therefore \( S = \bigoplus_{\mu \in \mathbb{Z}^3_2} S_\mu \), with \( \dim S_\mu = 1 \) for any nonzero \( \mu \in \mathbb{Z}^3_2 \), \( S_0 = 0 \), and \( S_\mu \ast S_\nu \subseteq S_{\mu + \nu} \). Moreover, for any nonzero \( \mu \in \mathbb{Z}^3_2 \) and any nonzero \( z \in S_\mu \), we have

\[
\begin{align*}
- n_S(z) &= 0, \\
n_{S'}(z) &= Fz \ast z, \\
S_\mu \oplus S_{-\mu} &= \ker \text{ad}_z^* = \{ t \in S \mid t \ast z = z \ast t \}, \text{ and this coincides with the subalgebra } \text{alg}(z) \text{ generated by } z, \text{ and} \\
\bigoplus_{\nu \neq 0, \pm \mu} S_\nu &= \text{the subspace orthogonal to } \text{alg}(z) \text{ relative to } n_S, \\
\text{and coincides with the image of } \text{ad}_z^*.
\end{align*}
\]

Let \( \varphi : (S, *)^− \to (S', *)^− \) be an isomorphism of Lie algebras. Then the Lie algebra \((S', *)^−\) inherits a grading by \( \mathbb{Z}^3_2' \), \( S' = \bigoplus_{\nu \in \mathbb{Z}^3_2} S'_\nu \), with \( S'_\mu = \varphi(S_\mu) \) for any \( \mu \in \mathbb{Z}^3_2 \). Since \( \varphi \) is an isomorphism of Lie algebras, for any nonzero \( \mu \in \mathbb{Z}^3_2 \) and any nonzero \( u \in S'_\mu \), we have

\[
\begin{align*}
- S'_\mu \oplus S'_{-\mu} &= \ker \text{ad}_u^*, \\
\bigoplus_{\nu \neq 0, \pm \mu} S'_\nu &= \text{ad}_u^*(S').
\end{align*}
\]

In particular we have \( (\ker \text{ad}_u^*) \cap \text{ad}_u^*(S') = 0 \). Since \( n_{S'} \) is associative, we get

\[
\begin{align*}
S'_\mu \oplus S'_{-\mu} &= \ker \text{ad}_u^*, \\
\bigoplus_{\nu \neq 0, \pm \mu} S'_\nu &= \text{ad}_u^*(S').
\end{align*}
\]

Now, \( \text{alg}(u) = Fu + Fu \ast u \) is contained in \( \ker \text{ad}_u^* = S'_\mu \oplus S'_{-\mu} \). If \( u \ast u \) were a scalar multiple of \( u \), we would have \( u \ast u = \lambda u \) for some \( 0 \neq \lambda \in F \). But

\[
\begin{align*}
(\text{ad}_u^*)^3 &= (l_u^* - r_u^*)^3 \\
&= (l_u^*)^3 - (r_u^*)^3 \\
&= \lambda^3 l_u^* r_u^* - \lambda^3 r_u^* l_u^* = n_{S'}(u) \text{ Id.}
\end{align*}
\]

But for any \( v \) orthogonal to \( u \),

\[
(l_u^*)^3(v) = u \ast (u \ast (u \ast v)) = -u \ast (v \ast (u \ast u)) = -\lambda u \ast (v \ast u) = -\lambda n_{S'}(u)v.
\]
and also \((r_\nu^*)^3(v) = -\nu n_{S'}(u)v\). Thus \((\text{ad}_u^*)^3(v) = 0\) for any \(v\) orthogonal to \(u\), but this contradicts the fact that \(\ker\text{ad}_u^* \cap \text{ad}_u^*(S') = 0\). We conclude that \(u\) is not a scalar multiple of \(u\) and hence \(\text{alg}(u) = Fu + Fu \ast u = S'_u \oplus S'_{-\mu}\). This shows that \(\text{alg}(u)\) is a regular subalgebra, as we may consider the subalgebra \(S'_\nu \oplus S'_{-\nu}\) for \(\nu \neq 0, \pm \mu\) and use the characterization of regular subalgebras in item (iii) of Lemma 12.9. Finally, for \(0 \neq u \in S'_\mu\) and \(0 \neq v \in S'_{-\mu}\), \(\text{ad}_u^*\text{ad}_u^*(S'_\nu) \subseteq S'_{\mu - \nu + \nu} = S'_\nu\) for any \(\nu \in \mathbb{Z}_3^2\), and this implies that \(\text{ad}_u^*\text{ad}_u^*\) is diagonalizable. By Lemma 12.10 we conclude that \(n_{S'}(u) = 0\), \(v \in Fu \ast u\), and the eigenvalues of \((\text{ad}_u^*)^3\) are \(0\), with multiplicity 2, and \(\pm g_{S'}(u)\), each with multiplicity 3. Take now \(u = \varphi(x) \in S'_0(1,0)\) and \(v = \varphi(y) \in S'_0(0,1)\). Then \(n_{S'}(u) = 0 = n_{S'}(v)\), and
\[
n_{S'}(\text{alg}(u), \text{alg}(v)) = n_{S'}(S'_0(1,0) \oplus S'_0(0,1), S'_0(1,0) \oplus S'_0(0,1)) = 0.
\]
Then \(0 \neq g_{S'}(u) = n_{S'}(u, u \ast u)\) is an eigenvector of \((\text{ad}_u^*)^3\), and hence also of \((\text{ad}_u^*)^3\), so \(g_{S'}(u)\) is either \(g_{S'}(x)\) or \(-g_{S'}(x)\). Also \(g_{S'}(v) = \pm g_{S'}(y)\). Changing \(u\) by \(-u\) and \(v\) by \(-v\) if necessary, we get elements \(u \in S'_0(1,0), v \in S'_0(0,1),\) with \(g_{S'}(u) = g_{S'}(x)\) and \(g_{S'}(v) = g_{S'}(y)\). Now, [Eld09, Theorem 3.12] shows that either \(u \ast v = 0\) or \(v \ast u = 0\), and that \((S'_0, n_{S'})\) and \((S'_0, n_{S'})\) are isomorphic if \(u \ast v = 0\), and anti-isomorphic if \(v \ast u = 0\).

REMARK 12.12. With the notation in Section 12, if \(\dim_{F^3} g_{\alpha, \beta}(C_{\alpha, \beta})\) equals 1 or 3, then \((C_{\alpha, \beta}, \Delta, n)\) and its opposite algebra \((C_{\beta, \alpha}, \Delta, n)\) are isomorphic (see Theorem 12.7). However, if \(\dim_{F^3} g_{\alpha, \beta}(C_{\alpha, \beta})\) is 8, then \((C_{\alpha, \beta}, \Delta, n)\) is not isomorphic to its opposite algebra.

13. Groups admitting triality over arbitrary fields

In this section we classify all simple adjoint groups \(G\) of classical type \(\frac{1}{2}D_4\) which admit triality automorphisms over an arbitrary field \(F\). The first reduction is to groups of type \(\text{PGO}^+(n)\) and \(\text{Spin}(n)\), where \(n\) is a 3-Pfister form. More precisely:

THEOREM 13.1. Let \(F\) be an arbitrary field. Let \(G\) be an adjoint (resp. simply connected) simple group of type \(\frac{1}{2}D_4\) which admits a triality automorphism \(\phi\). There exists a symmetric composition \((S, \ast, n)\) such that the pair \((G, \phi)\) is isomorphic to a pair \((\text{PGO}^+(n), \rho_+)\) (resp. \((\text{Spin}(n), \rho_+))\).

Proof. It suffices to consider the adjoint case. The group \(G\) is a twisted form of \(G_0 = \text{PGO}_3^+(n)\), i.e., there exists a finite field extension \(L/F\) and a cocycle \(\xi \in (G_0 \times \mathbb{Z}/2)(L \otimes L)\) such that
\[
G(F) = \{x \in G_0(L) \mid \xi x_1 = x_2 \xi\}
\]
where \(\mathbb{Z}/2 \hookrightarrow \mathbb{S}_3\) is a fixed embedding and \(x_1, x_2\) are images of \(x\) under two natural mappings \(\pi_i: G_0(L) \to G_0(L \otimes L)\). The isomorphism class of \(G\) is given by the image of the cohomology class \([\xi] \in H^1_{\text{fppf}}(F, G_0 \times \mathbb{Z}/2)\) in \(H^1_{\text{fppf}}(F, G_0 \times \mathbb{S}_3)\). We view \(\phi\) as an element of \((G_0 \times \mathbb{S}_3)(L)\) (as in the proof of Lemma 3.2). The fact that \(\phi\) is \(F\)-defined implies that \(\xi \phi_1 = \phi_2 \xi\).
Replacing if necessary $L$ by a bigger field extension of $F$, we may assume in view of Theorem 7.6 that $\phi$ is conjugate in $(G_0 \rtimes \mathfrak{S}_3)(L)$ to one of the standard trialtarian automorphisms $\beta = \rho_5$ or $\beta = \rho_3$. Let $\beta = \gamma \phi \gamma^{-1}$. Take the cocycle $\xi' = \gamma_2 \xi_1^{-1}$, which is cohomologous to $\xi$. We have

$$\xi' \beta_1 = \gamma_2 \xi_1^{-1} \gamma_1 \phi_1 \gamma_1 = \gamma_2 \xi_1 \gamma_1 = \beta_2 \xi'. $$

Thus replacing $\xi$ by $\xi'$ we may assume that $\phi = \beta$. We now note that by construction $\beta \in \langle G_0 \rtimes \mathfrak{S}_3 \rangle(F)$, hence $\beta_1 = \beta_2 = \beta$ and this implies that $\xi$ takes values in $C_{G_0 \rtimes \mathfrak{S}_3}(\beta) = C_{G_0}(\beta) \rtimes \langle \beta \rangle$. Furthermore, since $G$ has type $1,2, D_4$ the image of the class $[\xi]$ under the projection

$$H^1_{\text{fppf}}(F, G_0 \rtimes \mathfrak{S}_3) \to H^1_{\text{fppf}}(F, \mathfrak{S}_3)$$

takes values in a subgroup of order 2 and on the other hand in the subgroup $\langle \beta \rangle$ of order 3. It follows that $[\xi] \in H^1_{\text{fppf}}(F, H_0)$ where $H_0 = C_{G_0}(\beta)$ is the subgroup of $G_0$ fixed under $\beta$. In view of Lemma 4.7 and Proposition 8.1 the cohomology set $H^1_{\text{fppf}}(F, H_0)$ classifies isomorphism classes of symmetric compositions which over an algebraic closure of $F$ induce trialtarian automorphisms conjugate to $\beta$. The map $H^1_{\text{fppf}}(F, H_0) \to H^1_{\text{fppf}}(F, G_0 \rtimes \mathfrak{S}_3)$ induced by the embedding $H_0 \to G_0 \rtimes \mathfrak{S}_3$ maps the class of the symmetric composition $(S, \star, n)$ to the isomorphism class of the group $\text{PGO}^+(n)$. It follows that $G \cong \text{PGO}^+(n)$ and that our automorphism $\phi$ is of the form $\rho_\star$ for some symmetric composition $\star$ on $S$.

Let $n$ be a 3-Pfister form over $F$ and let $G$ be either $\text{PGO}^+(n)$ or $\text{Spin}(n)$. The next aim is to describe the conjugacy classes of trialtarian automorphisms of $G$. Let $\sigma \in \tilde{G}(F)$ be a trialtarian automorphism of $G$ of order 3. We proved above that $\sigma$ is of the form $\sigma = \rho_\star$ for a proper symmetric composition algebra $(S, \star, n)$ which is either a para-octonion algebra or an Okubo algebra. The fixed subgroup $H = C_G(\langle \sigma \rangle)$ in $G$ is isomorphic to the automorphism group $\text{Aut}(S, \star, n)$, by Proposition 8.1. The group $\tilde{G}$ acts on itself by conjugation and we denote by $X = \text{Cl}_G(\sigma) \subset \tilde{G}$ the orbit of $\sigma$. This is a quasi-projective variety defined over $F$. As we proved before an arbitrary $F$-defined outer automorphism $\phi$ of $G$ whose centralizer in $G$ has the same type as that of $H$ is conjugate to $\sigma$ over an algebraic closure $\overline{F}$ of $F$ and hence $\phi$ can be viewed as an $F$-point of $X$. We denote by $X(F)/\sim$ the set of conjugacy classes of $F$-defined outer automorphisms of $G$ of order 3 whose centralizers have the same type as that of $H$ (equivalently, the set of $\tilde{G}(F)$-orbits in $X(F)$). Let now $\overline{H} = H \times \langle \sigma \rangle \cong C_G(\sigma)$. The group $\tilde{G}$ acts in a natural way on the quotient space $\tilde{G}/\overline{H}$ and by the general formalism of cohomology (see [DG70, p. 372–373]) we have a natural bijection between the set of $\tilde{G}(F)$-orbits in $(\tilde{G}/\overline{H})(F)$ and the set $\text{Ker}[H^1_{\text{fppf}}(F, \overline{H}) \to H^1_{\text{fppf}}(F, \tilde{G})]$. Also, by the universal property of the quotient $\tilde{G}/\overline{H}$ we have a natural $\tilde{G}$-equivariant morphism $\tilde{G}/\overline{H} \to X$ defined over $F$ which induces a bijection $(\tilde{G}/\overline{H})(\overline{F}) \to X(\overline{F})$. 
Consider the commutative diagram

\[
\begin{array}{ccc}
(G/\tilde{H})(F) & \xrightarrow{(G/\tilde{H})(F) \otimes F} & (G/\tilde{H})(F) \\
\lambda & \simeq & \pi_1 \\
X(F) & \xrightarrow{\phi} & X(F) \\
\end{array}
\]

Here \(\pi_1\), \(\pi_2\) (resp. \(\pi'_1\), \(\pi'_2\)) are maps induced by \(a \mapsto a \otimes 1\) and \(a \mapsto 1 \otimes a\) respectively and the top and bottom lines are the diagrams appearing in descent theory. We want to show that \(\lambda\) is a bijection. Let \(x \in X(F)\). Since \(\phi\) is a bijection there exists \(g \in G(F)\) such that \(\phi(g\tilde{H}) = x\). Let \(\theta: (G/\tilde{H})(F) \otimes F \rightarrow (G/\tilde{H})(F)\) (resp. \(X(F) \otimes F \rightarrow X(F)\)) be the bijection corresponding to \(a \otimes b \mapsto b \otimes a\). This is the descent data for the variety \(G/\tilde{H}\) (resp. \(X\) and \(\tilde{G}\)). Since \(x \in X(F)\) we get \(\theta(\pi'_1(x)) = \pi'_2(x)\). Then \(\psi(\theta(\pi_1(g\tilde{H}))) = \psi(\pi'_2(g\tilde{H}))\) and hence \(\psi(\pi_2(g)^{-1}\theta(\pi_1(g))\tilde{H}) = \sigma\). This implies that \(\pi_2(g)^{-1}\theta(\pi_1(g)) \in \tilde{H}(F)\otimes F\) or \(\theta(\pi_1(g)\tilde{H}) = \pi_2(g\tilde{H})\). By descent theory it follows that \(g\tilde{H} \in (G/\tilde{H})(F)\). Thus we have a natural bijection between the set of \(G(F)\)-orbits in \((G/\tilde{H})(F)\) and the set \(X(F)/\sim\). Combining all these facts we obtain

**Theorem 13.2.** There exists a natural bijection between the set of conjugacy classes of \(F\)-defined outer automorphisms of \(G\) of order 3 whose centralizers have the same type as that of \(H\) and the set \(\text{Ker}[H^1_{\text{fppf}}(F, \tilde{H}) \rightarrow H^1_{\text{fppf}}(F, \tilde{G})]\).

Before proceeding with consequences of the theorem we compute the kernel of the last map in terms of the connected groups \(H\) and \(G\).

**Proposition 13.3.** The natural mapping

\[
\lambda: \text{Ker}[H^1_{\text{fppf}}(F, H) \rightarrow H^1_{\text{fppf}}(F, G)] \rightarrow \text{Ker}[H^1_{\text{fppf}}(F, \tilde{H}) \rightarrow H^1_{\text{fppf}}(F, \tilde{G})]
\]

induced by the embedding \(H \hookrightarrow \tilde{H}\) is a bijection.

**Proof.** The mapping \(\lambda\) is induced by the embedding \(H \hookrightarrow \tilde{H} = H \times \langle \sigma \rangle\). Since

\[
H^1_{\text{fppf}}(F, H) \rightarrow H^1_{\text{fppf}}(F, H \times \langle \sigma \rangle)
\]

is clearly injective so is \(\lambda\). As for surjectivity of \(\lambda\), let

\[
\xi \in \text{Ker}[H^1_{\text{fppf}}(F, \tilde{H}) \rightarrow H^1_{\text{fppf}}(F, \tilde{G})].
\]

Since \(\tilde{H} = H \times \langle \sigma \rangle\) we can write \(\xi\) in the form \(\xi = \xi_1 \cdot \xi_2\) where \(\xi_1 \in H^1_{\text{fppf}}(F, H)\) and \(\xi_2 \in H^1_{\text{fppf}}(F, \langle \sigma \rangle)\). The composition \(\tilde{H} \rightarrow \tilde{G} \rightarrow \mathfrak{S}_3\) induces the map

\[
H^1_{\text{fppf}}(F, H \times \langle \sigma \rangle) \rightarrow H^1_{\text{fppf}}(F, \mathfrak{S}_3)
\]

which can be factored through

\[
H^1_{\text{fppf}}(F, \langle \sigma \rangle) \rightarrow H^1_{\text{fppf}}(F, \mathfrak{S}_3)
\]
and the last map has trivial kernel. But it follows from (13.4) that \( \xi_2 \) is in its kernel. Then \( \xi_2 = 1 \) implies \( \xi = \xi_1 \in H_1^{ipf}(F, H) \). It remains to show that \( \xi \) viewed as an element in \( H_1^{ipf}(F, \tilde{G}) \) is trivial. To do this, we look at the exact sequence

\[
1 \rightarrow G \rightarrow \tilde{G} \rightarrow \mathfrak{S}_3 \rightarrow 1
\]

which induces

\[
\tilde{G}(F) \xrightarrow{\mu_1} \mathfrak{S}_3(F) \rightarrow H_1^{ipf}(F, G) \xrightarrow{\mu_2} H_1^{ipf}(F, \tilde{G}).
\]

Surjectivity of \( \mu_1 \) implies that the kernel of \( \mu_2 \) is trivial. Since by our construction the cocycle \( \xi \) viewed as an element in \( H_1^{ipf}(F, \tilde{G}) \) is trivial, it is also trivial as an element in \( H_1^{ipf}(F, G) \). Thus we proved that \( \xi \) is in \( \text{Ker}\left[H_1^{ipf}(F, H) \rightarrow H_1^{ipf}(F, G)\right] \) and we are done. \( \square \)

In the next corollaries we describe explicitly the set

\[ \text{Ker}\left[H_1^{ipf}(F, H) \rightarrow H_1^{ipf}(F, G)\right] \]

where \( F \) is an arbitrary field. In view of 13.2 and 13.3 this leads to a complete list of conjugacy classes of trialitarian automorphisms of \( G \). Clearly, the description depends on the structure of the algebraic group \( H \) in question. We recall that this group is the automorphism group of a symmetric composition algebra of dimension 8. Each corollary corresponds to one of the types of symmetric algebras described in Section 8. We recall that we have para-octonion algebras (Type I), Okubo algebras in characteristic different from 3 (Type IIa and Type IIb), which are obtained from central simple algebras of degree 3, and Okubo algebras in characteristic 3 (Type III). In each case the trialitarian automorphisms are those linked to symmetric compositions, as described in Theorem 4.8.

**Corollary 13.5.** Let \( F \) be an arbitrary field. Let \( n \) be a 3-Pfister form over \( F \) and let \( G = \text{PGO}^{-4+8}(n) \) or \( G = \text{Spin}(n) \). Then there is a unique conjugacy class of trialitarian \( F \)-automorphisms of \( G \) whose centralizers in \( G \) have type \( G_2 \). The class is represented by the trialitarian automorphism \( \rho_o \) associated to the isomorphism class of the para-octonion algebra with norm \( n \).

**Proof.** Let \( \sigma \) be an \( F \)-defined trialitarian automorphism of \( G \) with the property \( H = C_G(\sigma) = G_2(n) \) where \( G_2(n) \) is the automorphism group of an octonion algebra with norm \( n \). By Proposition 8.3 the map \( H_1^{ipf}(F, H) \rightarrow H_1^{ipf}(F, G) \) is injective. The result now follows from Theorem 13.2 and Proposition 13.3. \( \square \)

The next corollary corresponds to the case where the algebraic group \( H \) is the group of automorphisms of the split Okubo algebra over a field of characteristic not 3 containing a primitive cube root of unity, hence \( H \) is the split group \( \text{PGL}_3 \).

**Corollary 13.6.** Let \( F \) be an arbitrary field of characteristic different from 3 and assume that a primitive cube root of unity \( \omega \) is contained in \( F \). Let \( G = \text{PGO}_8^* \). Then the set of conjugacy classes of trialitarian \( F \)-automorphisms of
G whose centralizers in G have type A2 is in one-to-one correspondence with the set of isomorphism classes of central simple algebras over F of degree 3.

Proof. We apply Theorem 13.2 and Proposition 13.3 with σ the standardtrialitarian automorphism π, associated to the split Okubo multiplication. Its centralizer H in G is PGL3, hence \( H^1_{\text{fppf}}(F, H) \) classifies central simple algebras over F of degree 3. Also, it is clear that \( H^1_{\text{fppf}}(F, H) \to H^1_{\text{fppf}}(F, G) \) is a trivial mapping because on one side every element in \( H^1_{\text{fppf}}(F, H) \) is split by a cubic extension of F and, on the other side, no element in \( H^1_{\text{fppf}}(F, G) \) is split by a cubic extension, in view of Springer’s Theorem (see for instance [Sch85, Theorem 5.4]). The result now follows from Theorem 13.2 and Proposition 13.3. \( \square \)

Remark 13.7. Tracing through our constructions, we can make the correspondence of Corollary 13.6 explicit as follows, using Galois homology since the characteristic of F is not 3. Let \( \rho \) be the standard trialitarian automorphism associated to the split Okubo multiplication over F. Its centralizer in \( G = \text{PGO}^8_3 \) is a split group PGL3. Let \( \rho \) be another trialitarian automorphism over F whose centralizer is a group of type A2. We may view \( \rho_\gamma \) and \( \rho \) as elements of the group \( (\text{PGO}^8_3 \times \text{G}_3)(F) \). We know that they are conjugate over a separable closure \( F_{\text{sep}} \) of F by an element of \( \text{PGO}^8_3 \). Let \( g \in \text{PGO}^8_3(F_{\text{sep}}) \) be such that \( \rho = g\rho_\gamma g^{-1} \). For \( \gamma \in \text{Gal}(F_{\text{sep}}/F) \) we have \( \rho_\gamma = \rho \) and \( \rho_\gamma^1 = \rho_\gamma \) since \( \rho \) and \( \rho_\gamma \) are defined over F, hence \( (g_\gamma)^{-1}g \) lies in the centralizer of \( \rho_\gamma \) in \( \text{PGO}^8_3 \). Thus we get a cocycle \( a_\gamma = (g_\gamma)^{-1}g \in Z^1(F, \text{PGL}_3) \). This cocycle gives rise to a central simple algebra, say \( A \), of degree 3 over F. Note that it follows from the construction that the group \( \text{PGL}(A) \) is the centralizer of \( \rho \) in \( \text{PGO}^8_3 \).

Conversely, assume that we have a central simple algebra A of degree 3 over F. It is determined by a cocycle \( a_\gamma \) with coefficients in \( \text{PGL}_3 \). Since the map \( H^1(F, \text{PGL}_3) \to H^1(F, \text{PGO}^8_3) \) is trivial (as observed in the proof of Corollary 13.6), there exists \( g \in \text{PGO}^8_3(F_3) \) such that \( a_\gamma = (g_\gamma)^{-1}g \). Consider \( \rho = g\rho_\gamma g^{-1} \). Since \( \rho_\gamma \) is defined over F and \( (g_\gamma)^{-1}g \) centralizes \( \rho_\gamma \) for all \( \gamma \), it follows that \( \rho \) is defined over F.

Similar considerations apply to corollaries 13.8 and 13.9 below.

Now, suppose the field F does not contain \( \omega \), and char \( F \neq 3 \). Let \( K = F(\omega) \), which is a quadratic field extension of F. Before stating the next corollary we recall that on a given central simple K-algebra B of degree 3, unitary involutions \( \tau \) fixing F are classified by a quadratic form invariant \( \pi(\tau) \), which is a 3-Pfister form over F split by K, see [KMRT98, (19,6)]. This result is proved in [KMRT98] under the hypothesis that char \( F \neq 2 \), but the following observations show that it also holds when char \( F = 2 \).

If \( \text{char} F \neq 2 \), the 3-fold Pfister form \( \pi(\tau) \) is obtained by modifying the form \( Q_\tau(x) = \text{Trd}(x^2) \) defined on the F-vector space Sym(\( \tau \)) of \( \tau \)-symmetric elements: see [KMRT98, (19.4)]. The arguments in the proof of the classification theorem [KMRT98, (19.6)] can be used when char \( F = 2 \), substituting for \( Q_\tau \) the restriction to Sym(\( \tau \)) of the quadratic form \( \text{Sr}(x) \), which is the
coefficient of the indeterminate $t$ in the reduced characteristic polynomial of $x$ (as in Proposition 8.5). Let $\varphi_\tau$ be the restriction of $Srd$ to the $F$-vector space $\text{Sym}(\tau)^0$ of $\tau$-symmetric elements of trace zero, and let $[1,1]$ denote the quadratic form $X^2 + XY + Y^2$ over $F$. If $B$ is split and $\tau$ is adjoint to a hermitian form $h$ with diagonalization $\langle \delta_1, \delta_2, \delta_3 \rangle_K$, computation shows that $\varphi_\tau \cong [1,1] \perp \langle \delta_1 \delta_2, \delta_2 \delta_3, \delta_3 \delta_1 \rangle \cdot n_K$, where $n_K$ is the norm form of $K/F$. Therefore, if $\tau'$ is also a unitary involution on the split algebra $B$, and $\tau'$ is adjoint to a hermitian form $h'$, the arguments on [KMRT98, p. 305] show that $\varphi_{\tau'} \cong \varphi_\tau$ implies $h'$ is similar to $h$, hence $\tau'$ and $\tau$ are isomorphic. Therefore, the 3-fold Pfister form $\pi(\tau) = \langle 1, \delta_1 \delta_2, \delta_2 \delta_3, \delta_3 \delta_1 \rangle \cdot n_K$ determines the involution $\tau$ up to isomorphism. The case where $B$ is not split reduces to the split case by an odd-degree scalar extension. The arguments on p. 305 of [KMRT98] apply in characteristic 2, since [KMRT98, (6.17)] relies on a result of Bayer-Lenstra that also holds in characteristic 2 (see [BFT07, Theorem 1.13] for a discussion of the Bayer-Lenstra result in characteristic 2).

**Corollary 13.8.** Let $F$ be an arbitrary field of characteristic not 3 and assume that $F$ does not contain a primitive cube root of unity $\omega$. Let $n$ be a 3-Pfister form split by $K = F(\omega)$ and let $G$ be $\text{PGO}^+(n)$ or $\text{Spin}(n)$. Then the set of conjugacy classes of trialitarian $F$-automorphisms of $G$ of order 3 whose centralizers in $G$ have type $A_2$ is in one-to-one correspondence with the set of $F$-isomorphism classes of pairs $(B, \tau)$ where $B$ is a central simple $K$-algebra of degree 3 and $\tau$ is a unitary involution on $B$ fixing $F$ such that $\pi(\tau) = n$.

**Proof.** Let $\sigma$ be an outer $F$-automorphism of $G$ of order 3 whose centralizer $H$ in $G$ is an outer form of type $^2A_2$, of inner type over $K$. Its existence follows from our previous results. Indeed, take the standard trialitarian automorphism $\rho_0$ of $G$ corresponding to the para-octonion algebra $(C, \cdot, n)$ with norm $n$. Its automorphism group has type $G_2$ and splits over $K = F(\omega)$. Hence it contains a subtorus of the form $R_K^{(1)}(\mathbb{G}_m)$. Such a torus contains an element of order 3 over $F$, namely $\omega$. Twisting the multiplication $\cdot$ by $\omega$ we get an Okubo algebra $(C, \cdot_\omega, n)$. The corresponding automorphism $\rho_{\omega}$ of $G$ is as required.

Let $H_0$ be an adjoint quasi-split $F$-group of type $^2A_2$, of inner type over $K$. It is known that $H$ is a twisted form of $H_0$, i.e. $H = \xi H_0$ where $\xi \in Z^1(F,H_0)$. Also, we know that the pointed set $H^{1}_{\text{ppf}}(F,H_0)$ classifies pairs $(B, \tau)$ where $B$ is a central simple algebra over $K$ of degree 3 and $\tau$ is a unitary involution on $B$ (see [KMRT98, (30.21)]) and that there exists a natural bijection $H^1(F,H_0) \to H^1(F,H)$ which takes the class $[\xi]$ into the trivial class. Thus the pointed set $H^{1}_{\text{ppf}}(F,H)$ also classifies the same pairs $(B, \tau)$. The mapping $H^{1}_{\text{ppf}}(F,H) \to H^{1}_{\text{ppf}}(F,G)$ takes a pair $(B, \tau)$ to the class in $H^{1}_{\text{ppf}}(F,G)$ corresponding to the 3-Pfister form $\pi(\tau)$ (see [KMRT98, §30.C]). So the result follows from Theorem 13.2 and Proposition 13.3. \qed
Finally, suppose \( \text{char } F = 3 \). If the centralizer \( H \) of a trialitarian automorphism \( \phi \) is not a form of type \( G_2 \), this centralizer is the automorphism group of an Okubo algebra, as computed in Section 10.

**Corollary 13.9.** Let \( G = \mathbf{PGO}_8^+ \) be defined over an arbitrary field \( F \) of characteristic 3. The set of conjugacy classes of trialitarian \( F \)-automorphisms of \( G \) of Okubo type in \( G \) is in one-to-one correspondence with the set of isomorphism classes of Okubo algebras over \( F \).

**Proof.** Taking Theorem 12.6 into account, the proof is along the same lines as the proof of Corollary 13.6. \( \square \)

**References**


Abstract. A collection of \( n \) distinct hyperplanes \( L_i = \{ l_i = 0 \} \subset \mathbb{P}^{n-1} \), the \((n-1)\)-dimensional projective space over an algebraically closed field of characteristic not equal to 2, is a polar simplex of a smooth quadric \( Q^{n-2} = \{ q = 0 \} \), if each \( L_i \) is the polar hyperplane of the point \( p_i = \bigcap_{j \neq i} L_j \), equivalently, if \( q = l_1^2 + \ldots + l_n^2 \) for suitable choices of the linear forms \( l_i \). In this paper we study the closure \( VPS(Q, n) \subset \text{Hilb}_n(\mathbb{P}^{n-1}) \) of the variety of sums of powers presenting \( Q \) from a global viewpoint: \( VPS(Q, n) \) is a smooth Fano variety of index 2 and Picard number 1 when \( n < 6 \), and \( VPS(Q, n) \) is singular when \( n \geq 6 \).

2010 Mathematics Subject Classification: 14J45, 14M

Keywords and Phrases: Fano n-folds, Quadric, polar simplex, syzygies

Contents

1. Introduction 470
2. Apolar subschemes of length \( n \) 474
3. A rational parameterization 477
4. The Mukai form 480
5. An open affine subvariety 487
6. Global invariants of \( VPS(Q, n) \) 501
References 504

The first author was partially supported by an UCM-EEA grant under the NILS program, and both authors were supported by Institut Mittag-Leffler.
1. Introduction

Let \( Q = \{ q = 0 \} \) be a \((n - 2)\)-dimensional smooth quadric defined over the complex numbers, or any algebraically closed field of characteristic not equal to 2. We denote the projective space containing \( Q \) by \( \mathbb{P}^{n-1} \) because its dual space \( \mathbb{P}^{n-1} \) plays the major role in this paper. A collection \( L_1 = \{ l_1 = 0 \}, \ldots, L_n = \{ l_n = 0 \} \) of \( n \) hyperplanes is a polar simplex iff each \( L_i \) is the polar of the point \( p_i = \bigcap_{j \neq i} L_j \), equivalently, iff the quadratic equation

\[
q = \sum_{i=1}^{n} l_i^2
\]

holds for suitable choices of the linear forms \( l_i \) defining \( L_i \). In this paper we study the collection of polar simplices, or equivalently, the variety of sums of powers presenting \( q \) from a global viewpoint.

We may regard a polar simplex as a point in \( \text{Hilb}_n(\mathbb{P}^{n-1}) \). Let \( VPS(Q, n) \subset \text{Hilb}_n(\mathbb{P}^{n-1}) \) be the closure of the variety of sums of \( n \) squares presenting \( Q \).

The first main result is:

**Theorem 1.1.** If \( 2 \leq n \leq 5 \), then \( VPS(Q, n) \) is a smooth rational \( \binom{n}{2} \)-dimensional Fano variety of index 2 and Picard number 1. If \( n \geq 6 \), then \( VPS(Q, n) \) is a singular rational \( \binom{n}{2} \)-dimensional variety.

If \( n = 2 \), then \( VPS(Q, n) = \mathbb{P}^1 \), and if \( n = 3 \), then \( VPS(Q, n) \) is a rational Fano threefold of index 2 and degree 5 (cf. [Muk92]).

The quadratic form defines a collineation \( q : \mathbb{P}^{n-1} \to \mathbb{P}^{n-1} \), let \( q^{-1} : \mathbb{P}^{n-1} \to \mathbb{P}^{n-1} \) be the inverse collineation, and \( Q^{-1} = \{ q^{-1} = 0 \} \subset \mathbb{P}^{n-1} \) the corresponding quadric. Consider the double Veronese embedding \( Q^{-1} \to \mathbb{P}(\binom{n+1}{2} - 2) \), and let \( TQ^{-1} \) be the image by the Gauss map of tangent spaces \( Q^{-1} \to \mathbb{G}(n - 1, \binom{n+1}{2} - 1) \). Our second main result is:

**Theorem 1.2.** \( VPS(Q, n) \) has a natural embedding in the Grassmannian variety \( \mathbb{G}(n - 1, \binom{n+1}{2} - 1) \) and contains the image \( TQ^{-1} \) of the Gauss map of the quadric \( Q^{-1} \) in its Veronese embedding. When \( n = 4 \) or \( n = 5 \), the restriction of the Plicker bundle generates the Picard group of \( VPS(Q, n) \), and the degree is 310, resp. 395780.

We denote the coordinate ring of \( \mathbb{P}^{n-1} \) by \( S = \mathbb{C}[x_1, \ldots, x_n] \) and the coordinate ring of the dual \( \mathbb{P}^{n-1} \) by \( T = \mathbb{C}[y_1, \ldots, y_n] \). In particular \( S_1 = (T_1)^* \), so we may set \( \mathbb{P}^{-1} = \mathbb{P}(T_1) \), the projective space of 1-dimensional subspaces of \( T_1 \) with coordinate functions in \( S \), and \( \mathbb{P}^{-1} = \mathbb{P}(S_1) \) with coordinate functions in \( T \). Let \( q \in T = \mathbb{C}[y_1, \ldots, y_n] \) be a quadratic form defining the smooth \((n - 2)\)-dimensional quadric \( Q \subset \mathbb{P}^{n-1} = \mathbb{P}(S_1) \). Regard \( [q] \) as a point in \( \mathbb{P}(T_2) \) and consider the Veronese variety \( V_2 \subset \mathbb{P}(T_2) \) of squares,

\[
V_2 = \{ [l^2] \in \mathbb{P}(T_2) | l \in \mathbb{P}(T_1) \}.
\]

Then a polar simplex to \( Q \) is simply a collection of \( n \) points on \( V_2 \) whose linear span contains \([q]\). Any length \( n \) subscheme \( \Gamma \subset V_2 \) whose span in \( \mathbb{P}(T_2) \)
contains \([q]\) is called an apolar subschemes of length \(n\) to \(Q\). The closure \(VPS(Q, n)\) of the polar simplices in \(\text{Hilb}_n(\mathbb{P}(T_1))\) consists of apolar subschemes of length \(n\). We denote by \(VAPS(Q, n)\) the subset of \(\text{Hilb}_n(\mathbb{P}(T_1))\), with reduced scheme structure, parameterizing all apolar subschemes of length \(n\) to \(Q\). Our third main result is:

**Theorem 1.3.** The algebraic set \(VAPS(Q, n)\) is isomorphic to the complete linear section

\[ VAPS(Q, n) = \langle TQ^{-1} \rangle \cap G(n - 1, T_2/q) \subset \mathbb{P}({\wedge}^{n-1}(T_2/q)) \]

in the Plücker space. For \(n \leq 6\) the two subschemes \(VPS(Q, n)\) and \(VAPS(Q, n)\) coincide. For \(n \geq 24\), the scheme \(VAPS(Q, n)\) has more than one component.

Notice that we do not claim that the linear section \(\langle TQ^{-1} \rangle \cap G(n - 1, T_2/q)\) is reduced, only that its reduced structure coincides with \(VAPS(Q, n)\). The linear span \(\langle TQ^{-1} \rangle\) has dimension \(\binom{2n-1}{n-1} - \binom{2n-3}{n-2} - 1\), while the Grassmannian has dimension \((n - 1)\binom{n}{2} - (\binom{n+1}{2} - 1)\)-dimensional Plücker space. So this linear section is far from a proper linear section when \(n \geq 4\), i.e. the codimension of \(VAPS(Q, n)\) in the Grassmannian is much less than the codimension of its linear span in the Plücker space.

We find a covering of \(VAPS(Q, n)\) by affine subschemes \(V^{{\text{aff}}}h(n)\) that are contractible to a point \([p] \in VPS(Q, n)\) (Lemma 5.3). Therefore the apolar subschemes \(\Gamma_p\) play a crucial point. Let us explain what they are: The projection of the Veronese variety \(V_2 \subset \mathbb{P}(T_2)\) from \([q] \in \mathbb{P}(T_2)\) is a variety \(V_{2,q} \subset \mathbb{P}(T_2/q)\). Since a polar simplex to \(Q\) is a collection on \(n\) points on \(V_2\) whose span contains \([q]\), the variety \(V^{SP}(Q, n)\) is naturally embedded in and in fact coincides with the variety of \((n - 2)\)-secant spaces of the projected Veronese variety \(V_{2,q}\). The double Veronese embedding of \(Q^{-1}\) is a linearly normal subvariety in \(V_{2,q}\) that spans \(\mathbb{P}(T_2/q)\). For each point \(p \in Q^{-1}\) consider the tangent space to \(Q^{-1}\) in this embedding. This tangent space intersects \(V_{2,q}\) along the subscheme \(\Gamma_p\), and belong to the boundary of variety of \((n - 2)\)-secant spaces of \(V_{2,q}\).

The affine subscheme \(V_h^{{\text{aff}}}n\) is contractible to \(\Gamma_p\), but depend only on a hyperplane: It consists of the apolar subschemes that do not intersect a tangent hyperplane \(h\) to \(Q^{-1}\). The point \(p\) is simply a point on \(Q^{-1}\) that does not lie in this hyperplane.

Our computations show that the affine scheme \(V_h^{{\text{aff}}}n\) and certain natural subschemes has particularly interesting structure: \(V_h^{{\text{aff}}}n\) is isomorphic to an affine space when \(n < 6\) while \(V_h^{{\text{aff}}}(6)\) is isomorphic to a 15-dimensional spinor variety (Corollary 5.16). Why this spinor variety appears is quite mysterious to us. Recall that Mukai showed that a general canonical curve of genus 7 is a linear section of the spinor variety. Let \(V^{{loc}}_p(n) \subset VAPS(Q, n)\) be the subscheme of apolar subschemes in \(VAPS(Q, n)\) with support at a single point \(p \in Q^{-1}\). The subscheme \(V^{{loc}}_p(n)\) is naturally contained in \(V^{{sec}}_p(n)\), the variety of apolar subschemes in \(V_h^{{aff}}(n)\) that contains
the point \( p \). We compute these subschemes with Macaulay2 [GS] when \( n < 6 \) and find that \( V_{\text{loc}}(5) \) is isomorphic to a 3-dimensional cone over the tangent developable of a rational normal sextic curve. This cone is a codimension 3 linear section of the scheme \( V_{\text{sec}}(5) \), which is isomorphic to a 6-dimensional cone over the intersection of the Grassmannian \( G(2, 5) \) with a quadric. Mukai showed that a general canonical curve of genus 6 is a linear section of the intersection of \( G(2, 5) \) with a quadric. The appearances in the cases \( n = 5, 6 \) of a natural variety whose curve sections are canonical curves is both surprising and unclear to us. The computational results are summarized in Table 1 in Section 5.

By the very construction of polar simplices, it is clear that \( \text{VPS}(Q, n) \) has dimension \( \binom{n}{2} \). On the other hand, the special orthogonal group \( SO(n, q) \) that preserves the quadratic form \( q \), acts on the set of polar simplices: If we assume that the symmetric matrix of \( q \) with respect to the variables in \( T \) is the identity matrix, then regarding \( SO(n, q) \) as orthogonal matrices the rows define a polar simplex. Matrix multiplication therefore defines a transitive action of \( SO(n, q) \) on the set of polar simplices. By dimension count, this action has a finite stabilizer at a polar simplex. This stabilizer is the subgroup \( H \subset SO(n, q) \) of rotational symmetries of the hypercube \([-1, 1]^n \subset \mathbb{R}^n\) of order \( 2^{n-1} \cdot n! \) as suggested by an anonymous referee. We get

**Proposition 1.4.** \( \text{VPS}(Q, n) \) is a compactification of the group \( SO(n, q)/H \).

The linear representation of \( SO(n, q) \) on \( T_2 \) decomposes

\[
T_2 = \langle q \rangle \oplus T_{2,q},
\]

where the hyperplane \( \text{P}(T_{2,q}) \) intersect the Veronese variety \( V_2 \) along the Veronese image of \( Q^{-1} \). Therefore we may identify \( T_2/q = T_{2,q} \) and the projection from \( [q]; \text{P}(T_2) \rightarrow \text{P}(T_{2,q}) \) is an \( SO(n, q) \)-equivariant projection. \( Q^{-1} \subset \text{P}(T_{2,q}) \) is a closed orbit, and similarly the image \( TQ^{-1} \) of the Gauss map is a closed orbit for the induced representation on the Plücker space of \( G(n - 1, T_{2,q}) \). The linear span of this image is therefore the projectivization of an irreducible representation of \( SO(n, q) \). The set of polar simplices form an orbit for the action of \( SO(n, q) \), so the linear span of \( \text{VPS}(Q, n) \) is also the projectivization of an irreducible representation of \( SO(n, q) \). Therefore

\[
\text{VPS}(Q, n) \subset \langle TQ^{-1}\rangle \cap G(n - 1, T_{2,q}).
\]

We show that the intersection \( \langle TQ^{-1}\rangle \cap G(n - 1, T_{2,q}) \) parameterizes all apolar subschemes of length \( n \), hence Theorem 1.1.

The organization of the paper follows distinct approaches to \( \text{VPS}(Q, n) \). To start with we introduce the classical notion of apolarity and regard polar simplices as apolar subschemes in \( \text{P}(T_1) \) of length \( n \) with respect to \( q \). We use syzygies to characterize these subschemes among elements of the Hilbert scheme. In fact, polar simplices are characterized by their smoothness, the Betti numbers of their resolution, and their apolarity with respect to \( q \). Allowing singular subschemes, we consider all apolar subschemes of length \( n \). We show in Section 2 that these subschemes naturally appear in the closure \( \text{VPS}(Q, n) \) of the
set of polar simplices in the Hilbert scheme. For \( n > 6 \) there may be apolar subschemes of length \( n \) that do not belong to the closure \( VPS(Q, n) \) of the smooth ones. In fact, we show in Section 2 that at least for \( n \geq 24 \), there are nonsmoothable apolar subschemes of length \( n \), i.e. that \( VPS(Q, n) \) is not the only component of \( VAPS(Q, n) \).

The variety \( VPS(Q, n) \), in its embedding in \( G(n - 1, T_{2,q}) \), has order one, i.e. through a general point in \( P(T_{2,q}) \) there is a unique \((n - 2)\)-dimensional linear space that form the span of an apolar subscheme \( \Gamma \) of length \( n \). This is a generalization of the fact that a general symmetric \( n \times n \) matrix has \( n \) distinct eigenvalues. In Section 3 we use a geometric approach to characterize the generality assumption.

The fact that \( VPS(Q, n) \) has order one, means that it is the image of a rational map

\[
\gamma : P(T_{2,q}) \to G(n - 1, T_{2,q}).
\]

In Section 4 we use a trilinear form introduced by Mukai to give equations for the map \( \gamma \). With respect to the variables in \( T \) we may associate a symmetric matrix \( A \) to each quadratic form \( q' \in T_{2,q} \). The Mukai form associates to \( q' \) a space of quadratic forms in \( S_{2} \) that vanish on all the projectivized eigenspaces of the matrix \( A \). For general \( q' \) these quadratic forms generate the ideal of the unique common polar simplex of \( q \) and \( q' \). This is Proposition 4.2. The Mukai form therefore defines the universal family of polar simplices, although it does not extend to the whole boundary. Common apolar subschemes to \( q \) and \( q' \), when \( q' \) has rank at most \( n - 2 \), form the exceptional locus of the map \( \gamma \).

We do not compute the image of \( \gamma \) in \( G(n - 1, T_{2,q}) \). Instead we compute affine perturbations of \([\Gamma_{p}]\) in \( G(n - 1, T_{2,q}) \) that correspond to apolar subschemes to \( Q \). These perturbations form the affine subschemes \( V_{h}^{n\text{aff}}(n) \) that cover \( VAPS(Q, n) \). In Section 5 we make extensive computations of these affine subschemes. Each one of them is contractible to a point \( [\Gamma_{p}] \) on the subvariety \( TQ^{-1} \subset VPS(Q, n) \). The question of smoothness of \( VPS(Q, n) \) is reduced to a question of smoothness of the affine scheme \( V_{h}^{n\text{aff}}(n) \) at the point \( [\Gamma_{p}] \). For \( n \leq 5 \) we show that such a point is smooth, while for \( n \geq 6 \), it is singular. The main result of Section 5 is however Theorem 1.3, that \( VAPS(Q, n) \) is a linear section of the Grassmannian.

In the final Section 6 we return to the geometry of \( VPS(Q, n) \) and compute the degree by a combinatorial argument for any \( n \). The Fano-index is computed using the natural \( P^{n-2} \)-bundle on \( VPS(Q, n) \), obtained by restricting the incidence variety over the Grassmannian, and its birational morphism to \( P(T_{2,q}) \).

We thank Tony Iarrobino for sharing his insight on Artinian Gorenstein rings with us, Francesco Zucconi for valuable comments on a previous version of this paper.

Let us briefly summarize the notation:

- \( \mathbb{C} \) denotes the field of complex numbers.
• $q \in T_2$ is a non-degenerate quadratic form, and defines a collineation $q : S_1 \to T_1$ and a linear form $q : S_2 \to \mathbb{C}$.
• $Q$ is the quadratic hypersurface \{ $q = 0$ \} $\subseteq \mathbb{P}(S_1)$.
• $q^{-1} \in S_2$ is a quadratic form, that defines the collineation $q^{-1} : T_1 \to S_1$ inverse to $q$ and a linear form $q^{-1} : T_2 \to \mathbb{C}$.
• $Q^{-1}$ is the quadratic hypersurface \{ $q^{-1} = 0$ \} $\subseteq \mathbb{P}(T_1)$.
• $q^2 \subseteq S_2$ is the kernel of the linear form $q : S_2 \to \mathbb{C}$.
• $T_{2,q}$ is the kernel $(q^{-1})^2$ of the linear form $q^{-1} : T_2 \to \mathbb{C}$.
• $\pi_q : \mathbb{P}(T_2) \dashrightarrow \mathbb{P}(T_{2,q})$ is the projection from $[q] \in \mathbb{P}(T_2)$, and $V_{2,q} \subseteq \mathbb{P}(T_{2,q})$ is the image under this projection of the Veronese variety $V_2 \subseteq \mathbb{P}(T_2)$.

2. Apolar subschemes of length $n$

We follow the approach of [RS00]: The apolarity action is defined as the action of $S = \mathbb{C}[x_1, \ldots, x_n]$ as polynomial differential forms on $T = \mathbb{C}[y_1, \ldots, y_n]$ by setting $x_i = \frac{\partial}{\partial y_i}$. This makes the duality between $S_1$ and $T_1$ explicit and, in fact, defines a natural duality between $T_1$ and $S_1$. The form $q \in T_2$ define the smooth $(n - 2)$-dimensional quadric hypersurface

$$Q = \{ \left( \sum a_i x_i \right) | \left( \sum a_i \frac{\partial}{\partial y_i} \right)^2(q) = 0 \} \subseteq \mathbb{P}(S_1).$$

Apolarity defines a graded Artinian Gorenstein algebra associated to $Q$:

$$A^Q = \mathbb{C}[x_1, \ldots, x_n]/(q^\perp)$$

where

$$q^\perp = \{ D \in S_2 = \mathbb{C}[x_1, \ldots, x_n]_2 | D(q) = 0 \}.$$

A subscheme $Y \subseteq \mathbb{P}(T_1)$ is apolar to $Q$, or equivalently apolar to $q$, if the space of quadratic forms in its ideal $I_{Y,2} \subseteq q^\perp$. The apolarity lemma (cf. [RS00] 1.3) says that any smooth $\Gamma$, $[\Gamma] \in \text{Hilb}_n(\mathbb{P}(T_1))$ is a polar simplex with respect to $Q \subseteq \mathbb{P}(S_1) = \mathbb{P}^{n-1}$ if and only if $I_{\Gamma,2} \subseteq q^\perp \subseteq S_2$, i.e. $\Gamma$ is apolar to $Q$. We drop, for the moment, the smoothness criterion and consider any $[\Gamma] \in \text{Hilb}_n(\mathbb{P}(T_1))$, such that $\Gamma$ is apolar to $Q$. Notice that since $Q$ is nonsingular, $\Gamma$ is nondegenerate. But more is known: The following are the graded Betti numbers of $A^Q$ and $\Gamma$, given in Macaulay2 notation [GS].

**Proposition 2.1.** a) For a smooth quadric $Q \subseteq \mathbb{P}^{n-1}$ the syzygies of the apolar Artinian Gorenstein ring $A^Q$ are

$$1 \quad - \quad - \quad - \quad - \quad - \quad - \quad -$$

$$- \quad - \quad \frac{n-1}{n+1} \binom{n+2}{2} \quad \cdots \quad \frac{k(n-k)}{n+1} \binom{n+2}{k+1} \quad \cdots \quad \frac{n-1}{n+1} \binom{n+2}{n} \quad -$$

$$- \quad - \quad \cdots \quad - \quad \cdots \quad - \quad - \quad 1$$

b) A zero-dimensional nondegenerate scheme $\Gamma \subseteq \mathbb{P}^{n-1}$ of length $n$ has syzygies

$$1 \quad - \quad - \quad - \quad - \quad - \quad -$$

$$- \quad - \quad \binom{n}{2} \quad \cdots \quad k \binom{n}{k+1} \quad \cdots \quad (n-1) \binom{n}{n} \quad -$$

**Proof.** Eg. [Beh81] and [ERS81]
Corollary 2.2. The natural morphism

\[ VAPS(Q, n) \to \mathcal{G}(\binom{n}{2}, q^\perp); \quad \Gamma \mapsto I_{\Gamma, 2} \subset q^\perp \]

is injective. Equivalently, there is a natural injective morphism

\[ VAPS(Q, n) \to \mathcal{G}(n-1, T_{2,q}); \quad \Gamma \mapsto I_{\Gamma, 2}^2 \subset T_{2,q} \]

into the variety of \((n-2)\)-dimensional subspaces of \(\mathbf{P}(T_{2,q})\) that intersect the projected Veronese variety \(V_{2,q}\) in a scheme of length \(n\). In particular, the Hilbert scheme and Grassmannian compactification in \(\mathcal{G}(n-1, T_{2,q})\) of the variety of polar simplices coincide.

Proof. Apolarity defines a natural isomorphism \(q^\perp \cong T_{2,q}^\perp\). Therefore the subspace \(I_{\Gamma, 2} \subset q^\perp\) defines a \((n-1)\)-dimensional subspace \(I_{\Gamma, 2}^2 \subset T_{2,q}\). The intersection \(P(I_{\Gamma, 2}^2) \cap V_{2,q}\) with the projected Veronese variety is precisely \(\pi_q(\Gamma)\).

The variety \(V_{PS}(Q, n) \subset \text{Hilb}_n(P^{n-1})\) is the closure of the set of polar simplices inside the set of apolar subschemes of length \(n\). The former set is irreducible, while the latter set is a closed variety defined by the condition that the generators of the ideal of the subscheme lie in \(q^\perp\). By Proposition 2.1, the map \(\Gamma \mapsto I_{\Gamma, 2} \subset q^\perp\) extends to all of \(V_{PS}(Q, n)\) as an injective morphism. □

We relate apolarity to polarity with respect to a quadric hypersurface. The classical notion of polarity is the composition of the linear map \(q^{-1}\) with apolarity: The polar to a point \([l] \in P(T_1)\) with respect to \(Q^{-1}\) is the hyperplane \(h_1 = P(q^{-1}(l^\perp)) \subset P(T_1)\), where

\[ (q^{-1}(l^\perp))^\perp = \{ l' \in T_1 | l'(q^{-1}(l)) = q^{-1}(l \cdot l') = 0 \}. \]

In particular, the polar hyperplane to \(l\) contains \(l\) if and only if \(q^{-1}(l^2) = 0\), i.e. the point \([l]\) lies on the hypersurface \(Q^{-1}\).

Let \(\Gamma \subset P(T_1)\) be a length \(n\) subscheme that contains \([l]\) and is apolar to \(Q\). The subscheme \(\Gamma' \subset \Gamma\) residual to \([l]\) is defined by the quotient \(I_{\Gamma'} = I_{\Gamma} : (l^\perp)\). Since \(\Gamma\) is non degenerate, \(\Gamma'\) spans a unique hyperplane. This hyperplane is defined by a unique linear form \(u' \in S_1\), and is characterized by the fact that \(u' \cdot u(q) = u'q(u) = 0\) for all \(u \in l^\perp\), so it is the hyperplane \(P(q(l^\perp))\). But

\[ l' \in q(l^\perp) \iff 0 = q^{-1}(l')l = q^{-1}(l \cdot l') = q^{-1}(l)l' \iff l' \in (q^{-1}(l))^\perp, \]

so \(P(q(l^\perp))\) is the polar hyperplane \(P(q^{-1}(l^\perp))\) to \([l]\) with respect to \(Q^{-1}\). Thus the subscheme \(\Gamma'\) residual to \([l]\) in \(\Gamma\) spans the polar hyperplane to \([l]\) with respect to \(Q^{-1}\).

Lemma 2.3. A component of an apolar subscheme has support on \(Q^{-1}\) if and only if this component is nonreduced.

Proof. If a component is a reduced point, the residual is contained in the polar hyperplane to this point, so by nondegeneracy the polar hyperplane cannot contain the point. If a component is nonreduced, the residual to the point supporting the component lies in the polar hyperplane to this point, so the point is on \(Q^{-1}\). □
Each component $\Gamma_0$ of an apolar subscheme to $q$ is apolar to a quadratic form $q_0$ defined on the span of $\Gamma_0$ and uniquely determined as a summand $q$. This is the content of the next proposition.

**Proposition 2.4.** Let $\Gamma = \Gamma_1 \cup \Gamma_2$ be an apolar subscheme of length $n$ to $q$ that decomposes into two disjoint subschemes $\Gamma_1$ and $\Gamma_2$ of length $n_1$ and $n_2$. Let $U_1 \subset T_1$ and $U_2 \subset T_1$ be subspaces such that $\Gamma_1$ spans $P(U_1)$. Then there is a unique decomposition $q = q_1 + q_2$ with $q_i \in (U_i)^\perp$.

Furthermore, subschemes $\Gamma_1 \subset P(U_1)$ and $\Gamma_2 \subset P(U_2)$ of length $n_1$ and $n_2$ are apolar to $q_1$ and $q_2$ respectively, if and only if $\Gamma_1 \cap \Gamma_2$ is apolar to $q$. 

**Proof.** Since $\Gamma$ is nondegenerate, $T_1 = U_1 \oplus U_2$. Let $U_1^\perp \subset S_1$ be the space of forms vanishing on $U_1$ via apolarity. Then $U_1^\perp$ are natural coordinates on $P(U_2)$ and likewise, $U_2^\perp$ are natural coordinates on $P(U_1)$. Let $I_1 \subset (U_2^\perp)^2$ be the quadratic forms generating the ideal of $\Gamma_1$ in $P(U_1)$, and likewise $I_2$ the quadratic forms generating the ideal of $\Gamma_2$ in $P(U_2)$. Then $I_1 \oplus I_2 \oplus (U_1^\perp) \cdot (U_2^\perp) \subset S_2$ is the space of quadratic forms in the ideal of $\Gamma$.

Consider the intersections, $q_2^\perp = q^\perp \cap (U_1^\perp)^2$ and $q_1^\perp = q^\perp \cap (U_2^\perp)^2$. Since $q$ is non degenerate, $q^\perp$ does not contain either of the subspaces $(U_1^\perp)^2$. Therefore $q_2^\perp$ is a codimension one subspace in $(U_1^\perp)^2$ and is apolar to a quadratic form $q_2 \in (U_2^\perp)^2$, unique up to scalar. Similarly, $q_1^\perp$ is apolar to a unique quadratic from $q_1 \in (U_1^\perp)^2$. The space of quadratic forms $q_1^\perp \oplus q_2^\perp \oplus (U_1^\perp) \cdot (U_2^\perp)$ is contained in $q^\perp$ and is apolar to the subspace $S(q_1,q_2) \subset T_2$. Therefore, there are unique nonzero coefficients $c_1$ and $c_2$ such that $q = c_1 q_1 + c_2 q_2$. Furthermore, each $\Gamma_1$ is apolar to $q_i$, $i = 1, 2$.

It remains only to show the last statement. Assume $\Gamma_1$ and $\Gamma_2$ are apolar to $q_1$ and $q_2$ respectively. Then $\Gamma_1 \cup \Gamma_2$ is non degenerate of length $n$. Let $I_1 \subset (U_2^\perp)^2$ be the generators of the ideal of $\Gamma_1$ and $I_2 \subset (U_1^\perp)^2$ be the generators of the ideal of $\Gamma_2$. Then the quadratic forms in 

$$I_1 \oplus I_2 \oplus (U_1^\perp) \cdot (U_2^\perp)$$

all lie in the ideal of $\Gamma_1 \cup \Gamma_2$. The dimension of this space of quadratic forms is 

$$\binom{n_1}{2} + \binom{n_2}{2} + n_1 \cdot n_2 = \binom{n}{2},$$

so they generate the ideal of $\Gamma_1 \cup \Gamma_2$. Since all these forms are apolar to $q = q_1 + q_2$, the subscheme $\Gamma_1 \cup \Gamma_2$ is apolar to $q$. \qed

**Remark 2.5.** By Proposition 2.4, the orbits of $SO(n,q)$ in $\text{VASP}(Q,n)$ are characterized by their components.

We shall return to the set of local apolar subschemes $V_p^{loc}(n)$ supported at a point $p \in Q^{-1}$ in section 5. Here we show that apolar subschemes of length $n$ to $q$ are all locally Gorenstein.

**Lemma 2.6.** Let $B$ be a local Artinian $C = B/m_B$-algebra of length $n$ and $\Phi : \text{Spec}B \rightarrow A^{n-1} \subset P^{n-1}$ the reembedding given by $C$-basis of $m_B$. The subscheme $\text{Im} \Phi$ is apolar to a full rank quadric if and only if $B$ is Gorenstein.
Proof. Let $\phi : A = \mathbb{C}[x_1, \ldots, x_{n-1}] \to B$ be the ring homomorphism corresponding to $\Phi$. Thus $\phi$ is defined by a linear $k$-isomorphism $\phi_1 = A_{\leq 1} = (1, x_1, \ldots, x_{n-1}) \to B$. Let $\pi : B \to (0 : m_B)$ be the projection onto the socle of $B$, let $\psi : (0 : m_B) \to \mathbb{C}$ be a linear form and consider the bilinear form

$$A \times A \overset{\phi \cdot \phi}{\longrightarrow} B \overset{\pi}{\longrightarrow} (0 : m_B) \overset{\psi}{\longrightarrow} \mathbb{C},$$

where the first map is the composition of $\phi$ with multiplication. This map extends to the tensor product $A \otimes A$, and the restriction then to the symmetric part $(A_{\leq 1})^2 \subset A_{\leq 1} \otimes A_{\leq 1}$ defines a linear form

$$\beta_\psi : (A_{\leq 1})^2 \to \mathbb{C}$$

and an associated quadratic form

$$q_\psi : A_{\leq 1} \to \mathbb{C}.$$

Clearly the kernel of $\beta_\psi$ generate an ideal in $A$ that is apolar to $q_\psi$. On the other hand, $B$ is Gorenstein if and only if the socle is 1-dimensional. So for the lemma, it suffices to prove that $q_\psi$ is non degenerate, i.e. has rank $n$, if and only if the linear form $\psi$ is an isomorphism.

But $q_\psi$ is degenerate if and only if the kernel of $\beta_\psi$ contains $x \cdot A_{\leq 1}$ for some nonzero element $x \in A_{\leq 1}$. Now, $\beta_\psi(x \cdot A_{\leq 1}) = 0$ if and only if $\phi(x) \cdot B \cap (0 : m_B) \subset \ker \psi$. Since $B$ is Artinian, $\phi(x) \cdot B \cap (0 : m_B)$ is a nonzero subspace of $(0 : m_B)$, so it suffices to consider elements $x$, which map to the socle. But then the kernel of $\beta_\psi$ contains $x \cdot A_{\leq 1}$ precisely when $x$ is in the kernel of $\psi$ and the lemma follows. \qed

Corollary 2.7. $VAPS(Q, n)$ is reducible for $n \geq 24$

Proof. Consider a general graded Artinian Gorenstein algebra $B$ of embedding dimension $e$ and socle in degree 3. The length of $B$ is $2e + 2$. By the Macaulay correspondence [Mac16], such algebras are in bijection with homogeneous forms, up to scalars, of degree 3 in $e$ variables, hence depends on $\binom{e-1+3}{3} = (e + 2)(e + 1)e/6 - 1$ variables. The family of smoothable algebras have dimension at most $e(2e + 2) - 1$ So for $e + 2 > 2 \cdot 6$ a general algebra $B$ cannot be smoothable, for trivial reason. In particular, $e = 11$ hence $n = 24$ is enough. \qed

We do not believe the bound $n \geq 24$ is sharp.

3. A rational parameterization

In this section we show that through a general point in $\mathbb{P}(T_{2,q})$ there is a unique $n$-secant $(n - 2)$-space to the projected Veronese variety $V_{2,q}$. Furthermore, we give a characterization of the points for which there are more than one, i.e. infinitely many $n$-secant $(n - 2)$-spaces to $V_{2,q}$.

If we choose basis $a$ for $T_1$ such that the symmetric matrix associated to $q$ is the identity matrix, then the eigenvectors of the symmetric matrix associated
to a general quadric $q'$ are distinct. Thus, the symmetric matrices associated to $q$ and $q'$ have a unique set of $n$ common 1-dimensional eigenspaces. Hodge & Pedoe [HP52, XIII.8,Theorem II] and Gantmacher [Gan59, Chapter XII, Theorem 3] found canonical forms for any pair of quadratic forms $q, q'$ as soon as one of them is nonsingular. Here we are concerned with the simplex formed by the set of common eigenspaces and give a geometric formulation and proof.

**Proposition 3.1.** Let $q, q' \in T_2$ be two general quadrics. Then there exists a unique $n$-simplex $\{L_1, \ldots, L_n\}$ polar to both $q$ and $q'$.

**Proof.** By the above, it suffices to show the relation between the collection of common eigenspaces of the associated symmetric matrices and the common simplex. So we assume that $q, q'$ are quadrics of rank $n$ and that

$$q = \sum_{i=1}^{n} l_i^2, \quad q' = \sum_{i=1}^{n} \lambda_i l_i^2,$$

where the $\lambda_i$ are pairwise distinct coefficients and $L_i = \{l_i = 0\}$, $i = 1, \ldots, n$. Let

$$q_i = \lambda_i q - q', \quad i = 1, \ldots, n.$$

Then the $q_i$ are precisely the quadratic forms of the pencil generated by $q$ and $q'$ that have rank less than $n$. The rank of $q_i$ is exactly $n - 1$ since $\lambda_i \neq \lambda_j$ for $i \neq j$, so $q_i \in (U_i)^2$ for a unique rank $n - 1$ subspace $U_i \subset T_1$. The intersection $\cap_{i \neq j} U_i$ is the 1-dimensional subspace generated by the nonzero linear form $l_j$. These forms are therefore determined uniquely by the pencil generated by $q$ and $q'$. □

A precise condition for generality in the proposition is given by rank:

**Lemma 3.2.** A pencil of quadratic forms in $n$ variables have a unique common apolar subscheme of length $n$ if and only if every quadric in the pencil have rank at least $n - 1$ and some, hence the general quadric has rank $n$. Furthermore the unique apolar subscheme is curvilinear, i.e. embeddable in a smooth curve.

**Proof.** Let

$$(q' + \lambda q)_{\lambda \in \mathbb{A}_k^1}$$

be a pencil with discriminant $\Delta \subset \mathbb{A}_k^1$, a scheme of length $n$. Consider the incidence

$$\{(D, \lambda)|D(q' + \lambda q) = 0\} \subset \mathbb{P}(T_1) \times \mathbb{A}_k^1$$

with projections $p_T$ and $p_C$. Clearly the fibers of each projection are all linear. Now as in the proof of the proposition, a general length $n$ subscheme of $p_T(\mathbb{P}^1_c(\Delta))$ is a common apolar subscheme to the pencil of quadratic forms. Therefore, the common apolar subscheme is unique if and only if $p_T(\mathbb{P}^1_c(\Delta))$ is finite, i.e. the corank of any quadric in $L$ is at most 1. In this case both projections restricted to the incidence are isomorphisms onto their images. In particular the apolar subscheme is isomorphic to $\Delta$, so it is curvilinear. □
Remark 3.3. The ideal of the curvilinear image $\Gamma$ of the map

$$\text{Spec } \mathbb{C}[t]/(t^n) \to \mathbb{P}^{n-1}$$

$$t \mapsto (1 : t : t^2 : \ldots : t^{n-1}),$$

is generated by the $2 \times 2$ minors of

$$\begin{pmatrix} x_1 & x_2 & \ldots & x_{n-1} & x_n \\ x_2 & x_3 & \ldots & x_n & 0 \end{pmatrix},$$

so the $\Gamma$ is apolar to the maximal rank quadric

$$\sum_{k=1}^{n} y_k y_{n+1-k}.$$  

This remark generalizes to a partial converse of Lemma 3.2.

Lemma 3.4. Any curvilinear nondegenerate zero-dimensional subscheme $\Gamma \subset \mathbb{P}^{n-1}$ of length $n$ is apolar to a quadric $Q \subset \mathbb{P}^{n-1}$ of maximal rank.

Proof. Let $\Gamma$ be a nondegenerate curvilinear subscheme with $r$ components of length $n_1, \ldots, n_r$ such that $n_1 + \ldots + n_r = n$. Then $\Gamma$ is projectively equivalent to $\Gamma' = \Gamma_1 \cup \ldots \cup \Gamma_r$, where $\Gamma_i$ is the image of

$$\text{Spec } \mathbb{C}[t]/(t^{n_i}) \to \mathbb{P}(\mathbb{C}^{n_1} \oplus \ldots \oplus \mathbb{C}^{n_r})$$

$$t \mapsto (1 : t : t^2 : \ldots : t^{n_i-1}),$$

where the nonzero coordinates in the image are $x_{i,1}, \ldots, x_{i,n_i}$. The ideal of $\Gamma'$ is generated by the $2 \times 2$ minors of the $r$ matrices

$$\begin{pmatrix} x_{(1,1)} & \ldots & x_{(1,n_i-1)} & x_{(1,n_i)} \\ x_{(1,2)} & \ldots & x_{(1,n_i)} & 0 \end{pmatrix} \ldots \begin{pmatrix} x_{(r,1)} & \ldots & x_{(r,n_i-1)} & x_{(r,n_i)} \\ x_{(r,2)} & \ldots & x_{(r,n_i)} & 0 \end{pmatrix},$$

and the products

$$x_{(i,k)}x_{(j,k)} \quad \text{for} \quad 1 \leq i < j \leq r, 1 \leq k_i \leq n_i, 1 \leq k_j \leq n_j.$$  

So $\Gamma'$ is apolar to the maximal rank quadric

$$\sum_{i=1}^{r} \sum_{k=1}^{n_i} y_{(i,k)} y_{(i,n_i+1-k)}.$$  

□

More important to us will be that rank $n$ quadrics have apolar subschemes of length $n$ that are not curvilinear (when $n > 3$).

Remark 3.5. Consider the rank $n$ quadric

$$q = 2y_1y_n + y_2^2 + \ldots + y_{n-1}^2.$$  

The subscheme $\Gamma_q \subset \mathbb{P}^{n-1}$ defined by

$$(x_1^2, x_1x_2, x_2^2 - x_1x_n, x_1x_3, \ldots, x_{n-1}^2 - x_1x_n),$$

has degree $n$ and is apolar to $q$, but it is clearly not curvilinear when $n > 3$. It contains the tangency locus of the quadric $q^{-1} = \frac{1}{2}x_1x_n + \frac{1}{4}x_2^2 + \ldots + \frac{1}{n}x_{n-1}^2$.  

Documenta Mathematica 18 (2013) 469–505
0) at the point $[0 : 0 : \ldots : 1]$. The tangency locus has length $n - 1$ and is defined by
\[(x_1, (x_2^2 - x_1x_n, x_2x_3, \ldots, x_{n-1}^2 - x_1x_n)).\]
The subscheme $\Gamma_p$ is itself not contained in the tangent hyperplane $\{x_1 = 0\}$, but it is the unique apolar subscheme to $q$ that contains the first order neighborhood of $[0 : 0 : \ldots : 1]$ on $\{q^{-1} = 0\}$. It will be the focus of our attention in Section 5.

It follows immediately from Proposition 3.1 that there is a rational and dominant map
\[\gamma : P(T_{2,q}) \to VPS(Q,n) \subset G(n-1,T_{2,q})\]
whose general fiber is a $n$-secant $(n-2)$-space to the projected Veronese variety $V_{2,q}$. In the next section we find equations for this map.

4. The Mukai form

Mukai introduced in [Muk92] a trilinear form in his approach to varieties of sums of powers of conics in particular, and to forms of even degree in general (see also [Dol12, Sections 1.4 and 2.1.3]). In this section we show how this form naturally gives equations for the map $\gamma$ and for the universal family of polar simplices. The main result of this section, Proposition 4.2, gives the equations for the common apolar subscheme of length $n$ of a pencil of quadrics in $n$ variables, whenever this subscheme is unique, cf. Lemma 3.2.

Both the quadratic form $q \in T_2$ and the inverse $q^{-1} \in S_2$ play a crucial role in the definition of the Mukai form. Recall that the form $q$ defines an invertible linear map $q : S_1 \to T_1$, and $q^{-1}$ defines the inverse map: $q^{-1} : T_1 \to S_1$. In coordinates, if $q = (a_1 y_1^2 + \ldots + a_n y_n^2)$, then $q^{-1} = (\frac{1}{4a_1} x_1^2 + \ldots + \frac{1}{4a_n} x_n^2)$.

We will arrive at Mukai’s form from
\[\tau \in \text{Hom}(\wedge^2 S_1 \otimes T_2 \otimes T_2 \otimes S_2, C)\]
defined by
\[z_1 \wedge z_2 \otimes q_1 \otimes q_2 \otimes \alpha \mapsto (z_1(q_1))z_2(q_2) - z_2(q_1)z_1(q_2))(\alpha)\]
where $f(g)$, as above, means $f$ viewed as differential operator applied to $g$. Interpreting $\omega = z_1 \wedge z_2 \in \wedge^2 S_1 \subset \text{Hom}(T_1, S_1)$, $q_j \in T_2 \subset \text{Hom}(S_1, T_1)$ and $\alpha \in S_2 \subset \text{Hom}(T_1, S_1)$ the expression
\[(\omega \otimes q_1 \otimes q_2 \otimes \alpha) \mapsto \frac{1}{2} \text{trace} \ (\alpha \circ q_2 \circ \omega \circ q_1 - \alpha \circ q_1 \circ \omega \circ q_2)\]
\[= \frac{1}{2} \text{trace} \ (\omega \circ q_1 \circ \alpha \circ q_2 - \alpha \circ q_1 \circ \omega \circ q_2)\]
gives an alternative description of $\tau$. In fact $f(g) = \frac{1}{2} \text{(trace } f \circ g \text{)}$ holds for $f \otimes g \in S_2 \otimes T_2 \subset \text{Hom}(T_1, S_1) \otimes \text{Hom}(S_1, T_1)$ and $\text{trace} ((\alpha \circ q_2) \circ (\omega \circ q_1)) = \text{trace} ((\omega \circ q_1) \circ (\alpha \circ q_2))$. We now substitute $\alpha = q^{-1}$. Then $\frac{1}{2} \text{trace} \ (\omega \circ q_1 \circ q^{-1} \circ q_2 - q^{-1} \circ q_1 \circ \omega \circ q_2) = 0$ for $q_1 = q$, and, since the first expression for $\tau$ is
alternating on $T_2 \otimes T_2$, we have $\frac{1}{2} \text{trace } (\omega \circ q_1 \circ q^{-1} \circ q_2 - q^{-1} \circ q_1 \circ \omega \circ q_2) = 0$ for $q_2 = q$ as well. Thus $\tau$ induces a well defined trilinear form

$$
\tau_q \in \text{Hom}(\wedge^2 S_1 \otimes T_{2,q} \otimes T_{2,q}, \mathbb{C})
$$
on the quotient space $T_{2,q} = T_2 / \langle q \rangle$.

Since $T_{2,q}^* = q^* \subset S_2 = T_2^*$ and

$$
\text{Hom}(\wedge^2 S_1 \otimes T_{2,q} \otimes T_{2,q}, \mathbb{C}) \cong \text{Hom}(T_{2,q}, \text{Hom}(\wedge^2 S_1, q^*))
$$
we have a second interpretation of $\tau_q$. With this interpretation, the image of $\tau_q(q_1) \in \text{Hom}(\wedge^2 S_1, q^*) \subset \text{Hom}(\wedge^2 S_1, S_2)$ is defined by

$$
\omega \mapsto [\omega \circ q_1 \circ q^{-1} - q^{-1} \circ q_1 \circ \omega] \in q^* \subset S_2 \subset \text{Hom}(T_1, S_1).
$$
The form $\tau$ is alternating on $T_2 \otimes T_2$, so $\tau(\omega, q', q'^{-1}) = 0$ for every $\omega \in \wedge^2 S_1$. Therefore

$$
\tau_q(q')(\wedge^2 S_1) \subset (q')^\perp.
$$
If $Q'$ is the quadric $\{q' = 0\} \subset \mathbb{P}(S_1)$, we may therefore conclude:

**Lemma 4.1.** Any quadratic form in $\tau_q(q')(\wedge^2 S_1)$ is apolar to both $Q$ and $Q'$:

$$
\tau_q(q')(\wedge^2 S_1) \subset q^\perp \cap (q')^\perp.
$$

□

Notice that the linear space of quadratic forms $\tau_q(q')(\wedge^2 S_1)$ is not all of $q^\perp \cap (q')^\perp$. It is a special subspace of the intersection. Since $\tau_q(q) = 0$, we have $\tau_q(q') = \tau_q(q' + \lambda q)$ for any $\lambda$, so the space $\tau_q(q')(\wedge^2 S_1)$ of quadratic forms depends only on the pencil $\langle q, q' \rangle$.

If the pencil of quadratic forms $\langle q, q' \rangle \subset T_2$ contains no forms of corank at least 2, then, by Lemma 3.2, there is a unique common apolar subscheme $\Gamma_{q'}$ of length $n$ to $q$ and $q'$. The significance of the form $\tau_q$ is

**Proposition 4.2.** Let $q' \in T_{2,q}$. Then the linear map

$$
\tau_q(q') : \wedge^2 S \rightarrow q^\perp
$$
is injective if and only if $q$ and $q'$ have a unique common apolar subscheme of length $n$. Furthermore, in this case the image generates the ideal in $S$ of this subscheme.

**Proof.** Our argument depends on several lemmas, in which we study $\text{Im } \tau_q(q') \subset S_2$ by considering the symmetric matrices associated to these quadratic forms with respect to a suitable basis. Thus, we choose coordinates such that $q = \frac{1}{2}(y_1^2 + y_2^2 + \ldots + y_n^2)$ and hence $q^{-1} = \frac{1}{2}(x_1^2 + x_2^2 + \ldots + x_n^2)$. The symmetric matrices of these quadratic forms with respect to the coordinate basis of $T_1$ and $S_1$ are both the identity matrix. We denote by $A$ the symmetric matrix of $q'$, i.e. $q' = \frac{1}{2}(y_1, \ldots, y_n)A(y_1, \ldots, y_n)^t$. For a form $\omega \in \wedge^2 S_1$ there is similarly an associated skew symmetric matrix $\Lambda_\omega$. For a form $l \in T_1$ we denote by $v_l$ the column vector of its coordinates. The quadratic forms in the image $\tau_q(q')$ are the forms associated to the symmetric bilinear forms

$$
\{\omega \circ q' \circ q^{-1} - q^{-1} \circ q' \circ \omega | \omega \in \wedge^2 S_1\}.
$$

**Documenta Mathematica 18 (2013) 469–505**
so their associated symmetric matrices are

\[ \{ \lambda \omega A - A \lambda \omega | \omega \in \wedge^2 S_1 \} \]

**Lemma 4.3.** Let \([l] \in P(T_1)\), then every quadric in \(\tau_q(q')(\wedge^2 S) \subset q^+\) vanishes at the point \([l]\) if and only if there is a quadric \(q_\lambda = q' + \lambda q\) for some \(\lambda \in \mathbb{C}\), such that \(l\) lies in the kernel of the linear transformation \(q_\lambda \circ q^{-1} : T_1 \to T_1\).

Equivalently, in terms of matrices: If \(v_l\) is the column coordinate vector of \(l\), then \(v_l^t (\lambda \omega A - A \lambda \omega) v_l = 0\) for every \(\omega \in \wedge^2 S_1\) if and only if \(v_l\) is an eigenvector for the matrix \(A\).

**Proof.** Note first that the matrix of the linear transformation \(q_\lambda \circ q^{-1}\), with respect to the coordinate basis of \(T_1\), is simply \(A + \lambda I\). Hence, the two parts of the lemma are equivalent.

In the matrix notation, if \(v_l\) is an eigenvector for \(A\) with eigenvalue \(\lambda\), then

\[
v_l^t (\lambda \omega A - A \lambda \omega) v_l = v_l^t \lambda \omega A v_l - v_l^t A \lambda \omega v_l = v_l^t \lambda \omega v_l - \lambda v_l^t \lambda \omega v_l = 0,
\]

so the if part follows.

Conversely, assume that

\[
v_l^t (\lambda \omega A - A \lambda \omega) v_l = 0
\]

for every skew symmetric \(n \times n\) matrix \(\lambda \omega\). Again

\[
v_l^t (\lambda \omega A - A \lambda \omega) v_l = v_l^t \lambda \omega A v_l - v_l^t \lambda \omega A v_l,
\]

and since \(A\) is symmetric and \(\lambda \omega\) is skewsymmetric, \((v_l^t \lambda \omega A v_l)^t = -v_l^t A \lambda \omega v_l\), so we deduce that

\[
v_l^t \lambda \omega A v_l = 0.
\]

But \(v_l^t \lambda \omega u = 0\) for every skew symmetric matrix \(\lambda \omega\) only if \(u\) is proportional to \(v_l\), so we conclude that \(A(v_l) = \lambda v_l\) for some \(\lambda\).

**Remark 4.4.** A point \(l \in T_1\) lies in the kernel of \(q_\lambda \circ q^{-1}\) if and only if \(q^{-1}(l)\) lies in the kernel of \(q_\lambda : S_1 \to T_1\). Equivalently, \(\{q_\lambda = 0\} \subset P(S_1)\) is a singular quadric and \([q^{-1}(l)] \in P(S_1)\) lies in its singular locus.

**Corollary 4.5.** \(\tau_q(q')\) is injective only if \(\langle q, q' \rangle\) contains no quadratic form of rank less than \(n - 1\).

**Proof.** If the quadratic form \(q_\lambda = q' + \lambda q\) has rank less than \(n - 1\), then there are independent forms \(l, l' \in T_1\) such that \((q^{-1}(l), q^{-1}(l'))\) is contained in the kernel of \(q_\lambda : S_1 \to T_1\). In particular, viewed as differential operators applied to \(q_\lambda\),

\[
q^{-1}(l)(q_\lambda) = q^{-1}(l')(q_\lambda) = 0 \in T_1.
\]

Let

\[
\omega = q^{-1}(l) \wedge q^{-1}(l') \in \wedge^2 S_1 \subset \text{Hom}(T_1, S_1).
\]

Then

\[
\omega \otimes q_\lambda \otimes q_2 \otimes q^{-1} \to [q^{-1}(l)(q_\lambda) \cdot q^{-1}(l')(q_2) - q^{-1}(l')(q_\lambda) \cdot q^{-1}(l)(q_2)](q^{-1}) = 0
\]

for every \(q_2 \in T_2\), so \(\tau_q(q')(\omega) = 0\) and \(\tau_q(q')\) is not injective. \(\square\)
To complete the proof of Proposition 4.2, we assume that $q$ and $q'$ have a unique common apolar subscheme $\Gamma$ of length $n$, i.e. by Lemma 3.2, no quadratic form in $\langle q, q' \rangle$ has rank less than $n - 1$. We want to show that $\tau_q(q')$ is injective and that the image generates the ideal of $\Gamma$.

Let $\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_r$ be a decomposition of $\Gamma$ into its connected components. Then each $\Gamma_i$ is a finite local curvilinear scheme. Let $n_i$ be the length of $\Gamma_i$. By Proposition 2.4, there is a decomposition $T_1 = \oplus_i U_i$ such that each $\Gamma_i \subset P(U_i)$. Furthermore $U_i$ has dimension $n_i$ and $q$ and $q'$ have unique decompositions $q = q_1 + \ldots + q_r$ and $q' = q'_1 + \ldots + q'_r$ with $q_i, q'_i \in (U_i)^2 \subset T_2$. Denote by $U'_i = \oplus_{j \neq i} U_j$, and let $(U'_i)' \subset$ be the orthogonal subspace of linear forms in $S_1$. Then

$$\sum_i U_i \cdot (U'_i)' \subset S_2$$

generate the ideal of $U_i P(U_i) \subset P(T_1)$.

The linear forms $q^{-1}(U_i) \subset S_1$ are natural coordinates on $P(U_i)$. Denote by $I_{\Gamma_i, 2}$ the quadratic forms in these coordinates in the ideal of $\Gamma_i$. Then $I_{\Gamma_i, 2} \subset (q^{-1}(U_i))^2 \subset S_2$ and the space of quadratic forms in the ideal of $\Gamma$ is

$$I_{\Gamma, 2} = \sum I_{\Gamma_i, 2} + \sum_i U_i \cdot (U'_i)' \subset S_2.$$ 

We have

**Claim 4.6.**

$$\text{Im } \tau_q(q') \supset \sum_i I_{\Gamma_i, 2} + \sum_i U_i \cdot (U'_i)'.$$ 

If the claim holds, $\tau_q(q')$ is injective, since $\dim \wedge^2 S_1 = \dim I_{\Gamma, 2} = \binom{n}{2}$, so the equality $\text{Im } \tau_q(q') = I_{\Gamma, 2}$ holds and the proof of Proposition 4.2 is complete.

We use matrices to prove the claim. To interpret the decomposition of $q$ and $q'$ in terms of matrices, we choose a basis for each $U_i$ such that the symmetric matrix associated to each $q_i$ is the $n_i \times n_i$ identity matrix. Let $A_i$ be the symmetric $n_i \times n_i$ matrix associated to $q'_i$. The union of the bases for the $U_i$ form a basis for $T_1$ with respect to which the symmetric matrix $A$ of $q'$ has $r$ diagonal blocks $A_i$ and zeros elsewhere.

The matrices $A_i$ each have a unique eigenvalue $\lambda_i$, and these eigenvalues are pairwise distinct. Furthermore, each $A_i$ has a 1-dimensional eigenspace, so their Jordan form has a unique Jordan block, and we may write $A_i = \lambda_i I_{n_i} + B_i$ with $B_i$ a nilpotent symmetric matrix. (See ([DZ04, Theorem 2.3]) for a nice normal form for the matrices $B_i$.)

By extending each $A_i$ with zeros to $n \times n$ matrices, we may write $A = \sum A_i$. The decomposition $T_1 = \oplus_i U_i$ is then defined by $U_i = \ker(\lambda_i I - A)^{n_i} \subset T_1$.

Denote by $U'_i = \oplus_{j \neq i} U_j$. Then $P(U_i)$ and $P(U'_i)$ have complementary dimension in $P(T_1)$. We shall use the techniques applied by Gantmacher in the analysis of commuting matrices ([Gan59, Chapter VIII]) to show

**Lemma 4.7.** Let $A$ be the symmetric matrix of the quadratic form $q' \in T_2, q$ as above. Let $T_1 = U_i \oplus U'_i$ be the decomposition associated to the eigenvalue $\lambda_i$. 

\[
\text{Im } \tau_q(q') \supset \sum_i I_{\Gamma_i, 2} + \sum_i U_i \cdot (U'_i)'.
\]
Set $A$ a symmetric matrix and let $\Lambda_{\lambda, \mu}$ an $d, j > d$.

Then \(U^\perp \cdot (U')^\perp \subset \text{Im} \tau_q(q') \subset S_2\).

**Proof.** Set $d = n_i$ and $\lambda = \lambda_i$ and choose coordinates such that $U_\lambda = \langle y_1, \ldots, y_d \rangle$ and $U'_\lambda = \langle y_{d+1}, \ldots, y_n \rangle$. Then $(U')^\perp = \langle x_1, \ldots, x_d \rangle$ and $(U^\perp)^\perp = \langle x_{d+1}, \ldots, x_n \rangle$.

Consider the matrix $B$ of the quadratic form $\tau_q(q')(x_i \wedge x_j)$ with $i \leq d$ and $j > d$.

The skew symmetric matrix $\Lambda_{ij}$ of $x_i \wedge x_j$ has $(ij)$-th entry 1, consequently $(ji)$-th entry $-1$, and 0 elsewhere, and

\[ B = \Lambda_{ij} A - AA_{ij} . \]

The nonzero entries in $\Lambda_{ij} A$ are in positions $(i, k)$ with $k > d$ and $(j, k)$ with $k \leq d$, while the nonzero entries in $A \Lambda_{ij}$ are in positions $(k, i)$ with $k > d$ and $(k, j)$ with $k \leq d$. Therefore the quadratic form $\tau_q(q')(x_i \wedge x_j)$ lies in the space $\langle x_a x_b \rangle_{a \leq d < b} = (U')^\perp \cdot (U^\perp)^\perp$.

A linear relation between these quadratic forms would correspond to a skew symmetric matrix $A$ with nonzero entries only in the rectangular block $(ij)$, $i \leq d, j > d$, such that $AA - AA = 0$. Write $A$ as a sum $A = A_\lambda + A_{\mu_1} + \ldots + A_{\mu_n}$ where the $\mu_i$ are the eigenvalues of $A$ distinct from $\lambda$. Let $\Lambda$ be a skew symmetric matrix and let $\Lambda_{\lambda, \mu_i}$ be the rectangular submatrix with rows equal to the nonzero rows of $A_\lambda$ and columns equal to the nonzero columns of $A_{\mu_i}$.

Then the corresponding submatrix

\[(\Lambda A - AA)_{\lambda, \mu_i} = \Lambda_{\lambda, \mu_i} A_{\mu_i} - A_\lambda \Lambda_{\lambda, \mu_i} . \]

So $\Lambda A - AA = 0$ only if $\Lambda_{\lambda, \mu_i} A_{\mu_i} - A_\lambda \Lambda_{\lambda, \mu_i} = 0$ for each $\mu_i$.

Let $\mu$ be one of the $\mu_i$, for simplicity $U_\mu = \langle y_{d+1}, \ldots, y_{d+e} \rangle$. Let $I_d$ be the diagonal matrix with 1 in the $d$ first entries and 0 elsewhere, and let $I_e$ be the diagonal matrix with 1 in the entries $d + 1, \ldots, d + e$ and 0 elsewhere. Then the special summand $A_\lambda$ of $A$ can be written as a sum $A_\lambda = M_d + B_d$ where $B_d$ is nilpotent of order $d$. Likewise, $A_{\mu} = \mu I_e + B_e$ where $B_e$ is nilpotent of order $e$. So we may write $A = M_d + B_d + \mu I_e + B_e + A'$, where $A' = A - A_\lambda - A_{\mu}$.

But then $(\Lambda A - AA)_{\lambda, \mu} = 0$ only if

\[ A_{\lambda, \mu} A_{\mu} - A_\lambda A_{\lambda, \mu} = 0 , \]

i.e. when

\[ A_{\lambda, \mu}(\mu I_e + B_e) - (\lambda I_d + B_d) A_{\lambda, \mu} = 0 . \]

This is equivalent to

\[(\lambda - \mu) A_{\lambda, \mu} = A_{\lambda, \mu} B_e - B_d A_{\lambda, \mu} . \]

Multiplying both sides by $(\lambda - \mu)$ and substituting on the right hand side $(\lambda - \mu) A_{\lambda, \mu}$ with $A_{\lambda, \mu} B_e - B_d A_{\lambda, \mu}$ we get

\[ (\lambda - \mu)^2 A_{\lambda, \mu} = (A_{\lambda, \mu} B_e - B_d A_{\lambda, \mu}) B_e - B_d (A_{\lambda, \mu} B_e - B_d A_{\lambda, \mu}) \]

\[ = (A_{\lambda, \mu} B_e)^2 - 2 B_d A_{\lambda, \mu} B_e + (B_d)^2 A_{\lambda, \mu} . \]

Iterating $m = d + e - 1$ times we get
Let $\Lambda$ be a skew symmetric matrix. Hence $\Lambda$ has rank $\leq d, j > d$ are linearly independent. The corresponding quadratic forms are linearly independent in the space $\langle x_1, \ldots, x_d \rangle \times \langle x_{d+1}, \ldots, x_n \rangle$. Since the dimensions coincide, the quadratic forms span this space, and the lemma follows. 

Next, we consider the case when the symmetric matrix $A$ has only one eigenvalue and up to scalars only one nonzero eigenvector. Hence $q$ has rank $n - 1$ i.e. that the eigenvalue is 0. Then $A$ is nilpotent, and since $A$ is a one-dimensional eigenvector space, $A^n = 0$ and $A^i \neq 0$ for any $i < n$.

**Lemma 4.8.** Let $q' \in T_{2,q}$ be a quadratic form whose associated $n \times n$ matrix $A$ is symmetric, nilpotent and has rank $n - 1$. Then the ideal generated by the quadratic forms $\tau_q(q') \subset q^c$ is the ideal of the unique common apolar subscheme $\Gamma$ of length $n$ of $q$ and $q'$. Moreover $\Gamma$ is a local curve linear subscheme.

**Proof.** Let $A$ be a skew symmetric $n \times n$ matrix and think of $A$ and $\Lambda$ as the matrices of linear endomorphisms of a $n$-dimensional vector space $V$. Then we may choose a basis $v_1, \ldots, v_n \in V$ such that $Av_1 = 0$ and $Av_i = v_{i-1}$ for $i = 2, \ldots, n$. Let $\rho : \text{Spec}(\mathbb{C}[t]/t^n) \to \mathbf{P}(V) : t \mapsto [v_1 + tv_2 + \ldots + t^{n-1}v_n]$ and set $\Gamma = \text{Im} \rho$. Then $I_{\Gamma}$ is generated by $\binom{n}{2}$ quadratic forms. We shall show that the symmetric matrices of these forms coincide with the matrices $\Lambda A - A$ as $\Lambda$ varies. We evaluate the quadratic form associated to $\Lambda A - A$ on the vector $v = v_1 + tv_2 + \ldots + t^{n-1}v_n$:

$$v^t(\Lambda A - A)v = v^t\Lambda A v - v^t A v.$$ 

But

$$v^t\Lambda A v = (v_1 + tv_2 + \ldots + t^{n-1}v_n)^t \Lambda (v_1 + tv_2 + \ldots + t^{n-1}v_n)$$

$$= (v_1 + tv_2 + \ldots + t^{n-1}v_n)^t \Lambda (v_1 + \ldots + t^{n-1}v_{n-1})$$

$$= (v_1 + \ldots + t^{n-2}v_{n-1})^t \Lambda (v_1 + \ldots + t^{n-2}v_{n-1}) t$$

$$+ t^n v_n^t \Lambda (v_1 + \ldots + t^{n-2}v_{n-1})$$

$$= 0$$

since $\Lambda$ is skew symmetric and $t^n = 0$. Therefore the quadratic forms with matrices $\Lambda A - A$ are in the ideal of $\Gamma$. They are independent and therefore
generate the ideal unless AA - ΛΛ = 0 for some nontrivial Λ. But then Λ and A commute, hence have common eigenvectors. Λ is nontrivial and skew symmetric so it has at least 2 independent eigenvectors, while A has only one, so this is impossible. Clearly, Γ is curvilinear, and any non degenerate local curvilinear subscheme of length n in P(V) is projectively equivalent to it, so the lemma follows. □

To complete the proof of the claim 4.6 and the proof of Proposition 4.2, we consider the common apolar subscheme Γ = Γ₁ ∪ ... ∪ Γᵣ to q and q', and the corresponding decompositions q = ∑ qᵢ and q' = ∑ q'ᵢ as above. By Lemma 4.7,

\[ \sum_i U_i ^⊥ \cdot (U_i ') ^⊥ \subset \text{Im } τ(q). \]

Furthermore, applying Lemma 4.8 to each component qᵢ and qᵢ', the image of τᵢ(qᵢ) in (q⁻¹(Uᵢ))² is I₊,i. But τᵢ(qᵢ') is the restriction of τᵢ(q') to ∧²(q⁻¹(Uᵢ)), so

\[ I₊,i \subset \text{Im } τ(q') \quad i = 1, ..., r \]

and the claim and Proposition 4.2 follows. □

By Lemma 4.3, the quadratic forms in Im τᵢ(q') vanish in every point on any common apolar subscheme of length n to q and q'. Combined with Proposition 4.2 it may be reasonable to guess that Im τᵢ(q') is precisely the quadratic forms in the intersection of the ideals of these common apolar subschemes. We do not have a clear answer and leave this as an open question.

We are now ready to analyze our main object VPS(Q,n) in its embedding in G(n - 1, T₂,q), i.e. as the image of the rational map

\[ γ : P(T₂,q) \rightarrow G(n - 1, T₂,q). \]

We identify the restriction of the Plücker divisor to VPS(Q,n).

Let h ⊂ P(T₁) be a hyperplane, and denote by Hₕ ⊂ VSQ(Q,n) the set

\[ Hₕ = \{ [Γ] \in VSQ(Q,n)|Γ \cap h \neq 0 \}. \]

Lemma 4.9. Hₕ is the restriction to VPS(Q,n) of a Plücker divisor on G(n - 1, T₂,q).

Proof. The hyperplane h ⊂ P(T₁) is defined by some l ∈ S₁. Let V(l) = {q' ∈ T₂|l(q') = 0}, then V(l) = l ∩ S₁ = {l' | l' ∈ S₁} ⊂ S₂.

For any nondegenerate subscheme Γ ⊂ P(T₁) of length n, the ideal Iₚ ⊂ S contains a reducible quadric l₁.l₂ only if Γ intersects both hyperplanes {l₁ = 0} and {l₂ = 0}. On the other hand the subspace of quadrics I₊,₂ ⊂ S₂ has codimension n, which coincides with the dimension of l ∩ S₁. Therefore

\[ I₊,₂ ∩ l ∩ S₁ \neq 0 \subset S₂ \] if and only if \( (I₊,₂) ^⊥ \cap V(l) \neq 0 \subset T₂. \]

Notice that P((I₊,₂) ^⊥) equals the span (Γ) ⊂ P(T₂) of Γ in the Veronese embedding.

For the lemma we now consider apolar subschemes to q and the projection from P(T₂) to P(T₂,q). Since q has maximal rank, l(q) ≠ 0, i.e. q ∉ V(l). Thus
The orthogonal space of quadratic forms in $S_T$ and intersect $q_P$: 1]

We choose coordinates such that $P$ irreducible and therefore coincides with $VPS_Q$ The quadric $Q$ is spanned by \[ \langle x_1^2, x_1x_2, x_2^2, x_1x_3, x_2x_3, \ldots, x_{n-1}^2, x_n \rangle \]

n intersection of the projected Veronese variety $V$ and consider the apolar subscheme $\Gamma$ of length $n$. But the set of $(n - 2)$-dimensional subspaces in $P(T_2,q)$ that intersect a linear space of codimension $n$ form a Plücker divisor, so the lemma follows. □

In the next section we use the special Plücker divisors $H_h$ of this lemma to give a local affine description of $VPS(Q,n)$, or better, the variety $VAPS(Q,n)$ of all apolar subschemes of length $n$.

5. An open affine subvariety

We use a standard basis approach to compute an open affine subvariety of $VAPS(Q,n)$, the variety of all apolar subschemes of length $n$ to $Q$. Of course this will include our primary object of interest, namely $VPS(Q,n)$. For small $n$ there will be no difference, but for larger $n$ we have already seen that they do not coincide. The distinction between the two will eventually be the main concern in our analysis. The computations in this section extensively use Macaulay2 [GS]. In particular when we show, by direct computation, that $VAPS(Q,6)$ is irreducible and therefore coincides with $VPS(Q,6)$ (Corollary 5.16).

We choose coordinates such that $Q = \{ q = 2y_1y_n + y_2^2 + \ldots + y_{n-1}^2 = 0 \}$, and consider the apolar subscheme $\Gamma_p$ to $q$ defined by

\[ \langle x_1^2, x_1x_2, x_2^2 - x_1x_n, x_1x_3, x_2x_3, \ldots, x_{n-1}^2 - x_1x_n \rangle. \]

It is of length $n$ and corresponds in the setting of the previous section to the intersection of the projected Veronese variety $V_2,q$ with the tangent space $T_p$ to $v_2(Q^{-1}) \subset P(T_2,q)$ at the point $v_2(p) = [y_2] \in P(T_2,q)$, where $p = [y_n] = [0 : \ldots : 1] \in P(T_1)$. The tangent space to the Veronese variety $V_2 \subset P(T_2)$ at $[y_2]$ is spanned by

\[ \langle y_1y_n, y_2y_n, \ldots, y_{n-1}y_n, y_{n-1} \rangle. \]

The quadric $Q^{-1}$ is defined by $\frac{1}{2}x_1x_n + \frac{1}{4}(x_2^2 + \ldots + x_{n-1}^2)$. Its tangent space in $P(T_1)$ at $[y_n]$ is defined by $x_1$, so its tangent space in $P(T_2)$ at $[y_2]$ is defined by $x_1$ inside the tangent space to the Veronese variety. Therefore, the tangent space $T_p$ to $v_2(Q^{-1})$ is spanned by

\[ \langle y_2y_n, \ldots, y_{n-1}y_n, y_{n-1} \rangle. \]

The orthogonal space of quadratic forms in $S_2$ is spanned by

\[ \langle x_1^2, x_1x_2, x_2^2, x_1x_3, x_2x_3, \ldots, x_{n-1}^2, x_1x_n \rangle \]

and intersect $q^{-1}$ precisely in the ideal of $\Gamma_p$ given above.
We therefore consider the open subvariety $V \in \text{Aff}(\mathbb{P}^n)$ for $\Gamma$ the tangent hyperplane to any apolar scheme $\Gamma$ that does not intersect the hyperplane $\{x_n = 0\}$. In fact, if the initial ideal of $\Gamma$ contains $x_i x_n$, then $x_n$ divides a quadratic form in the ideal of $\Gamma$. But if $\Gamma$ does not intersect $\{x_n = 0\}$, then $\Gamma$ would be degenerate.

We therefore consider the open subvariety $V_{\text{aff}}(n)$ containing $[\Gamma_p]$ in $V_{\text{APS}}(Q, n)$, parametrizing apolar subschemes $\Gamma$ of length $n$ with support in $D(x_n)$. This is the complement of the divisor $H_n$ defined by $h = [x_n = 0]$, the tangent hyperplane to $Q^{-1}$ at $[y_1] = [1 : 0 : \ldots : 0] \in \mathbb{P}(T_1)$.

For $\Gamma \in V_{\text{aff}}(n)$ the initial terms of the generators of the ideal $I_{\Gamma}$ coincide with those of $I_{\Gamma_p}$. More precisely, the generators of $I_{\Gamma}$ may be obtained by adding suitable multiples of the monomials $x_i x_n, i \geq 1$ to these initial terms. We may therefore write these generators in the form

$$x_i^2 - a_{(1,1)} x_1 x_n - a_{(1,2)} x_2 x_n - \ldots - a_{(1,n)} x_n^2,
$$

$$x_i x_j - a_{(i,j,1)} x_1 x_n - a_{(i,j,2)} x_2 x_n - \ldots - a_{(i,j,n)} x_n^2, \quad 1 \leq i < j \leq n - 1,
$$

$$x_i^2 - x_1 x_n - a_{(i,2)} x_2 x_n - \ldots - a_{(i,n)} x_n^2, \quad 2 \leq i \leq n - 1.$$  

Analyzing these equations of $\Gamma$ further, we see that the apolarity condition, i.e., that $I_{\Gamma,2} \subset q^1$, means that $a_{(1,1)} = 0$ and that $a_{(i,j,1)} = 0$ when $i \neq j$. Therefore they take the form

$$f_{11} = x_1^2 - a_{(1,2)} x_2 x_n - \ldots - a_{(1,n)} x_n^2,
$$

$$f_{12} = x_1 x_2 - a_{(2,2)} x_2 x_n - \ldots - a_{(2,n)} x_n^2,
$$

$$f_{22} = (x_2^2 - x_1 x_n) - a_{(2,2)} x_2 x_n - \ldots - a_{(2,n)} x_n^2,
$$

$$f_{13} = x_1 x_3 - a_{(3,2)} x_2 x_n - \ldots - a_{(3,n)} x_n^2,
$$

$$f_{23} = x_2 x_3 - a_{(3,3)} x_2 x_n - \ldots - a_{(3,n)} x_n^2,
$$

$$f_{33} = (x_3^2 - x_1 x_n) - a_{(3,3)} x_2 x_n - \ldots - a_{(3,n)} x_n^2,
$$

$$\vdots
$$

$$f_{(n-1)(n-1)} = (x_{n-1}^2 - x_1 x_n) - a_{((n-1)(n-1),2)} x_2 x_n - \ldots - a_{((n-1)(n-1),n)} x_n^2.$$  

To insure that these perturbed equations actually define length $n$ subschemes, we ask that the first order relations or syzygies among the generators of $I_{\Gamma_p}$ lift to the entire family. This is in fact precisely the requirement for the perturbation to define a flat family [Art76, Proposition 3.1], and will be pursued below when we find equations for $V_{\text{aff}}(n)$.

Here, we introduce weights and a torus action on this family: We give

- $x_n$ and $a_{(i,j)}$, where $2 \leq i, j, k \leq n - 1$, weight $1$,
- $x_i$, where $2 \leq i \leq n - 1$, and $a_{(i,j,1)}$, where $2 \leq i, j \leq n - 1$, weight $2$,
- $x_1$ and $a_{(1,n)}$ and $a_{(1,1)}$, where $2 \leq i \leq n - 1$, weight $3$. 

\[\text{Documenta Mathematica} 18 (2013) 469–505\]
Notice that with these weights each generator $f_{ij}$ is homogeneous. A $\mathbb{C}^*$-action defined by multiplying each parameter with a constant $\lambda^w$ to the power of its weight, acts on each generator by a scalar multiplication, i.e. on the total family in $\mathbf{P}(T_1) \times \mathcal{V}^{\text{aff}}_h(n)$. This $\mathbb{C}^*$-action induces an action on the family $\mathcal{V}^{\text{aff}}_h(n)$. In particular, if $[a] = [a_{(i,j,k)}] \in \mathcal{V}^{\text{aff}}_h(n)$ defines a subscheme $\Gamma[a]$, then $[\lambda^w(a)] = [\lambda^w(a_{(i,j,k)})] \in \mathcal{V}^{\text{aff}}_h(n)$ and defines a subscheme $\Gamma[\lambda^w a]$, such that $p' \in \Gamma[a]$ if and only if $\lambda^w(p') \in \Gamma[\lambda^w a]$.

Since $\lim_{\lambda \to 0} \lambda^w(a_{(i,j,k)}) = 0$, the limit when $\lambda \to 0$ of the $\mathbb{C}^*$-action is the point in $\mathcal{V}^{\text{aff}}_h(n)$ representing $\Gamma_p$. Thus we have shown

**Lemma 5.1.** The affine algebraic set $\mathcal{V}^{\text{aff}}_h(n)$ of apolar subschemes of length $n$ contained in $D(x_n)$ coincides with the apolar schemes of length $n$ whose equations are affine perturbations of the equations of $\Gamma_p$.

Furthermore, the family $\mathcal{V}^{\text{aff}}_h(n)$ is contractible to the point $[\Gamma_p]$.

An immediate consequence is the

**Corollary 5.2.** The apolar subscheme $\Gamma_p$ belongs to $\mathcal{V}^{\text{PS}}_Q(n)$. In particular, the variety of tangent spaces $TQ^{-1} \subset G(n-1,T_{q,2})$ to the Veronese embedding of the quadric $Q^{-1} \subset \mathbf{P}(T_{q,2})$ is a subvariety of $\mathcal{V}^{\text{PS}}_Q(n)$.

Notice that $\mathcal{V}^{\text{aff}}_h(n)$ depends only on $h$, and not on $p$. Only the coordinates on $\mathcal{V}^{\text{aff}}_h(n)$ depend on $p$. On the other hand, the contractible varieties $\mathcal{V}^{\text{aff}}_h(n)$ form a covering of $\mathcal{V}^{\text{APS}}_Q(n)$:

**Lemma 5.3.** If $h_j = \{l_j = 0\}, j = 1, \ldots, n^2$ is a collection of tangent hyperplanes to $Q^{-1}$, so that no subset of $n$ of them have a common point, then the open subvarieties $\mathcal{V}^{\text{aff}}_{h_j}(l_j)$ parametrizing apolar subschemes $Z$ of length $n$ with support in $D(l_j)$ form a covering of $\mathcal{V}^{\text{APS}}_Q(n)$ of isomorphic varieties.

**Proof.** If an apolar subscheme $\Gamma$ has $k \leq n$ components, then the collection of hyperplanes among the $\{l_j = 0\}$ that intersect $\Gamma$ is at most $k(n-1) < n^2$, so the $\mathcal{V}^{\text{aff}}_{h_j}(l_j)$ form a covering. The last part follows from the homogeneity. \qed

To find equations for the family $\mathcal{V}^{\text{aff}}_h(n)$, we use the parameters for the generators in (5.1), i.e.

$$a_{(i,j,k)} \quad i,j \in \{1,\ldots,n-1\}, \ 2 \leq k \leq n,$$

where we read the first index $(ij)$ as an unordered pair.
It will be useful to write the generators with matrices:

\[
\begin{pmatrix}
  f_{11} \\
  f_{12} \\
  f_{22} \\
  \vdots \\
  f_{(n-1)(n-1)}
\end{pmatrix} =
\begin{pmatrix}
  1 & 0 & \ldots & 0 & -a_{(1,2)} & \ldots & -a_{(1,n)} \\
  0 & 1 & \ldots & 0 & -a_{(2,2)} & \ldots & -a_{(2,n)} \\
  0 & 0 & \ldots & 0 & -a_{(2,2)} & \ldots & -a_{(2,n)} \\
  \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \ldots & 1 & \vdots & \ldots & -a_{((n-1)(n-1),n)}
\end{pmatrix}
\begin{pmatrix}
  x_1 \\
  x_1 x_2 \\
  x_1^2 - x_1 x_n \\
  \vdots \\
  x_{n-1}^2 - x_1 x_n \\
  x_{n-1} x_n \\
  \vdots \\
  x_n^2 
\end{pmatrix}.
\]

We denote by $A_F$ the $(n^2) \times ((n^2) + n - 1)$-dimensional coefficient matrix of these generators. The maximal minors of $A_F$ are, of course, precisely the Plücker coordinates for $V_{h}^{\text{aff}}(n)$ in $G((n^2), q^{-})$, or equivalently in $G(n - 1, T_{2,q})$.

We find the equations of the family by asking that the first order syzygies among the generators of $I_{F_p}$ lift to the entire family. By [Art76, Proposition 3.1], this is precisely the requirement for the perturbation to define a flat family. We use a standard basis approach (cf. [Sch91]). The syzygies for a subscheme $Z$ in the family are all linear, and the initial terms are inherited from $\Gamma_p$.

Therefore, the difference between syzygies of $Z$ and syzygies of $\Gamma_p$ are only multiples of $x_n$. By the division theorem ([Sch91, Theorem A.3]), every syzygy has the initial term $x_k(x_i x_j)$, where $k > j \geq i$, and has the form

\[x_k f_{ij} = \sum_{st} g_{ij}^{st} f_{st}\]

where $f_{ij}$ is the generator with initial term $x_i x_j$ and $g_{ij}^{st}$ is a linear form such that $g_{ij}^{st} f_{st}$ has higher order than $x_k(x_i x_j)$. More precisely, we therefore consider products of the generators ($f_{ij}$) with a first order syzygy for $\Gamma_p$ and add precisely those multiples of $x_n$ in the syzygy that eliminates monomials $x_k x_i x_n$ with $k \leq l < n$ in the product. The relations among the parameters required for the lifting of the syzygies can then be read off as the coefficients of the monomials $x_i x_n^2$.

**Theorem 5.4.** The equations defining $V_{h}^{\text{aff}}(n)$ all lie in the linear span of the $2 \times 2$ minors of the coefficient matrix $A_F$ of the family of equations $f_{ij}$. In particular $V_{\text{APS}}(Q, n)$ is a linear section of the Grassmannian $G(n - 1, T_{2,q})$.

**Remark 5.5.** This generalizes the result of Mukai in case $n = 3$, cf. [Muk92]. For an exposition of different approaches to the $n = 3$ case, see [Dol12, Section 2.1.3].

**Proof.** Consider the following first order syzygies of $\Gamma_p$ of rank 2 and 3:

\[R_i \cdot S_i(m), 1 < i < n, m = 1, 2:\]

\[
\begin{pmatrix}
  x_i^2 & x_1 x_i & x_i^2 - x_1 x_n \\
  -x_i & x_n & x_i \\
  0 & x_1 & x_i
\end{pmatrix} = 0,
\]

where $S_i(m)$ is the $m$-th column vector in the syzygy matrix $S_i$. 
$R_{ij} \cdot S_{ij}(m), 1 < i < j < n, m = 1, \ldots, 4$:

$$(x_i^2 - x_1 x_n \ \ x_1 x_i \ \ x_1 x_j \ \ x_i x_j \ \ x_j^2 - x_1 x_n) \cdot \begin{pmatrix} x_j & 0 & 0 & 0 \\ 0 & x_n & x_j & 0 \\ x_n & 0 & -x_i & -x_i \\ -x_i & -x_j & 0 & x_1 \\ 0 & x_i & 0 & 0 \end{pmatrix} = 0,$$

and

$R_{ijk} \cdot S_{ijk}(m), 1 < i < j < k < n, m = 1, 2$:

$$(x_i x_j \ \ x_j x_k \ \ x_i x_k) \cdot \begin{pmatrix} -x_k & -x_k \\ x_i & 0 \\ 0 & x_j \end{pmatrix} = 0.$$
Similarly
\[ R_i \cdot S_i(2) = (a_{(i,j)} - a_{(i,n)}) x_i x_n^2 + a_{(i,n)} x_i x_n^2 + \sum_{j=2}^{n-1} (-a_{(i,j)} + \sum_{k=2}^{n-1} (a_{(i,k)} a_{(i,k,j)} - a_{(i,k)} a_{(i,j,k)})) x_i x_n^2, \]
\[ R_j \cdot S_j(1) = (a_{(j,i)} - a_{(j,n)}) x_j x_n^2 + a_{(j,n)} x_j x_n^2 + \sum_{i=2}^{n-1} (-a_{(j,i)} + \sum_{k=2}^{n-1} (a_{(j,k)} a_{(j,m,k)} - a_{(j,k)} a_{(j,m,i)})) x_j x_n^2, \]
\[ R_k \cdot S_k(1) = (a_{(k,i)} - a_{(k,n)}) x_k x_n^2 + a_{(k,n)} x_k x_n^2 + \sum_{i=2}^{n-1} (-a_{(k,i)} + \sum_{j=2}^{n-1} (a_{(k,j)} a_{(j,m,i)} - a_{(j,m)} a_{(j,m,i)})) x_k x_n^2, \]
\[ R_{ijk} \cdot S_{ijk}(1) = (a_{(i,k)} - a_{(i,j)}) x_i x_j x_k^2 + a_{(i,j)} x_i x_j x_k^2 + \sum_{m=2}^{n-1} (-a_{(i,k)} + \sum_{j=2}^{n-1} (a_{(j,m)} a_{(j,m,k)} - a_{(j,m)} a_{(j,m,i)})) x_i x_j x_k^2, \]
\[ R_{ijk} \cdot S_{ijk}(2) = (a_{(i,k)} - a_{(i,j)}) x_i x_j x_k^2 + a_{(i,j)} x_i x_j x_k^2 + \sum_{m=2}^{n-1} (-a_{(i,k)} + \sum_{j=2}^{n-1} (a_{(j,m)} a_{(j,m,k)} - a_{(j,m)} a_{(j,m,i)})) x_i x_j x_k^2. \]

The linear relations in the parameters of the family \( V_h^{\text{aff}}(n) \) are precisely the coefficients of \( x_i x_j x_k^2 \) in these products:

**Lemma 5.6.** The space of linear forms in the ideal of \( V_h^{\text{aff}}(n) \) is generated by the following forms, where \( \{i, j, k\} \) is any subset of distinct elements in \( \{2, \ldots, n-1\} \)

\[ a_{(i,j)} - a_{(i,n)}, a_{(i,n)} - a_{(i,n)} \]
\[ a_{(i,j)} - a_{(i,n)}, a_{(i,j)} - a_{(i,j,n)}, a_{(i,j)} - a_{(j,j,i)}, a_{(i,j,k)} - a_{(j,k,i)}. \]

Notice that only the first two occur when \( n = 3 \), and only the first four occur when \( n = 4 \).

Using these linear relations, the quadratic ones all become linear in the \( 2 \times 2 \) minors of the matrix \( A_F \) of coefficients \( a_{(i,j,k)} \), i.e. linear in the Pl"ucker coordinates. In fact, by a straightforward but tedious derivation from the above presentation, we may write the generators of the ideal \( V_h^{\text{aff}}(n) \) as linear combinations of \( 2 \times 2 \) minors in the coefficient matrix \( A_F \) (cf. the documented computer algebra code to perform the computation of ideal generators [RS11]):

**Lemma 5.7.** Modulo the linear forms the ideal of \( V_h^{\text{aff}}(3) \) is generated by
\[ a_{(1,3)} + a_{(2,3)} a_{(22,2)} - a_{(12,2)} a_{(22,3)}. \]

**Lemma 5.8.** Modulo the linear forms the ideal of \( V_h^{\text{aff}}(4) \) is generated by
\[ -a_{(12,2)} - a_{(13,3)} + (a_{(23,2)} a_{(22,3)} - a_{(22,2)} a_{(23,3)}) + (a_{(33,2)} a_{(23,3)} - a_{(23,3)} a_{(33,3)}), \]
\[ -a_{(13,3)} + (a_{(12,3)} a_{(23,3)} - a_{(23,3)} a_{(23,2)}) + (a_{(13,3)} a_{(33,3)} - a_{(33,3)} a_{(13,2)}), \]
\[ -a_{(11,3)} + (a_{(12,2)} a_{(22,2)} - a_{(22,2)} a_{(22,3)}) + (a_{(13,2)} a_{(23,3)} - a_{(23,3)} a_{(13,2)}), \]
\[ (a_{(11,3)} a_{(22,3)} - a_{(12,2)} a_{(23,3)} + a_{(13,3)} a_{(22,2)} - a_{(12,2)} a_{(23,3)}), \]
\[ (a_{(11,3)} a_{(22,3)} - a_{(12,2)} a_{(23,3)} + a_{(13,3)} a_{(22,2)} - a_{(12,2)} a_{(23,3)}), \]
\[ a_{(11,4)} + (a_{(12,4)} a_{(22,2)} - a_{(22,4)} a_{(22,2)}) + (a_{(13,4)} a_{(22,3)} - a_{(22,4)} a_{(13,2)}), \]
\[ a_{(11,4)} + (a_{(12,4)} a_{(23,3)} - a_{(23,4)} a_{(12,3)}) + (a_{(13,4)} a_{(33,3)} - a_{(33,4)} a_{(13,3)}). \]
LEMMA 5.9. Modulo the linear forms the ideal of $V_h^{aff}(n)$, when $n > 4$, is generated by the following forms:

For $i \in \{2, \ldots, n-1\}$,

$$a_{(1,1),n} - \sum_{m=2}^{n-1} (a_{(im,n)}a_{(1m,i)} - a_{(1m,n)}a_{(im,i)}),$$

for any subset $\{i,j\} \subset \{2, \ldots, n-1\}$,

$$a_{(1,1),i} - \sum_{m=2}^{n-1} (a_{(jm,n)}a_{(im,j)} - a_{(im,n)}a_{(jm,j)}),$$

and

$$n-1 \sum_{m=2}^{n-1} (a_{(1m,n)}a_{(im,j)} - a_{(im,n)}a_{(1m,j)})$$

and

$$a_{(1i,j)} + a_{(1j,i)} - \sum_{m=2}^{n-1} (a_{(jm,i)}a_{(im,j)} - a_{(im,i)}a_{(jm,j)}),$$

for any subset $\{i,j,k\} \subset \{2, \ldots, n-1\}$,

$$a_{(1j,k)} - \sum_{m=2}^{n-1} (a_{(jm,i)}a_{(im,k)} - a_{(im,i)}a_{(jm,k)}),$$

and

$$n-1 \sum_{m=2}^{n-1} (a_{(jm,n)}a_{(im,k)} - a_{(im,n)}a_{(jm,k)}),$$

and for any subset $\{i,j,k,l\} \subset \{2, \ldots, n-1\}$,

$$n-1 \sum_{m=2}^{n-1} (a_{(im,j)}a_{(km,l)} - a_{(km,j)}a_{(im,l)}).$$

Since the open affine sets $V_h^{aff}(n)$ cover $VAPS(Q,n)$, we conclude that $VAPS(Q,n)$ is a linear section of the Grassmannian $\mathbb{G}((\mathbb{C}^n)^q, q^-$). Equivalently, $VAPS(Q,n)$ is projectively equivalent to a linear section of $\mathbb{G}(n-1, T_{2,q})$ in its Plücker embedding. This concludes the proof of Theorem 5.4. □

Using the linear relations we may reduce the number of variables, when $n > 4$, and use as indices the following unordered three element sets:

$$\mathcal{I} = \{\{1k\}|1 < k \leq n\} \cup \{\{1j\}|1 < j < n\} \cup \{\{ijk\}|1 < i < j < k < n\}. $$

Let $R = \mathbb{C}[a_I | I \in \mathcal{I}]$. We substitute $a_{11k} = a_{11,k}, a_{1jk} = a_{1j,k}, a_{ijk} = a_{ij,k}$ and get:
Lemma 5.10. The ideal of $V^\text{aff}_h(n)$, when $n > 4$, is generated by the following polynomials in $R$: For $i \in \{2, \ldots, n-1\}$,
\[
a_{11n} - \sum_{m=2}^{n-1} (a_{1im}^2 - a_{1ln} a_{im}),
\]
for any subset $\{i, j\} \subset \{2, \ldots, n-1\}$,
\[
a_{11i} - \sum_{m=2}^{n-1} (a_{1jm} a_{ijm} - a_{1im} a_{jjm}),
\]
\[
\sum_{m=2}^{n-1} (a_{11m} a_{ijm} - a_{1jm} a_{1jm}),
\]
for any subset $\{i, j, k\} \subset \{2, \ldots, n-1\}$,
\[
a_{1jk} - \sum_{m=2}^{n-1} (a_{ijm} a_{ikm} - a_{ijm} a_{jkm}),
\]
\[
\sum_{m=2}^{n-1} (a_{1jm} a_{ikm} - a_{1im} a_{jkm}),
\]
\[
\sum_{m=2}^{n-1} (a_{ijm} a_{kjm} - a_{ijm} a_{kjm}).
\]
\[\square\]

Notice that these generators are all homogeneous in the weights introduced above. The linear parts of the ideal generators define the tangent space of the family $V^\text{aff}_h(n)$ at $[\Gamma_p]$, so another consequence of our computations is the tangent space dimension.

Proposition 5.11. Let $L(n)$ be the space of linear forms spanned by the linear parts of the generators in the ideal of $V^\text{aff}_h(n)$. Then $L(3)$ is spanned by
\[
a_{(11,3)}, \quad a_{(11,2)} - a_{(12,3)} \quad \text{and} \quad a_{(12,2)} - a_{(22,3)},
\]
$L(4)$ is spanned by
\[
a_{(11,4)}, \quad a_{(11,2)}, \quad a_{(12,4)}, \quad a_{(11,3)}, \quad a_{(13,4)}, \quad a_{(12,2)} - a_{(22,4)}, \quad a_{(13,3)} - a_{(33,4)}, \quad a_{(12,2)} + a_{(13,3)},
\]
\[
a_{(12,3)} - a_{(23,4)}, \quad a_{(13,2)} - a_{(12,3)}, \quad a_{(23,3)} - a_{(33,2)} \quad \text{and} \quad a_{(22,3)} - a_{(23,2)},
\]
$L(n)$, when $n > 4$, is spanned by $a_{(11,n)}$ and for any $i \in \{2, \ldots, n-1\}$,
\[
a_{(11,i)}, a_{(11,1)}, a_{(11,1)}, a_{(ii,1)}, a_{(ii,i)},
\]
for any subset $\{i, j\} \subset \{2, \ldots, n-1\}$,
\[
a_{(1j,1)}, a_{(ij,1)}, a_{(ij,i)} - a_{(ij,i)},
\]
and for any subset $\{i, j, k\} \subset \{2, \ldots, n-1\}$,
\[
a_{(ij,k)} - a_{(jk,i)}.
\]
In particular
\[ V_{h}^{\text{aff}}(3) \cong \mathbf{A}^3, \quad V_{p}^{\text{aff}}(4) \cong \mathbf{A}^6, \quad V_{h}^{\text{aff}}(5) \cong \mathbf{A}^{10}. \]

Proof. The linear parts of the generators can be read off Lemma 5.6 and Lemmas 5.7, 5.8 and 5.9. Notice only that the two term forms
\[ a_{(12,2)} - a_{(22,4)}, \quad a_{(13,3)} - a_{(33,4)}, \quad a_{(12,2)} + a_{(13,3)} \]
span a three dimensional space, while
\[ a_{(1i,i)} - a_{(ii,n)}, \quad a_{(jj,j)} - a_{(jj,n)}, \quad a_{(1i,i)} + a_{(1j,j)}, \quad \text{for } 1 < i < j < n \]
span the space generated by
\[ a_{(1i,i)}, \quad a_{(ii,n)}, \quad 1 < i < n \]
when \( n > 4 \).

For \( n = 3 \), the family \( V_{h}^{\text{aff}}(3) \) has 6 parameters, while there are three independent linear forms in the relations so the tangent space at \([\Gamma_p]\) has dimension \( 6 - 3 = 3 \) as expected. In fact \( V_{h}^{\text{aff}}(3) \cong \mathbf{A}^3 \) with parameters
\[ a_{(11,2)}, a_{(12,2)}, a_{(22,2)}. \]

For \( n = 4 \) the family \( V_{h}^{\text{aff}}(4) \) has 18 parameters, while the linear forms in the relations are generated by 12 independent forms, so the tangent space at \([\Gamma_p]\) has dimension \( 18 - 12 = 6 \). In fact \( V_{h}^{\text{aff}}(4) \cong \mathbf{A}^6 \) with parameters
\[ a_{(12,2)}, a_{(12,3)}, a_{(23,3)}, a_{(22,3)}, a_{(22,2)}, a_{(33,3)}. \]

For \( n > 4 \) we see that all parameters with a 1 or an \( n \) in the index are independent forms in the space of linear parts of ideal generators in \( V_{h}^{\text{aff}}(n) \). Furthermore, the other linear parts, simply expresses that \( \{(ijk)|1 < i < j < k < n\} \) form a natural index set for representatives of the parameters. The cardinality of this index set is simply the cardinality of monomials of degree 3 in \( n - 2 \) variables, i.e. \( \binom{n}{3} \). In case \( n = 5 \) we again conclude that \( V_{h}^{\text{aff}}(5) \cong \mathbf{A}^{10} \) with parameters
\[ \{a_{ijk}|2 \leq i \leq j \leq k \leq 4\}. \]

\[ \square \]

Corollary 5.12. The tangent space dimension of \( \text{VAPS}(Q,n) \) at \([\Gamma_p]\) is \( \begin{pmatrix} n \\ 3 \end{pmatrix} \) when \( n > 5 \). When \( n \leq 5 \), \( \text{VAPS}(Q,n) \) has a finite cover of affine spaces, in particular \( \text{VAPS}(Q,n) \) is smooth and coincides with \( \text{VPS}(Q,n) \).

Remark 5.13. Let \( \Gamma \) be a smooth apolar subscheme to \( Q \) consisting of \( n \) distinct points. Any subset of \( n - 2 \) points in \( \Gamma \) is contained in a pencil of apolar subschemes that form a line in \( \text{VPS}(Q,n) \) through \([\Gamma]\). Thus \( \begin{pmatrix} n \\ 2 \end{pmatrix} \) lines in \( \text{VPS}(Q,n) \) through \([\Gamma]\) is contained in the tangent space at \([\Gamma]\).

We extend this remark and give a conceptual reason for the dimension of the tangent space to \( \text{VAPS}(Q,n) \) at \([\Gamma_p]\).
Proposition 5.14. Let $[\Gamma_p] \in VPS(Q,n) \subset G(n-1, T_{2,q})$ be a point on the subvariety $TQ^{-1}$ in its Grassmannian embedding. Then $VPS(Q,n)$ contains the cone over a 3-uple embedding of $P^{n-3}$ with vertex at $[\Gamma_p]$.\\

Proof. We first identify a cone over a 3-uple embedding of $P^{n-3}$ inside $VAPS(Q,n)$, and then give an explicit description of the apolar subschemes parameterized by this cone in order to show that the cone is contained in $VPS(Q,n)$.\\

Consider the subvariety $V_{p \text{vero}}(n) \subset V_{\text{aff}}^{\text{geo}}(n)$ parameterizing ideals $I_{\Gamma}$ with coefficient matrix $A_{p}(G) = (I A)$ where the submatrix $A = (a_{ij,k})$ has rank at most 1 and has nonzero entries only in the submatrix $A_0 \subset A$ with entries $\{a_{ij,k} | 1 \leq i < j < n, 1 < k < n\}$. As above, using the linear relations, we may substitute the parameters $a_{ij,k}$ with parameters $a_{ijk}$ whose indices are unordered triples $(ijk)$. In these new parameters the matrix $A_0$ takes the form:

$$
A_0 = \begin{pmatrix}
a_{222} & a_{223} & \cdots & a_{22(n-1)} \\
a_{223} & a_{333} & \cdots & a_{23(n-1)} \\
\vdots & \vdots & \ddots & \vdots \\
a_{22(n-1)} & a_{23(n-1)} & \cdots & a_{2(n-1)(n-1)} \\
a_{2(n-1)(n-1)} & a_{3(n-1)(n-1)} & \cdots & a_{(n-1)(n-1)(n-1)}
\end{pmatrix}.
$$

By Theorem 5.4, the equations of $V_{\text{aff}}^{\text{geo}}(n)$ are linear in the 2 × 2 minors of the coefficient matrix $A$, so any rank 1 matrix $A_0$ defines a point on $V_{p \text{vero}}(n)$. The symmetry in the indices explains why the 2 × 2 minors of the matrix define the 3-uple embedding of $P^{n-3}$. Since the ideal of $\Gamma_p$ correspond to the zero matrix, we conclude that the subvariety $V_{p \text{vero}}(n)$ in $VAPS(Q,n)$ is the cone over this 3-uple embedding.\\

To see that $V_{p \text{vero}}(n)$ is contained in $VPS(Q,n)$ we show that a general point on $V_{p \text{vero}}(n)$ lies in the closure of smooth apolar subscheme to $Q$. For this, we describe for each general point $s \in P^{n-3}$ an apolar subscheme $\Gamma_s$ belonging to $V_{p \text{vero}}(n)$. It has two components $\Gamma_s = \Gamma_{s,0} \cup p_s$, the first one $\Gamma_{s,0}$ of length $n-1$ and supported at $p$, while the second component $p_s$ is a closed point. We shall show that $q$ has a decomposition $q = q_l + q(l)^2 \in T_2$ where $[q(l)] = p_l \in P(T_1)$ and $q_l \in (l^2)^2$. The subscheme $\Gamma_{s,0}$ is apolar to $q_l$ and contains the first order neighborhood of $p$ inside the quadric $\{q_l^{-1} = 0\} \subset P(l^2)$ in the hyperplane polar to $p_s$. Then $\Gamma_{s,0}$ lies in the closure of smooth apolar subschemes to $q_l$.\\

We conclude by applying Proposition 2.4.\\
Let $s = [s_2 : \ldots : s_{n-1}] \in P^{n-3}$ and let

$$
||s||^2 = s_2^2 + \ldots + s_{n-1}^2, \quad \langle s, x \rangle = \sum_{i=2}^{n-1} s_i x_i, \quad \langle s, y \rangle = \sum_{i=2}^{n-1} s_i y_i,
$$

then

$$
x_i^2 - x_1 x_n - s_i^2 \langle s, x \rangle x_n \quad 1 < i < n, \quad x_i x_j - s_i s_j \langle s, x \rangle x_n \quad 1 < i < j < n
$$
defines a subscheme \( \Gamma_s \) that belongs to \( V_{p_{\text{vero}}}^n(\mathbb{P}) \). When \( ||s||^2 \neq 0 \), then \( \Gamma_s \) contains the point

\[ p_s = [q(||s||^2(s, x) + x_1)] = [||s||^2(s, y) + y_n] \in \mathbb{P}(T_1) \]

Consider the linear subspace \( L_s = \{x_1 = 0\} \cap \{s, x = 0\} \). The intersection \( \Gamma_s \cap L_s \) is the subscheme defined by

\[ x_2^2 = x_2x_3 = \ldots = x_{n-1}^2 = 0. \]

This subscheme has length \( n - 2 \). The union \( p_s \cup (\Gamma_s \cap L_s) \) spans the hyperplane \( \{x_1 = 0\} \), so the residual point in \( \Gamma_s \) is the pole, with respect to \( Q^{-1} \), of this hyperplane, i.e. the point \( p \). Therefore the subscheme \( \Gamma_{s,0} = \Gamma_s \setminus p_s \) has length \( n - 1 \), is supported in \( p \), and contains the first order neighborhood of \( p \) in the codimension two linear space \( L_s \).

The subscheme \( \Gamma_{s,0} \) is apolar to the quadric

\[ q_s = (||s||^2(s, y) + y_n)^2 - ||s||^9(2y_1y_n + y_2^2 + \ldots + y_{n-1}^2). \]

Let \( l = ||s||^2(s, x) + x_1 \). Then \( p_s = [q(l)] \), while

\[ l_1^2 = \langle y_2 - ||s||^2s_1y_1, \ldots, y_{n-1} - ||s||^2s_1y_1, y_n \rangle. \]

Then \( q_s \in (l_1^2)^2 < 2 \) and

\[ (q(l))^2 - q_s = ||s||^9 \cdot q \in T_2. \]

According to Proposition 2.4 a subscheme \( \Gamma_0 \) in \( \mathbb{P}(l_1^2) \) of length \( n - 1 \) is apolar to \( q_s \) if and only if \( \Gamma = \Gamma_0 \cup p_s \) is apolar to \( q \). Now, \( \Gamma_{s,0} \) is apolar to \( q_s \) and contains a first order neighborhood of a point on the smooth quadric \( \{q_s^{-1} = 0\} \) in \( \mathbb{P}(l_1^2) \subset \mathbb{P}(T_1) \). By Remark 3.5, the subscheme \( \Gamma_{s,0} \) is a subscheme like \( \Gamma_p \), with respect to \( q_s \). Therefore \( \Gamma_{s,0} \) lies in the closure of smooth apolar subschemes to \( q_s \). But then \( \Gamma_s \) must lie in the closure of smooth apolar subschemes to \( q \). Hence \( [\Gamma_s] \in VPS(q, n) \).

\[ \square \]

**Corollary 5.15.** \( VPS(Q, n) \) is singular when \( n \geq 6 \).

**Proof.** The cone with vertex at \( [\Gamma_p] \in TQ^{-1} \) over the 3-uple embedding of \( \mathbb{P}^{n-3} \) is contained in the tangent space of \( VPS(Q, n) \) at \( [\Gamma_p] \), i.e. also in the tangent space of \( VAPS(Q, n) \). Since the span of the cone and the tangent space of the latter have the same dimension, they coincide. In particular the tangent space of \( VPS(Q, n) \) at \( [\Gamma_p] \) has dimension \( \binom{n}{3} \). When \( n \geq 6 \), then \( \binom{n}{3} > \binom{6}{3} = \dim VPS(Q, n) \) so \( VPS(Q, n) \) is singular.

\[ \square \]

We pursue the case \( n = 6 \) a bit further and show that \( VAPS(Q, 6) \) and \( VPS(Q, 6) \) coincide. We use the symmetric variables

\[ a_{ijk} = a_{(i,j,k)}, \ 1 \leq i, j, k \leq 6 \]

for any permutation of the letters \( i, j, k \). According to Lemma 5.10 we may list the generators explicitly. This list is however not minimal. In fact, a minimal set of generators is given by the following twenty generators in weight 2, four generators in weight 3 and one generator in weight 4. The twenty generators
of weight 2 are the generators of weight 2 in Lemma 5.10: For each \(1 < k < 6\), and each pair \(\{i, j\} \subset \{2, 3, 4, 5\} \setminus \{k\}\) the generator

\[-a_{ij6} + \sum_{m=2}^{5} (a_{ikm}a_{jkm} - a_{ijm}a_{km})\]

for each pair \(\{i, j\} \subset \{2, 3, 4, 5\}\) the generator

\[-a_{ii6} - a_{jj6} + \sum_{m=2}^{5} (a_{ijm}a_{ijm} - a_{iim}a_{jjm})\]

and additionally the two generators

\[\sum_{m=2}^{5} (a_{23m}a_{45m} - a_{24m}a_{35m})\]

and

\[\sum_{m=2}^{5} (a_{23m}a_{45m} - a_{25m}a_{34m}).\]

The last five generators are computed from the list of Lemma 5.10 using Macaulay2 \([GS]\), see the documented code in \([RS11]\).

Of weight 3 we find, for \(i = 2, 3, 4\):

\[a_{11i} - \sum_{m=2}^{5} (a_{im6}a_{m55} - a_{m56}a_{im5})\]

and

\[a_{115} - \sum_{m=2}^{5} (a_{m46}a_{m45} - a_{m56}a_{m44}).\]

The generator of weight 4 is

\[a_{116} - \sum_{m=2}^{5} a_{m56}^2 + \sum_{m=2}^{5} a_{11m}a_{m55}.\]

The ten parameters with 6 in the index appear linearly in the 20 generators of weight 2, while the five parameters with 11 in the index appear linearly in the five generators of weights 3 and 4. The remaining 10 generators of weight 2 therefore depend only on 20 parameters \(a_I\). In fact they depend only on 16 linear forms. It is a remarkable fact that these ten quadratic forms define the 10-dimensional spinor variety. To see this we choose and rename the following 16 forms:

\[x_{1234} = -a_{353} + a_{252}, \quad x_{15} = -a_{555} + a_{454} + a_{353} + a_{252}, \quad x_{34} = a_{453},\]

\[x_{1235} = a_{554} - a_{444} + a_{343} + a_{242}, \quad x_{14} = a_{343} - a_{242}, \quad x_{35} = a_{553} - a_{232},\]

\[x_{1245} = -a_{553} - a_{443} + a_{333} - a_{232}, \quad x_{13} = a_{553} - a_{443}, \quad x_{24} = a_{453},\]

\[x_{1345} = a_{552} + a_{442} + a_{332} - a_{222}, \quad x_{12} = a_{552} - a_{442}, \quad x_{23} = a_{352},\]

\[x_{2345} = a_{454} - a_{353}, \quad x_{45} = a_{554} - a_{242}, \quad x_{25} = -a_{442} + a_{332}, \quad x_{0} = a_{342}.\]

In these variables the ten quadratic generators takes the form

\[q_0 = x_{25}x_{34} - x_{35}x_{24} + x_{45}x_{23} + x_{2345}x_{0},\]

\[q_1 = -x_{45}x_{13} + x_{14}x_{35} - x_{15}x_{34} + x_{1345}x_{0}.\]
We will do the explicit computation in the cases where \( Pfaffians \). The ten quadratic forms satisfy the following quadratic relation

\[
q_2 = x_{45}x_{12} + x_{14}x_{25} + x_{15}x_{24} + x_{1245}x_0,
q_3 = -x_{35}x_{12} + x_{13}x_{25} - x_{15}x_{23} + x_{1235}x_0,
q_4 = x_{12}x_{34} - x_{13}x_{24} + x_{14}x_{23} + x_{1234}x_0,
q_5 = x_{1345}x_{12} + x_{1245}x_{13} + x_{1235}x_{14} + x_{15}x_{1234},
q_6 = -x_{2345}x_{12} + x_{1245}x_{23} + x_{1235}x_{24} + x_{1234}x_{25},
q_7 = -x_{2345}x_{13} - x_{1345}x_{23} + x_{1235}x_{34} + x_{1234}x_{35},
q_8 = -x_{2345}x_{14} - x_{1345}x_{24} - x_{1245}x_{34} + x_{1234}x_{45},
q_9 = -x_{15}x_{2345} - x_{1345}x_{25} - x_{1245}x_{35} - x_{1235}x_{45}.
\]

The first five express (when \( x_0 = 1 \)) the variables \( x_{ijkl} \) as quadratic Pfaffians in the \( x_{st} \), while the last five quadrics express the linear syzygies among these Pfaffians. The ten quadratic forms satisfy the following quadratic relation

\[
q_9q_5 + q_1q_6 + q_2q_7 + q_3q_8 + q_4q_9 = 0.
\]

In fact the ten quadratic forms generate the ideal of the 10-dimensional spinor variety embedded in \( \mathbb{P}^{15} \) by its spinor coordinates [RS00, Section 6],[Muk95].

**Corollary 5.16.** \( V_{h}^{af}(6) \) is isomorphic to a cone over the ten-dimensional spinor variety embedded in \( \mathbb{P}^{15} \) by its spinor coordinates. In particular \( VAPS(Q,6) \) is singular, irreducible and coincides with \( VPS(Q,6) \).

We end this section summarizing some computational results, for small \( n \), of some natural subschemes of \( VAPS(Q,n) \). The first is the punctual part \( V_{p}^{loc}(n) \) of \( VAPS(Q,n) \), i.e. the variety of apolar subschemes in \( VAPS(Q,n) \) with support at a single point \( p \). The support \( p \) of a local apolar subscheme must lie on \( Q^{-1} \) by Lemma 2.3. Therefore we may assume that \( p = [0:0:...:1] \), and use the equations 5.1. Of course, \( [\Gamma] \) is then in \( V_{p}^{loc}(n) \). Furthermore, \( V_{p}^{loc}(n) \) is naturally contained in a second natural subscheme of \( VAPS(Q,n) \), namely \( V_{p}^{sec}(n) \), the variety of apolar subschemes in \( V_{h}^{af}(n) \) that contains the point \( p \).

We will do the explicit computation in the cases where \( VAPS(Q,n) = VPS(Q,n) \) is smooth, i.e. when \( n < 6 \). An apolar subscheme in \( V_{h}^{af}(n) \) lies in \( V_{p}^{sec}(n) \) if and only if the term \( x_{n}^2 \) does not appear in any equation, so \( V_{p}^{sec}(n) \) is defined by the equations \( a_{i,j,n} = 0 \) for \( 1 \leq i \leq j < n \) in \( V_{h}^{af}(n) \). The linear relations then imply that \( a_{1,i,j} = 0 \) for all \( i,j \), and as before that each parameter \( a_{i,j,k} \) with \( 1 < i,j,k < n \) may be represented by a parameter

\[
a_{ijk} \text{ with } 1 < i \leq j \leq k < n.
\]

In particular for \( n = 3 \) the only parameter left is \( a_{222} \), and \( V_{p}^{sec}(3) \) is isomorphic to the affine line. The equations \( x_{1}^2 = x_{1}x_{2} = x_{2}^2 - x_{1}x_{3} = a_{222}x_{2}x_{3} \) define a scheme supported at \( p \) only if \( a_{222} = 0 \), so \( V_{p}^{loc}(n) \) is a point in the case \( n = 3 \). The computation of \( V_{p}^{sec}(n) \) follow the same procedure for every \( n \). For a local scheme \( \Gamma \) in \( V_{p}^{loc}(n) \) we may set \( x_{n} = 1 \) in the equations.

**Lemma 5.17.** A local scheme \( \Gamma \), supported at \( p \), that belongs to \( V_{p}^{loc}(n) \), is Gorenstein. The maximal ideal of its affine coordinate ring is spanned by \( x_{2}, \ldots, x_{n-1}, x_{1} \), and its socle is generated by \( x_{1} \).
The scheme $\Gamma$ is Gorenstein by Lemma 2.6. The maximal ideal is certainly generated by $x_1, x_2, \ldots, x_{n-1}$, and since $\Gamma$ is nondegenerate these are linearly independent. Finally, $x_1 x_i = 0$ for all $i$ by the apolarity condition as soon as $p \in \Gamma$, so the socle is generated by $x_1$.

We may now get explicit equations for $V_p^{loc}(n)$. If $[\Gamma] \in V_p^{loc}(n)$, then by definition $m_p^n = 0$. But the maximal ideal is generated by $x_1, x_2, \ldots, x_{n-1}$, so this means that any monomial of degree $n$ in the $x_i$ must vanish in the coordinate ring of $\Gamma$.

On the other hand, the equations for $\Gamma$ define the products $x_i x_j = \sum_{k=2}^{n-1} a_{ijk} x_k$ and $x_i^2 = x_1 + \sum_{k=2}^{n-1} a_{iik} x_k$ in this ring. Therefore, by iteration, we get polynomial relations in the parameters $a_{ijk}$. Imposing the apolarity condition, symmetrizing the parameters and adding the equations for $V_p^{sec}(n)$, then after, possibly, saturation we get set theoretic equations for $V_p^{loc}(n)$.

When $n = 4$ we have the parameters $a_{222}, a_{223}, a_{233}, a_{333}$ for $V_p^{sec}(4)$ and the relation

$$a_{223}^2 - a_{222} a_{233} + a_{233}^2 - a_{222} a_{333}.$$

Thus $V_p^{sec}(4)$ is a quadric hypersurface in $\mathbb{A}^4$.

For $V_p^{loc}(4)$ we first get the subscheme defined by the equations

$$x_1^2 = x_1 x_2 = x_1 x_3 = 0$$

and

$$x_2^2 = x_1 + a_{222} x_2 + a_{223} x_3, x_2 x_3 = a_{223} x_2 + a_{233} x_3; x_3^2 = x_1 + a_{233} x_2 + a_{333} x_3.$$

The coefficient of $x_1$ using these relations iteratively to compute $x_2^2, \ldots, x_4^2$, must vanish, so it yields the equations $a_{222} + a_{233} = a_{223} + a_{333} = a_{233}^2 + a_{223}^2 = 0$. The other coefficients give no additional relations, and neither does the equations for $V_p^{sec}(4)$, so $V_p^{loc}(4)$ is 1-dimensional and consists of a pair of affine intersecting lines.

When $n = 5$, the computation becomes a bit more involved. There are ten parameters $a_{ijk}$. The equations of $V_p^{sec}(5)$ are

$$a_{234}^2 - a_{233} a_{244} + a_{334}^2 - a_{333} a_{344} + a_{344}^2 - a_{334} a_{444} = 0,$$

$$a_{224} a_{234} - a_{223} a_{244} + a_{234} a_{334} - a_{233} a_{344} + a_{244} a_{334} - a_{234} a_{444} = 0,$$

$$a_{224} a_{233} - a_{223} a_{234} + a_{234} a_{333} - a_{233} a_{334} + a_{244} a_{334} - a_{234} a_{344} = 0,$$

$$a_{234}^2 + a_{224}^2 - a_{222} a_{244} + a_{244}^2 - a_{223} a_{334} - a_{224} a_{444} = 0,$$

$$a_{223} a_{224} - a_{223} a_{234} + a_{234} a_{344} - a_{233} a_{344} = 0,$$

$$a_{223}^2 - a_{222} a_{233} + a_{233}^2 - a_{222} a_{333} - a_{223} a_{333} + a_{224} a_{334} + a_{224}^2 = 0.$$

They define in $\mathbb{A}^{10}$ the affine cone over the intersection of the Grassmannian variety $G(2, 5)$ with a quadric. For $V_p^{loc}(5)$ there are additional equations defining the cone over the tangent developable of a rational normal sextic curve, a
codimension 3 linear section of $V^\sec_p(5)$. The cone over the rational normal curve parameterizes local apolar subschemes that are not curvilinear. For the computations in Macaulay2 [GS], see the documented code in [RS11]. The findings are summarized in Table 1.

6. Global invariants of $VPS(Q,n)$

We consider $VPS(Q,n)$ as a subscheme of $\mathbb{G}(n-1,T_2,q)$, and the incidence

$I_{Q}^{VPS} = \{(\gamma',\Gamma) | (\gamma') \subset (\Gamma) \} \subset \mathbb{P}(T_2,q) \times VPS(Q,n)$.

The incidence is a projective bundle,

$I_{Q}^{VPS} = \mathbb{P}(E_Q) \rightarrow VPS(Q,n)$,

while the first projection is birational (the rational map $\gamma : \mathbb{P}(T_2,q) \dashrightarrow VPS(Q,n)$ factors through the inverse of this projection). Denote by $L$ the tautological divisor on $\mathbb{P}(E_Q)$. It is the pullback of the hyperplane divisor on $\mathbb{P}(T_2,q)$. When $VPS(Q,n)$ is smooth,

$Pic(I_{Q}^{VPS}) \cong Pic(VPS(Q,n)) \oplus \mathbb{Z}[L]$.

Recall from Lemma 4.9, that the set $H_h \subset VPS(Q,n)$ of subschemes $\Gamma$ that intersects a hyperplane $h \subset \mathbb{P}(T_1)$ form a Plücker divisor restricted to $VPS(Q,n)$. Therefore the class of the Plücker divisor coincides with the first Chern class $c_1(E_Q)$.

**Theorem 6.1.**

i) $Pic(VPS(Q,4) \cong Pic(VPS(Q,5)) \cong \mathbb{Z}$.

ii) The ample generator $H$ is very ample, and $VPS(Q,4)$ and $VPS(Q,5)$ are Fano-mapsfolds of index 2.

iii) The boundary in $VPS(Q,n)$ consisting of singular apolar subschemes is, when $n \leq 5$, an anticanonical divisor.

**Proof.** i) Let $n = 4$ or $n = 5$. Then the Plücker divisor $H$ is very ample by the above. Furthermore, the complement $V^\text{aff}_p$ of the special Plücker divisor defined by a tangent hyperplane to $Q^{-1} \subset \mathbb{P}(T_1)$, the divisor $H_{(x_n=0)}$ in the

<table>
<thead>
<tr>
<th>$n$</th>
<th>$V^\text{loc}_p(n)$</th>
<th>$V^\text{sec}_p(n)$</th>
<th>$V^\text{aff}_h(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dim</td>
<td>degree</td>
<td>dim</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cone over tangent developable of a rational sextic curve</td>
<td>cone over $\mathbb{G}(2,5) \cap Q$</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cone over $S_{10}$</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.
above notation, is isomorphic to affine space by Proposition 5.11. Therefore the Picard group has rank 1 as soon as this special Plücker divisor is irreducible. The tangent hyperplanes to $Q^{-1}$ cover all of $P(T_1)$, so the corresponding Plücker divisors cover $VPS(Q,n)$. Furthermore, for any subscheme $\Gamma$ in $VPS(Q,n)$, there is a tangent hyperplane that does not meet $\Gamma$, so these special Plücker divisors have no common point on $VPS(Q,n)$. Assume that the special Plücker divisors are reducible, then we may write $H = H_1 + H_2$, where both $H_1$ and $H_2$ moves without base points on $VPS(Q,n)$. Since $H \cdot l = 1$ for every line on $VPS(Q,n)$, only one of the two components can have positive intersection with a line. The other, say $H_2$, must therefore contain every line that it intersects. But this is impossible, since $H_2$ must contain all of $VPS(Q,n)$, by the following lemma:

**Lemma 6.2.** Any two polar simplices $\Gamma$ and $\Gamma'$ are connected by a sequence of lines in $VPS(Q,n)$.

**Proof.** This is immediate when $n = 2$. For $n > 2$, let $[l] \in \Gamma$ and $[l'] \in \Gamma'$, and let $P(U) = h_1 \cap h_U \subseteq P(T_1)$ be the intersection of their polar hyperplanes. Then $q = l^2 + l_1^2 + qv = (l')^2 + (l_1')^2 + qv$ for $qv \in U^2$ and suitable $l_1$ and $l_1'$. Let $\Gamma_U$ be a polar simplex for $qv$. Then $\Gamma$ is line connected to $\Gamma_U \cup \{[l_1],[l]\}$ by induction hypothesis. Likewise $\Gamma$ is line connected to $\Gamma_U \cup \{[l],[l']\}$. Finally $\Gamma_U \cup \{[l],[l]\}$ and $\Gamma_U \cup \{[l_1],[l']\}$ span a line in $VPS(Q,n)$, which completes the induction. □

ii) Since Pic$(t^{VPS}_Q) \cong \text{Pic}(VPS(Q,n)) \oplus \mathbb{Z}[L]$ we deduce from i) that the birational morphism

$$\sigma : t^{VPS}_Q \rightarrow P(T_{2,q})$$

has an irreducible exceptional divisor. Let $E \in \text{Pic}(t^{VPS}_Q)$ be the class of this exceptional divisor. Then, since the map $\gamma : P(T_{2,q}) \dashrightarrow G(n-1,T_{2,q})$ is defined by polynomials of degree $\binom{n}{2}$, the size of the minors in the Mukai form, we have

$$\pi^*H = \left(\begin{array}{c} n \\ 2 \end{array}\right) L - E \text{ and } K^{VPS}_Q = -\left(\begin{array}{c} n + 1 \\ 2 \end{array}\right) - 1)L + E.$$ 

On the other hand $H = c_1(E_Q)$ where $I^{VPS}_Q = P(E_Q)$ is a projective bundle over $VPS(Q,n)$ so

$$-\left(\begin{array}{c} n + 1 \\ 2 \end{array}\right) - 1)L + E = K^{VPS}_Q = \pi^*K_{VPS} + \pi^*(c_1(E_Q)) - (n - 1)L.$$

Therefore $-K_{VPS(Q,n)} = 2H$. Finally, since $VPS(Q,n) \subseteq G(n-1,T_{2,q})$ contains lines, $H$ is not divisible.

iii) The boundary in $VPS(Q,n)$ consisting of singular apolar subschemes, coincides, by Lemma 2.3, with the set of subschemes $\Gamma \subset P(T_1)$ that intersect quadric $Q^{-1}$. The Plücker divisor $H$ is represented by the divisor of subschemes $\Gamma$ that intersect a hyperplane in $P(T_1)$, so $-K = 2H$ is represented by the boundary. □
Let $n > 2$ and let $VPS(Q, n) \subset G(n - 1, T_{2,q})$ be the variety of polar simplices in its Grassmannian embedding, with Plücker divisor $H$. The $VPS(Q, n)$ has degree

$$H^m = \sum_{\lambda \vdash m} \binom{m}{\lambda}/(\lambda!) \cdot d_{\lambda}$$

where the sum runs over all partitions $\lambda = (\lambda_1, \ldots, \lambda_n)$ of $m = \binom{n}{2} = \dim VPS(Q, n)$ into integers $n - 1 \geq \lambda_1 \geq \ldots \geq \lambda_n \geq 0$. Here $\lambda^* = (\lambda_1^*, \ldots, \lambda_{n-1}^*)$ denotes the sequence $\lambda_i^* = |\{ j \mid \lambda_j = i \}|$ and $\lambda^! = \prod \lambda_i^!$. Finally

$$d_{\lambda} = \prod_{1 \leq i < j \leq n} (D_i + D_j)$$

is the intersection number of $m$ divisors on the product

$$\mathbf{P}^{n-1-\lambda_1} \times \cdots \times \mathbf{P}^{n-1-\lambda_n}$$

with $D_i$ the pullback of the hyperplane class on the $i$th component.

**Proof.** We first show that for $\binom{n}{2}$ general hyperplanes $h_i \subset \mathbf{P}(T_1)$, the corresponding Plücker divisors $H_{h_i}$ has a proper transverse intersection on the smooth part of $VPS(Q, n)$. Therefore, by properness, the intersection is finite, and, by transversality, it is smooth, so it is a finite set of points. The cardinality is the degree of $VPS(Q, n)$.

First, let $\lambda = (\lambda_1, \ldots, \lambda_n)$ be a partition of $m$ and consider the partition $h_{11}, \ldots, h_{1\lambda_1}, \ldots, h_{n1}, \ldots, h_{n\lambda_n}$ of $m$ general hyperplanes into $n$ sets of size $\lambda_1, \ldots, \lambda_n$. Let $L_i = \cap_j h_{ij}$, it is a linear space of dimension $n - 1 - \lambda_i$. Consider the product of these linear spaces in the product $\mathbf{P}(T_1)^n$:

$$L_1 \times \cdots \times L_n \subset \mathbf{P}(T_1) \times \cdots \times \mathbf{P}(T_1).$$

Let $\Delta \in \mathbf{P}(T_1)^n$ be the union of all diagonals and let $L^\circ = L_1 \times \cdots \times L_n \setminus \Delta \subset \mathbf{P}(T_1)^n$. Then $L^\circ$ parameterizes $n$-tuples of points $\Gamma = \{p_1, \ldots, p_n\} \subset \mathbf{P}(T_1)$, with $p_i \in L_i$. Of course, $L^\circ$ has a natural map to the Hilbert scheme of $\mathbf{P}(T_1)$ that forgets the ordering, so we will identify elements in $L^\circ$ with their image in the Hilbert scheme.

Consider the incidence between subschemes $\Gamma \in L^\circ$ and quadratic forms $q \in T_2$:

$$I_L = \{ (\Gamma, [q]) | I_T \subset q^1 \} \subset L^\circ \times \mathbf{P}(T_2).$$

This variety is defined by the equations $h_{ij}(p_i) = 0$ and the apolarity, $q(I_T) = 0$. Clearly $L$ is a smooth scheme of dimension $\binom{n}{2}$. The fibers of the projection $I_L \rightarrow L$ are $(n - 1)$-dimensional projective spaces, so $I_L$ is a smooth variety of dimension equal to $\dim \mathbf{P}(T_2)$. The projection $I_L \rightarrow \mathbf{P}(T_2)$ is clearly onto, so the fibers are finite. Since both spaces are smooth, the general fiber is smooth. Now, $\Gamma \subset L^\circ$ lies in the fiber over $[q]$, precisely when $I_T \subset q^1$, i.e. $\Gamma \in VPS(Q, n)$ and $h_{ij}(p_i) = 0$, i.e. $\Gamma$ lies in the intersection of all the Plücker hyperplanes $H_{h_{ij}}$. Since the general fibers are smooth the divisors $H_{h_{ij}}$ intersect transversally in $VPS(Q, n)$, where $Q = \{ q = 0 \}$, and have an isolated intersection point at each point $[\Gamma]$. Turning the argument around and
considering all partitions, we get that for general hyperplanes \( h_1, \ldots, h_m \) in \( \mathbf{P}(T_1) \) the Plücker hyperplanes \( H_{h_i} \) has a transversal intersection at a finite number of points in \( VPS(Q, n) \) corresponding to smooth apolar subschemes.

We proceed to compute the cardinality of the intersection, i.e. the formula given in the theorem. Let \( [\Gamma] = \left\{ [p_1, \ldots, p_n] \right\} \in VPS(Q, n) \) be a point in the intersection of the hyperplanes \( H_{h_j} \). Then each \( h_j \) contains some \( p_i \in \Gamma \), by the definition of \( H_{h_j} \). For each \( i \) let \( \lambda_i \) be the number of hyperplanes \( h_j \) that contains \( p_i \). The set of positive integers \( \{\lambda_1, \ldots, \lambda_n\} \) must add up to \( m \):

It is at least \( m \) by definition, and at most \( m \) by the generality assumption discussed above. Therefore the point \([\Gamma]\) defines a unique partition of the set of hyperplanes \( \{h_j\}_{j=1}^m \) into subsets \( \{h_{i j}\}_{j=1}^m \) of cardinality \( \lambda_i \), as above.

The factor \( \binom{m}{\lambda} \) in the degree formula counts the number of ordered partitions of \( m \) hyperplanes into subsets of cardinality \( \lambda_i \), while \( \lambda! \) counts the permutations of the subsets of the same cardinality, i.e. the number of ordered partitions determined by \([\Gamma]\). Therefore the remaining factor \( d_\lambda \) for each partition should count the number of polar simplices \( \Gamma \) that intersect the \( n \) linear subspaces \( L_i = \cap_j h_{i j} \subset \mathbf{P}(T_1) \) of codimension \( \lambda_i \), \( i = 1, \ldots, n \).

Let \( [\Gamma] = \left\{ [p_1, \ldots, p_n] \right\} \in VPS(Q, n) \) and assume that \( (p_1, \ldots, p_n) \in L_1 \times \cdots \times L_n \).

For each pair of linear spaces \( L_i, L_j \) the bilinear form associated to the quadratic form \( q \) restricts to a linear form on the product \( L_i \times L_j \) that vanishes on \( (p_i, p_j) \).

This linear form defines a divisor \( H_{ij} \) in the divisor class \( L_i + L_j \), where \( L_i \) is the pullback to the product of the hyperplane class on \( L_i \). If \( D_{ij} \) and \( D_i \) are the pullbacks of \( H_{ij} \), respectively \( H_i \), to the product \( \Pi_i L_i \), then \( \Gamma \subset \Pi_i L_i \) lies in the intersection \( \cap_{i<j} D_{ij} \).

Conversely, consider a point \( (p_1, \ldots, p_n) \in L_1 \times \cdots \times L_n \) that lies in the intersection of the divisors \( \cap_{i<j} D_{ij} \). The projection of this point into \( \mathbf{P}(T_1) \) is a collection of \( n \) points \( \Gamma = \{ p_1, \ldots, p_n \} \). Let \( p_i = [v_i], v_i \in T_1 \), then the bilinear form \( q : T_1 \times T_1 \to \mathbf{C} \), \( q^{-1}(v_i)(q)(v_j) = 0 \) for every \( i \neq j \), so the hyperplanes \( p_i^\perp \subset \mathbf{P}(S_i) \) form a polar simplex to \( Q \). Hence \([\Gamma]\) is a point in \( VPS(Q, n) \).

Thus \( d_\lambda \) counts the number of polar simplices \( \Gamma \) that intersect the \( n \) linear subspaces \( L_i \) and the degree formula follows.

\( \square \)

The Theorem 1.1 in the introduction follows from Corollary 5.12, Corollary 5.15 and Theorem 6.1. Theorem 1.2 follows from Corollary 2.2, Corollary 5.2, Theorem 6.1 and the degree is computed from Theorem 6.3. Theorem 1.3 follows from Theorem 5.4, Corollary 5.16 and Corollary 2.7.

REFERENCES


Variety of Polar Simplices


Kristian Ranestad
Matematisk institutt
Universitetet i Oslo
PO Box 1053
Blindern
NO-0316 Oslo
Norway
ranestad@math.uio.no

Frank-Olaf Schreyer
Mathematik und Informatik
Universität des Saarlandes
Campus E2 4
Blindern
D-66123 Saarbrücken
Germany
schreyer@math.uni-sb.de
Stable Maps and Chow Groups

D. Huybrechts and M. Kemeny

Received: April 26, 2012
Revised: May 15, 2012

Communicated by Thomas Peternell

Abstract. According to the Bloch–Beilinson conjectures, an automorphism of a K3 surface $X$ that acts as the identity on the transcendental lattice should act trivially on $\text{CH}^2(X)$. We discuss this conjecture for symplectic involutions and prove it in one third of all cases. The main point is to use special elliptic K3 surfaces and stable maps to produce covering families of elliptic curves on the generic K3 surface that are invariant under the involution.

2010 Mathematics Subject Classification: 14J28 (14J50)

0.1. Let $X$ be a complex projective K3 surface with an automorphism $f : X \to X$. According to the general philosophy of the Bloch–Beilinson conjectures, the induced action of $f$ on the kernel of the cycle map $\text{CH}^*(X) \to H^*(X, \mathbb{Z})$ should be determined by the action of $f$ on the cokernel of the cycle map. More precisely, one expects the following to be true:

Conjecture 0.1. $f^* = \text{id}$ on $\text{CH}^2(X)_0$ if and only if $f^* = \text{id}$ on $T(X)$.

Here, $\text{CH}^2(X)_0 \subset \text{CH}^2(X)$ is the degree zero part, i.e. the kernel of the cycle map $\text{CH}^2(X) \to H^4(X, \mathbb{Z}) \cong \mathbb{Z}$, and $T(X) \subset H^2(X, \mathbb{Z})$ is the transcendental lattice which can be described as the orthogonal complement of the Néron–Severi group $\text{NS}(X) \subset H^2(X, \mathbb{Z})$. Alternatively, $T(X) \subset H^2(X, \mathbb{Z})$ is the smallest sub-Hodge structure such that $H^{2,0}(X) \subset T(X) \otimes \mathbb{C}$. Thus, $f^* = \text{id}$ on $T(X)$ if and only if $f$ acts trivially on $H^{2,0}(X)$. The latter is spanned by the unique (up to scaling) regular two-form $\sigma \in H^{0,2}(X, \Omega^2_X)$, which we think of as a holomorphic symplectic structure. For this reason, an automorphism $f : X \to X$ with $f^* = \text{id}$ on $T(X)$ is called a symplectomorphism.

It is well known that $f^* = \text{id}$ on $\text{CH}^2(X)_0$ implies that $f$ acts trivially on $T(X)$ (see e.g. [16 Ch. 23]). Appropriately rephrased, this holds for arbitrary

\footnote{This work was supported by the SFB/TR 45 ‘Periods, Moduli Spaces and Arithmetic of Algebraic Varieties’ of the DFG (German Research Foundation). The second author is supported by a PhD-scholarship of the Bonn International Graduate School in Mathematics.}
smooth projective varieties and for arbitrary correspondences. It is the converse of the statement that is difficult and that shall be discussed here for symplectic involutions of K3 surfaces, i.e. automorphisms $f$ of order two with $f^*\sigma = \sigma$.

0.2. K3 surfaces $X$ endowed with a symplectic involution $f : X \cong X$ come in families. As shown by van Geemen and Sarti in [4], the moduli space of such $(X, f)$ has one resp. two connected components in each degree $2 d > 0$ depending on the parity of $d$. To be more precise, let $\Lambda_d$ be the lattice $\mathbb{Z} \ell \oplus E_8(-2)$ with $(\ell, \ell) = 2d$ and denote for $d \equiv 0(2)$ by $\tilde{\Lambda}_d$ the unique even lattice containing $\Lambda_d$ with $\tilde{\Lambda}_d/\Lambda_2 \cong \mathbb{Z}/2\mathbb{Z}$ and such that $E_8(-2) \subset \tilde{\Lambda}_d$ is primitive (see [4, Prop. 2.2]).

Then for generic $(X, f)$ one has $\text{NS}(X) \cong \Lambda_d$ or $\text{NS}(X) \cong \tilde{\Lambda}_d$. The class $\ell$ corresponds under this isomorphism to an ample line bundle $L$ on $X$ which spans the $f$-invariant part of $\text{NS}(X)$. If $(X, f)$ is not generic, then one still finds $\Lambda_d$ or $\tilde{\Lambda}_d$ as a primitive sublattice in $\text{NS}(X)$ with $E_8(-2)$ as the orthogonal complement of the invariant part. Conversely, by the Global Torelli theorem any $X$ parametrized by the (non-empty and in fact 11-dimensional) connected moduli spaces $\mathcal{M}_{\Lambda_d}$ or $\mathcal{M}_{\tilde{\Lambda}_d}$ of $\Lambda_d$ resp. $\tilde{\Lambda}_d$-lattice polarized K3 surfaces comes with a symplectic involution $f$ that is determined by its action $= -\text{id}$ on $E_8(-2)$ and $= \text{id}$ on its orthogonal complement.

In other words, for each $d \equiv 1(2)$ the moduli space of K3 surfaces $X$ with a symplectic involution $f$ and an invariant polarization of degree $2d$ has one connected component $\mathcal{M}_{\Lambda_d}$, whereas for $d \equiv 0(2)$ it has two connected components, $\mathcal{M}_{\Lambda_d}$ and $\mathcal{M}_{\tilde{\Lambda}_d}$. Thus the following theorem, the main result of the present paper, proves Conjecture 0.1 in one third of all possible cases.

**Theorem 0.2.** Let $d \equiv 0(2)$ and $(X, f) \in \mathcal{M}_{\tilde{\Lambda}_d}$. Then $f^* = \text{id}$ on $\text{CH}^2(X)$.

0.3. For $d = 1$ (double covers of $\mathbb{P}^2$) and $d = 2$ (quartics in $\mathbb{P}^3$) the conjecture is known to hold, see [2, 14, 15]. For $d = 3$ (complete intersection of a cubic and a quadric in $\mathbb{P}^4$) an interesting approach is outlined in [9]. Theorem 3.2 in [17] proves the conjecture for equivariant complete intersections in varieties with trivial Chow groups.

In [7] the conjecture has been proven for $(X, f)$ in dense subsets of $\mathcal{M}_{\Lambda_d}$ and $\mathcal{M}_{\tilde{\Lambda}_d}$. The proof there relies on Fourier–Mukai equivalences of the bounded derived category of coherent sheaves on $X$ and it is not clear how to push the techniques further to cover generic and hence arbitrary $(X, f)$. The techniques to prove Theorem 0.2 can be applied to symplectic automorphisms $f : X \cong X$ of order $> 2$. If the order of $f$ is a prime $p$, then $p = 2, 3, 5,$ or 7 (cf. [13]), and the results of [1] have in [3] been successfully generalized to cover also the cases $p = 3, 5,$ and 7. Our methods prove Conjecture 0.1 for many components of the moduli space of $(X, f)$ in these cases too, see Section 5.
Theorem 0.3. For $p = 3, 5, \text{ or } 7$ and $d = ep$, there exists one component of the moduli space of polarized K3 surfaces $(X, L)$ with a symplectic automorphism $f : X \to X$ of order $p$ and $L^2 = 2d$ such that Conjecture 0.1 holds true.

0.4. The proof of Theorem 0.2 neither uses derived categories as in [7] nor any deep cycle arguments as e.g. in [6]. As we shall explain in Section 1, it is enough to find a dominating family of integral genus one curves on $X$ that are invariant under $f$ and avoid the fixed points of $f$. The conjecture is then deduced from the absence of torsion in $\text{CH}^2(X)$. It is not clear whether the existence of such a family should be expected in general, but it will be shown here for generic K3 surfaces parametrized by points in $\mathfrak{M}_{\Lambda}$. This is done in two steps. Firstly, we construct a family of genus one (reducible) curves on a particular elliptic K3 surface for which $f$ is given by translation by a two-torsion section, see Section 3. Then, the theory of stable maps is applied to obtain the desired family for generic $X$.

The missing piece to prove Conjecture 0.1 in full generality, or at least for symplectic involutions, is the lack of special K3 surfaces in $\mathfrak{M}_{\Lambda}$ for which appropriate families of genus one curves can be described explicitly.

Acknowledgments: We thank Richard Thomas for a useful discussion concerning Section 2.

Addendum: The results of this article have meanwhile been improved by C. Voisin and the first author. In [18] Voisin replaces singular elliptic curves in the quotient $X/(f)$ and their étale cover in $X$, as used in this paper, by smooth ample curves of high genus. A dimension estimate for the corresponding Prym variety yields finite dimensionality of the anti-invariant part of the Chow group and eventually $f^* = \text{id}$ on $\text{CH}^2(X)$ for symplectic involutions. In particular, neither degeneration to special elliptic K3 surfaces nor the deformation theory of stable maps, which we believe to be of independent interest, enter her arguments. In [8] we use instead arguments involving the derived category $D^b(\text{Coh}(X))$ and lattice theory as developed in [7] to prove the result for symplectic automorphisms of prime order $p \neq 2$. Both results taken together prove $f^* = \text{id}$ on $\text{CH}^2(X)$ for symplectic automorphisms of arbitrary finite order.

1. Covering families of elliptic curves

Let $f : X \to X$ be a symplectic automorphism of finite order and denote its quotient by $\bar{X} := X/(f)$, which is a singular K3 surface. If $f$ has prime order $p$, then $p = 2, 3, 5, \text{ or } 7$ (see [13]). In order to prove Conjecture 0.1 for symplectic automorphisms of finite order (and we do not have anything to say for automorphisms of infinite order), one can restrict to those. The number of fixed points of $f$, all isolated, can be determined by the Lefschetz fixed point formula. E.g. for a symplectic involution, i.e. $p = 2$, there are exactly eight fixed points.

In the following, a family $\mathcal{C}_t \subset X$ of curves given by $\mathcal{C} \subset S \times X$ is called dominating if the projection $\mathcal{C} \to X$ is dominant, i.e. if the curves $\mathcal{C}_t$ parametrized by the closed point $t \in S$ cover a Zariski open subset of $X$. 
Proposition 1.1. Let \( f : X \rightarrow X \) be a symplectic automorphism. Assume there exists a dominating family of integral \( f \)-invariant curves \( C \subset X \) of geometric genus one with \( C_t \cap \text{Fix}(f) = \emptyset \) for generic \( t \). Then \( f^* = \text{id} \) on \( \text{CH}^2(X) \).

Proof. It suffices to prove that for generic \( x \in X \) the points \( x \) and \( y := f(x) \) are rationally equivalent, i.e. \( [x] = [y] \) in \( \text{CH}^2(X) \). Since by Roitman’s theorem \( \text{CH}^2(X) \) is torsion free (see e.g. [10, Ch. 22]), the latter is equivalent to \( [x] = [y] \) being torsion. For any morphism \( g : C \rightarrow X \) from a smooth irreducible curve \( C \) the induced \( g_* : \text{Pic}(C) = \text{CH}^1(C) \rightarrow \text{CH}^2(X) \) is a group homomorphism. Thus, if there exist lifts \( \tilde{x}, \tilde{y} \in C \) of \( x \) resp. \( y \) such that \( O(\tilde{x} - \tilde{y}) \in \text{Pic}^0(C) \) is a torsion line bundle, then automatically \( [x] = [y] \) in \( \text{CH}^2(X) \).

By assumption, a generic closed point \( x \in X \) lies on one of the curves \( C_t \). Since the curves \( C_t \) are assumed to be \( f \)-invariant, \( y = f(x) \) is contained in the same curve and \( f \) lifts to an automorphism \( \tilde{f} \) of the normalization \( C := \tilde{C}_t \), which is a smooth integral curve of genus one. As \( C_t \) avoids the fixed points of \( f \), the automorphism \( \tilde{f} : C \rightarrow C \) is fixed point free and hence \( D := C/\langle \tilde{f} \rangle \) is also smooth of genus one. After choosing origins for \( C \) and \( D \) appropriately, \( C \rightarrow D \) is a morphism of elliptic curves which can be viewed as a quotient of \( C \) by a finite subgroup \( \Gamma \subset C \cong \text{Pic}^0(C) \). Hence points in the same fibre of \( C \rightarrow D \) differ by elements of \( \Gamma \). In particular, \( O_C(\tilde{x} - \tilde{y}) \in \Gamma \subset \text{Pic}^0(C) \) is a torsion line bundle. \( \square \)

The problem now becomes to construct a family of genus one curves as required. We do not know how to do this directly. On the special elliptic surface considered in Section 3 a family of genus one curves is constructed, but the curves are not integral. They become integral only after deformations to the generic case.

2. Stable maps to K3 surfaces

Let \( X \rightarrow S \) be an irreducible family of K3 surfaces with a global line bundle \( \mathcal{L} \). Consider the moduli stack \( \mathcal{M}_g(X, \mathcal{L}) \rightarrow S \) of stable maps \( h : D \rightarrow X_t \) to fibres of \( X \rightarrow S \) such that \( D \) is of arithmetic genus \( g \) with \( h_* (D) \in |\mathcal{L}| \). As the stack structure of \( \mathcal{M}_g(X, \mathcal{L}) \) is of no importance to us, we shall ignore it and treat \( \mathcal{M}_g(X, \mathcal{L}) \) as a moduli space. If we do not want to fix the linear equivalence class of the image curves, we will simply write \( \mathcal{M}_g(X) \).

The following fact has been used in various contexts in the literature, but mostly for \( g = 0 \) (see e.g. [11, 12]). We shall need the following statement for \( g = 1 \).

Proposition 2.1. Every irreducible component of \( \mathcal{M}_g(X, \mathcal{L}) \) is of dimension at least \( g + \dim(S) \).

Proof. The starting point is [10, Thm. 2.17]: For simplicity let \( \pi : X \rightarrow S \) be a smooth projective family over an irreducible base \( S \) and let \( \mathcal{D} \rightarrow S \) be a flat and projective family of curves. Every irreducible component of \( \text{Mor}_S(D, X) \)
containing a morphism \( h : D := D_0 \to X := X_0 \) is of dimension at least
\[
\chi(D, h^* T_X) + \dim(S).
\]

The first term \( \chi(D, h^* T_X) = h^0(D, h^* T_X) - h^1(D, h^* T_X) \) reflects the usual deformation-obstruction theory for the morphism \( h : D \to X \). A priori, the obstructions to deforming the morphism \( h : D \to X \) are contained in \( H^1(D, h^* T_X) \), which is part of an exact sequence
\[
\ldots \to H^1(D, h^* T_X) \to H^1(D, h^* T_X) \to H^1(D, h^* \pi^* T_S) \to 0.
\]

Since the morphism \( D \to X \subset X \to S \) is constant, there are no obstructions to deforming it sideways at least when \( S \) is smooth. In other words, the obstructions to deforming \( h : D \to X \) are contained in the image of \( H^1(D, h^* T_X) \) which leads to the stronger bound in (1).

A similar argument allows one to treat the case of varying domain \( D \). The usual obstruction theory for stable maps shows that \( M_g(X, L) \) is of dimension at least \( \chi(D, (h^* \Omega_X \to \Omega_D)^*) \), where the two term complex \( h^* \Omega_X \to \Omega_D \) is concentrated in degree \(-1\) and \( 0 \), see \([5]\). For \( X = X_0 \) in a family \( X \to S \), the analogue of (1) then says that \( M_g(X) \) in a point corresponding to a stable map \( h : D \to X \) is of dimension at least
\[
\chi(D, (h^* \Omega_X \to \Omega_D)^*) + \dim(S) = g - 1 + \dim(S).
\]

The last equation follows from a standard Riemann–Roch calculation.

The remaining issue is to increase the bound by restricting to families \( X \to S \) which come with a deformation \( L \) of \( L := O(h_*(D)) \). One can either invoke reduced deformation theory for K3 surfaces as developed recently in \([11]\) or use the following trick.

Any given family \( (X, L) \to S \) with a polarization \( L \) can be thickened to a family \( \tilde{X} \to \tilde{S} \) with \( \dim \tilde{S} = \dim S + 1 \) such that transversally to \( S \subset \tilde{S} \) the line bundle \( L \) is obstructed (even to first order). More precisely, for \( t \in S \) the line bundle \( L_t \) on \( X_t \) deforms to first order in the direction of \( v \in T_{\tilde{S}, t} \), if and only if \( v \in T_{\tilde{S}, t} \subset T_{\tilde{S}, t} \). If \( L \) is fibrewise ample, then the thickening \( \tilde{X} \to \tilde{S} \) can be explicitly described by using the twistor space construction for each fibre \( X_t \) and the Kähler class given by \( c_1(L_t) \). (Note that in particular, \( \tilde{X} \to \tilde{S} \) will in general not be projective.) Otherwise, one uses the standard deformation theory of K3 surfaces to produce such a family at least locally, which is enough for the following dimension count.

By the discussion above, \( M_g(\tilde{X}) \) is in \([h : D \to X]\) of dimension
\[
g - 1 + \dim(\tilde{S}) = g + \dim(S).
\]

On the other hand, \( h : D \to X \) cannot deform sideways in a tangent direction \( v \in T_{\tilde{S}, 0} \) that is not contained in \( T_{\tilde{S}, 0} \), because \( O(h_*(D)) = L_0 \). This shows that the two moduli spaces \( M_g(\tilde{X}) \) and \( M_g(\tilde{X}) \) coincide near the point given by \([h : D \to X]\). \( \square \)
Suppose the fibre $\mathcal{M}_0$ of an irreducible component $\mathcal{M} \subset \mathcal{M}_g(X, \mathcal{L})$ is of dimension $\leq g$ for some $0 \in S$. Then $\mathcal{M}$ dominates $S$. \hfill \square

In other words, if the moduli space $\mathcal{M}_g(X_0, \mathcal{L}_0)$ of stable maps to one fibre $X_0$ has the expected dimension $g$ in $[h : D \rightarrow X_0]$, then $h$ can be deformed to a stable map $h_2 : D_2 \rightarrow X_2$ to the generic fibre. To ensure that the condition is met, we shall later use the following criterion, c.f. [9, Cor. 1.2.5] and [12, Lem. 2.6].

**Proposition 2.3.** Suppose the stable map $h : D \rightarrow X$ satisfies the following conditions:

i) If $D_1, D_2, \ldots, D_n$ are the components of $D$, then $D_2, \ldots, D_n$ are smooth and rational.

ii) The first component $D_1$ is smooth of genus $g$ and $h|_{D_1} : D_1 \rightarrow X$ is an embedding.

iii) The morphism $h$ is unramified.

iv) Two components $D_i$ and $D_j$ intersect transversally in one point if $|i - j| = 1$ and not at all otherwise.

Then $\mathcal{M}_g(X)$ is of dimension $g$ in $[h : D \rightarrow X]$.

**Proof.** We copy the argument from [11, Lem. 2.7]. First of all, since $h$ is unramified, the complex $h^*\Omega_X \rightarrow \Omega_D$ is a locally free sheaf of rank one concentrated in degree $-1$, the dual of which is denoted $N_h$. Then, one proceeds by induction over $n$ and uses the exact sequence

$$0 \rightarrow N_h(-x)|_{D'} \rightarrow N_h \rightarrow N_h|_{D_n} \rightarrow 0,$$

where $D' := D_2 \cup \ldots \cup D_{n-1}$ and $\{x\} = D_{n-1} \cap D_n$. From the exact sequence

$$0 \rightarrow N_h|_{D_n} \rightarrow h^*\Omega_X|_{D_n} \rightarrow \Omega_D|_{D_n} \rightarrow 0$$

and $\Omega_D|_{D_n} \cong \mathcal{O}(-1)$, one deduces $N_h|_{D_n} \cong \mathcal{O}(-1)$. Thus, $H^i(N_h) \cong H^i(N_h(-x)|_{D'})$. On the other hand, $N_h(-x)|_{D'} = N_{h'}$, where $h' := h|_{D'} : D' \rightarrow X$. By induction this eventually yields $H^i(N_h) \cong H^i(N_{D_1/X})$. But clearly, $h^0(N_{D_1/X}) = h^0(D_1, \omega_{D_1}) = g$ and the deformations of $D_i \subset X$ are unobstructed. \hfill \square

**Remark 2.4.** Maybe more geometrically, the arguments show that deformations of $h : D \rightarrow X$ are all given by deforming $D_1 \subset X$.

**3. Special elliptic surfaces**

We follow [11, Sect. 4] for the construction of an elliptic K3 surface $X \rightarrow \mathbb{P}^1$ with a symplectic involution given by a two-torsion section. Deformations of $X$ will lead to K3 surfaces with Néron–Severi group $\tilde{\Lambda}_{2d}$ with $d = 2e > 2$.

The elliptic K3 surface $X \rightarrow \mathbb{P}^1$ is described by an equation of the form

$$y^2 = x(x^2 + a(t)x + b(t))$$

with general $a(t)$ and $b(t)$ of degree 4 resp. 8. Then the fibration has two obvious sections: The section at infinity $\sigma$ given by $x = z = 0$, which will serve...
us as the zero section, and a disjoint section \( \tau \) given by \( x = y = 0 \). Using the explicit equation, one finds that \( \tau \) has order two. Thus, translation by \( \tau \) defines an involution \( f : X \to X \) which is symplectic. Still following [4], one computes the singular fibres of \( X \to \mathbb{P}^1 \): There are eight fibres of type \( I_1 \) (a rational curve with one node) and eight fibres of type \( I_2 \) (the union of two copies of \( \mathbb{P}^1 \) intersecting transversally in two points). They can be found over the zeroes of \( b \in H^0(\mathbb{P}^1, \mathcal{O}(8)) \) resp. \( a^2 - 4b \in H^0(\mathbb{P}^1, \mathcal{O}(8)) \). The fixed points of \( f \) are the nodes of the eight \( I_1 \)-fibres which are all avoided by \( \sigma \) and \( \tau \). Moreover, \( f \) interchanges the two components of each \( I_2 \)-fibre.

The components of the \( I_2 \)-fibres not meeting \( \sigma \) are denoted \( N_1, \ldots, N_8 \). Then \( \hat{N} = (1/2) \sum N_i \in \text{NS}(X) \). Moreover, if \( F \) denotes the class of a generic fibre (and by abuse also a generic fibre itself), then \( \sigma \) and \( F \) span a hyperbolic plane and \( \tau = \sigma + 2F - \hat{N} \). The Néron–Severi group of \( X \) (for general \( a \) and \( b \)) is thus \( \langle \sigma, F \rangle \oplus \langle N_1, \ldots, N_8, \hat{N} \rangle \), which is of rank 10. Next consider a curve of the form \( C = eN + F + \sigma + \tau \), where \( N \) is one of the \( I_2 \)-fibres, and let \( L := \mathcal{O}(C) \). Then \( L \) is big and nef. Indeed, \( (L.L) = 4e > 0 \) and \( C \) intersects all its irreducible components positively, e.g. \( (C.\sigma) = e - 1 > 0 \). In fact, \( L \) is ample as it clearly intersects all horizontal curves positively and has also positive intersection with all \((-2)\)-curves (e.g. the two components of the \( I_2 \)-fibres). Moreover, \( L \) is primitive, as \( (C.N_i) = 1 \). Since \( f \) respects the fibration and interchanges \( \sigma \) and \( \tau \), the curve \( C \) is \( f \)-invariant and disjoint from \( \text{Fix}(f) \).

Let us now consider the quotient \( \tilde{X} := X/\langle f \rangle \) which is a singular K3 surface with eight ordinary double points. Its minimal resolution \( Y \to \tilde{X} \) comes with a natural elliptic fibration \( Y \to \mathbb{P}^1 \). Note that the fibres of type \( I_1 \) and \( I_2 \) are interchanged when passing from \( X \) to \( Y \).
The quotient $\overline{C} := C / \langle f \rangle \subset \overline{X}$ avoids the singular locus of $\overline{X}$ and thus can also be viewed as a curve in $Y$. For the same reason, the line bundle $L$ descends to an ample line bundle $\overline{L}$ on $\overline{Y}$. Note that $\overline{C}$ decomposes as $\overline{C} = e \overline{N} + \overline{F} + \overline{\sigma}$, where $\overline{N}$ is an $I_1$-fibre of $Y \dashrightarrow \mathbb{P}^1$, $\overline{F}$ is a smooth fibre, and $\overline{\sigma}$ is a section.

**Lemma 3.1.** There exists a stable map $h : D \rightarrow \overline{X}$ of arithmetic genus one with image $\overline{C}$ and such that $\mathcal{M}_1(\overline{X}, \overline{L})$ is one-dimensional in $h$.

**Proof.** Since $\overline{C}$ avoids the singularities of $\overline{X}$, we can equally work with $\overline{C} \subset Y$. The curve $D$ shall have components $D_1, D_2, \ldots, D_n$, $n = e + 2$, with $D_1 \rightarrow \overline{F}$, $D_2 \rightarrow \overline{\sigma}$, and $D_i \rightarrow \overline{N}$, $i \geq 2$, being the normalization. The gluing is defined according to the picture (cf. [1, 9, 12]):

Obviously, $D$ is of arithmetic genus one and $h_*(D) = \overline{C}$. Moreover, the assumptions of Proposition 2.3 are satisfied and hence $\mathcal{M}_1(Y)$ is of dimension one in $[h : D \rightarrow Y]$. □

Now consider a generic deformation

\[
(\mathcal{X}, \mathcal{L}) \rightarrow S
\]

of $(X, L, f)$, i.e. $(\mathcal{X}_0, f_0) = (X, f)$ for a distinguished $0 \in S$ and for generic $t \in S$ the fibre $\text{NS}(\mathcal{X}_t)$ has rank $\rho = 9$ with $f_t$-invariant part spanned by $\mathcal{L}_t$. Taking quotients, one obtains a family of singular K3 surfaces $\mathcal{X} \rightarrow S$. Clearly, $L = \mathcal{O}(C)$ descends to the quotient $\overline{X}$, for $C$ is $f$-invariant and avoids the fixed points. Hence the line bundle $\mathcal{L}$ also descends to a relatively ample line bundle $\overline{\mathcal{L}}$. (The obstructions to deforming $L$ resp. $\mathcal{L}$ sideways are the same.) Note that for generic $t \in S$ the line bundle $\overline{\mathcal{L}}_t$ generates $\text{Pic}(\mathcal{X}_t)$.

Let us apply the discussion of Section 2 to $h : D \rightarrow \mathcal{X}_0 = \overline{X}$. So, consider the relative moduli space of stable maps of genus one $\mathcal{M}_1(\mathcal{X}, \mathcal{L}) \rightarrow S$.  

**Documenta Mathematica 18 (2013) 507–517**
Corollary 3.2. The stable map $h : D \to X$ thus constructed deforms sideways to stable maps $h_t : D_t \to \mathcal{X}_t$. Moreover, for generic $t \in S$ the curve $h_{t*}(D_t) \subset \mathcal{X}_t$ and its preimage in $\mathcal{X}_t$ are integral and disjoint from the singular locus resp. the fixed point set of $f_t$.

Proof. The existence of the deformation to the nearby fibres follows directly from Proposition 2.3 and Corollary 2.2. Since $C = h_*(D)$ avoids the singularities of $X$, this will hold for generic $t$. Clearly, $h_{t*}(D_t) \in |\mathcal{L}_t|$. Therefore, since $\mathcal{L}_t$ generates the invariant part of $\text{NS}(\mathcal{X}_t)$ and hence $\mathcal{L}_t$ generates NS$(\mathcal{X}_t)$, the curve $h_{t*}(D_t)$ must be integral. Suppose the preimage $C_t$ of $h_{t*}(D_t)$ were not integral for $t$ generic, i.e. $C_t = C_t' + C_t''$ with $f_t(C_t') = C_t''$. (Use that $f_t$ is an involution.) The two components would then specialize to $C'$ resp. $C''$ on $X$ with $C = C' + C''$ and $f(C') = C''$. We may assume that $F \subset C'$. But then also $F = f(F) \subset f(C') = C''$ which eventually yields the contradiction that $F$ appears with multiplicity at least two in $C$.

Remark 3.3. In fact, since the stable map $h : D \to X$ deforms with the fibre component $F$ in a one-dimensional family, also the deformations $h_t : D_t \to \mathcal{X}_t$ come in a family dominating $\mathcal{X}_t$. Thus, one obtains a dominating family of integral genus one curves in the generic deformation $\mathcal{X}_t$ that are $f_t$-invariant and avoid the fixed points of $f_t$.

4. Proof of the main theorem

The outcome of the above construction are generic K3 surfaces $\mathcal{X}_t \in \mathcal{M}_d$ with a symplectic involution $f_t$ such that $\mathcal{X}_t = \mathcal{X}_t/(f_t)$ contains a one-dimensional family of integral curves of geometric genus one that avoid the singular locus. This immediately leads to a proof of our main result.

Theorem 4.1. For all $(X, f) \in \mathcal{M}_d$, the symplectic involution $f : X \to X$ acts as id on $\text{CH}^2(X)$.

Proof. The case $d = 2$ follows from [14]. So we assume $d = 2e > 2$, i.e. $e > 1$. We first show that the above discussion combined with Proposition 1.1 proves the assertion for generic $(X, f) \in \mathcal{M}_d$. Consider a deformation $\mathcal{X}_t$ of the special elliptic K3 surface $\mathcal{X}$. Then for generic $t \in S$ one has $\text{NS}(\mathcal{X}_t) = \tilde{\Lambda}_d$. Indeed, by [4] Prop. 2.7 only in this case do all $f_t$-invariant line bundles actually descend to the quotient $\tilde{\mathcal{X}}_t$. Hence the elliptic K3 surfaces described by (3) can be connected to the generic K3 surface parameterized by $\mathcal{M}_d$. Here we use that $\mathcal{M}_d$ is connected.

The generic fibre of the family $\mathcal{X}_t$ satisfies the assumption of Proposition 1.1. Indeed, by Corollary 3.2 and Remark 3.3 there exists a dominating family of integral curves of arithmetic genus one on the generic fibre $\mathcal{X}_t$ that are invariant under the involution and avoid the fixed points.

Now consider an arbitrary $(X, f) \in \mathcal{M}_d$. Then any $x \in X$ can be viewed as a specialization of points $x_t$ in generic deformations $(\mathcal{X}_t, f_t) \in \mathcal{M}_d$. Clearly, the
points $f_t(x_t)$ then specialize to $f(x)$. For generic $X_t$ we have already shown $[x_t] = [f_t(x_t)]$ in $\text{CH}^2(X_t)$ and thus specialization yields $[x] = [f(x)]$ in $\text{CH}^2(X)$ for all $x \in X$. □

5. Further comments

We briefly outline how to adapt our techniques to the case of symplectic automorphisms of prime order. For $p = 3, 5,$ and $7$, Garbagnati and Sarti describe in [3] Thm. 4.1 lattices $\Omega_p$ of rank 12, 16, resp. 18 that are isomorphic to the anti-invariant part of $f^*$ acting on $H^2(X, \mathbb{Z})$. Similar to the case $p = 2$, the generic polarized K3 surface $(X, L)$ of degree 2$d$ with a symplectic automorphism $f : X \sim X$ of order $p$ leaving $L$ fixed has Picard group isomorphic to $\Lambda_{p,d} := \mathbb{Z}L \oplus \Omega_p$ or possibly, if $d \equiv 0(p)$, isomorphic to a lattice $\tilde{\Lambda}_{p,d}$ that contains $\Lambda_{p,d}$ as a primitive sublattice of index $p$. In fact, the case $\Lambda_{7,d}$ is not realized if $d \equiv 0(7)$ (cf. [3] Prop. 5.2), but unfortunately it is not known whether the lattices $\tilde{\Lambda}_{p,d}$ are unique for given $p$ and $d \equiv 0(p)$ (see [3] Sec. 6). The moduli spaces are of dimension 7, 3, resp. 1.

Examples of symplectic automorphisms of order 3, 5, and 7 have been described in [3] Sec. 3.1. They are again given by translation by a torsion section. The Picard numbers in these examples are 14, 18, resp. 20 and in each case they correspond to points in (at least) one component of the moduli space of polarized K3 surfaces $(X, L)$ with a symplectic automorphism $f$ of degree $L^2 = 2d$. This leads to the following result:

**Theorem 5.1**. For $p = 3, 5,$ or $7$ and $d \equiv 0(p)$, there exists one component of the moduli space of polarized K3 surfaces $(X, L)$ with a symplectic automorphism $f : X \sim X$ of order $p$ and $L^2 = 2d$ such that Conjecture [11] holds true. □

It is very likely that for $p = 7$ and $d \equiv 0(7)$ the result can be strengthened to cover all K3 surfaces, as we would expect that $\tilde{\Lambda}_{7,d}$ is in fact unique.

References


*Documenta Mathematica* 18 (2013) 507–517
Stable Maps and Chow Groups


Daniel Huybrechts  Michael Kemeny
Mathematisches Institut  Mathematisches Institut
Universität Bonn  Universität Bonn
Endenicher Allee 60  Endenicher Allee 60
53115 Bonn  53115 Bonn
Germany  Germany
huybrech@math.uni-bonn.de  michael.kemeny@gmail.com
Integration of Vector Fields on Smooth and Holomorphic Supermanifolds

Stéphane Garnier and Tilmann Wurzbacher

Received: September 19, 2012
Revised: December 12, 2012
Communicated by Christian Bär

Abstract. We give a new and self-contained proof of the existence and unicity of the flow for an arbitrary (not necessarily homogeneous) smooth vector field on a real supermanifold, and extend these results to the case of holomorphic vector fields on complex supermanifolds. Furthermore we discuss local actions associated to super vector fields, and give several examples and applications, as, e.g., the construction of an exponential morphism for an arbitrary finite-dimensional Lie supergroup.

2010 Mathematics Subject Classification: Primary 58A50, 37C10; Secondary 57S20, 32C11
Keywords and Phrases: Supermanifolds; Vector fields; Flows; Group actions

1 Introduction

The natural problem of integrating vector fields to obtain appropriate “flow maps” on supermanifolds is considered in many articles and monographs (compare, e.g., \cite{17}, \cite{2}, \cite{19}, \cite{14} and \cite{3}) but a “general answer” was to our knowledge only given in the work of J. Monterde and co-workers (see \cite{12} and \cite{13}). Let us consider a supermanifold $\mathcal{M} = (\mathcal{M}, \mathcal{O}_\mathcal{M})$ together with a vector field $X$ in $\mathcal{T}_\mathcal{M}(\mathcal{M})$, and an initial condition $\phi$ in $\text{Mor}(\mathcal{S}, \mathcal{M})$, where $\mathcal{S} = (\mathcal{S}, \mathcal{O}_\mathcal{S})$ is an arbitrary supermanifold and $\text{Mor}(\mathcal{S}, \mathcal{M})$ denotes the set of morphisms from $\mathcal{S}$ to $\mathcal{M}$. The case of classical, ungraded, manifolds leads one to consider the following question: does there exist a “flow map” $F$ defined on an open sub supermanifold $\mathcal{V} \subset \mathbb{R}^{1|1} \times \mathcal{S}$ and having values in $\mathcal{M}$ and an appropriate
derivation on $\mathbb{R}^{1|1}$, $D = \partial_t + \partial_\tau + \tau(a\partial_t + b\partial_\tau)$, where $\partial_t = \frac{\partial}{\partial t}$, $\partial_\tau = \frac{\partial}{\partial \tau}$ and $a$, $b$ are real numbers, such that the following equations are fulfilled

$$D \circ F^* = F^* \circ X$$
$$F \circ \text{inj}_{V(0) \times S}^V = \phi.$$  \hspace{1cm} (1)

Of course, $V$ should be a “flow domain”, i.e. an open sub supermanifold of $\mathbb{R}^{1|1} \times S$ such that $\{0\} \times S$ is contained in the body $V$ of $\mathcal{V}$ and for $x$ in $S$, the set $I_x \subset \mathbb{R}$ defined by $I_x \times \{x\} = (\mathbb{R} \times \{x\}) \cap V$ is an open interval. Furthermore $\text{inj}_{V(0) \times S}^V$ denotes the natural injection morphism of the closed sub supermanifold $\{0\} \times S$ of $\mathcal{V}$ into $\mathcal{V}$. Of course, we could concentrate on the case $S = M$ and $\phi = \text{id}_M$, but it will be useful for our later arguments to state all results in this (formally) more general setting.

Though for homogeneous vector fields ($X = X_0$ or $X = X_1$) system (11) does always have a solution, in the general case ($X = X_0 + X_1$ with $X_0 \neq 0$ and $X_1 \neq 0$) the system is overdetermined. A simple example of an inhomogeneous vector field such that (11) is not solvable is given by $X = X_0 + X_1 = \left(\frac{\partial}{\partial x} + \xi \frac{\partial}{\partial \xi}\right) + \left(\frac{\partial}{\partial \xi} + \xi \frac{\partial}{\partial x}\right)$ on $\mathcal{M} = \mathbb{R}^{1|1}$. The crucial novelty of [13] is to consider instead of (11) the following modified, weakened, problem

$$\left(\text{inj}_{\mathbb{R}^{1|1}}^\mathbb{R}\right)^* \circ D \circ F^* = \left(\text{inj}_{\mathbb{R}^{1|1}}^\mathbb{R}\right)^* \circ F^* \circ X$$
$$F \circ \text{inj}_{V(0) \times S}^V = \phi,$$  \hspace{1cm} (2)

where $\text{inj}_{\mathbb{R}^{1|1}}^\mathbb{R} = \text{inj}_{\mathbb{R}^{1|1} \times S}^\mathbb{R}$ is again the natural injection (and where the above more general derivation $D$ could be replaced by $\partial_t + \partial_\tau$ since $(\text{inj}_{\mathbb{R}^{1|1}}^\mathbb{R})^*$ annihilates germs of superfunctions of the type $\tau \cdot f, f \in \mathcal{O}_{\mathbb{R} \times S}$).

In [13] (making indispensable use of [12]) it is shown that in the smooth case (2) has a unique maximal solution $F$, defined on the flow domain $\mathcal{V} = (V, \mathcal{O}_{\mathbb{R}^{1|1} \times S|V})$, where $V \subset \mathbb{R} \times S$ is the maximal flow domain for the flow of the reduced vector field $\tilde{X} = \tilde{X}_0$ on $\mathcal{M}$ with initial condition $\tilde{\phi}$. Since the results of [12] are obtained by the use of a Batchelor model for $\mathcal{M}$, i.e. a real vector bundle $E \to M$ such that $\mathcal{M} \cong (M, \Gamma^\infty_{\Lambda E})$, and a connection on $E$, we follow here another road, closer to the classical, ungraded, case and also applicable in the case of complex supermanifolds and holomorphic vector fields.

Our new method of integrating smooth vector fields on a supermanifold in Section 2 consists in first locally solving a finite hierarchy of ordinary differential equations, and is here partly inspired by the approach of [3], where the case of homogeneous super vector fields on compact supermanifolds is treated. We then show existence and unicity of solutions of (2) on smooth supermanifolds and easily deduce the results of [13] from our Lemmata 2.1 and 2.2.
A second beautiful result of [13] (more precisely, Theorem 3.6 of that reference) concerns the question if the flow $F$ solving (2) fulfills “flow equations”, as in the ungraded case. Hereby, we mean the existence of a Lie supergroup structure on $\mathbb{R}^{1|1}$ such that $F$ is a local action of $\mathbb{R}^{1|1}$ on $\mathcal{M}$ (in case $S = \mathcal{M}, \phi = id_M$).

Again, the answer is a little bit unexpected: in general, given $X$ and its flow $F : \mathbb{R}^{1|1} \times \mathcal{M} \supset \mathcal{Y} \to \mathcal{M}$, there is no Lie supergroup structure on $\mathbb{R}^{1|1}$ such that $F$ is a local $\mathbb{R}^{1|1}$-action (with regard to this structure). The condition for the existence of such a structure on $\mathbb{R}^{1|1}$ is equivalent to the condition that (2) holds without the post-composition with $(\text{inj}_{\mathbb{R}^{1|1}})^*$, i.e. the overdetermined system (1) is solvable. Furthermore, both conditions cited are equivalent to the condition that $RX_0 \oplus RX_1$ is a sub Lie superalgebra of $\mathcal{T}_M(M)$, the Lie superalgebra of all vector fields on $\mathcal{M}$.

After discussing Lie supergroup structures and right invariant vector fields on $\mathbb{R}^{1|1}$, as well as local Lie group actions in the category of supermanifolds in general, we show in Section 3 the equivalence of the above three conditions, already given in [13]. We include our proof here notably in order to be able to apply it in the holomorphic case in Section 5 (see below) by simply indicating how to adapt it to this context. Let us nevertheless observe that our result is slightly more general since we do not need to ask for any normalization of the supercommutators between $X_1$ and $X_0$ resp. $X_1$, thus giving the criterion some extra flexibility in applications.

In Section 4, we give several examples of vector fields on supermanifolds, homogeneous and inhomogeneous, and explain their integration to flows. Notably, we construct an exponential morphism for an arbitrary finite-dimensional Lie supergroup, via a canonically defined vector field and its flow. We comment here also on the integration of what are usually called “(infinitesimal) supersymmetries” in physics, i.e., purely odd vector fields having non-vanishing self-commutators.

Finally, in Section 5 we adapt our method to obtain flows of vector fields (compare Section 2 and notably Lemma 2.1) to the case of holomorphic vector fields on holomorphic supermanifolds. To avoid monodromy problems one has, of course, to take care of the topology of the flow domains, and maximal flow domains are -as already in the ungraded holomorphic case- no more unique. Otherwise the analogues of all results in Section 2 and 3 continue to hold in the holomorphic setting.

Throughout the whole article we will work in the ringed space-approach to supermanifolds (see, e.g., [9], [10], [11] and [13] for detailed accounts of this approach). Given two supermanifolds $\mathcal{M} = (M, \mathcal{O}_M)$ and $\mathcal{N} = (N, \mathcal{O}_N)$, a “morphism” $\phi = (\tilde{\phi}, \phi^*) : \mathcal{M} \to \mathcal{N}$ is thus given by a continuous map $\tilde{\phi} : M \to N$ between the “bodies” of the two supermanifolds and a sheaf
homomorphism \( \phi^* : \mathcal{O}_N \to \tilde{\phi} \mathcal{O}_M \). The topological space \( M \) comes canonically with a sheaf \( C_M^\infty = \mathcal{O}_M / J \), where \( J \) is the ideal sheaf generated by the germs of odd superfunctions, such that \((M, C_M^\infty)\) is a smooth real manifold. Then \( \tilde{\phi} \) is a smooth map from \((M, C_M^\infty)\) to \((N, C_N^\infty)\). Let us recall that a (super) vector field on \( M = (M, \mathcal{O}_M) \) is, by definition, an element of the Lie superalgebra \( \mathcal{T}_M(M) = (\text{Der}(\mathcal{O}_M))(M) \) and that \( X \) always induces a smooth vector field \( \tilde{X} \) on \((M, C_M^\infty)\). For \( p \in M \) and \( f + J_p \in (C_M^\infty)_p = (\mathcal{O}_M / J)_p \), one defines \( \tilde{X}_p(f + J_p) = X_0(f)(p) \), where \( X_0 \) is the even part of \( X \) and for \( g \in (\mathcal{O}_M)_p \), \( g(p) \in \mathbb{R} \) is the value of \( g \) in the point \( p \) of \( M \).

2 Flow of a vector field on a real supermanifold

In this section we give our main result on the integration of general (i.e. not necessarily homogeneous) vector fields by a new method, avoiding auxiliary choices of Batchelor models and connections, as in [12]. Our more direct approach is inspired, e.g., by [3], where the case of homogeneous vector fields on compact manifolds is treated, and it can be adapted to the holomorphic case (see Section 4).

For the sake of readability we will often use the following shorthand: if \( P \) is a supermanifold, we write \( \text{inj}^{R|1}_R \) for \( \text{inj}^{R|1}_R \times P \). Furthermore, the canonical coordinates of \( R|1 \) will be denoted by \( t \) and \( \tau \), with ensuing vector fields \( \partial_t = \frac{\partial}{\partial t} \) and \( \partial_\tau = \frac{\partial}{\partial \tau} \).

**Lemma 2.1.** Let \( U \subset \mathbb{R}^{m|n} \) and \( W \subset \mathbb{R}^{p|q} \) be superdomains, \( X \in \mathcal{T}_W(W) \) be a super vector field on \( W \) (not necessarily homogeneous) and \( \phi \in \text{Mor}(U, W) \), and \( t_0 \in \mathbb{R} \). Let furthermore \( H : V \to W \) be the maximal flow of \( \tilde{X} \in \mathcal{X}(W) \), i.e. \( \partial_t \circ H^* = H^* \circ \tilde{X} \), subject to the initial condition \( H(t_0, \cdot) = \tilde{\phi} : U \to W \). Let now \( \mathcal{V} \) be \( (V, \mathcal{O}_{U \times V}) \) and \( (t, \tau) \) the canonical coordinates on \( R|1 \), then there exists a unique \( \tilde{F} : V \to W \) such that

\[
\left( \text{inj}^{R|1}_R \right)^* \circ \left( \partial_t + \partial_\tau \right) \circ F^* = \left( \text{inj}^{R|1}_R \right)^* \circ F^* \circ X \quad \text{and} \quad (3)
\]

\[
F \circ \text{inj}^{V}_{\{t_0\} \times U} = \phi. \quad (4)
\]

Moreover, \( \tilde{F} : V \to W \) equals the underlying classical flow map \( H \) of the vector field \( \tilde{X} \) with initial condition \( \phi \).

**Proof.** Let \((u_i) = (x_i, \xi_i)\) and \((w_j) = (y_j, \eta_j)\) denote the canonical coordinates on \( \mathbb{R}^{m|n} \) and \( \mathbb{R}^{p|q} \), respectively. Then there exist smooth functions \( a^j_i \in C^\infty_N(W) \) such that

\[
X = \sum_{j=1}^{p+q} \left( \sum_{j} a^j_i(y) \eta^j \right) \partial_{w_j},
\]

**Documenta Mathematica** 18 (2013) 519–545
where $J = (\beta_1, \ldots, \beta_q)$ runs over the index set $\{0, 1\}^q$ and $\eta^I = \prod_{s=1}^q \eta^\beta_s$. We then have, of course,

$$X_0 = \sum_j \left( \sum_{|J|=|w_j|} a_j^I(y) \eta^I \right) \partial_{w_j} \text{ resp. } X_1 = \sum_j \left( \sum_{|J|=|w_j|+1} a_j^I(y) \eta^I \right) \partial_{w_j}.$$ 

Here, $|J|$ equals $\beta_1 + \cdots + \beta_q \mod 2$ and $|w_j|$ is the parity of the coordinate function $w_j$. The morphism $F$ determines and is uniquely determined by functions $f^I_j, g^I_j \in \mathcal{C}^{\infty}_{R\times R^n}(V)$ fulfilling for each $j \in \{1, \ldots, p+q\}$

$$F^*(w_j) = \sum_{|I|=|w_j|} f^I_j(t, x) \xi^I + \sum_{|I|=|w_j|+1} g^I_j(t, x) \tau \xi^I$$

(and $f^I_j = 0$ if $|I| \neq |w_j|$, $g^I_j = 0$ if $|I| \neq |w_j| + 1$) as is well-known from the standard theory of supermanifolds (compare, e.g., Thm. 4.3.1 in [20]). Here and in the sequel $I = (\alpha_1, \ldots, \alpha_n)$ is an element of the set $\{0, 1\}^n$ and $\xi^I$ stands for the product $\xi^{I_1} \cdot \xi^{I_2} \cdots \xi^{I_n}$. The notation $|I|$ again denotes the parity of $I$, i.e. $|I| = \alpha_1 + \cdots + \alpha_n \mod 2$.

Equation (8) is equivalent to the following equations:

$$\left(\text{inj}_{R}^{R^{[1]}_\mathbb{R}}\right)^* \circ \partial_t \circ F^* = \left(\text{inj}_{R}^{R^{[1]}_\mathbb{R}}\right)^* \circ F^* \circ X_0 \quad (5)$$

$$\left(\text{inj}_{R}^{R^{[1]}_\mathbb{R}}\right)^* \circ \partial_x \circ F^* = \left(\text{inj}_{R}^{R^{[1]}_\mathbb{R}}\right)^* \circ F^* \circ X_1 \quad (6)$$

Applying (3) to the canonical coordinate functions on $W$, we get the following system, which is equivalent to (5):

$$\sum_{|I|=|w_j|} \partial_t f^I_j : \xi^I = \sum_{|J|=|w_j|} \tilde{F}^* (a^I_j) \tilde{F}^* (\eta^J) \text{ for all } j \in \{1, \ldots, p+q\}, \quad (7)$$

and (6) is equivalent to

$$\sum_{|I|=|w_j|+1} g^I_j : \xi^I = \sum_{|J|=|w_j|+1} \tilde{F}^* (a^I_j) \tilde{F}^* (\eta^J) \text{ for all } j \in \{1, \ldots, p+q\}, \quad (8)$$

where $\tilde{F} := F \circ \text{inj}_{R}^{R^{[1]}_\mathbb{R}} : \mathcal{V} := (V, \mathcal{O}_{R\times \ell}(V)) \rightarrow W$. Let us immediately observe that the underlying smooth map of $\tilde{F}$ equals $\tilde{F}$, the smooth map underlying the morphism $F$.

Moreover the initial condition (9) is equivalent to

$$\sum_{|I|=|w_j|} f^I_j(t_0, x) \xi^I = \phi^* (w_j) \text{ for all } j \in \{1, \ldots, p+q\}. \quad (9)$$

We are going to show that (8) and (9) uniquely determine the functions $f^I_j$ on $V$, i.e. the morphism $F$. Then the functions $g^I_j$ are unambiguously given by
Let us first describe the coefficients for the sequel of this proof. For fixed $I$, $\sum_{\mu=1}^p (\partial_{y_{\mu}} a^j_I)(\hat{F}(t,x)) \cdot f^\mu_K + R \left( a^j_I, f^\nu_I, \deg(I) < \deg(K) \right)$ (13)

if $\deg(K) > 0$.

Here for $I = (\alpha_1, \ldots, \alpha_n)$, $\deg(I) = \alpha_1 + \cdots + \alpha_n$, and -more importantly- $R = R_{j,I,K}$ is a polynomial function in $a^j_I$ and its derivatives in the $y$-variables up to order $q$ included, and in the functions $\{f^\nu_I \mid 1 \leq \nu \leq p+q, 0 \leq \deg(I) < \deg(K)\}$.

Equation (12) is obvious since $a^j_I$ is an even function, whereas equation (13) can be deduced from standard analysis on superdomains. More precisely, let $\hat{a}$ be a smooth function on $\mathbb{R}^p$ and $\hat{\psi} : \mathbb{R}^{n+1}\mathbb{R}^q \to \mathbb{R}^{p+q}$ a morphism (of course to be applied to $a = a^j_I, \hat{\psi} = \hat{F}$). Then we can develop $\hat{\psi}^*(a)$ as follows (compare the proof of Theorem 4.3.1 in [20]):

$$\hat{\psi}^*(a) = \sum_{\gamma} \frac{1}{\gamma!} (\partial_{\gamma_1} a)(\hat{\psi}^*(y_1), \ldots, \hat{\psi}^*(y_p)) \cdot \prod_{\mu=1}^p (\hat{\psi}^*(y_\mu) - \hat{\psi}^*(y_\mu))^{\gamma_\mu}$$

$$= a(\hat{\psi}(t,x)) + \sum_{\mu=1}^p (\partial_{y_{\mu}} a)(\hat{\psi}(t,x)) \cdot \left( \sum_{\nu \neq 0} f^\nu_M \cdot \xi^\nu \right) + \sum_{\mu=1}^P (\partial_{\gamma_1} a)(\hat{\psi}(t,x)) \cdot \prod_{\nu=1}^p (\hat{\psi}^*(y_\nu) - \hat{\psi}^*(y_\nu))^{\gamma_\nu}$$

For fixed $j$ this is an equation of Grassmann algebra-valued maps in the variables $t$ and $x$ that can be split in a system of scalar equations as follows. For $K = (\alpha_1, \ldots, \alpha_n) \in \{0,1\}^n$, we will denote the coefficient $h_K$ in front of $\xi^K$ of a superfunction $h = \sum_M h_M(t,x)\xi^M \in \mathcal{O}_{\mathbb{R}^{n+1}}$ compactly by $(h|\xi^K)$ in the sequel of this proof.

Let us first describe the coefficients for $\tilde{F}^*(a^j_I)$ in (11):

$$\tilde{F}^*(a^j_I)|\xi^K) = 0 \text{ if } |K| = 1$$

and, if $|K| = 0$,

$$(\tilde{F}^*(a^j_I)|\xi^K) = a^j_I \circ \tilde{F} \text{ if } K = (0, \ldots, 0),$$

and

$$\tilde{F}^*(a^j_I)|\xi^K) = \sum_{\mu=1}^P (\partial_{y_{\mu}} a^j_I)(\tilde{F}(t,x)) \cdot f^\mu_K + R \left( a^j_I, f^\nu_I, \deg(I) < \deg(K) \right) \text{ if } \deg(K) > 0.$$
\[ \frac{1}{2} \sum_{\mu',\mu''=1}^{p} (\partial_{\nu_{\mu}} \partial_{\nu_{\mu'}} a)(\tilde{\psi}(t,x)) \cdot \left( \sum_{M' \neq 0} f_{M'}^\mu \cdot \xi^{M'} \right) \cdot \left( \sum_{M'' \neq 0} f_{M''}^\mu \cdot \xi^{M''} \right) + \ldots, \]

where \( \sum_{M \neq 0} f_{M}^\mu \cdot \xi^{M} = \psi^s(y_{\mu}) - \tilde{\psi}^s(y_{\mu}) \) with \( f_{M}^\mu \) depending on \( t \) and \( x \). We observe that the last RHS is a finite sum since we work in the framework of finite-dimensional supermanifolds.

In order to get a contribution to \((\psi^s(a))^{\xi^K}\) we can either extract \( f_{K}^\mu \) from the “linear term” or from products coming from the higher order terms in the above development. Thus

\[
(\psi^s(a))^{\xi^K} = \sum_{\mu=1}^{p} (\partial_{\nu_{\mu}} a)(\tilde{\psi}(t,x)) \cdot f_{K}^\mu + R(a, (f_{J}^\mu)_{K, \deg(J) < \deg(K)}),
\]

where \( R \) is a polynomial as described after Equation (13).

Furthermore, for an element \( J = (\beta_{1}, \ldots, \beta_{q}) \) with \( |J| = 0 \) we have for \( \deg(K) > 0 \)

\[
\left( \prod_{s=1}^{q} \left( \sum_{|L|=1} f_{L}^{p+s} \xi^{L} \right)^{\beta_{s}} \right) \xi^{K} = R \left( (f_{J}^l)_{j, \deg(l) < \deg(K)} \right). \tag{14}
\]

And for an element \( J = (\beta_{1}, \ldots, \beta_{q}) \) with \( |J| = 1 \) we get for \( \deg(K) > 0 \)

\[
\left( \prod_{s=1}^{q} \left( \sum_{|L|=1} f_{L}^{p+s} \xi^{L} \right)^{\beta_{s}} \right) \xi^{K} = \begin{cases} f_{K}^{p+l} + R \left( (f_{J}^l)_{j, \deg(l) < \deg(K)} \right) & \text{if } \deg(J) = 1 \text{ and } l \in \{1, \ldots, q\} \\ \text{such that } \beta_{s} = \delta_{s,t} & \forall s, \end{cases} \tag{15}
\]

Obviously, the coefficient of \( \xi^{K} \) of the LHS of Equation (2) is given by

\[
\left( \sum_{|I|=|w_{j}|} \partial_{h} f_{I}^h \cdot \xi^{l} \right) \xi^{K} = \begin{cases} \partial_{h} f_{K}^h \xi^{l} & \text{if } |K| = |w_{j}| \\ 0 & \text{if } |K| = |w_{j}| + 1 \end{cases} \quad \text{for } 1 \leq j \leq p + q.
\]

Taking into account the above descriptions of the \( \xi^{K} \)-coefficients, we will show the existence (and uniqueness) of the solution functions \( \{f_{J}^l|1 \leq j \leq p + q, I \in \{0,1\}^n\} \) for \( (t,x) \in V \) by induction on \( \deg(I) \) and upon observing that all ordinary differential equations occuring are (inhomogeneous) linear equations for the unknown functions.
Let us start with \( \deg(I) = 0 \) that is \( I = (0, \ldots, 0) \). The “0-level” of the equations (11) and (9) is \( \partial_t f^I_j(0, \ldots, 0) = a^j(0, \ldots, 0) \circ \tilde{F} \) and \( f^I_j(0, \ldots, 0)(t_0, x) = y_j \circ \tilde{\phi}(x) \) for all \( j \) such that \( |w_j| = 0 \). We remark that \( f^I_j(0, \ldots, 0) \) is simply \( y_j \circ \tilde{F} \) and \( a^j(0, \ldots, 0) \) is \( \tilde{X}(y_j) \). Thus \( \tilde{F} \) is the flow of \( \tilde{X} \) with initial condition \( \tilde{\phi} \) at \( t = t_0 \), i.e., \( \tilde{F} = H \) on \( V \). Thus the claim is true for \( I = (0, \ldots, 0) \).

Suppose \( k > 0 \) and that the functions \( f^j_I \) are uniquely defined on \( V \) for all \( j \) and all \( I \) such that \( \deg(I) < k \). Let \( K \) be such that \( \deg(K) = k \). Let us distinguish the two possible parities of \( k \) in order to determine \( f^j_K \) for all \( j \).

Recall that \( f^j_K = 0 \) if the parities of \( K \) and \( j \) are different.

If \( k \) is even, i.e., \( |K| = 0 \), we only have to consider \( j \) such that \( |w_j| = 0 \). Putting (12) and (11) together, we find in this case

\[
\partial_t f^I_K = \left( \sum_{|J|=0} \tilde{F}^*(a^j_J) \prod_{s=1}^{|J|} \left( \sum_{|L| = 1} f^{I+s}_L \xi^L \right)^{\beta_s} \right) \xi^K \\
= \left( \sum_{\deg(J) = 0} \tilde{F}^*(a^j_J) \prod_{s=1}^{|J|} \left( \sum_{|L| = 1} f^{I+s}_L \xi^L \right)^{\beta_s} \right) + \sum_{\deg(J) > 0} \tilde{F}^*(a^j_J) \prod_{s=1}^{|J|} \left( \sum_{|L| = 1} f^{I+s}_L \xi^L \right)^{\beta_s} \xi^K \\
= \left( \tilde{F}^*(a^j(0, \ldots, 0)) + \sum_{\deg(J) > 0} \tilde{F}^*(a^j_J) \prod_{s=1}^{|J|} \left( \sum_{|L| = 1} f^{I+s}_L \xi^L \right)^{\beta_s} \right) + \sum_{\mu=1}^p \left( \partial_{y_\mu} a^j(0, \ldots, 0) \circ \tilde{F} \right) f^j_K + R \left( (a^j_J)_J, (f^j_L)_L, \deg(J) < \deg(K) \right).
\]

Moreover, the initial condition gives \( f^j_K(t_0, x) = (\phi^*(y_j))(\xi^K) \), for all \( j \) in \( \{1, \ldots, p\} \). Since the \( a^j_J \) are the (given) coefficients of the vector field \( X \) and the functions \( f^j_I \) with \( \deg(I) < k \) are known by the induction hypothesis, we have a unique local solution function \( f^j_K \). Since the ordinary differential equation for \( f^j_K \) is linear its solution is already defined for all \( (t, x) \in V \). Thus in the case that \( k \) is even \( f^j_K \) is unambiguously defined on \( V \) for all \( j \in \{1, \ldots, p + q\} \) and for all \( K \) with \( \deg(K) = k \).
Now, if \( k \) is odd, i.e., \(|K| = 1\), we only have to consider \( j \) such that \(|w_j| = 1\). Using (13) and (15), we find in this case:

\[
\partial_t f_K^j = \left( \sum_{|J|=1} F^*(a_j^I) \prod_{s=1}^q \left( \sum_{|L|=1} f_L^{p+s} \xi \right)^{\beta_s} \left| \xi^K \right) \right)
\]

\[
= \left( \sum_{\text{deg}(J)=1} F^*(a_j^I) \prod_{s=1}^q \left( \sum_{|L|=1} f_L^{p+s} \xi \right)^{\beta_s} \left| \xi^K \right) \right)
\]

\[
+ \sum_{\text{deg}(J)>1} F^*(a_j^I) \prod_{s=1}^q \left( \sum_{|L|=1} f_L^{p+s} \xi \right)^{\beta_s} \left| \xi^K \right)
\]

\[
= \sum_{s=1}^q F^*(a_j^I) \left( \sum_{|L|=1} f_L^{p+s} \xi \right)
\]

\[
+ \sum_{\text{deg}(J)>1} F^*(a_j^I) \prod_{s=1}^q \left( \sum_{|L|=1} f_L^{p+s} \xi \right)^{\beta_s} \left| \xi^K \right)
\]

\[
= \sum_{s=1}^q \left( a_j^I \right) \circ F \sum_{|L|=1} f_L^{p+s} + R \left( (a_j^I)_I, (f_I^*)_{\nu, \text{deg}(I)<\text{deg}(K)} \right).
\]

Moreover, the initial condition gives

\[
f_K^j(t_0, x) = (\phi^*(w_j))^{\xi^K} \text{ for all } j \text{ in } \{p+1, \ldots, p+q\}.
\]

It follows as in the case of \(|K| = 0\), that \(f_K^j\) exists uniquely for all \((t, x) \in V\), for all \(j \in \{1, \ldots, p+q\}\) and for all \(K\) with \(\text{deg}(K) = k\).

We conclude that the functions \(\{f_I^j\}_{1 \leq j \leq p+q, I \in \{0,1\}^n}\) are uniquely defined on the whole of \(V\). Since the \(\{g_I^j\}_{1 \leq j \leq p+q, I \in \{0,1\}^n}\) are determined by Equation (18) from the \(\{f_I^j\}_{1 \leq j \leq p+q}\) via comparison of coefficients, the morphism \(F : \mathcal{V} \to \mathcal{V}\) is uniquely determined. \(\Box\)

We now consider the global problem of integrating a vector field on a supermanifold. In order to prove that there exists a unique maximal flow of a vector field, the following lemma will be crucial.
**Lemma 2.2.** Let $\mathcal{M} = (M, \mathcal{O}_M)$ and $S = (S, \mathcal{O}_S)$ be supermanifolds, $X$ a vector field in $T_M(M)$ and $\phi$ in $\text{Mor}(S, \mathcal{M})$. Then

(i) there exists an open sub supermanifold $V = (V, \mathcal{O}_{\mathbb{R}^{1|1} \times S}|_V)$ of $\mathbb{R}^{1|1} \times S$ with $V$ open in $\mathbb{R} \times S$ such that $\{0\} \times S \subset V$ and for all $x$ in $S$, $(\mathbb{R} \times \{x\}) \cap V$ is an interval, and a morphism $F : V \to \mathcal{M}$ satisfying:

\[
\text{inj}_{\mathbb{R}^{1|1}}^* \circ (\partial_t + \partial_x) \circ F^* = \text{inj}_{\mathbb{R}^{1|1}}^* F^* \circ X \quad \text{and} \quad \tag{16}
\]

\[
F \circ \text{inj}_{\{0\} \times S} = \phi . \quad \tag{17}
\]

(ii) Let furthermore $F_1 : V_1 \to \mathcal{M}$ and $F_2 : V_2 \to \mathcal{M}$ be morphisms satisfying (17) and (14) where $V_i = (V_i, \mathcal{O}_{\mathbb{R}^{1|1} \times S}|_{V_i})$ with $V_i$ open in $\mathbb{R} \times S$ such that $\{0\} \times S \subset V_i$, and for all $x$ in $S$, $(\mathbb{R} \times \{x\}) \cap V_i$ is an interval, for $i = 1, 2$. Then $F_i|_{V_{12}} = F_2|_{V_{12}}$ on $V_{12} = (V_{12}, \mathcal{O}_{\mathbb{R}^{1|1} \times S}|_{V_{12}})$, where $V_{12} = V_1 \cap V_2$.

**Proof:** (i) Let $\tilde{\phi} : S \to M$ denote the induced map of the underlying classical manifolds. Given now $s$ in $S$ and coordinate domains $U_s$ of $\tilde{\phi}(s)$, isomorphic to superdomains $U_s \subset \mathbb{R}^{m|n}$ resp. $W_s \subset \mathbb{R}^{p|q}$, by Lemma 2.1 we get solutions of (15) and (17) near $s$ (upon reducing the size of $U_s$ if necessary): $\mathbb{R}^{1|1} \times S \supset \mathbb{R}^{1|1} \times U_s \supset V_s \xrightarrow{F^s} W_s \subset \mathcal{M}$. If $V_{s_1} \cap V_{s_2} \neq \emptyset$ (compare Figure 1) we know, again by Lemma 2.1 that $F^{s_1}$ and $F^{s_2}$ coincide on this intersection. Thus, by taking the union $\mathcal{V}$ of $V_s$ for all $s$ in $S$, we get a morphism $F : \mathbb{R}^{1|1} \times S \supset \mathcal{V} \to \mathcal{M}$ such that $F|_{V_s} = F^s$ for all $s$, and fulfilling (16) and (17).

![Figure 1](image)

(ii) We define $A$ as the set of points $(t, x) \in V_{12}$ such that there exists $\epsilon = \epsilon_{(t,x)} > 0$ and $U = U_{(t,x)}$ an open sub supermanifold of $S$, such that its body $U$ contains $x$ and for $\mathcal{V} = V_{(t,x)} = (V_{(t,x)}, \mathcal{O}_{\mathbb{R}^{1|1} \times S}) = (\{0\} \times U, \mathcal{O}_{\mathbb{R}^{1|1} \times S})$ we have $F_1|_{\mathcal{V}} = F_2|_{\mathcal{V}}$. Of course, if $t < 0$ the interval will be of the type $[t - \epsilon, \epsilon]$ (See Figure 2). The claim of the Lemma is now equivalent to $A = V_{12}$. The set $A$ is obviously open.
By an easy application of Lemma 2.1, $A$ contains $\{0\} \times S$. The assumptions imply that for all $x \in S$, the set $I_x \subset \mathbb{R}$, defined by $(\mathbb{R} \times \{x\}) \cap \mathcal{V}_{12} = I_x \times \{x\}$, is an open interval containing 0. The definition of $A$ implies that the set $J_x \subset CA$ defined by $(\mathbb{R} \times \{x\}) \cap \mathcal{V}_{12} = J_x \times \{x\}$, is an open interval containing 0 as well.

Assuming now that $A \neq \mathcal{V}_{12}$, then there exists a point $(t, x_0) \in \mathcal{V}_{12} \setminus A$ such that $J_{x_0} \neq I_{x_0}$. Without loss of generality we can assume that $t > 0$ and that for $0 \leq t' < t$, $(t', x_0) \in A$. Let $U_0$ be an open coordinate neighborhood of $x_0$ in $S$ and $\delta > 0$ such that, with $V_0 := ]t - \delta, t + \delta[ \times U_0 \subset \mathcal{V}_{12}$, $H(V_0) \subset W$, where $W = (W, \mathcal{O}_M |_W)$ is a coordinate patch of $M$ and $H$ is the maximal flow of $\tilde{X}$ as in Lemma 2.1. Choose $t_0 \in ]t - \delta, t[$. Then $(t_0, x_0) \in A$ and thus there exists $\epsilon > 0$ and $U$ an open sub supermanifold of $U_0 = (U_0, \mathcal{O}_S |_{U_0})$ containing $x_0$ such that

$$F_1 |_{V} = F_2 |_{V}, \quad \text{where } V = ]t - \epsilon, t_0 + \epsilon[ \times \mathbb{R}^{0|1} \times U \subset \mathcal{V}_{12}. \quad (18)$$

On $V' = ]t - \delta, t + \delta[ \times \mathbb{R}^{0|1} \times U \subset \mathcal{V}_{12}$, $F_1$ and $F_2$ are defined and for $i = 1, 2$ the maps $F_i \circ \text{inj} |_{[t_0] \times U}$ coincide by (13) (Compare Figure 3 for the relative positions of the underlying topological spaces of these open sub supermanifolds of $\mathbb{R}^{0|1} \times S$).
By Lemma 2.1 we have $F_1|_{V'} = F_2|_{V'}$. Thus $F_1 = F_2$ on $V \cup V'$, and we conclude that $(t, x_0) \in A$. This contradiction shows that $V_{12} = A$.

Remarks. (1) Obviously, Lemma 2.2 holds true for an arbitrary $t_0 \in \mathbb{R}$ replacing $t_0 = 0$.

(2) Let us call a “flow domain for $X$ with initial condition $\phi \in \text{Mor}(\mathcal{S}, \mathcal{M})$ (with respect to $t_0 \in \mathbb{R}$)“ a domain $V \subset \mathbb{R}^{1|1} \times \mathcal{S}$ such that $\{t_0\} \times \mathcal{S} \subset V$ and for all $s$ in $\mathcal{S}$, $(\mathbb{R} \times \{s\}) \cap V$ is connected, i.e. an interval (times $\{s\}$) and such that a solution $F$ (a “flow”) of (16) and (17) exists on $V$. By the preceding lemma there exists such “flow domains”.

Theorem 2.3. Let $\mathcal{M}$ and $\mathcal{S}$ be supermanifolds, $X$ be a vector field in $\mathcal{T}_\mathcal{M}(\mathcal{M})$, $\phi \in \text{Mor}(\mathcal{S}, \mathcal{M})$ and $t_0$ in $\mathbb{R}$. Then there exists a unique map $F : V \to \mathcal{M}$ such that

$$(\text{inj}_{\mathbb{R}})^* \circ (\partial_t + \partial_x) \circ F^* = (\text{inj}_{\mathbb{R}})^* \circ F^* \circ X \quad \text{and} \quad F \circ \text{inj}_{\{t_0\} \times \mathcal{S}} = \phi,$$

where $V = (V, \mathcal{O}_{\mathbb{R}^{1|1} \times \mathcal{U}}|_V)$ is the maximal flow domain for $X$ with the given initial condition.

Moreover, $\tilde{F} : V \to \mathcal{M}$ is the maximal flow of $\tilde{X} \in \mathcal{X}(\mathcal{M})$ subject to the initial condition $\tilde{\phi}$ at $t = t_0$.

Proof. The proof of the theorem follows immediately from the Lemmata 2.1 and 2.2 upon taking the union of all flow domains and flows for $X$ as defined in the preceding remark.

3 Supervector fields and local $\mathbb{R}^{1|1}$-actions

Given a vector field on a classical, ungraded, manifold, the flow map $\tilde{F}$ (for $S = \mathcal{M}, \tilde{\phi} = \text{id}_{\mathcal{M}}$) is always a local action of $\mathbb{R}$ with its usual (and unique up to isomorphism) Lie group structure, the standard addition. The flow maps for vector fields described in the preceding section (taking here $S = \mathcal{M}, \phi = \text{id}_{\mathcal{M}}$), do not always have the analogous property of being local actions of $\mathbb{R}^{1|1}$ with an appropriate Lie supergroup structure. Two characterizations of those vector fields $X = X_0 + X_1$ that generate a local $\mathbb{R}^{1|1}$-action were found by J. Monterde and O. A. Sánchez-Valenzuela. We will give in this section a short proof of a slightly more general result, whose condition (iii) seems to be more easily verified in practice than those given in (13) (compare Thm. 3.6 and its proof there).

Let us begin by giving a useful two-parameter family of Lie supergroup structures on the supermanifold $\mathbb{R}^{1|1}$ and their right invariant vector fields.

Lemma 3.1. Let $a$ and $b$ be real numbers such that $a \cdot b = 0$ and $\mu_{a,b} = \mu : \mathbb{R}^{1|1} \to \mathbb{R}^{1|1}$
\( R^{1|1} \times R^{1|1} \to R^{1|1} \) be defined by

\[
\widetilde{\mu}(t_1, t_2) = t_1 + t_2, \\
\mu^*(t) = t_1 + t_2 + a \tau_1 \tau_2, \\
\mu^*(\tau) = \tau_1 + e^{bt_1} \tau_2.
\]

Then

(i) there exists a unique Lie supergroup structure on \( R^{1|1} \) such that the multiplication morphism is given by \( \mu_{a,b} \),

(ii) the right invariant vector fields on \( (R^{1|1}, \mu_{a,b}) \) are given by the graded vector space \( RD_0 \oplus RD_1 \), where

\[
D_0 := \partial_t + b \cdot \tau \partial_\tau \quad \text{and} \quad D_1 := \partial_\tau + a \cdot \tau \partial_t.
\]

and they obey \([D_0, D_0] = 0, [D_0, D_1] = -bD_1 \) and \([D_1, D_1] = 2aD_0\).

PROOF. Both assertions follow by straightforward verifications. \( \square \)

REMARKS. (1) It can easily be checked that the above family yields only three non-isomorphic Lie supergroup structures on \( R^{1|1} \), since \( (R^{1|1}, \mu_{a,0}) \) with \( a \neq 0 \) is isomorphic to \( (R^{1|1}, \mu_{1,0}) \) and \( (R^{1|1}, \mu_{0,b}) \) with \( b \neq 0 \) is isomorphic to \( (R^{1|1}, \mu_{0,1}) \) and the three multiplications \( \mu_{0,0}, \mu_{1,0} \) and \( \mu_{0,1} \) correspond to non-isomorphic Lie supergroup structures on \( R^{1|1} \). Nevertheless it is very convenient to work here with the more flexible two-parameter family of multiplications.

(2) In fact, all Lie supergroup structures on \( R^{1|1} \) are equivalent to \( \mu_{0,0}, \mu_{1,0} \) or \( \mu_{0,1} \). See, e.g., [4] for a direct approach to the classification of all Lie supergroup structures on \( R^{1|1} \).

DEFINITION 3.2. Let \( G = (G, O_G) \) resp. \( M = (M, O_M) \) be a Lie supergroup with multiplication morphism \( \mu \) and unit element \( e \) resp. a supermanifold. A “local action of \( G \) on \( M \)” is given by the following data:

a collection \( \Pi \) of pairs of open subsets \( \pi = (U_\pi, W_\pi) \) of \( M \), where \( U_\pi \) is relatively compact in \( W_\pi \), such that \( \{U_\pi, \pi \in \Pi\} \) is an open covering of \( M \), and for all \( \pi \) in \( \Pi \) an open sub supermanifold \( \mathcal{G}_\pi \subset G \), containing the neutral element \( e \) and a morphism

\[
\Phi_\pi : \mathcal{G}_\pi \times U_\pi \to W_\pi
\]

fulfilling

(1) \( \Phi_\pi \circ (e \times id_{U_\pi}) = id_{U_\pi}, \) where \( e : \{pt\} \to G \) is viewed as a morphism,

(2) \( \Phi_\pi \circ (\mu \times id_M) = \Phi_\pi \circ (id_G \times \Phi_\pi), \) where both sides are defined,

(3) if \( U_\pi \cap U_{\pi'} \neq \emptyset, \Phi_\pi = \Phi_{\pi'} \) on \( (\mathcal{G}_\pi \cap \mathcal{G}_{\pi'}) \times (U_\pi \cap U_{\pi'}). \)

DOCUMENTA MATHEMATICA 18 (2013) 519–545
Proposition 3.3. Let $G = (G, O_G)$ resp. $M = (M, O_M)$ be a Lie supergroup resp. a supermanifold. Then

(i) a local $G$-action on $M$, specified by a set $\Pi$ and morphisms $\{(U_\pi, \omega_\pi, G_\pi, \Phi_\pi) | \pi \in \Pi\}$, gives rise to an open sub supermanifold $V \subset G \times M$ containing $\{e\} \times M$ and a morphism $\Phi_V : V \to M$ such that
\[ \Phi_V \circ (\mu \times id_M) = \Phi_V \circ (id_G \times \Phi_V), \]
where both sides are defined and such that $\Phi_\pi = \Phi_V$ on $(G_\pi \times U_\pi) \cap V, \forall \pi \in \Pi$.

(ii) an open sub supermanifold $V \subset G \times M$ containing $\{e\} \times M$ and a morphism $\Phi_V : V \to M$ such that $(\ast)$ is fulfilled, where it makes sense, yields a local $G$-action on $M$ such that $(\ast\ast)$ holds.

Proof. As in the classical case of ungraded manifolds and Lie groups. □

Theorem 3.4. Let $M$ be a supermanifold, $X$ a vector field on $M$ and $V \subset \mathcal{R}^{1|1} \times M$ the domain of the maximal flow $F : V \to M$ satisfying
\[ (\text{inj}_{\mathcal{R}^{1|1}})^* \circ (\partial_t + \partial_\tau) \circ F^* = (\text{inj}_{\mathcal{R}^{1|1}})^* \circ F^* \circ X \text{ and} \]
\[ F \circ \text{inj}_{\mathcal{R}^{1|1}} \mid_{\{0\} \times M} = id_M. \]

Let $a$ and $b$ be real numbers such that $a \cdot b = 0$. Then the following assertions are equivalent:

(i) the map $F$ fulfills
\[ (\partial_t + \partial_\tau + \tau (a\partial_t + b\partial_\tau)) \circ F^* = F^* \circ X, \]

(ii) the map $F$ is a local $(\mathcal{R}^{1|1}, \mu_{a,b})$-action on $M$,

(iii) $\mathcal{R}X_0 \oplus \mathcal{R}X_1$ is a sub Lie superalgebra of $\mathcal{T}_M(M)$ with commutators $[X_0, X_1] = -bX_1$ and $[X_1, X_1] = 2aX_0$.

Proof. Recall that $F$ fulfills
\[ (\text{inj}_{\mathcal{R}^{1|1}})^* \circ \partial_t \circ F^* = (\text{inj}_{\mathcal{R}^{1|1}})^* \circ F^* \circ X_0 \text{ and} \]
\[ (\text{inj}_{\mathcal{R}^{1|1}})^* \circ \partial_\tau \circ F^* = (\text{inj}_{\mathcal{R}^{1|1}})^* \circ F^* \circ X_1. \]

Denoting the projection from $\mathcal{R}^{1|1}$ to $\mathcal{R}$ by $p$, we have
\[ \text{id}_{\mathcal{R}^{1|1}}^* = p^* \circ (\text{inj}_{\mathcal{R}^{1|1}})^* + \tau \cdot p^* \circ (\text{inj}_{\mathcal{R}^{1|1}})^* \circ \partial_\tau \]
which we will write more succinctly as
\[ \text{id}_{\mathcal{R}^{1|1}}^* = \tau \cdot (\text{inj}_{\mathcal{R}^{1|1}})^* \circ \partial_\tau. \]
Using relation (19) and the equations fulfilled by $F^*$ we get

\[ F^* \circ X = (\text{inj}_R^{\mathbb{R}^{11}})^* \circ F^* \circ X + \tau \cdot (\text{inj}_R^{\mathbb{R}^{11}})^* \circ \partial_t \circ F^* \circ X \]

\[ = (\text{inj}_R^{\mathbb{R}^{11}})^* \circ (\partial_t + \partial_r) \circ F^* + \tau \cdot (\text{inj}_R^{\mathbb{R}^{11}})^* \circ F^* \circ X_1 \circ X \]

\[ = (\text{inj}_R^{\mathbb{R}^{11}})^* \circ (\partial_t + \partial_r) \circ F^* + \tau \cdot (\text{inj}_R^{\mathbb{R}^{11}})^* \circ F^* \]

\[ + \tau \cdot (\text{inj}_R^{\mathbb{R}^{11}})^* \circ F^* \circ \left( [X_1, X_0] + X_0 \circ X_1 + \frac{1}{2} [X_1, X_1] \right). \]

Since

\[ (\text{inj}_R^{\mathbb{R}^{11}})^* \circ F^* \circ X_0 \circ X_1 = (\text{inj}_R^{\mathbb{R}^{11}})^* \circ \partial_t \circ F^* \circ X_1 \]

\[ = \partial_t \circ (\text{inj}_R^{\mathbb{R}^{11}})^* \circ F^* \circ X_1 \]

\[ = \partial_t \circ (\text{inj}_R^{\mathbb{R}^{11}})^* \circ \partial_r \circ F^* \]

\[ = \partial_t \circ \partial_r \circ F^* \]

we arrive at

\[ F^* \circ X = (\text{inj}_R^{\mathbb{R}^{11}})^* \circ (\partial_t + \partial_r) \circ F^* + \]

\[ \tau \cdot F^* \circ \left( [X_1, X_0] + \frac{1}{2} [X_1, X_1] \right) + \tau \cdot \partial_r \circ \partial_t \circ F^*. \quad (20) \]

On the other hand, if $a$ and $b$ are real numbers, we have, again using (19)

\[ (\partial_t + \partial_r + \tau (a \partial_t + b \partial_r)) \circ F^* \]

\[ = \left( (\text{inj}_R^{\mathbb{R}^{11}})^* + \tau \cdot (\text{inj}_R^{\mathbb{R}^{11}})^* \circ \partial_r \right) \circ (\partial_t + \partial_r) \circ F^* \]

\[ + \tau \cdot \left( a \circ (\text{inj}_R^{\mathbb{R}^{11}})^* \circ \partial_t + b \cdot (\text{inj}_R^{\mathbb{R}^{11}})^* \circ \partial_r \right) \circ F^* \]

\[ = (\text{inj}_R^{\mathbb{R}^{11}})^* \circ (\partial_t + \partial_r) \circ F^* + \tau \cdot \partial_r \circ \partial_t \circ F^* \]

\[ + \tau \cdot F^* \circ (a X_0 + b X_1). \]

Thus we have

\[ (\partial_t + \partial_r + \tau (a \partial_t + b \partial_r)) \circ F^* = F^* \circ X \]

\[ = \tau \cdot F^* \circ \left( a X_0 - \frac{1}{2} [X_1, X_1] + b X_1 - [X_1, X_0] \right). \quad (21) \]

Since $F$ satisfies the initial condition $(\text{inj}_R^{\mathbb{R}^{11}})^* \circ F^* = \text{id}_M$, $\tau \cdot F^*$ is injective and thus Equation (21) easily implies the equivalence of (i) and (iii).

We remark that, in this case, we automatically have $a \cdot b = 0$ since the Jacobi identity implies that $[X_1, [X_1, X_1]] = [[X_1, X_1], X_1] + (-1)^{[1]} \cdot [X_1, [X_1, X_1]]$, i.e., $2a \cdot b \cdot X_1 = [X_1, [X_1, X_1]] = 0$. 

\[ \text{Documenta Mathematica 18 (2013) 519–545} \]
Assume now that \( a \) and \( b \) are real numbers such that \((i)\) satisfied, and let \( \mu = \mu_{a,b} \) be as in Lemma 5.1. We have to show that \( F \) is a local action of \((\mathbb{R}^{1|1}, \mu)\).

Let us define
\[
G := F \circ (\text{id}_{\mathbb{R}^{1|1}} \times F) : \mathbb{R}^{1|1} \times (\mathbb{R}^{1|1} \times \mathcal{M}) \to \mathcal{M}
\]
and
\[
H := F \circ (\mu \times \text{id}_{\mathcal{M}}) : (\mathbb{R}^{1|1} \times \mathbb{R}^{1|1}) \times \mathcal{M} \cong (\mathbb{R}^{1|1} \times \mathcal{M}) \to \mathcal{M}.
\]

In order to prove that \( F \) is a \( \mathbb{R}^{1|1} \)-action on, we have to show that \( G = H \).

We observe that \( G \) is the integral curve of \( X \) subject to the initial condition \( F \in \text{Mor}(\mathbb{R}^{1|1} \times \mathcal{M}, \mathcal{M}) \).

Let us prove that the morphism \( H \) satisfies the following conditions:
\[
\left( \text{inj}_{\mathbb{R}^{1|1} \times (\mathbb{R}^{1|1} \times \mathcal{M})} \right)^{*} \circ (\partial_{t_1} + \partial_{t_1}) \circ H^{*} = \left( \text{inj}_{\mathbb{R}^{1|1} \times (\mathbb{R}^{1|1} \times \mathcal{M})} \right)^{*} \circ H^{*} \circ X \quad (22)
\]
\[
H \circ \text{inj}_{\{0\} \times (\mathbb{R}^{1|1} \times \mathcal{M})}^{\mathbb{R}^{1|1} \times (\mathbb{R}^{1|1} \times \mathcal{M})} = F. \quad (23)
\]

Then by the unicity of integral curves we have \( H = G \).

Equation (23) holds true since \( \mu \circ \text{inj}_{\{0\} \times \mathbb{R}^{1|1}} = \text{id}_{\mathbb{R}^{1|1}} \).

Defining \( D := D_0 + D_1 = \partial_{t_1} + \partial_{t_1} + \tau_1(a \partial_{t_1} + b \partial_{t_1}) \) and writing \( \text{inj}_{[t_1]} \) for \( \text{inj}_{\mathbb{R}^{1|1} \times (\mathbb{R}^{1|1} \times \mathcal{M})} \) and using right invariance of \( D \), we arrive at equation (22) as follows
\[
\left( \text{inj}_{\mathbb{R}^{1|1} \times (\mathbb{R}^{1|1} \times \mathcal{M})} \right)^{*} \circ (\partial_{t_1} + \partial_{t_1}) \circ H^{*} = (\text{inj}_{[t_1]})^{*} \circ D \circ H^{*} = (\text{inj}_{[t_1]})^{*} \circ (D \otimes \text{id}_{\mathbb{R}^{1|1}}) \circ \mu^{*} \times \text{id}_{\mathcal{M}} \circ F^{*} = (\text{inj}_{[t_1]})^{*} \circ (\mu^{*} \circ D) \times \text{id}_{\mathcal{M}} \circ F^{*} = (\text{inj}_{[t_1]})^{*} \circ (\mu^{*} \times \text{id}_{\mathcal{M}}) \circ F^{*} \circ X = (\text{inj}_{[t_1]})^{*} \circ H^{*} \circ X.
\]

Thus we obtain that \((i)\) implies \((ii)\).

Assume now that \((ii)\) is satisfied, i.e., there exists a Lie supergroup structure on \( \mathbb{R}^{1|1} \) with multiplication \( \mu \) such that
\[
F \circ (\text{id}_{\mathbb{R}^{1|1}} \times F) = F \circ (\mu \times \text{id}_{\mathcal{M}}). \quad (24)
\]

Since \( F \) is a flow for \( X \), with initial condition \( \phi = \text{id}_{\mathcal{M}} \), the LHS of the preceding equality is a flow for \( X \) with initial condition \( \phi = F \), (24) implies
\[
\left( \text{inj}_{[t_1]} \right)^{*} \circ (\partial_{t_1} + \partial_{t_1}) \circ (\mu^{*} \times \text{id}_{\mathcal{M}}) \circ F^{*} = \left( \text{inj}_{[t_1]} \right)^{*} \circ (\mu^{*} \times \text{id}_{\mathcal{M}}) \circ F^{*} \circ X. \quad (25)
\]

Proof completed.
By Equation (20), the RHS gives for $t_1 = 0$:

$$\begin{align*}
\left(\text{inj}_{|t_1=0}\right)^* \circ (\mu^* \times \text{id}_M^*) \circ F^* \circ X &= F^* \circ X \\
&= \left(\text{inj}_{\mathbb{R}_{\mu}}\right)^* \circ (\partial_t + \partial_{\tau}) \circ F^* \\
&+ \tau \cdot \left[ F^* \circ \left( [X_1, X_0] + \frac{1}{2}[X_1, X_1] \right) + \partial_{\tau} \circ \partial_t \circ F^* \right].
\end{align*}$$

Moreover, we have by direct comparison

$$(\partial_t + \partial_{\tau}) \circ \mu^* = (\partial_t + \partial_{\tau})(\mu^*(t)) \cdot (\mu^* \circ \partial_t) + (\partial_t + \partial_{\tau})(\mu^*(\tau)) \cdot (\mu^* \circ \partial_{\tau}).$$

Thus, if $\mu : \mathbb{R}^{1|1} \times \mathbb{R}^{1|1} \to \mathbb{R}^{1|1}$ is given by

$$\begin{align*}
\mu^*(t) &= \tilde{\mu}(t_1, t_2) + \alpha(t_1, t_2)\tau_1 \tau_2 \\
\mu^*(\tau) &= \beta(t_1, t_2)\tau_1 + \gamma(t_1, t_2)\tau_2,
\end{align*}$$

and upon using $\left(\text{inj}_{\{0\} \times \mathbb{R}^{1|1}}\right)^* \circ \mu^* = \text{id}_{\mathbb{R}^{1|1}}^*$, we have

$$\begin{align*}
\left(\text{inj}_{\{0\} \times \mathbb{R}^{1|1}}\right)^* (\partial_t + \partial_{\tau}) \circ \mu^* &= ((\partial_t, \tilde{\mu})(0, t) + \alpha(0, t)\tau) \cdot \partial_t \\
&+ (\beta(0, t) + (\partial_t, \gamma)(0, t)\tau) \cdot \partial_{\tau}.
\end{align*}$$

Using again (19), we have

$$\begin{align*}
\partial_t \circ F^* &= \left(\text{inj}_{\mathbb{R}^{1|1}}\right)^* \circ F^* \circ X_0 + \tau \cdot \partial_{\tau} \circ \partial_t \circ F^* \quad \text{and} \\
\partial_{\tau} \circ F^* &= \left(\text{inj}_{\mathbb{R}^{1|1}}\right)^* \circ F^* \circ X_1.
\end{align*}$$

Then the LHS of (25) at $t_1 = 0$ is

$$\begin{align*}
\left(\text{inj}_{|t_1=0}\right)^* \circ (\partial_t + \partial_{\tau}) \circ (\mu^* \times \text{id}_M^*) \circ F^* \\
= (\partial_t, \tilde{\mu})(0, t) \cdot \left(\text{inj}_{\mathbb{R}^{1|1}}\right)^* \circ F^* \circ X_0 \\
+ \beta(0, t) \cdot \left(\text{inj}_{\mathbb{R}^{1|1}}\right)^* \circ F^* \circ X_1 \\
+ \tau \cdot \left( (\partial_t, \tilde{\mu})(0, t) \cdot \partial_{\tau} \circ \partial_t \circ F^* \\
+ \alpha(0, t) \cdot F^* \circ X_0 \\
+ (\partial_t, \gamma)(0, t) \cdot F^* \circ X_1 \right).
\end{align*}$$

Using the obtained identities for its LHS and RHS, the “$\tau$-part” of Equation (25) at $t_1 = 0$ gives us:

$$\begin{align*}
\tau \cdot \left[ F^* \circ \left( [X_1, X_0] + \frac{1}{2}[X_1, X_1] \right) + \partial_{\tau} \circ \partial_t \circ F^* \right] \\
= \tau \cdot [ (\partial_t, \tilde{\mu})(0, t) \cdot \partial_{\tau} \circ \partial_t \circ F^* + \alpha(0, t) \cdot F^* \circ X_0 + (\partial_t, \gamma)(0, t) \cdot F^* \circ X_1 ].
\end{align*}$$
Since \( \tilde{\mu}(t_1,0) = t_1 \), we have \( (\partial_{t_1}\tilde{\mu})(0,0) = 1 \) and therefore the preceding equation evaluated at \( t = 0 \) yields
\[
[X_1,X_0] + \frac{1}{2}[X_1,X_1] = (\partial_{t_1}\gamma)(0,0) \cdot X_1 + \alpha(0,0) \cdot X_0
\]
finishing the proof that \((ii)\) implies \((iii)\).

\( \square \)

### 4 Examples and applications

(4.1) If \( X = X_0 \) is an even vector field, the fact that it integrates to a (local) action of \( \mathbb{R} = \mathbb{R}^{1|0} \) is almost folkloristic. The relatively recent proof of \( \mathbb{R} \) - in the case of compact supermanifolds - is close to our approach. A non-trivial (local) action of \( \mathbb{R}^{1|1} \) can obviously be extended to a (local) action of \( (\mathbb{R}^{1|1},\mu_{a,b}) \) if and only if \( a = 0 \). Of course, the ensuing action of \( \mathbb{R}^{0|1} \) will not even be almost-effective, since the positive-dimensional sub Lie supergroup \( \mathbb{R}^{0|1} \) acts trivially.

(4.2) Our preferred example of an even vector field gives rise to the exponential map on Lie supergroups.

Let us first recall that an even vector field \( X \) on a supermanifold \( M \) corresponds to a section \( \sigma_X \) of the tangent bundle \( TM \to M \) (see, e.g., Sections 7 and 8 of \( \mathbb{R} \) for a construction of \( TM \) and a proof of this statement, and compare also the remark after Thm. 2.19 in \( \mathbb{R} \)). Given an auxiliary supermanifold \( S \) and a morphism \( \psi : S \to M \), one calls for \( i \in \{0,1\} \)
\[
\text{Der}_\psi(\mathcal{O}_M(M),\mathcal{O}_S(S)) := \{ D : \mathcal{O}_M(M) \to \mathcal{O}_S(S) | D \text{ is } \mathbb{R}-\text{linear and } \forall f,g \in \mathcal{O}_M(M) \text{ homogeneous,} \}
\]
\[
D(f \cdot g) = D(f) \cdot \psi^*(g) + (\pm 1)^i [\psi^*(f) \cdot D(g)],
\]
the “space of derivations of parity \( i \) along \( \psi \)”. In category-theoretical terms the tangent bundle \( TM \) represents then the functor from supermanifolds to sets given by \( S \mapsto \{ (\psi,D) | \psi \in \text{Mor}(S,M) \text{ and } D \in \text{Der}_\psi(\mathcal{O}_M(M),\mathcal{O}_S(S)) \} \) (compare, e.g., Section 3 of \( \mathbb{R} \)).

Let now \( G = (G,\mathcal{O}_G) \) be a Lie supergroup with multiplication \( \mu = \mu_G \) and neutral element \( e \). We define \( X \) in \( \text{Der}(\mathcal{O}_{T_eG}(G \times T_eG)) \) to be the even vector field on \( G \times T_eG \) corresponding to the following section \( \sigma_X \) of \( T\tilde{G} \times T\tilde{T}_e\tilde{G} \cong T(\tilde{G} \times T_e\tilde{G}) \to \tilde{G} \times T_e\tilde{G} \). We denote the zero-section of \( T\tilde{G} \to \tilde{G} \) by \( \sigma_0 \) and the canonical inclusion \( T_e\tilde{G} \to \tilde{G} \) by \( i_e \). Then \( \sigma_X := (T\mu \circ (\sigma_0 \times i_e),0) \), where \( T\mu : T\tilde{G} \times T\tilde{G} \cong T(\tilde{G} \times \tilde{G}) \to T\tilde{G} \) is the tangential morphism associated to the multiplication morphism. (For simplicity, we write 0 for the zero-section of \( T(T_e\tilde{G}) \to T_e\tilde{G} \) here and in the sequel.)
Let us recall that for $\mathcal{S}$ an arbitrary supermanifold, and $\phi : \mathcal{M} \to \mathcal{N}$ a morphism between supermanifolds, we have an induced map $\phi(S) : \mathcal{M}(S) = \text{Mor}(\mathcal{S}, \mathcal{M}) \to \text{Mor}(\mathcal{S}, \mathcal{N}) = \mathcal{N}(\mathcal{S})$, $\phi(S)(\psi) := \phi \circ \psi$. Given a finite-dimensional Lie superalgebra $\mathfrak{g}$ or a Lie supergroup $\mathcal{G}$, one easily checks that for all $k \geq 0$, $\mathfrak{g}(\mathbb{R}^{0|k})$ resp. $\mathcal{G}(\mathbb{R}^{0|k})$ is a finite-dimensional classical (i.e. even) Lie algebra resp. Lie group. Furthermore, $T_e(\mathcal{G}(\mathbb{R}^{0|k}))$ is canonically isomorphic to $(T_e\mathcal{G})(\mathbb{R}^{0|k})$, where the first $e$ is the obvious constant morphism from $\mathbb{R}^{0|k}$ to $\mathcal{G}$ and the second $e$ denotes the neutral element of $\mathcal{G}$. (Compare, e.g., [16] for more information on the superpoint approach to Lie supergroups.)

**Lemma 4.1.** Let $\mathcal{G}$ be a Lie supergroup with multiplication $\mu^{\mathcal{G}}$, and the vector field $X$ as above. Then

(i) the induced vector field $\tilde{X}$ on the underlying manifold $G \times T_eG$ is given as

$$\tilde{X}(g,\xi) = (\xi^k(g), 0) \quad \forall (g, \xi) \in G \times T_eG,$$

(ii) the (even) vector fields $\tilde{X}$ and $X$ are complete.

**Proof.** (i) For $k \geq 0$, let $\sigma_X(\mathbb{R}^{0|k})$ be the section of $T(\mathcal{G}(\mathbb{R}^{0|k})) \times T(T_e\mathcal{G}(\mathbb{R}^{0|k})) \to \mathcal{G}(\mathbb{R}^{0|k}) \times T_e\mathcal{G}(\mathbb{R}^{0|k})$ induced by $\sigma_X$, and let $X^k$ be the corresponding derivation on $\mathcal{G}(\mathbb{R}^{0|k}) \times T_e\mathcal{G}(\mathbb{R}^{0|k})$. Since $\mathcal{G}(\mathbb{R}^{0|k}) \times (T_e\mathcal{G})(\mathbb{R}^{0|k})$ is an ungraded manifold,

$$\sigma_X(\mathbb{R}^{0|k})(g, \xi) = (T\mu^{\mathcal{G}} \circ (\sigma_0 \times i_\xi) \circ (g \times \xi), 0)$$

$$= (T\mu^{\mathcal{G}} \circ (0 \times \xi), 0)$$

$$= (T\mu^{\mathcal{G}})(0, \xi), 0)$$

$$= (T\mu^{\mathcal{G}})(0, (\xi), 0),$$

where $\mu^{\mathcal{G}}$ is the multiplication on $\mathcal{G}(\mathbb{R}^{0|k})$ and $\mu^{\mathcal{G}}$ is the left-multiplication by the element $g$ of the group $\mathcal{G}(\mathbb{R}^{0|k})$. We conclude that $\sigma_X(\mathbb{R}^{0|k})(g, \xi)$ (or equivalently $X^k_{(g, \xi)}$) corresponds to $(\xi^L(g), 0)$, where $\xi^L$ is the unique left-invariant vector field on $\mathcal{G}(\mathbb{R}^{0|k})$ such that its value in $e$ is $\xi$. (Observe that $\mathcal{G}(\mathbb{R}^{0|k})$ is a classical Lie group and not only a group object in the category of supermanifolds, allowing us to argue “point-wise”.)

(ii) The flows of $X^k$ are simply given by $F^{X^k} : \mathbb{R} \times \mathcal{G}(\mathbb{R}^{0|k}) \times (T_e\mathcal{G})(\mathbb{R}^{0|k}) \to \mathcal{G}(\mathbb{R}^{0|k}) \times (T_e\mathcal{G})(\mathbb{R}^{0|k})$, $(t, g, \xi) \mapsto (g \exp^{\mathcal{G}(\mathbb{R}^{0|k})}(t\xi), \xi)$. All fields $X^k$ are thus complete, in particular this holds for $\tilde{X} = X^0$, the induced vector field on $G = \mathcal{G}(\mathbb{R}^{0|0})$. By Theorem 2.3, the flow $F^{X^k} : \mathbb{R} \times \mathcal{G} \times T_e\mathcal{G} \to \mathcal{G} \times T_e\mathcal{G}$ is then global as well, i.e. $X$ is complete.

**Definition 4.2.** Let $\mathcal{G} = (G, \mathcal{O}_{\mathcal{G}})$ be a Lie supergroup with multiplication $\mu$ and neutral element $e$, and with the even vector field $X$ and its flow morphism...
F = F^X as above. Then the “exponential morphism of G” is given by \( \exp^G = \text{proj}_1 \circ F \circ \text{inj}_{(1)} \times \{e\} \times T_eG : T_eG \to G \), where \( \text{proj}_1 : G \times T_eG \to G \) is the projection on the first factor. Diagrammatically, one has

\[
\begin{array}{ccc}
\mathbb{R} \times G \times T_eG & \cong & \{1\} \times \{e\} \times T_eG \\
\downarrow & & \downarrow \text{proj}_1 \\
T_eG & \xrightarrow{\exp^G} & G.
\end{array}
\]

**Theorem 4.3.** The exponential morphism \( \exp^G : T_eG \to G \) for a Lie supergroup \( G \) fulfills and is uniquely determined by the following condition: for all \( k \geq 0 \), \( \exp^G(\mathbb{R}^{0|k}) : T_eG(\mathbb{R}^{0|k}) \to G(\mathbb{R}^{0|k}) \) is the exponential map \( \exp^{\mathbb{G}}(\mathbb{R}^{0|k}) \) of the finite-dimensional, ungraded Lie group \( \mathbb{G}(\mathbb{R}^{0|k}) \).

**Proof.** Using the notations of Lemma 4.1, a straightforward calculation shows that the flow \( F^{X^k} \) of \( X^k \) on \((G \times T_eG)(\mathbb{R}^{0|k})\) is given as follows \( (t, (g, \xi)) \mapsto F^X \circ \text{inj}_k \circ F^X \circ \text{inj}_k \circ (g \times \xi) \) and, notably, we have \( F^{X^k} \circ \text{inj}_{(1,e)}(\xi) = F^X \circ \text{inj}_1 \circ (e \times \xi) = F^X \circ \text{inj}_{(1,e)} \circ \xi \). Hence

\[
\exp^{\mathbb{G}}(\mathbb{R}^{0|k})(\xi) = \text{proj}_1 \circ F^X \circ \text{inj}_{(1,e)}(\xi) = \text{proj}_1 \circ F^X \circ \text{inj}_1 \circ (e \times \xi) = \exp^G \circ \xi = (\exp^G(\mathbb{R}^{0|k}))(\xi).
\]

On the other hand, it is clear that the subcategory of superpoints with objects \( \{\mathbb{R}^{0|k} | k \geq 0\} \) generates the category of supermanifolds in the following sense: given two different morphisms \( \phi_1, \phi_2 : \mathcal{M} \to \mathcal{N} \) between supermanifolds, there exists a \( k \geq 0 \) and a morphism \( \psi : \mathbb{R}^{0|k} \to \mathcal{M} \) such that \( \phi_1 \circ \psi \neq \phi_2 \circ \psi \). Thus it follows that the family \( \{\exp^G(\mathbb{R}^{0|k}) | k \geq 0\} \) uniquely fixes \( \exp^G \). \( \square \)

(4.3) Part of our interest in the integration of supervector fields stemmed from the construction of a geodesic flow in \( \mathbb{M} \). Given a homogeneous (i.e., even or odd) Riemannian metric on a supermanifold \( \mathcal{M} \), the associated geodesic flow is, in fact, defined as the flow of an appropriate Hamiltonian vector field on its (co-)tangent bundle.

(4.4) An odd vector field \( X_1 \) on a supermanifold \( \mathcal{M} \) is called “homological” if \( X_1 \circ X_1 = \frac{b}{2}[X_1, X_1] = 0 \). Its flow is given by the following \( \mathbb{R}^{0|1} \)-action \( \Phi : \mathbb{R}^{0|1} \times \mathcal{M} \to \mathcal{M} \), \( \Phi^t(f) = f + \tau \cdot X_1(f) \), \( \forall f \in \mathcal{O}_{\mathcal{M}}(\mathcal{M}) \). This action can, of course, be extended to a (not almost-effective) \( \mathbb{R}^{1|1}, \mu_{a,b}) \)-action if and only if \( b = 0 \).

Typical examples arise as follows: let \( \mathbb{E} \to \mathcal{M} \) be a vector bundle over a classical manifold and \( T \) be an “appropriate” \( \mathbb{R} \)-linear operator on sections of \( \wedge \mathbb{E} \), then \( T \) yields a vector field on \( \Pi \mathbb{E} := (\mathcal{M}, \mathbb{E}^\infty_{\mathbb{E}}) \), the supermanifold...
associated to \( E \to M \) by the Batchelor construction. If \( E = TM \to M \), we have \( \Gamma_{\infty E^*} = \Omega^*_M(M) \), the sheaf of differential forms on \( M \), with its natural \( \mathbb{Z}/2\mathbb{Z} \)-grading. Taking \( T = d \), we get an odd vector field that is obviously homological. Taking \( T = t_\xi \), the contraction of differential forms with a vector field \( \xi \) on the (here classical) base manifold \( M \), we again get a homological vector field on \( \Pi TM \). Since \( t_\xi \circ t_\eta + t_\eta \circ t_\xi = 0 \), the vector space of all vector fields on \( M \) is realized as a commutative, purely odd sub Lie superalgebra of all vector fields on \( \Pi TM \). More generally, a section \( s \) of \( E \to M \) always gives rise to a contraction \( t_s : \Gamma_{\infty E^*}(M) \to \Gamma_{\infty E^*}(M) \) that is an odd derivation (i.e. an anti-derivation of degree -1 in more classical language). Furthermore, given two sections \( s \) and \( t \) of \( E \), the associated odd vector fields commute. In the article \( \mathbb{I} \) this construction is studied in the special case that \( E \) is the spinor bundle over a classical spin manifold \( M \).

(4.5) If \( G \) is a Lie group acting on a classical manifold \( M \), then the action can of course be lifted to an action on the total space of the tangent and the cotangent bundle of \( M \). The induced vector fields on \( \Pi TM \) are even and \( \xi \) in \( g = \text{Lie}(G) \), the Lie algebra of \( G \), acts on \( \mathcal{O}_{\Pi TM} \) by \( \mathcal{L}_\xi \), the Lie derivative with respect to the fundamental vector field on \( M \) associated to \( \xi \). Putting together these fields and the contractions constructed in Example (4.4), we get a Lie superalgebra with underlying vector space \( g \oplus g \), the first resp. second summand being the even resp. odd part. The commutators in \( g \oplus g \) are given as follows: \( [\mathcal{L}_\xi, \mathcal{L}_\eta] = \mathcal{L}_{[\xi, \eta]} \), \( [t_\xi, t_\eta] = 0 \) and \( [\mathcal{L}_\xi, t_\eta] = t_{[\xi, \eta]} \) for all \( \xi, \eta \in g \). In fact, the above can be interpreted as an action of the Lie supergroup \( \Pi TG \) on \( \Pi TM \).

The Lie algebra \( g \oplus g \) can be extended by a one-dimensional odd direction generated by the exterior derivative \( d \). The extended algebra \( g \oplus (g \oplus \mathbb{R} \cdot d) \), still a sub Lie superalgebra of \( \mathcal{T}_{\Pi TM}(M) \), has the following additional commutators:

\[
[\mathcal{L}_\xi, d] = 0 \quad \text{and} \quad [d, t_\xi] = \mathcal{L}_\xi \quad \text{for all} \quad \xi \in g.
\]

(4.6) If \( \mathcal{M} = \mathbb{R}^{1|1} \) with coordinates \((x, \xi)\), the vector field \( X_1 = \partial_x + \xi \partial_\xi \) is obviously odd and non-homological since \( X_1 \circ X_1 = \partial_x \). Direct inspection shows that the map \( \Phi : \mathbb{R}^{1|1} \times \mathcal{M} \to \mathcal{M}, \Phi^*(f) = f + \tau \cdot X_1(f) \) (compare Example (4.4)) does not fulfill \( \partial_\tau \circ \Phi^* = \Phi^* \circ X_1 \). Nevertheless, the trivial extension of \( \Phi \) to a morphism \( F : \mathbb{R}^{1|1} \times \mathcal{M} \to \mathcal{M} \) is the flow of \( X_1 \) in the sense of Theorem 3.3 fulfilling the initial condition \( \phi = \text{id}_\mathcal{M} \). We underline that this map is not an action of \( \mathbb{R}^{1|1} \). Upon extending \( X_1 \) to \( X := X_0 + X_1 \) with \( X_0 := \frac{1}{2}[X_1, X_1] = X_1 \circ X_1 \), we obtain by Theorem 3.3 an action of \((\mathbb{R}^{1|1}, \mu_{1,0})\) as the flow map of \( X \). Let us observe that the above vector field \( X_1 \) (and not \( X \)) is the prototype of what is called a “supersymmetry” in the physics literature (compare, e.g., [21] and the other relevant texts in these volumes.). More recently, the associated Lie supergroup structure on \( \mathbb{R}^{1|1} \) (and an analogous structure on \( \mathbb{R}^{2|1} \)) were introduced by S. Stolz and P. Teichner into their program to geometrize the cocycles of elliptic cohomology (compare

Documenta Mathematica 18 (2013) 519–545
Obviously, one can generalize this construction to $\mathbb{R}^{m|n}$ ($m, n \geq 1$) with coordinates $(x_1, \ldots, x_m, \xi_1, \ldots, \xi_n)$ by setting for $1 \leq k \leq m$, $1 \leq \alpha \leq n$

$$D_{\alpha,k} := \partial_{\xi_{\alpha}} + \xi_{\alpha} \cdot \partial_{x_k}.$$  

We then have $[D_{\alpha,k}, D_{\beta,l}] = \delta_{\alpha,\beta} \cdot (\partial_{x_k} + \partial_{x_l})$ and $[D_{\alpha,k}, D_{\beta)] = 0$. Taking $X_1 = D_{\alpha,k}$ and $X_0 = \partial_{x_k}$ we reproduce a copy of the preceding situation.

(4.7) The vector field $X = X_0 + X_1$ on $\mathcal{M} = \mathbb{R}^{1|1}$ with $X_0 = \partial_x + \xi \cdot \partial_{\xi}$ and $X_1 = \partial_{\xi} + \xi \cdot \partial_x$, already mentioned in the introduction, is a very simple example of an inhomogeneous vector field not generating any local $\mathbb{R}^{1|1}$-action, since, e.g., condition (iii) in Theorem 3.4 is violated. Thus integration of $X$ is only possible in the sense of Theorem 2.3, i.e. upon using the evaluation map. Let us observe that the sub Lie superalgebra $g$ of $T_M(M)$ generated by $X_0$ and $X_1$ since sub Lie superalgebras are by definition graded sub vector spaces, is four-dimensional with two even generators $Z, W$ and two odd generators $D, Q$ such that:

$Z$ is central, $[W, D] = Q$, $[W, Q] = D$, $D^2 = -Q^2 = Z$ and $[D, Q] = 0$. (This amounts in physical interpretation to the presence of two commuting supersymmetries $D$ and $Q$, generating the same supersymmetric Hamiltonian $Z$ plus an even symmetry commuting with the Hamiltonian and exchanging the supersymmetries $D$ and $Q$.)

5 Flow of a holomorphic vector field on a holomorphic supermanifold

In this section we extend our results to the holomorphic case. We will always denote the canonical coordinates on $\mathbb{C}^{1|1}$ by $z$ and $\xi$ and write $\partial_z$ resp. $\partial_{\xi}$ for $\frac{\partial}{\partial z}$ resp. $\frac{\partial}{\partial \xi}$. All “auxiliary” supermanifolds $S$ and morphisms having these as sources will be assumed to be holomorphic in this section.

**Definition 5.1.** Let $\mathcal{M} = (\mathcal{M}, \mathcal{O}_\mathcal{M})$ be a holomorphic supermanifold and $X$ a holomorphic vector field on $\mathcal{M}$ and $S$ a supermanifold with a morphism $\phi \in \text{Mor}(S, \mathcal{M})$ and $z_0$ in $\mathbb{C}$.

(1) A “flow for $X$ (with initial condition $\phi$ and with respect to $z_0$)” is an open sub supermanifold $V \subset \mathbb{C}^{1|1} \times S$, such that $\{z_0\} \times S \subset V$ (the bodies of $S$ and $V$) and such that for all $s$ in $S$, $(\mathbb{C} \times \{s\}) \cap V$ is connected, together with a morphism of holomorphic supermanifolds $F : V \to \mathcal{M}$ such that

$$(\text{inv}_{\mathbb{C}^{1|1}})^* \circ (\partial_{\xi} + \partial_z) \circ F^* = (\text{inv}_{\mathbb{C}^{1|1}})^* \circ F^* \circ X$$ and

$$F \circ \text{inv}_{\{z_0\} \times S} = \phi.$$  

Sometimes we call the supermanifold $V$ (or abusively its body $V$) the “flow domain” of $X$.
(2) A flow domain $V$ of a flow $(\mathcal{V}, F)$ for $X$ is called “fibrewise 1-connected (relative to the projection $\mathcal{V} \to S$)” (or “fibrewise 1-connected over $S$”) if for all $s$ in $S$, $(\mathcal{C} \times \{s\}) \cap V$ is connected and simply connected.

**Remark.** We avoid the term “complex supermanifold” here, since it is often used to describe supermanifolds that are, as ringed spaces, locally isomorphic to open sets $D \subset \mathcal{C}^{k}$ with structure sheaf $\mathcal{O}_{D} \otimes \mathcal{C}^{k}$. “Holomorphic supermanifolds” are of course locally isomorphic to open sets $D \subset \mathcal{C}^{k}$ with structure sheaf $\mathcal{O}_{D} \otimes \mathcal{C}^{k}$, where $\mathcal{O}_{D}$ denotes the sheaf of holomorphic functions on $D$.

Let us first give the holomorphic analogue of Lemma 2.1.

**Lemma 5.2.** Let $U \subset \mathcal{C}^{m|n}$ and $W \subset \mathcal{C}^{p|q}$ be superdomains, $X$ a holomorphic vector field on $W$ (not necessarily homogeneous), $\phi$ in $\text{Mor}(U, W)$ and $z_{0}$ in $\mathcal{C}$. Then

(i) it exists a holomorphic flow $(\tilde{V}, \tilde{F})$ for the reduced holomorphic vector field $\tilde{X}$ on $U$ with initial condition $\tilde{\phi}$ with respect to $z_{0}$ such that the flow domain $V \subset \mathcal{C} \times U$ is fibrewise 1-connected over $U$. Furthermore on every flow domain in the sense of Definition 5.1 the holomorphic flow is unique.

(ii) Let now $(\tilde{V}, \tilde{F})$ be a fibrewise 1-connected flow domain for $\tilde{X}$ over $U$. Then there exists a unique holomorphic flow $F : V \to W$ for $X$, with $V$ the open sub supermanifold of $\mathcal{C}^{1|1} \times U$ with body equal to $V$.

**Remark.** The example of the holomorphic vector field $X = (w^{2} + w^{3}\xi_{1}\xi_{2})\frac{\partial}{\partial w}$ on $W = \mathcal{C}^{1|2}$ with coordinates $(w, \xi_{1}, \xi_{2})$ shows that the condition of fibrewise 1-connectedness of $V$ is not only a technical assumption to our proof. The underlying vector field $\tilde{X} = w^{2}\frac{\partial}{\partial w}$ on $\mathcal{C}$, with initial condition $\tilde{\phi} = id : \mathcal{C} \to \mathcal{C}$ with respect to $z_{0} = 0$, can be integrated to the flow $\tilde{F} : V = \mathcal{C}^{2}\backslash\{z \cdot w = 1\} \to \mathcal{C}$, $\tilde{F}(z, w) = \frac{1}{1-zw}$ for $w \neq 0$ and $\tilde{F}(z, 0) = 0$. Obviously, for $w \neq 0$, $(\mathcal{C} \times \{w\}) \cap V$ is connected, but not simply connected. Direct inspection now shows that the flow $F$ of $X$ with initial condition $\phi = id$ and with respect to $z_{0} = 0$ cannot be defined on the whole of $V = (V, \mathcal{O}_{\mathcal{C}^{1|1} \times \mathcal{C}^{1|2}}|_{V})$.

**Proof of Lemma 5.2.** (i) The existence (and the stated unicity property) of a flow $(\tilde{V}, \tilde{F})$ for $\tilde{X}$, with $\{z_{0}\} \times U \subset \tilde{V} \subset \mathcal{C} \times U$ fulfilling the initial condition $\tilde{\phi}$ with respect to $z_{0} \in \mathcal{C}$ is of course a classical application of the existence of solutions of holomorphic ordinary differential equations (see, e.g., [5]). Upon reducing the size of $\tilde{V}$ we always find flow domains that are fibrewise 1-connected.

(ii) The induction procedure of the proof of Lemma 2.1 can be applied here upon recalling the following standard facts from the theory of holomorphic linear ordinary differential equations (compare, e.g., [5]):
Fact 1. Let \( \Omega \subset \mathbb{C} \) be open and 1-connected (i.e. connected and simply connected), and \( z_0 \in \Omega \). If \( A : \Omega \to \text{Mat}(N \times N, \mathbb{C}) \) and \( b : \Omega \to \mathbb{C}^N \) are holomorphic and \( \psi_0 \in \mathbb{C}^N \), then there exists a unique holomorphic map \( \psi : \Omega \to \mathbb{C}^N \) fulfilling
\[
\frac{\partial}{\partial z}\psi(z) = A(z)\psi(z) + b(z)
\]
such that \( \psi(z_0) = \psi_0 \).

Fact 2. Let \( \Omega \) and \( z_0 \) be as in Fact 1, and let \( P \) be a holomorphic manifold ("a parameter space"), and let \( A : \Omega \times P \to \text{Mat}(N \times N, \mathbb{C}) \), \( b : \Omega \times P \to \mathbb{C}^N \), as well as \( \psi_0 : P \to \mathbb{C}^N \) be holomorphic maps. Then there exists a unique holomorphic map \( \psi : \Omega \times P \to \mathbb{C}^N \) fulfilling
\[
\frac{\partial}{\partial z}\psi(z,x) = A(z,x)\psi(z,x) + b(z,x)
\]
such that \( \psi(z_0, x) = \psi_0(x) \), \( \forall x \in P \).

Obviously, to apply these facts in our context, we need the fibrewise 1-connectivity of the “underlying flow domain” \( V \) for \( \tilde{X} \).

Before stating and proving our central result in the holomorphic case, we give the following useful shorthand.

**Definition 5.3.** Let \( S \) be a supermanifold, \( z_0 \) in \( \mathbb{C} \) and \( N \subset \mathbb{C}^{1|1} \times S \) be an open sub supermanifold containing \( \{z_0\} \times S \). Then \( N^{z_0} \) is defined as the open sub supermanifold of \( N = (N, \mathcal{O}_N) \) whose body equals \( \bigcup_{s \in S} ((\mathbb{C} \times \{s\}) \cap N)^{(z_0, s)} \), where \( ((\mathbb{C} \times \{s\}) \cap N)^{(z_0, s)} \) is the connected component of \( (\mathbb{C} \times \{s\}) \cap N \) containing \( (z_0, s) \).

**Remark.** A flow domain \( V \) in the sense of Definition 5.1 is always open and contains \( \{z_0\} \times S \). Furthermore for all \( s \) in \( S \) the section \( (\mathbb{C} \times \{s\}) \cap V \) is connected. The preceding definition will in fact be useful for discussing intersections of flow domains in the next theorem.

**Theorem 5.4.** Let \( M \) be a holomorphic supermanifold and \( X \) a holomorphic vector field on \( M \), and let \( S \) be a holomorphic supermanifold with a holomorphic morphism \( \psi : S \to M \), and \( z_0 \in \mathbb{C} \). Then

(i) there exists a flow \((V, \tilde{F})\) for the reduced vector field \( \tilde{X} \) with initial condition \( \phi \) with respect to \( z_0 \) such that the flow domain \( V \subset \mathbb{C} \times S \) is fibrewise 1-connected over \( S \),

(ii) if \((V, \tilde{F})\) is as in (i), then there exists a unique flow for \( X \) with initial condition \( \phi \) with respect to \( z_0 \), \( F : V \to M \), where \( V \subset \mathbb{C}^{1|1} \times S \) is the open sub supermanifold with body \( V \).
(iii) if \((V_1, F_1)\) and \((V_2, F_2)\) are two flows for \(X\), both with initial condition \(\phi\) with respect to \(z_0\), then \(F_1 = F_2\) on the flow domain \((V_1 \cap V_2)^{z_0}\).

(iv) there exists maximal flow domains for \(X\) and the germs of their flows coincide on \(\{z_0\} \times S\).

**Proof.** (i) and (ii). It easily follows from Lemma 5.2 that \(S\) can be covered by open sub supermanifolds \(\{U^\alpha | \alpha \in A\}\) such that \(X|_{U^\alpha}\) has a holomorphic flow with initial condition \(\phi|_{U^\alpha}\) with respect to \(z_0\), \(F^\alpha : V^\alpha = \Delta_{r_\alpha}(z_0) \times C^{01} \times U^\alpha \to M\), where \(r_\alpha > 0\), and for \(r > 0\), \(\Delta_r(z_0)\) is the open disc of radius \(r\) centred in \(z_0\). Since \(F^\alpha = F^\beta\) on \(V^\alpha \cap V^\beta\) by the unicity part of Lemma 5.2, we can glue these flows to obtain \(\bigcup_{\alpha \in A} V^\alpha \to M\), a flow for \(X\) on \(M\) with initial condition \(\phi\) with respect to \(z_0\). Obviously, the “fibres” \((C \times \{s\}) \cap V\) are 1-connected for all \(s\) in \(S\), i.e., the flow domain \(V\) is fibrewise 1-connected over \(S\).

Note that if we have a flow \(\tilde{F}\) for the reduced vector field \(\tilde{X}\) on a flow domain \(V\) that is fibrewise 1-connected over \(S\), then part (ii) of Lemma 5.2 yields a flow for \(X\) defined on \(V = C^{11} \times S\).

(iii) The body of \((V_1 \cap V_2)^{z_0}\) has as a strong deformation retract the body of \(\{z_0\} \times S\). Without loss of generality we can assume that \(S\) and thus \((V_1 \cap V_2)^{z_0}\) are connected. The local unicity in Lemma 5.2 together with the identity principle for holomorphic morphisms of holomorphic supermanifolds imply that \(F_1 = F_2\) on \((V_1 \cap V_2)^{z_0}\).

(iv) By Zorn’s lemma we get maximal flow domains and by part (iii) the corresponding flows coincide near \(\{z_0\} \times S\). □

**Remarks.** (1) The non-unicity of maximal flow domains for holomorphic vector fields is a well-known phenomenon already in the ungraded case. A simple example for this is the vector field \(X\) on \(C^*\) such that \(X(w) = \frac{1}{w} \frac{dw}{dx}\) for all \(w\) in \(C^*\).

(2) Given the above theorem, the analogues of Lemma 3.1, Proposition 3.3, and Theorem 3.4 can now without difficulty be proven to hold for holomorphic supermanifolds.

**Acknowledgements.** We gratefully acknowledge the support of the SFB/TR 12 Symmetries and Universality in Mesoscopic Systems program of the Deutsche Forschungsgemeinschaft. Furthermore, we thank the referee for very helpful comments.
References


Stéphane Garnier  
Fakultät für Mathematik  
Ruhr-Universität Bochum  
D-44780 Bochum, Germany  
Stephane.Garnier@rub.de

Tilmann Wurzbacher  
LMAM-UMR 7122  
Université de Lorraine et C.N.R.S.  
F-57045 Metz, France  
Tilmann.Wurzbacher@rub.de
PROJECTIVE VARIETIES
WITH BAD SEMI-STABLE REDUCTION AT 3 ONLY

TO I. R. SHAFAREVICH, ON THE OCCASION OF HIS 90TH BIRTHDAY

VICTOR ABRASHKIN

Received: April 30, 2013
Communicated by Peter Schneider

Abstract. Suppose \( F = W(k)[1/p] \) where \( W(k) \) is the ring of Witt vectors with coefficients in algebraically closed field \( k \) of characteristic \( p \neq 2 \). We construct integral theory of \( p \)-adic semi-stable representations of the absolute Galois group of \( F \) with Hodge-Tate weights from \( [0, p) \). This modification of Breuil’s theory results in the following application in the spirit of the Shafarevich Conjecture. If \( Y \) is a projective algebraic variety over \( \mathbb{Q} \) with good reduction modulo all primes \( l \neq 3 \) and semi-stable reduction modulo 3 then for the Hodge numbers of \( Y_\mathbb{C} = Y \otimes_\mathbb{Q} \mathbb{C} \), one has \( h^2(Y_\mathbb{C}) = h^{1,1}(Y_\mathbb{C}) \).

2010 Mathematics Subject Classification: 11S20, 11G35, 14K15
Keywords and Phrases: \( p \)-adic semi-stable representations, Shafarevich Conjecture

Introduction
Everywhere in the paper \( p \) is a fixed prime number, \( p \neq 2 \), \( k \) is algebraically closed field of characteristic \( p \), \( F \) is the fraction field of the ring of Witt vectors \( W(k) \), \( \bar{F} \) is a fixed algebraic closure of \( F \) and \( \Gamma_F = \text{Gal}(\bar{F}/F) \) is the absolute Galois group of \( F \).
Suppose \( Y \) is a projective algebraic variety over \( \mathbb{Q} \). Denote by \( Y_\mathbb{C} \) the corresponding complex variety \( Y \otimes_\mathbb{Q} \mathbb{C} \). For integers \( n, m \geq 0 \), set \( h^n(Y_\mathbb{C}) = \dim_\mathbb{C} H^n(Y_\mathbb{C}, \mathbb{C}) \) and \( h^{n,m}(Y_\mathbb{C}) = \dim_\mathbb{C} H^n(\Omega^m_{Y_\mathbb{C}}) \).
The main result of this paper can be stated as follows.

Theorem 0.1. If \( Y \) has semi-stable reduction modulo 3 and good reduction modulo all primes \( l \neq 3 \) then \( h^2(Y_\mathbb{C}) = h^{1,1}(Y_\mathbb{C}) \).
Remind that a generalization of the Shafarevich Conjecture about the non-existence of non-trivial abelian varieties over $\mathbb{Q}$ with everywhere good reduction was proved by Fontaine [16] and the author [2], and states that

\[ h^1(Y_C) = h^3(Y_C) = 0, \quad h^2(Y_C) = h^{1,1}(Y_C) \]

if $Y$ has everywhere good reduction. (The Shafarevich Conjecture appears then as the equality $h^1(Y_C) = 0$.) This result became possible due to the following two important achievements of Fontaine’s theory of $p$-adic crystalline representations:

— the Fontaine-Messing theorem relating etale and de Rham cohomology of smooth proper schemes over $W(k)$ in dimensions $[0, p)$, [15] (it was later proved by Faltings in full generality, [12]);

— the Fontaine-Laffaille integral theory of crystalline representations of $\Gamma_F$ with Hodge-Tate weights from $[0, p-2]$, [13].

Note that the Fontaine-Laffaille theory works essentially for Hodge-Tate weights from $[0, p)$ but does not give all Galois invariant lattices in the corresponding crystalline representations. Nevertheless, this theory admits improvement developed by the author in [1]. As a result, there was obtained a suitable integral theory for the case of Hodge-Tate weights from $[0, p)$, which allowed us to prove some extras to statements (0.1), in particular, that modulo the Generalized Riemann Hypothesis one has $h^4(Y_C) = h^{2,2}(Y_C)$.

Since that time there was a huge progress in the study of semi-stable $p$-adic representations. Tsuji [23] proved a semi-stable case of the relation between etale and crystalline cohomology and Breuil [5, 6] developed an analogue of the Fontaine-Laffaille theory in the context of semi-stable representations (even for ramified basic fields). The papers [4] and [21] studied the problem of the existence of abelian varieties over $\mathbb{Q}$ with only one prime of bad semi-stable reduction. Note that the progress in this direction is quite restrictive because our knowledge of algebraic number fields with prescribed ramification at a given prime number $p$ (and unramified outside $p$) is very far from to be complete.

Theorem 0.1 represents an exceptional situation where the standard tools: the Odlyzko estimates of the minimal discriminants of algebraic number fields and the modern computing facilities (SAGE) are sufficient to resolve upcoming problems. In addition, the proof of this theorem requires a modification of Breuil’s theory to work with semi-stable representations of $\Gamma_F$ with Hodge-Tate weights from $[0, p)$.

The structure of this paper can be described as follows.

In Section 1 we introduce the category $\mathcal{L}_\psi^*$ of filtered ($\varphi, N$)-modules over $\mathcal{W}_1 := k[[u]]$. This is a special pre-abelian category, that is an additive category with kernels, cokernels and sufficiently nice behaving short exact sequences. Note that such categories play quite appreciable role in all our constructions. In Section 2 we construct the functor $\mathcal{V}^*$ from $\mathcal{L}_\psi^*$ to the category of $\mathbb{F}_p[\Gamma_F]$-modules $\mathbb{M}_\psi$ by introducing a “truncated” version of Fontaine’s ring of semi-stable periods $A_{st}$. The functor $\mathcal{V}^*$ is not fully faithful but by taking into account the maximal etale subobjects of filtered modules from $\mathcal{L}_\psi^*$ we define a
modification $CV^*$ of $V^*$. This functor gives already a fully faithful functor from $L^*$ to the category of cofiltered $\Gamma_F$-modules $CM\Gamma_F$. In Section 3 we give an interpretation of Breuil’s theory in terms of $W := W(k[[u]])$-modules (Breuil worked with modules over the divided powers envelope of $W$) by introducing the category of filtered $(\varphi,N)$-modules $L^{ft}$ over $W$. The advantage of this construction is that the objects of this category appear as strict subquotients of $p$-divisible groups in suitable pre-abelian category. This allows us to use devissage despite that all involved categories are not abelian. We also introduce the subcategories $L^{u,ft}$ and, resp., $L^{m,ft}$ of unipotent and, resp., multiplicative objects in $L^{ft}$ and prove that any $L \in L^{ft}$ is a canonical extension

\[(0.2) \quad 0 \to L^u \to L \to L^m \to 0\]

of a multiplicative object $L^m$ by a unipotent object $L^u$. In Section 4 we study Breuil’s functor $V^{ft} : L^{ft} \to M\Gamma_F$ in the situation of Hodge-Tate weights from $[0,p)$. We show that on the subcategory $L^{u,ft}$ this functor is still fully faithful by proving that on the subcategory of killed by $p$ unipotent objects the functors $V^{ft}$ and $V^*$ coincide. Then we show that for any killed by $p$ object $L$ of $L^{ft}$, the functor $V^{ft}$ transforms the standard short exact sequence (0.2) into a short exact sequence in $M\Gamma_F$, which admits a functorial splitting. This splitting is used then to construct a modified version $\tilde{CV}^{ft} : L^{ft} \to CM\Gamma_F$ of $V^{ft}$, which is already fully faithful. This gives us an efficient control on all Galois invariant lattices of semi-stable representations with weights from $[0,p)$. Especially, we have an explicit description of all killed by $p$ subquotients of such lattices and the corresponding ramification estimates. Finally, in Section 5 we give a proof of Theorem 0.1 following the strategy from [2].

Essentially, we obtain the following result: if $V$ is a 3-adic representation of $\Gamma_{\overline{Q}} = Gal(\overline{Q}/Q)$ which is unramified outside 3 and is semi-stable at 3 then there is a $\Gamma_{\overline{Q}}$-equivariant filtration by $\mathbb{Q}_3$-subspaces $V = V_0 \supset V_1 \supset V_2 \supset V_3 = 0$ such that for $0 \leq i \leq 2$, the $\Gamma_{\overline{Q}}$-module $V_i/V_{i+1}$ is isomorphic to the product of finitely many copies of the Tate twist $\mathbb{Q}_3(i)$. If $V = H^2_{et}(Y_{\mathbb{C}}, \mathbb{Q}_3)$ then looking at the eigenvalues of the Frobenius morphisms of reductions modulo $l \neq 3$, we obtain that $V = V_1$ and $V_2 = 0$, and this implies that $h^2(Y_{\mathbb{C}}) = h^{1,1}(Y_{\mathbb{C}})$

Note that our construction of the modification of Breuil’s functor gives automatically the modification of the Fontaine-Laffaille functor, which essentially coincides with the modification constructed in [1]. It is worth mentioning that switching from Breuil’s $S$-modules to $W$-modules means moving in the direction of Kisin’s approach [18] and recent approach to integral theory of $p$-adic representations by Liu [19, 20]. It would be also very interesting to study the opportunity to modify Breuil’s functor over ramified base [8, 9] to the case of Hodge-Tate weights from $[0,p)$. Finally, mention quite surprising matching of the ramification estimates for semi-stable representations and the Leopoldt conjecture for the field $Q(\sqrt[3]{3}, \zeta_9)$, cf. Section 5.

Acknowledgements. The author is very grateful to Shin Hattori for numerous and helpful discussions.
1. The categories $\mathcal{L}_0^*, \mathcal{L}_1^*, \mathcal{L}_2^*$

Remind that $k$ is algebraically closed field of characteristic $p > 2$. Let $W = W(k)[[x]]$, where $W(k)$ is the ring of Witt vectors with coefficients in $k$ and $u$ is an indeterminate. Denote by $\sigma$ the automorphism of $W(k)$ induced by the $p$-th power map on $k$ and agree to use the same symbol for its continuous extension to $W$ such that $\sigma(u) = u^p$. Denote by $N : W \to W$ the continuous $W(k)$-linear derivation such that $N(u) = -u$.

We shall often use below the following statement.

**Lemma 1.1.** Suppose $L$ is a finitely generated $W$-module and $A$ is a $\sigma$-linear operator on $L$. Then the operator $id_L - A$ is epimorphic. If, in addition, $A$ is nilpotent then $id_L - A$ is bijective.

**Proof.** Part b) is obvious. In order to prove a) notice first that we can replace $L$ by $L/uL$ and, therefore, assume that $L$ is a finitely generated $W(k)$-module. Clearly, it will be enough to consider the case $pL = 0$. Then there is a decomposition of $k$-vector spaces $L = L_1 \oplus L_2$, where $A$ is invertible on $L_1$ and nilpotent on $L_2$. It remains to note that $L_1 = L_0 \otimes_{\mathbb{F}_p} k$, where $L_0$ is a finite dimensional $\mathbb{F}_p$-vector space such that $A|_{L_0} = id$. The existence of $L_0$ is a standard fact of $\sigma$-linear algebra: if $s = \dim_k L_1$ and $A \in M_s(k)$ is a matrix of $A|_{L_1}$ in some $k$-basis of $L_1$ then $L_0 = \{(x_1, \ldots, x_s) \in k^s \mid (x_1^p, \ldots, x_s^p)A = (x_1, \ldots, x_s)\}$; the $\mathbb{F}_p$-linear space $L_0$ has dimension $s$ because the corresponding equations determine an etale algebra of rank $p^s$ over algebraically closed field $k$.

**Remark.** In above Lemma and everywhere below we use the following agreement: $A$ is nilpotent on $L$ iff it is “topologically nilpotent”, i.e. $\bigcap_n A^n(L) = 0$.

1.1. Definitions and general properties. Let $W_1 = W/pW$ with induced $\sigma$, $\varphi$ and $N$.

**Definition.** The objects of the category $\mathcal{L}_0^*$ are the triples $L = (L, F(L), \varphi)$, where

- $L$ and $F(L)$ are $W_1$-modules such that $L \supset F(L)$;
- $\varphi : F(L) \to L$ is a $\sigma$-linear morphism of $W_1$-modules; (Note that $\varphi(F(L))$ is a $\sigma(W_1)$-submodule in $L$.)

If $L_1 = (L_1, F(L_1), \varphi)$ is also an object of $\mathcal{L}_0^*$, then the morphisms $f \in \text{Hom}_{\mathcal{L}_0^*}(L_1, L)$ are given by $W_1$-linear maps $f : L_1 \to L$ such that $f(F(L_1)) \subset F(L)$ and $f\varphi = \varphi f$.

**Definition.** The objects of the category $\mathcal{L}_1^*$ are the quadruples $L = (L, F(L), \varphi, N)$, where

- $(L, F(L), \varphi)$ is an object of the category $\mathcal{L}_0^*$;
- $N : L \to L/u^nL$ is a $W_1$-differentiation, i.e. for all $w \in W_1$ and $l \in L$, $N(wl) = N(w)(l \mod u^nL) + wN(l)$;
• if \( L_1 = (L_1, F(L_1), \varphi, N) \) is another object of \( L^*_0 \) then the morphisms \( \text{Hom}_{L^*_0}(L_1, L) \) are given by \( f: (L_1, F(L_1), \varphi) \to (L, F(L), \varphi) \) from \( L^*_0 \) such that \( fN = NF \). (We use the same notation \( f \) for the reduction of \( f \) modulo \( u^{2p}L \).

The categories \( L^*_0 \) and \( L^*_1 \) are additive.

**Definition.** The category \( L^*_0 \) is a full subcategory of \( L^*_1 \) consisting of the objects \( L = (L, F(L), \varphi) \) such that

- \( L \) is a free \( W_1 \)-module of finite rank;
- \( F(L) \supset u^{p-1}L \);
- the natural embedding \( \varphi(F(L)) \subset L \) induces the identification \( \varphi(F(L)) \otimes_{\sigma(W_1)} W_1 = L \).

Note that \( \varphi \) induces a map \( F(L)/u^{2p}L \to L/u^{2p}L \); use that \( u^{2p}L \subset u^{p+1}F(L) \subset u^2F(L) \) and \( \varphi(uF(L)) \subset u^{2p}L \). We shall denote this map by the same symbol \( \varphi \).

**Definition.** The category \( L^*_1 \) is a full subcategory of \( L^*_0 \) consisting of the objects \( L = (L, F(L), \varphi, N) \) such that

- \( (L, F(L), \varphi) \in L^*_0 \);
- for all \( l \in F(L), uN(l) \in F(L) \mod u^{2p}L \) and \( N(\varphi(l)) = \varphi(uN(l)) \).

The categories \( L^*_0 \) and \( L^*_1 \) are additive.

In the case of objects \( (L, F(L), \varphi, N) \) of \( L^*_1 \) the morphism \( N \) can be uniquely recovered from the \( W_1 \)-differentiation \( N_1 = N \mod u^pL \) due to the following property.

**Proposition 1.2.** Suppose \( (L, F(L), \varphi) \in L^*_0 \) and \( N_1 : L \to L/u^pL \) is a \( W_1 \)-differentiation such that for any \( m \in F(L), uN_1(m) \in F(L) \mod u^pL \) and \( N_1(\varphi(m)) = \varphi(uN_1(m)) \). Then there is a unique \( W_1 \)-differentiation \( N : L \to L/u^{2p}L \) such that \( N \mod u^p = N_1 \) and \( (L, F(L), \varphi, N) \in L^*_1 \).

**Proof.** Choose a \( W_1 \)-basis \( m_1, \ldots, m_s \) of \( F(L) \). Then \( l_1 = \varphi(m_1), \ldots, l_s = \varphi(m_s) \) is a \( W_1 \)-basis of \( L \) and a \( \sigma(W_1) \)-basis of \( \varphi(F(L)) \).

Let \( N(l_i) := \varphi(uN_1(m_i)) \in L/u^{2p}L \), where \( N_1(m_i) \) are some lifts of \( m_i \) to \( L/u^{2p}L \). Clearly, the elements \( N(l_i) \in \varphi(F(L)) \subset L/u^{2p}L \) are well-defined (use that \( u^{p+1}L \subset u^{2p}L \).

For any \( l = \sum w_i l_i \in L \), let \( N(l) := \sum w_i N(l_i) + \sum w_i N(l_i) \). Then \( N : L \to L/u^{2p}L \) is a \( W_1 \)-differentiation and \( N \mod u^p = N_1 \). Clearly, \( N \) is the only candidate to satisfy the requirements of our Proposition.

Now suppose \( m = \sum w_i m_i \in F(L) \) with all \( w_i \in W_1 \). Then \( N(\varphi(m)) = \sum w_i \sigma(w_i) l_i + \sum w_i \sigma(w_i) uN(m_i) \).

On the other hand, \( \varphi(uN(m)) \) equals

\[
\sum w_i \sigma(N(w_i)) l_i + \sum w_i \varphi(w_i uN(m_i)) = \sum w_i \sigma(w_i) l_i \mod u^{2p}
\]

because all \( \sigma(N(w_i)) \in u^p \sigma(W_1) \).

The proposition is proved.
Remark. By above Proposition in the definition of objects of $\mathcal{L}^*$ one can replace $N : L \rightarrow L/u^p L$ by $N_1 = N \mod u^p L$ and use $N$ as a unique extension of $N_1$ if necessary. An example of the situation where we do need to extend $N_1$ is described in Proposition 1.3 below. Another situation is related to the definition of the truncated version $R^p_\ast$ of $A_\ast$ in Subsection 2. Here we need $N$ to be defined modulo some smaller module than $u^p L$, e.g. $u^{p+1} L$. Our choice was done in favour of the module $u^p L$ because it is the smallest possible module where the definition of $N$ makes sense.

Proposition 1.3. $\mathcal{L}_0^*$ and $\mathcal{L}^*$ are pre-abelian categories (cf. Appendix A for the concept of pre-abelian category).

Proof. Suppose $\mathcal{S}$ is an additive category and $f \in \text{Hom}_\mathcal{S}(A, B)$, $A, B \in \mathcal{S}$. Then $i \in \text{Hom}_\mathcal{S}(K, A)$ is a kernel of $f$ if for any $D \in \mathcal{S}$, the sequence of abelian groups

$$0 \rightarrow \text{Hom}_\mathcal{S}(D, K) \xrightarrow{i} \text{Hom}_\mathcal{S}(D, A) \xrightarrow{f} \text{Hom}_\mathcal{S}(D, B)$$

is exact. Similarly, $j \in \text{Hom}_\mathcal{S}(B, C)$, $B, C \in \mathcal{S}$, is a cokernel of $f$ if for any $D \in \mathcal{S}$, the sequence

$$0 \rightarrow \text{Hom}_\mathcal{S}(C, D) \xrightarrow{j} \text{Hom}_\mathcal{S}(B, D) \xrightarrow{f} \text{Hom}_\mathcal{S}(A, D)$$

is exact.

Let $FF_{\mathcal{W}_1}$ be the category of free $\mathcal{W}_1$-modules with filtration. This category is pre-abelian. More precisely, consider the objects $\mathcal{L} = (L, F(L))$ and $\mathcal{M} = (M, F(M))$ in $FF_{\mathcal{W}_1}$ and let $f \in \text{Hom}_{FF_{\mathcal{W}_1}}(\mathcal{L}, \mathcal{M})$. Then $\text{Ker}_{FF_{\mathcal{W}_1}}f$ is a natural embedding $i_C : K = (K, F(K)) \rightarrow \mathcal{L}$, where $K = \text{Ker}(f : L \rightarrow M)$ and $F(K) = K \cap F(L)$. The cokernel $\text{Coker}_{FF_{\mathcal{W}_1}}f$ of $\text{Ker}_{FF_{\mathcal{W}_1}}f$ appears as a natural projection $j_{\mathcal{L}} : \mathcal{L} \rightarrow \mathcal{L}' = (L', F(L'))$, where $L' = f(L)$ and $F(L') = f(F(L))$. Similarly, $\text{Ker}_{FF_{\mathcal{W}_1}}f$ is a natural projection $j_M : \mathcal{M} \rightarrow \mathcal{C} = (C, F(C))$, where $C = (M/L')/(M/L)_{\text{tor}}$ and $F(C) = j_M(F(M))$. Then the image $\text{Im}_{FF_{\mathcal{W}_1}}f = \text{Ker}_{FF_{\mathcal{W}_1}}(\text{Coker}_{FF_{\mathcal{W}_1}}f)$ is a natural embedding $M' = (M', F(M')) \rightarrow \mathcal{M}$, where $M'$ is the kernel of $j_M$ and $F(M') = F(M) \cap M'$. As usually, there is a natural map $\mathcal{L}' \rightarrow M'$ induced by $L' \subset M'$. Note that $M/M' = C$ is free and $M'/L'$ is torsion $\mathcal{W}_1$-modules and these properties completely characterize $M'$ as a $\mathcal{W}_1$-submodule of $M$.

Now suppose $\mathcal{L} = (L, F(L), \varphi)$, $\mathcal{M} = (M, F(M), \varphi)$ are objects of $\mathcal{L}_0^*$ and $f \in \text{Hom}_{\mathcal{L}_0^*}(\mathcal{L}, \mathcal{M})$. Use the obvious forgetful functor $\mathcal{L}_0^* \rightarrow FF_{\mathcal{W}_1}$ and the same notation for the corresponding images of $\mathcal{L}$, $\mathcal{M}$ and $f$. Show that $K = \text{Ker}_{FF_{\mathcal{W}_1}}f$ and $\mathcal{C} = \text{Coker}_{FF_{\mathcal{W}_1}}f$ have the natural structures of objects of $\mathcal{L}_0^*$ and with respect to this structure they become the kernel and, resp, cokernel of $f$ in $\mathcal{L}_0^*$. Indeed, $u^{p-1} K = u^{p-1} L \cap K \subset F(L) \cap K = F(K) = \text{Ker}(f : F(L) \rightarrow F(M))$. Therefore, $\varphi(F(K)) \subset K \cap \varphi(F(L))$ and there is a natural embedding $\iota : \varphi(F(K)) \otimes_{\mathcal{W}_1} \mathcal{W}_1 \subset K$. On the one hand,

$$\text{rk}_{\mathcal{W}_1}(\varphi(F(K))) = \text{rk}_{\mathcal{W}_1} F(K) = \text{rk}_{\mathcal{W}_1} K.$$
On the other hand, $F(L)/F(K) \subset L/K = L'$ have no $W_1$-torsion. This implies that the quotient $\varphi(F(L))/\varphi(F(K))$ has no $\sigma W_1$-torsion and the factor of $L = \varphi(F(L))$, $\mathcal{W}_1$ by $\varphi(F(K)) \otimes_{\sigma W_1} W_1$ also has no $W_1$-torsion. So, $i$ becomes the equality $\varphi(F(K)) \otimes_{\sigma W_1} W_1 = K$ and $K = (K, F(K), \varphi) = \text{Ker} F \varphi$. 

The above description of $\text{Ker} F \varphi$ implies that $u^{p-1} L' \subset F(L')$, $\varphi(F(L')) = \varphi(F(M))/\varphi(F(K))$ and $L' = \varphi(F(L')) \otimes_{\sigma W_1} W_1$. In other words, $L' = (L', F(L'), \varphi) \in \mathcal{L}_L$. 

Now note that for $M' = (M', F(M'))$, we have 

$$u^{p-1} M' = (u^{p-1} M) \cap M' \subset F(M') \cap M' = F(M')$$

and, therefore, $F(M')/F(L')$ is torsion $W_1$-module and 

- $(\varphi(F(M')) \otimes_{\sigma W_1} W_1)/L'$ is torsion $W_1$-module; 

On the other hand, $F(M)/F(M') = F(C)$ is torsion free $W_1$-module. This implies that $\varphi(F(M))/F(M')$ is torsion free $\sigma W_1$-module and, therefore, 

- $M/(\varphi(F(M')) \otimes_{\sigma W_1} W_1)$ is torsion free $W_1$-module.

The above two conditions completely characterize $M'$ as a submodule of $M$. Therefore, $\varphi(F(M')) \otimes_{\sigma W_1} W_1 = M'$, $M' = (M', F(M'), \varphi) \in \mathcal{L}_L$ and $(M/M', F(M))/F(M')$, $\varphi = (C, F(C), \varphi) = \mathcal{C} \in \mathcal{L}_L$. Now a formal verification shows that $\mathcal{C} = \text{Coker} F \varphi$.

Again $\text{Coim} F \varphi = (L', F(L'), \varphi)$ and $\text{Im} F \varphi = (M', F(M'), \varphi)$ together with their natural embedding $\text{Coim} F \varphi \hookrightarrow \text{Im} F \varphi$ in $\mathcal{L}_L$. As a matter of fact, these two objects of the category $\mathcal{L}_L$ do not differ very much due to the following Lemma.

**Lemma 1.4.** $\varphi(F(L')) \supset u^p \varphi(F(M'))$ (and, therefore, $L' \supset u^p M'$).

**Proof of Lemma.** Otherwise, there is an $l \in \varphi(F(L')) \setminus u^p \varphi(F(L'))$ such that $l \in u^{2p} \varphi(F(M'))$.

Form the sequence $l_n \in L'$ such that $l_1 = l$ and for all $n \geq 2$, $l_{n+1} = \varphi(u^n l_n)$, where $a_n \geq 0$ is such that $u^{a_n} l_n \in F(L') \setminus u F(L')$. Clearly, all $l_n \notin u F(L') \supset u^p L'$. 

On the other hand, $l \in u^{2p} \varphi(F(M')) \subset u^{p+1} F(M')$ and, therefore, for all $n \geq 1$, $l_n \in \varphi(u^{2p} M') \subset u^{p+1} u^p M'$. So, for $n \gg 0$, $l_n \in u^p L'$. The contradiction. \qed

Now suppose $\mathcal{L} = (L, F(L), \varphi, N)$ and $M = (M, F(M), \varphi, N)$ are objects of $\mathcal{L}_L$, and $f \in \text{Hom}_{\mathcal{L}_L} (L, M)$. Prove that the kernel $(K, F(K), \varphi)$ and the cokernel $(C, F(C), \varphi)$ of $f$ in the category $\mathcal{L}_L$ have a natural structure of objects of the category $\mathcal{L}_L^*$. 

Clearly, $\mathcal{N}(K) \subset \text{Ker} f \mod u^{2p} : L/u^{2p} L \rightarrow M/u^{2p} M$. The above Lemma 1.4 implies that $u^p L' \supset u^p M'$ and we obtain the following natural maps 

$$L'/u^p L' \xrightarrow{\alpha} L'/u^{2p} M' \xrightarrow{\beta} M'/u^{2p} M' \xrightarrow{\gamma} M/u^{2p} M.$$ 

Note that $\alpha$ is epimorphic but $\beta$ and $\gamma$ are monomorphic. This implies that $\mathcal{N}(K) \subset \text{Ker} F \varphi L/u^p L \rightarrow L'/u^p L'$ and $\mathcal{N}(K) \mod u^p L \subset \text{Ker} L/u^p L \rightarrow L'/u^p L' = K/u^p K$.
Therefore, by Proposition 1.2, \( N \) (as a unique lift of \( N_1 = N \ mod \ u^p \)) maps \( K \) to \( K/u^pK \) and \( (K, F(K), \varphi, N) \in \mathcal{L}^* \).

The above property of Ker\( L \), \( f \) implies that \( N(L') \subset L'/u^pL' \). Now use that \( u^pM' \subset L' \), \( u^pL' \subset u^pM' \) and \( N(u^pM') \subset u^pM/u^pM \) to deduce that

\[
N(u^pM') \subset (L'/u^pM') \cap (u^pM/u^pM) = u^pM'/u^pM'.
\]

So, \( N \ mod \ u^p \) maps \( M' \) to \( M'/u^pM' \) and again by Proposition 1.2 \( N(M') \subset M'/u^pM' \). This means that the kernel of the above constructed Coker\( f : (M, F(M), \varphi) \rightarrow (C, F(C), \varphi) \) is provided with the structure of object of the category \( \mathcal{L}^* \). Therefore, \( N \) induces the map \( N : C \rightarrow C/u^pC \) and \( (C, F(C), \varphi, N) \in \mathcal{L}^* \). The proposition is proved.

The above proof shows that the kernels and cokernels in the category \( \mathcal{L}^* \) appear on the level of filtered modules as the kernel and cokernel of the corresponding map of filtered modules \( (L_1, F(L_1)) \) to \( (L, F(L)) \) in the category of filtered \( \mathcal{W}_1 \)-modules. Therefore, the category \( \mathcal{L}^* \) is special, cf. Appendix A, and we can apply the corresponding formalism of short exact sequences. In particular, if we take another object \( \mathcal{L}_2 = (L_2, F(L_2), \varphi, N) \in \mathcal{L}^* \) then

- \( i \in \text{Hom}_{\mathcal{L}}(L_1, \mathcal{L}) \) is strict monomorphism iff \( i : L_1 \rightarrow L \) is injective and \( i(L_1) \cap F(L) = i(F(L_1)) \);

- \( j \in \text{Hom}_{\mathcal{L}}(\mathcal{L}, \mathcal{L}_2) \) is strict epimorphism iff \( j : L \rightarrow L_2 \) is epimorphic and \( j(F(L)) = F(L_2) \).

As usually, cf. Appendix A, if a monomorphism \( i \) is strict then the monomorphism \( i = \text{Ker} j \) is strict, and if an epimorphism \( j \) is strict then the monomorphism \( i = \text{Ker} j \) is strict, and under these assumptions \( 0 \rightarrow \mathcal{L}_1 \xrightarrow{i} \mathcal{L} \xrightarrow{j} \mathcal{L}_2 \rightarrow 0 \) is short exact sequence.

With relation to the above result that the categories \( \mathcal{L}_0^* \) and \( \mathcal{L}^* \) are pre-abelian, note that the situation with the categories \( \mathcal{L}_0^* \) and \( \mathcal{L}^* \) is different. Indeed, let \( F\mathcal{M}_{\mathcal{W}_1} \) be the category of filtered (not necessarily free) modules over \( \mathcal{W}_1 \). This category is pre-abelian: for \( \mathcal{M}_i = (M_i, F(M_i)), i = 1, 2, \) and \( f \in \text{Hom}_{F\mathcal{M}_{\mathcal{W}_1}}(\mathcal{M}_1, \mathcal{M}_2) \), we have the following equalities Ker\( F\mathcal{M}_{\mathcal{W}_1} f = (\text{Ker} f, \text{Ker} f \cap F(M_1)) \) (together with its natural embedding into \( \mathcal{M}_1 \)) and Coker\( F\mathcal{M}_{\mathcal{W}_1} f = (\text{Coker} f, F(M_2))/(f(M_1) \cap F(M_2)) \) (together with the natural projection from \( \mathcal{M}_2 \)).

Now suppose that \( \mathcal{M}_i = (M_i, F(M_i), \varphi) \in \mathcal{L}_0^*, i = 1, 2, \) and \( f \in \text{Hom}_{\mathcal{L}_0^*}(\mathcal{M}_1, \mathcal{M}_2) \). Then Ker\( F\mathcal{M}_{\mathcal{W}_1} f \) exists (and coincides on the level of filtered modules with Ker\( F\mathcal{M}_{\mathcal{W}_1} f \)) but Coker\( F\mathcal{M}_{\mathcal{W}_1} f \) exists (and coincides on the level of filtered modules with Coker\( F\mathcal{M}_{\mathcal{W}_1} f \)) only if we have \( f(F(M_1)) = f(M_1) \cap F(M_2) \). In particular, on the level of filtered modules the composition Coker\( F\mathcal{M}_{\mathcal{W}_1} (\text{Ker} F\mathcal{M}_{\mathcal{W}_1} f) \) always makes sense and coincides with the natural projection \( \mathcal{M}_1 \rightarrow (f(M_1), f(F(M_1))) \). Therefore, one can introduce the concept of strict epimorphism in \( \mathcal{L}_0^* \): \( f \) is strict epimorphism iff \( f(M_1) = M_2 \) and \( f(F(M_1)) = F(M_2) \).
The following situation will appear several times below.

**Lemma 1.5.** Suppose \( M_1, M_2 \in \mathcal{L}_0^* \), \( \iota \in \text{Hom}_{\mathcal{L}_0^*}(M_1, M_2) \) is a strict epimorphism and \( \text{Ker}_{\mathcal{L}_0^*}\iota = (K, K, \varphi) \). Then for any \( \mathcal{L} \in \mathcal{L}_0^* \),

\[
\iota^* : \text{Hom}_{\mathcal{L}_0^*}(\mathcal{L}, M_1) \to \text{Hom}_{\mathcal{L}_0^*}(\mathcal{L}, M_2)
\]

is epimorphic. In addition, if \( \varphi|_K \) is nilpotent then \( \iota^* \) is bijective.

**Proof.** The structure of \( \mathcal{L} = (L, F(L), \varphi) \) can be given by a vector \( \bar{l} = (l_1, \ldots, l_s) \) and a matrix \( C \in M_s(W_1) \) such that

- \( l_1, \ldots, l_s \) is a \( W_1 \)-basis of \( L \);
- if \( lC = \bar{m} = (m_1, \ldots, m_s) \) then \( m_1, \ldots, m_s \) is a \( W_1 \)-basis of \( F(L) \);
- \( \bar{m} = \varphi(\bar{m}) := (\varphi(m_1), \ldots, \varphi(m_s)) \).

Suppose \( M_1 = (M_1, F(M_1), \varphi) \) and \( M_2 = (M_2, F(M_2), \varphi) \). Any \( f \in \text{Hom}_{\mathcal{L}_0^*}(\mathcal{L}, M_2) \) is given by \( f(\bar{l}) \in M_2^* \) such that \( f(\bar{l})C \in F(M_2)^* \) and \( \varphi(f(\bar{l})C) = f(\bar{l}) \).

Choose an \( \tilde{f}(\bar{l}) \in M_2^* \) such that \( \tilde{f}(\bar{l}) \mod K = f(\bar{l}) \). Then \( \tilde{f}(\bar{l})C \mod K \) belongs to \( F(M_2)^* \) and, therefore, \( \tilde{f}(\bar{l})C \in F(M_2)^* \). Clearly, we have that \( \tilde{k}_0 := \varphi(\tilde{f}(\bar{l})C) - f(\bar{l}) \in K^* \). We must prove the existence of \( \tilde{k}_1 \in K^* \) such that \( \varphi((\tilde{f}(\bar{l}) + \tilde{k}_1)C) = \tilde{f}(\bar{l}) + \tilde{k}_1 \). This is equivalent to

\[
\tilde{k}_1 - \varphi(\tilde{k}_1C) = \tilde{k}_0
\]

and the existence of \( \tilde{k}_1 \) follows from Lemma 1.1. This proves that \( \iota^* \) is surjective. If \( \varphi|_K \) is nilpotent then the bijectivity of \( \iota^* \) follows in a similar way from part b) of Lemma 1.1. \( \square \)

1.2. **Standard exact sequences.** Suppose \( \mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}_0^* \). Introduce a \( \sigma \)-linear map \( \phi : L \to L \) by \( \phi : l \mapsto \varphi(u^{p-1}l) \).

**Definition.** The object \( \mathcal{L} \) is etale (resp., connected) if \( \phi \mod u \) is invertible (resp., nilpotent) on \( L/uL \).

Let \( \mathcal{L}(0) = (W_1, W_1 u^{p-1}, \varphi, N) \in \mathcal{L}_0^* \), where \( \varphi(u^{p-1}) = 1 \) and \( N(1) = u^p \mod u^{2p} \). Then \( \mathcal{L}(0) \) is etale. As a matter of fact, it is the simplest etale object of \( \mathcal{L}_0^* \) due to the following Lemma.

**Lemma 1.6.** Suppose \( \mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}_0^* \) is etale. Then \( \mathcal{L} \) is a product of finitely many copies of \( \mathcal{L}(0) \).

**Proof.** If \( \tilde{L}_0 = \{l \in L/uL \mid \phi(l) = l\} \) then \( L/uL = \tilde{L}_0 \otimes_{\mathbb{F}_p} k \). Then there is a unique \( \mathbb{F}_p \)-submodule \( L_0 \) of \( L \) such that \( \phi|_{L_0} = \text{id} \) and \( L = L_0 \otimes_{\mathbb{F}_p} W_1 \).

Suppose \( l \in L_0 \). Then \( \varphi(u^{p-1}l) = l \) and \( N(l) = N(\varphi(u^{p-1}l)) = \varphi(uN(u^{p-1}l)) = \varphi(u^p(l \mod u^{2p}) + u^pN(l)) = u^p l \mod u^{2p} L \). Therefore, if \( e_1, \ldots, e_s \) is an \( \mathbb{F}_p \)-basis of \( L_0 \) then all \( (W_1 e_i, W_1 u^{p-1} e_i) \) determine the subobjects \( \mathcal{L}_i \simeq \mathcal{L}(0) \) of \( \mathcal{L} \) and \( \mathcal{L} \simeq \mathcal{L}_1 \times \cdots \times \mathcal{L}_s \). \( \square \)
Proposition 1.7. Any \( \mathcal{L} \in \mathcal{L}^* \) contains a unique maximal etale subobject \( (\mathcal{L}^{et}, l^{et}) \) and a unique maximal connected quotient object \( (\mathcal{L}^c, f^c) \) and the sequence \( 0 \to \mathcal{L}^{et} \xrightarrow{l^{et}} \mathcal{L} \xrightarrow{f^c} \mathcal{L}^c \to 0 \) is short exact.

Proof. Let \( \mathcal{L} = (L, F(L), \varphi, N) \) and, as earlier, let \( \phi : L \to L \) be such that for any \( l \in L \), \( \phi(l) = \varphi(u^{p^{-1}}l) \). Then for \( \hat{L} = L/uL \), we have the \( k \)-linear subspaces \( \hat{L}^{et} \) and \( \hat{L}^c \) in \( \hat{L} \) such that \( \hat{\phi} := \phi \mod u \) is invertible on \( \hat{L}^{et} \) and nilpotent on \( \hat{L}^c \) and \( \hat{L} = \hat{L}^{et} \oplus \hat{L}^c \).

Then there is a unique \( W_0 \)-submodule \( L^{et} \) of \( L \) such that \( \phi|_{L^{et}} \) is invertible and \( L^{et}/uL^{et} = \hat{L}^{et} \). The filtered submodule \( (L^{et}, u^{p^{-1}}L^{et}) \) determines an etale subobject \( l^{et} : \mathcal{L}^{et} \to \mathcal{L} \). Clearly, \( u^{p^{-1}}L^{et} \subset L^{et} \cap F(L) \). If the inverse embedding does not take place then there is an \( l \in L^{et}\setminus uL^{et} \) such that \( u^{p^{-1}}l \in uF(L) \). Therefore, \( \phi(l) = \varphi(u^{p^{-1}}l) \notin uL \) but \( \phi|_{L^{et}} \) is invertible. So, \( l^{et} \) is strict monomorphism and we can consider \( \text{Coker} l^{et} = f^{c} : \mathcal{L} \to \mathcal{L}^{c} \). Clearly, \( \mathcal{L}^{c} \) is connected. The maximality properties of \( L^{et} \) and \( \mathcal{L}^{c} \) are formally implied by the following easy statement:

if \( L_1 \in \mathcal{L}^* \) is etale and \( L_2 \in \mathcal{L}^* \) is connected then \( \text{Hom}_{\mathcal{L}}(L_1, L_2) = 0 \). \( \square \)

Suppose \( \mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}^* \). Then \( \varphi(F(L)) \) is a \( \sigma(W_1) \)-module and \( L = \varphi(F(L)) \otimes \sigma(W_1) W_1 \). If \( l \in L \) and for \( 0 \leq i < p \), \( l^{(i)} \in F(L) \) are such that \( l = \sum_{0 \leq i < p} \varphi(l^{(i)}) \otimes u^i \), set \( V(l) = l^{(0)} \). Then \( V \mod u \) is a \( \sigma^{-1} \)-linear endomorphism of the \( k \)-vector space \( L/uL \).

Definition. The module \( \mathcal{L} \) is multiplicative (resp., unipotent) if \( \hat{V} := V \mod u \) is invertible (resp., nilpotent) on \( \hat{L} := L/uL \).

Let \( \mathcal{L}(1) = (W_1, W_1, \varphi, N) \in \mathcal{L}^* \), where \( \varphi(1) = 1 \) and \( N(1) = 0 \). Then \( \mathcal{L}(1) \) is multiplicative. As a matter of fact, it is the simplest multiplicative object of \( \mathcal{L}^* \) due to the following Lemma.

Lemma 1.8. Suppose \( \mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}^* \) is multiplicative, then \( \mathcal{L} \) is the product of finitely many copies of \( \mathcal{L}(1) \).

Proof. Clearly, the embedding \( F(L) \to L \) induces the identification \( F(L)/uF(L) = L/uL \) and, therefore, \( F(L) = L \).

Let \( \hat{L}_0 \subset \hat{L} \) be such that \( \hat{V}|_{\hat{L}_0} = \text{id} \). If \( l \in L \) is such that \( l \mod uL \in \hat{L}_0 \) then \( \varphi(l) \equiv l \mod uL \). This implies the existence of a unique \( l' \in L \) such that \( l' \equiv l \mod uL \) and \( \varphi(l') = l' \). In other words, there is an \( \mathbb{F}_p \)-submodule \( L_0 \) in \( L \) such that \( L = L_0 \otimes_{\mathbb{F}_p} W_1 \) and \( \varphi|_{L_0} = \text{id} \).

If \( l \in L_0 \) then \( N(l) = N(\varphi(l)) = \varphi(uN(l)) = u^p\varphi(N(l)) = 0 \). So, if \( e_1, \ldots, e_s \) is an \( \mathbb{F}_p \)-basis of \( L_0 \) then the filtered modules \( (W_1 e_i, W_1 e_i) \) determine the subobjects \( \mathcal{L}_i \simeq \mathcal{L}(1) \) of \( \mathcal{L} \) and \( \mathcal{L} \simeq \mathcal{L}_1 \times \cdots \times \mathcal{L}_s \). \( \square \)

Proposition 1.9. Any \( \mathcal{L} = (L, M, \varphi, N) \in \mathcal{L}^* \) contains a unique maximal multiplicative quotient object \( (\mathcal{L}_m, j^m) \) and a unique maximal unipotent subobject.
\((L^u, i^u)\) and the sequence

\[
0 \rightarrow L^u \xrightarrow{i^u} L \xrightarrow{j^m} L^m \rightarrow 0
\]
is exact.

Proof. Let \(\tilde{L} = L/uL, \tilde{M} = M/uM\) and \(\tilde{L} = \tilde{L}^m \oplus \tilde{L}^u\), where \(\tilde{V} := V \mod u\) is invertible on \(\tilde{L}^m\) and nilpotent on \(\tilde{L}^u\).

Note that \(\varphi\) induces a \(\sigma\)-linear isomorphism \(\tilde{\varphi} : \tilde{M} \rightarrow \tilde{L}\). Denote by \(\tilde{i} : \tilde{M} \rightarrow \tilde{L}\) the \(k\)-linear morphism induced by the embedding \(M \rightarrow L\). With this notation, for any \(l \in \tilde{L}\), \(\tilde{V}(l) = \tilde{i}(\tilde{\varphi}^{-1}(l))\).

Consider the filtration \(\tilde{L}^u \supset V\tilde{L}^u \supset \cdots \supset \tilde{V}\tilde{L}^u = \{0\}\) and set for \(1 \leq i \leq s + 1\), \(\tilde{M}_i = \tilde{\varphi}^{-1}(\tilde{V}^{i-1}\tilde{L}^u)\). Then \(\tilde{M}_1 \supset \tilde{M}_2 \supset \cdots \supset \tilde{M}_s \supset \tilde{M}_{s+1} = \{0\}\) and for \(1 \leq i \leq s\),

\[
(1.1) \quad \tilde{i}(\tilde{M}_i) = \tilde{V}^i\tilde{L}^u = \tilde{\varphi}(\tilde{M}_{i+1}).
\]

For \(1 \leq i \leq s + 1\), introduce the \(W_i\)-submodules \(M_i^{(0)}\) of \(M\) such that \(M_i^{(0)} \supset M_2^{(0)} \supset \cdots \supset M_s^{(0)} \supset M_{s+1}^{(0)} = 0\) and \(M_i^{(0)}/uM_i^{(0)} = \tilde{M}_i\) with respect to the natural projection \(M \rightarrow \tilde{M}\). Then conditions \((1.1)\) imply that for all \(i\),

\[
M_i^{(0)} \subset \varphi(M_{i+1}^{(0)}) \circ \sigma W_i W_1 + uL.
\]

Let \(\tilde{M}_m = \tilde{\varphi}^{-1}(\tilde{L}^m)\) and let \(M_m \subset M\) be a \(W_i\)-submodule such that \(M_m/uM_m = \tilde{M}_m\) with respect to the natural projection \(M \rightarrow \tilde{M}\). Then

\[
(1.2) \quad M_m + uL = \varphi(M_m) \circ \sigma W_i W_1 + uL
\]

and \(M = M_m \oplus M_i^{(0)}\).

Prove the existence of “more precise” lifts \(M_i^{(n)}\) of \(\tilde{M}_i\), where \(0 \leq i \leq s + 1\) and \(n \geq 1\).

Lemma 1.10. For all \(n \geq 1\) and \(0 \leq i \leq s + 1\), there are \(W_i\)-modules \(M_i^{(n)}\) such that

a) \(M_1^{(n)} \supset M_2^{(n)} \supset \cdots \supset M_s^{(n)} \supset M_{s+1}^{(n)} = \{0\}\) and \(M_i^{(n)}/uM_i^{(n)} = \tilde{M}_i\) with respect to the natural projection \(M \rightarrow \tilde{M}\);

b) \(M_i^{(n)} \subset \varphi(M_{i+1}^{(n)}) \circ \sigma W_i W_1 + u\varphi(M_{i+1}^{(n)}) \circ \sigma W_i W_1 + u^{n+1}L\);

c) \(M_i^{(n-1)} + u^n M = M_i^{(n)} + u^n M\).

Proof of Lemma. The modules \(M_i^{(0)}, 0 \leq i \leq s + 1\), satisfy the requirements a) and b) of our Lemma. Therefore, we can assume that the modules \(M_i^{(n)}\) satisfying the requirements a)-c) have been already constructed for \(n = N - 1\), where \(N \geq 1\).

Note that \(M = M_m \oplus M_i^{(N-1)}\) (it is known for \(N = 1\) and follows from c) for \(N > 1\)). Therefore, \((1.2)\) implies that

\[
L = \varphi(M_m) \circ \sigma W_i W_1 + \varphi(M_i^{(N-1)}) \circ \sigma W_i W_1
\]

\[
\subset M_m + \varphi(M_i^{(N-1)}) \circ \sigma W_i W_1 + uL.
\]
Therefore, for $1 \leq i \leq s$ (use b) for $n = N - 1$,
\[ M_{i}^{(N-1)} \subset \varphi(M_{i+1}^{(N-1)}) \otimes_{\sigma W_i} W_i + u \varphi(M_{i}^{(N-1)}) \otimes_{\sigma W_i} W_i + u^{N} M^m + u^{N+1} L \]
and we can define the submodules $M_i^{(N)}$ in such a way that the property c) holds for $n = N$
\[ M_i^{(N)} + u^{N} M^m = M_i^{(N-1)} + u^{N} M^m \]
and
\[ M_i^{(N)} \subset \varphi(M_{i+1}^{(N-1)}) \otimes_{\sigma W_i} W_i + u \varphi(M_{i}^{(N-1)}) \otimes_{\sigma W_i} W_i + u^{N+1} L. \]
Note that (1.3) implies that $\varphi(M_i^{(N)}) + u^{Np} L = \varphi(M_{i}^{(N-1)}) + u^{Np} L$ and, therefore, we can replace $\varphi(M_i^{(N-1)})$ and $\varphi(M_{i}^{(N-1)})$ by $\varphi(M_i^{(N)})$ and, resp. $\varphi(M_i^{(N)})$ in (1.4). The lemma is proved.

Let $M^u = \bigcap_{n > 0}(M^{1(n)} + u^{n+1} M)$. Then $M^u/uM^u = \tilde{M}^u$ with respect to the natural projection $M \twoheadrightarrow M$ and $M = M^m \oplus M^u$.
Let $L^u = \varphi(M^u) \otimes_{\sigma W_i} W_i$. Then $\text{rk}_{W_i} L^u = \text{rk}_{W_i} M^u$ and
\[ L^u = \bigcap_{n > 0} \left( \varphi(M_{i}^{(n)}) \otimes_{\sigma W_i} W_i + u^{(n+1)p} L \right) \supset M^u \]
(\text{use Lemma 1.10b}). On the other hand,
\[ L = \varphi(M^m \oplus M^u) \otimes_{\sigma W_i} W_i = M^m \oplus L^u \]
implies that $M^u \supset u^{p-1} L^u$ and $L^u \cap M = M^u$. Therefore, the filtered module $(L^u, M^u)$ defines a unipotent subobject $L^u$ of $L$ in the category $L^*$ and the natural embedding $L^u \rightarrow L$ is strict.
Suppose $l \in M^u$ and $N(l) \equiv l_0 + l_1 \mod u^p L$, where $l_0 \in M^m$ and $l_1 \in L^u$. Then $uN(l) \equiv ul_0 + ul_1 \in (M^m \oplus M^u) \mod u^p L$ and $N(\varphi(l)) \equiv \varphi(ul_1) \mod u^p L$ implies that $N(L^u) \subset L^u \mod u^p L$. Then from Proposition 1.2 it follows that $L^u$ is a subobject of $L$ in the category $L^*$. Clearly, the quotient $L/L^u := \tilde{L}^m$ is multiplicative.
The maximality of $L^u$ and $L^m$ are formally implied by the following easy property of objects $L_1, L_2 \in L^*$:
if $L_1$ is unipotent and $L_2$ is multiplicative then $\text{Hom}_{L^*}(L_1, L_2) = 0$.

Using the above results we can introduce the subcategories $L^{*, u}, L^{*, c}, L^{*, m}, L^{*, et}$ in $L^*$. They consist of, resp., etale, connected, multiplicative and unipotent objects of the category $L^*$. The correspondences $L \rightarrow L^{*, u}, L \rightarrow L^{*, c}, L \rightarrow L^{*, m}, L \rightarrow L^{*, et}$ determine the natural exact functors from $L^*$ to, resp., $L^{*, u}, L^{*, c}, L^{*, m}$ and $L^{*, et}$.
1.3. The category \( \mathcal{L}^*_r \).

**Proposition 1.11.** Suppose \( \mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}^*_r \). Then the following conditions are equivalent:

(a) \( N(F(L)) \subset F(L) \mod u^{2p}L \);

(b) \( N(\varphi(F(L))) \subset u^pL \mod u^{2p}L \).

**Proof.** (a) \( \Rightarrow \) (b): if for any \( l \in F(L) \), \( N(l) \in F(L) \mod u^{2p}L \) then \( N(\varphi(l)) = \varphi(uN(l)) = u^p\varphi(N(l)) \in u^pL \mod u^{2p}L \).

(b) \( \Rightarrow \) (a): for any \( l \in F(L) \), \( \varphi(uN(l)) = N(\varphi(l)) \in u^pL \mod u^{2p}L \); now use that \( \varphi \) induces embedding of \( F(L)/uF(L) \) into \( L/u^pL \) to deduce that \( N(l) \in uF(L) \mod u^{2p}L \), i.e. \( N(l) \in F(L) \mod u^{2p}L \) (use that \( u^{p-1}L \subset F(L) \)). □

**Definition.** The category \( \mathcal{L}^*_r \), is a full subcategory of \( \mathcal{L}^* \) consisting of \( (L, F(L), \varphi, N) \) such that \( N : L \to L \) satisfies the equivalent conditions from Proposition 1.11.

**Remark.** a) If \( \mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}^*_r \), then \( N_1 = N \mod u^p \) is a unique \( W_1 \)-differentiation \( N_1 : L \to L/u^pL \) whose restriction to \( \varphi(F(L)) \) is the zero map. Therefore, any \( \mathcal{L} \in \mathcal{L}^*_0 \) has at most one structure of object of the category \( \mathcal{L}^*_r \).

b) Any etale or multiplicative object from \( \mathcal{L}^*_r \) belongs to \( \mathcal{L}^*_r \).

c) If \( f \) is a morphism in \( \mathcal{L}^*_r \), then \( \ker \mathcal{L}^*_r f = \ker \mathcal{L}^*_r f \) and \( \coker \mathcal{L}^*_r f = \coker \mathcal{L}^*_r f \). In particular, we can introduce the full subcategories \( \mathcal{L}^*_e \), \( \mathcal{L}^*_c \), \( \mathcal{L}^*_m \), \( \mathcal{L}^*_u \) of, resp., etale, connected, multiplicative and unipotent objects of \( \mathcal{L}^*_r \).

**Proposition 1.12.** Suppose \( \mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}^*_r \). Then there is a \( \sigma(W_1) \)-basis \( l_1, \ldots, l_s \) of \( \varphi(F(L)) \) and integers \( 0 \leq c_i < p \), where \( 1 \leq i \leq s \), such that \( u^{c_i}l_1, \ldots, u^{c_s}l_s \) is a \( W_1 \)-basis of \( F(L) \).

**Proof.** Choose a \( W_1 \)-basis \( m_1, \ldots, m_s \) of \( L \) such that for suitable integers \( c_1, \ldots, c_s \), the elements \( u^{c_1}m_1, \ldots, u^{c_s}m_s \) form a \( W_1 \)-basis of \( F(L) \). Clearly all \( 0 \leq c_i < p \).

For \( 1 \leq i \leq s \) and \( j \geq 0 \), let \( l_{ij} \in \varphi(F(L)) \) be such that \( m_i = \sum_{j \geq 0} w^j l_{ij} \). Note that \( \{ l_{ij} \mid 1 \leq i \leq s \} \) is a \( \sigma(W_1) \)-basis of \( \varphi(F(L)) \) and it will be sufficient to prove that all \( u^{c_i}l_i \in F(L) \) because then the elements \( l_i := l_{0i} \) will satisfy the requirements of our proposition.

For all \( 1 \leq i \leq s \), the element

\[
N(u^{c_i}m_i) = -\sum_j (j + c_i)u^{j+c_i}(l_{ij}) \mod u^{2p}L \quad + \quad \sum_j u^{j+c_i}N(l_{ij})
\]

belongs to \( F(L) \mod u^{2p}L \) if and only if \( \sum_j (j + c_i)u^{j+c_i}l_{ij} \in F(L) \). (Use that \( u^{p}L \subset uF(L) \).) This implies that for all integers \( k \geq 0 \), \( \sum_j (j + c_i)^k u^{j+c_i}l_{ij} \in F(L) \). Therefore, for any \( \alpha \in \mathbb{Z}/p\mathbb{Z} \),

\[
\sum_{(j+c_i) \mod p = \alpha} u^{j+c_i}l_{ij} \in F(L).
\]
In particular, taking $a = c_i \mod p$ and using that $u^pl_{ij} \in F(L)$, we obtain that $u^c_l \in F(L)$. \hfill \Box

**Remark.** a) Suppose $\mathcal{L} = (L, F(L), \varphi) \in \mathcal{L}^*_p$ and satisfies the conclusion of Proposition 1.12. Define the $W_1$-differentiation $N_1 : L \to L/u^pL$ by setting $N_1(l_1) = \cdots = N(l_p) = 0$. If $N : L \to L/u^2pL$ is the extension of $N_1$ given by Proposition 1.2 then $(L, F(L), \varphi, N) \in \mathcal{L}^*_p$. In other words, Proposition 1.12 characterizes the objects of $\mathcal{L}^*_p$ coming from $\mathcal{L}^*_p$.

b) For an object $(L, F(L), \varphi, N) \in \mathcal{L}^*_p$, Proposition 1.12 implies that if $\sum_{0 \leq i < p} u^i l_i \in F(L)$, where all $l_i \in \varphi(F(L))$, then all $u^i l_i \in F(L)$.

Consider the category of filtered Fontaine-Laffaille modules $\text{MF}_{p^{-1}}$ from [13]. The objects of this category are finite dimensional $k$-vector spaces $M$ with decreasing filtration of length $p$ by subspaces $M = M^0 \supset M^1 \supset \cdots \supset M^{p-1} \supset M^p = 0$ and $\sigma$-linear maps $\varphi_i : M^i \to M$ such that $\text{Ker} \varphi_i \supset M^{i+1}$, where $0 \leq i < p$, and $\sum_i \text{Im} \varphi_i = M$. The morphisms in $\text{MF}_{p^{-1}}$ are the morphisms of filtered vector spaces which commute with the corresponding morphisms $\varphi_i$, $0 \leq i < p$.

The category $\text{MF}_{p^{-1}}$ is abelian. The object $M$ of $\text{MF}_{p^{-1}}$ is:

- étale (resp., multiplicative) if $M^1 = 0$ (resp., $M = M^{p-1}$);
- connected (resp., unipotent) if $M$ has no étale (resp., multiplicative) sub-quotient.

Introduce the full subcategories $\text{MF}^*_p$, $\text{MF}^m_p$, $\text{MF}^c_p$ and $\text{MF}^u_p$ of, resp., étale, multiplicative, connected and unipotent objects in $\text{MF}_{p^{-1}}$. These subcategories are closed under the operations of taking subobjects and quotient objects and, therefore, are also abelian. For any $M \in \text{MF}_{p^{-1}}$, there are standard exact sequences $0 \to M^e \to M \to M^c \to 0$ and $0 \to M^u \to M \to M^m \to 0$, where $M^e$ (resp., $M^u$) is the maximal étale (resp., unipotent) subobject and $M^c$ (resp., $M^m$) is the maximal connected (resp., multiplicative) quotient object.

The categories $\mathcal{L}^*_p$ and $\text{MF}_{p^{-1}}$ do not differ very much.

Indeed, introduce the functor $\text{Md} : \mathcal{L}^*_p \to \mathcal{L}^*_p$ induced on the level of filtered modules by $(L, F(L)) \mapsto (L/u^pL, F(L)/u^pL)$. Denote by $\text{Md}(\mathcal{L}^*_p)$ the full subcategory of $\mathcal{L}^*_p$ consisting of the objects $\text{Md}(L)$, where $L \in \mathcal{L}^*_p$.

Define the functor $\mathcal{F} : \text{MF}_{p^{-1}} \to \mathcal{L}^*_p$ as follows. Let $M \in \text{MF}_{p^{-1}}$ with the corresponding filtration $M^i$ and $\sigma$-linear morphisms $\varphi_i$, $0 \leq i < p$. Then on the level of objects, $\mathcal{F}(M) = (L, F(L), \varphi, N)$, where $L = M \otimes_k W_1/u^pW_1$, $F(L) = \sum_{0 \leq i < p} u^{p-1-i}W_1(M^i \otimes 1)$ and for any $m \in M_i$, $\varphi(u^{p-1-i}m) = \varphi_i(m)$. One can easily see that $\mathcal{F}$ is equivalence of the categories $\text{MF}_{p^{-1}}$ and $\text{Md}(\mathcal{L}^*_p)$.

Now the difference between the categories $\mathcal{L}^*_p$ and $\text{MF}_{p^{-1}}$ is described by the following Proposition.

**Proposition 1.13.** For $\mathcal{L}_1, \mathcal{L}_2 \in \mathcal{L}^*_p$, $\text{Md}$ induces a surjection from $\text{Hom}_{\mathcal{L}^*_p}(\mathcal{L}_1, \mathcal{L}_2)$ to $\text{Hom}_{\mathcal{L}^*_p}(\text{Md}(\mathcal{L}_1), \text{Md}(\mathcal{L}_2))$ and its kernel coincides with
(i_{L_2^*} \circ j_{L_2^*}) \text{Hom}_{L^*}(L_1^m, L_2^*)$, where $i_{L_2^*} : L_2^* \rightarrow L_2$ (resp., $j_{L_1^*} : L_1 \rightarrow L_1^m$) is the maximal etale subobject in $L_2$ (resp., multiplicative quotient object for $L_1$).

**Proof.** For $L_2 = (L_2, F(L_2), \varphi, N)$, let $\phi : L_2 \rightarrow L_2$ be such that $\phi(l) = \varphi(u^p - 1)l$ for any $l \in L_2$. Let $L_2^c = \{l \in L_2 \mid \phi(l)^n \rightarrow 0\}$ and let $L_2' \in \text{L}^*$ be the filtered module $(L_2/u^nL_2', F(L_2)/u^nL_2')$ with $\varphi$ and $N$ induced from $L_2$. Then there are natural strict epimorphisms

$$L_2 \rightarrow \alpha \rightarrow L_2' \rightarrow \text{Md}(L_2),$$

where Ker $\alpha$ is associated with the filtered module $(u^pL_2', u^pL_2''')$ and Ker $\beta$ — with $(u^pL_2'/u^pL_2', u^pL_2'/u^pL_2'')$. Clearly, $\varphi|_{u^pL_2'}$ is nilpotent and then by Lemma 1.5,

$$\alpha_* : \text{Hom}_{\text{L}^*}(L_1, L_2) \rightarrow \text{Hom}_{\text{L}^*}(L_1, L_2')$$

is bijective. Note that the natural embedding $L_2^c \rightarrow L_2$ induces the identification $u^pL_2'/u^pL_2''' = u^p L_2''' / u^p L_2'$. Let $L_2''' = (u^pL_2', u^pL_2'') \in \text{L}^\cdot$ with inducing $\varphi$ and $N$. Then $L_2'''$ is multiplicative and there is a natural projection $\gamma : L_2''' \rightarrow \text{Ker} \beta$ such that Ker $\gamma$ is associated with $(u^p+1L_2''', u^p+1L_2''')$. Note that $\varphi$ is nilpotent on $u^{p+1}L_2'''$. Applying Lemma 1.5 we obtain that

$$\beta_* : \text{Hom}_{\text{L}^*}(L_1, L_2') \rightarrow \text{Hom}_{\text{L}^*}(L_1, \text{Md}(L_2))$$

is surjective and

$$\text{Ker} \beta = \text{Hom}_{\text{L}^*}(L_1, \text{Ker} \beta) = \text{Hom}_{\text{L}^*}(L_1, L_2') \simeq \text{Hom}(L_1^m, L_2''').$$

It remains to note that $\text{Hom}_{\text{L}^*}(L_1^m, L_2''') = \text{Hom}_{\text{L}^*}(L_1^m, L_2''')$ via the natural embedding of $L''$ into $L_2^c$.

**Corollary 1.14.** The functor $\text{Md} \circ F^{-1}$ induces equivalence of the categories $\text{L}^c_r$ (resp., $\text{L}^c_{\text{cr}}$) and $\text{MF}^c_{p-1}$ (resp., $\text{MF}^c_{p-1}$).

**1.4. Simple objects in $\text{L}^*$.**

**Definition.** An object $L$ of $\text{L}^*$ is simple if any strict monomorphism $i : L_1 \rightarrow L$ in $\text{L}^*$ is either isomorphism or the zero morphism. Equivalently, $L$ is simple iff any strict epimorphism $j : L \rightarrow L_2$ is either isomorphism or the zero morphism.

All simple objects in $\text{L}^*$ can be described as follows.

Let $\{0, 1\}_p = \{r \in \mathbb{Q} \mid 0 \leq r \leq 1, v_\rho(r) = 0\}$, where $v_\rho$ is a $p$-adic valuation. Then any $r \in [0, 1]_p$ can be uniquely written as $r = \sum_{i \geq 1} a_ip^{-i}$, where the digits $0 \leq a_i = a_i(r) < p$ form a periodic sequence. The minimal positive period of this sequence will be denoted by $s(r)$.

Let $\tilde{r} = 1 - r$. Then $\tilde{r} \in [0, 1]_p$ and $\tilde{r} = \sum_{i \geq 1} \tilde{a}_ip^{-i}$, where for all $i \geq 1$, the digits $\tilde{a}_i = a_i(\tilde{r})$ are such that $a_i + \tilde{a}_i = p - 1$. The document is about varieties with bad reduction at 3, and the proof involves the study of objects in $\text{L}^*$ and their properties, including strict monomorphisms and epimorphisms, as well as the concept of simple objects. The proof includes a specific construction involving filtrations and filtered modules to establish equivalences between categories.
DEFINITION. For \( r \in [0,1)_p \), let \( \mathcal{L}(r) = (L(r), F(L(r)), \varphi, N) \) be the following object of the category \( \mathcal{L}^r \):
- \( L(r) = \oplus_{i \in \mathbb{Z}/s(r)} W_i l_i \);
- \( F(L(r)) = \sum_{i \in \mathbb{Z}/s(r)} W_i u^i l_i \);
- for \( i \in \mathbb{Z}/s(r) \), \( \varphi(u^i l_i) = l_{i+1} \).
- \( N \) is uniquely recovered from the condition \( N|_{\varphi(F(L))} = 0 \) mod \( u^p \), cf. Proposition 1.2.

REMARK. If \( r = 0 \) or \( r = 1 \) we obtain the objects \( \mathcal{L}(0) \) and \( \mathcal{L}(1) \) introduced in Subsection 1.2. Note also that \( \mathcal{L}(r) \) is connected iff \( r \neq 0 \) and unipotent iff \( r \neq 1 \).

For \( n \in \mathbb{N} \) and \( r \in [0,1)_p \), set \( r(n) = \sum_{i \geq 1} a_{i+n(r)} p^{-i} \). Extend this definition to any \( n \in \mathbb{Z} \) by setting \( r(n) := r(n + Ns(r)) \) for a sufficiently large \( N \in \mathbb{N} \).

**Proposition 1.15.**
(a) If \( r \in [0,1)_p \) then \( \mathcal{L}(r) \) is simple;
(b) if \( r_1, r_2 \in [0,1)_p \) then \( \mathcal{L}(r_1) \simeq \mathcal{L}(r_2) \) if and only if there is an \( n \in \mathbb{Z} \) such that \( r_1 = r_2(n) \);
(c) if \( \mathcal{L} \) is a simple object of the category \( \mathcal{L}^r \) then there is an \( r \in [0,1)_p \) such that \( \mathcal{L} \simeq \mathcal{L}(r) \).

**Proof.** Lemma 1.16 below implies that the simple objects in the categories \( \mathcal{L}^r \) and \( \mathcal{L}^r \) are the same. By Corollary 1.14, the functor \( \text{Md} \circ F^{-1} \) transforms simple objects of \( \mathcal{L}^r \) to simple objects in \( \mathcal{MF}_{p-1} \). It remains to note that an analogue of our Proposition for the category \( \mathcal{MF}_{p-1} \) is proved in [13].

**Lemma 1.16.** For any \( \mathcal{L} \in \mathcal{L}^r \), there is an \( \mathcal{L}' \in \mathcal{L}^r \) and a strict monomorphism \( \iota' : \mathcal{L} \hookrightarrow \mathcal{L}' \) such that \( \varphi' \in \text{Hom}_{\mathcal{L}^r}(\mathcal{L}', \mathcal{L}) \) is a strict monomorphism and \( \mathcal{L}' \in \mathcal{L}^r \) if there is a strict monomorphism \( \alpha : \mathcal{L}' \hookrightarrow \mathcal{L}^r \) such that \( \iota'' = \iota' \circ \alpha \).

**Proof of Lemma.** Suppose \( \mathcal{L} = (L, F(L), \varphi, N) \). Consider the \( k \)-linear space \( M := \varphi(F(L))/u^p \varphi(F(L)) \). Let \( \tilde{L} = M \otimes_k (W_i/u^i W_i) = L/u^p L, \tilde{F} = F(L)/u^p L \) and \( \tilde{\varphi} : \tilde{F} \to M \) be the map induced by \( \varphi \).

Proceed by induction to define for all \( i \geq 1 \), the subspaces \( M_i \subseteq M \) and the \( W_i \)-submodules \( \tilde{F}_i \subseteq \tilde{L} \) as follows.

From the definition of \( N : L \to L/u^p L \) it follows easily that \( N \) induces a \( k \)-linear map \( \tilde{N}_1 : M \to M \) and \( \tilde{N}_1^p = 0 \). Therefore, \( M_1 := \text{Ker} \tilde{N}_1 \) is a non-trivial subspace in \( M \).

Suppose \( i \geq 1 \) and \( M_i \) has been already defined. Let \( \tilde{F}_i \) be the submodule of the elements of the form \( u^i l \) in \( M \otimes_k (W_i/u^i W_i) \), where \( a \geq 0, l \in M_i \) and \( u^i l \in \tilde{F}_i \). Then set \( M_{i+1} = \tilde{\varphi}(\tilde{F}_i) \).

Verify that for all indices \( i \), \( M_{i+1} \subseteq M_i \). If \( i = 1 \) we must prove that \( \tilde{N}_1(M_2) = 0 \). Indeed, \( M_2 \) is spanned by \( \varphi(u^l) \), where \( l \in M_1 \) and \( u^l \in \tilde{F}_1 \). But \( N(\varphi(u^l)) = \varphi(uN(u^l)) = \varphi(-u^{l+1} + u^{l+1} N(l)) \in u^p L \). If \( i > 1 \) then we can assume by induction that \( M_{i-1} \subseteq M_i \). This implies that \( \tilde{F}_{i-1} \subseteq \tilde{F}_i \) and \( M_i \subseteq M_{i+1} \).
We obtained a decreasing sequence of non-trivial finite dimensional $k$-linear spaces $\{M_i \mid i \geq 1\}$. For $i \gg 1$, these spaces become a non-trivial constant space $M^{cr} \subset M$ such that if $F^{cr} = \{u^l \in F \mid a \geq 0, l \in M^{cr}\}$ then $\tilde{\varphi}(F^{cr}) = M^{cr}$.

This subspace $M^{cr}$ has the maximality property: if $M' \subset M$ is such that for $F' = \{u^l \in F \mid a \geq 0, l \in M'\}$, $\tilde{\varphi}(F') = M'$ then $M' \subset M^{cr}$. Indeed, show as earlier that $M' \subset M_1$ and then proceed by induction proving that $M' \subset M_i$ for all $i \geq 1$.

Now in notation from Subsection 1.3, there is an $L^{cr} \in L^{cr}_r$ such that $\text{Md}(L^{cr}) = (M^{cr} \otimes_k (\mathcal{W}_1/u\mathcal{W}_1), \tilde{F}^{cr}, \tilde{\varphi}, \tilde{N})$, where $\tilde{N}|_{M^{cr}} = 0$. Then from Proposition 1.13 it follows the existence of a strict monomorphism $s^{cr} : L^{cr} \hookrightarrow L$. If $L' = (L, F(L'), \varphi, N) \rightarrow L$ is strict monomorphism and $L' \in L^{cr}_r$ then $L_1 = \text{Md}(L')$ is associated with the filtered module $(M' \otimes_k (\mathcal{W}_1/u\mathcal{W}_1), \tilde{F}')$ and by the above maximality property of $M^{cr}$, $M'$ is a subspace in $M^{cr}$ and $\text{Md}(L')$ is a strict subobject of $\text{Md}(L^{cr})$. This gives the required strict embedding $\alpha$.

The Lemma follows.

1.5. Extensions in $L^{cr}_r$. Suppose $r_1, r_2 \in [0, 1]_s$. Choose an $s \in \mathbb{N}$ which is divisible by $s(r_1)$ and $s(r_2)$ and introduce the objects $L_1 = (L_1, F(L_1), \varphi, N)$ and $L_2 = (L_2, F(L_2), \varphi, N)$ of the category $L^{cr}_r$ as follows:

$L_1 = \oplus_{i \in \mathbb{Z}/s} \mathcal{W}_i \gamma^{1, 0}, F(L_1) = \sum_{i \in \mathbb{Z}/s} \mathcal{W}_i u_i \gamma^{1, 0}$, where $r_1 = \sum_{i \in \mathbb{Z}/s} a_i \gamma^{1, 0}$, with the digits $0 \leq a_i < p, a_i = (p - 1) - a_i$ and for all $i \in \mathbb{Z}/s$, $\varphi(u_i \gamma^{1, 0}) = i^{(1)}$

$L_2 = \oplus_{j \in \mathbb{Z}/s} \mathcal{W}_j \gamma^{2, 0}, F(L_2) = \sum_{j \in \mathbb{Z}/s} \mathcal{W}_j u_j \gamma^{2, 0}$, where $r_2 = \sum_{j \in \mathbb{Z}/s} b_j \gamma^{2, 0}$, with the digits $0 \leq b_j < p, b_j = (p - 1) - b_j$ and for all $j \in \mathbb{Z}/s$, $\varphi(u_j \gamma^{2, 0}) = j^{(2)}$

**Lemma 1.17.** For $\kappa = 1, 2$, $L_\kappa$ is isomorphic to the product of $s/s(r_\kappa)$ copies of the (simple) object $L(r_\kappa)$.

**Proof.** Take $\kappa = 1$. For $\gamma \in \mathbb{F}_{p^\nu}$ and $\tilde{i} \in \mathbb{Z}/s(r_1)$, let $m_i(\gamma) = \sum_{i \equiv \tilde{i} \mod S(r_1)} a_i \gamma^{1, 0}$ and $M(\gamma) = \sum_{i \equiv \tilde{i} \mod S(r_1)} \mathcal{W}_i m_i(\gamma) \subset L_1$. Then all $M(\gamma) = (M(\gamma), M(\gamma) \cap F(L_1), \varphi, N)$ with induced $\varphi$ and $N$ are subobjects of $L_1$ isomorphic to $L(r_1)$. If $\gamma_1, \ldots, \gamma_d$ is an $\mathbb{F}_{p^\nu(r_1)}$-basis of $\mathbb{F}_{p^\nu}$ then $M(\gamma_1) \times \cdots \times M(\gamma_d)$ is isomorphic to $L_1$. (Use that $d = s/s(r_1)$ and $\det(\sigma(\gamma_\lambda)) \neq 0$, where for a given $\tilde{i}$, $i$ is such that $i \equiv \tilde{i} \mod S(r_1)$ and $1 \leq i \leq d$).

If $L = (L, F(L), \varphi, N) \in L^{cr}_r$ then we shall use the same notation $L$ for the image $(L, F(L), \varphi)$ of $L$ under the forgetful functor from $L^{cr}_r$ to $L^{cr}_r$. Clearly, this forgetful functor induces a group homomorphism $\text{Ext}_{L^{cr}_r}(L_2, L_1) \rightarrow \text{Ext}_{L^{cr}_r}(L_2, L_1)$.

Suppose $L = (L, F(L), \varphi) \in \text{Ext}_{L^{cr}_r}(L_2, L_1)$. Consider a $\sigma(\mathcal{W}_1)$-linear section $S : l^{(2)}_j \rightarrow l^{(2)}_j, j \in \mathbb{Z}/s$, of the corresponding epimorphic map $\varphi(F(L)) \rightarrow \varphi(F(L_2))$. Then

a) $L = L_1 \oplus (\oplus_{j \in \mathbb{Z}/s} \mathcal{W}_j l^{(2)}_j)$

b) for all indices $j \in \mathbb{Z}/s$, there are unique elements $v_j \in L_1$, such that $F(L) = F(L_1) + \sum_{j \in \mathbb{Z}/s} \mathcal{W}_j (u_j \gamma^{2, 0} + v_j)$ and $\varphi(u_j \gamma^{2, 0} + v_j) = l^{(2)}_j$. 

**Documenta Mathematica 18 (2013) 547–619**
c) $F(L) \supset u^{p-1} L$ if and only if for all $j \in \mathbb{Z}/s$, $u^jv_j \in F(L_1)$;
d) if $S^\mathbf{t}(l_{ij}) \rightarrow l'_{ij} = l_j + \varphi(w_{j-1})$, where $j \in \mathbb{Z}/s$ and $w_{j-1} \in F(L_1)$, is another section of the epimorphism $\varphi(F(L)) \rightarrow \varphi(F(L_2))$ then for the corresponding elements $v'_j \in L_1$, $v'_j - v_j = w_j - u^b_j \varphi(w_{j-1})$.

The constructions from above items a)-d) can be summarized as follows.

**Lemma 1.18.** Let $Z(L_2, L_1) = \{(v_j)_{j \in \mathbb{Z}/s} \in L_1^a | u^b_jv_j \in F(L_1)\}$ be a subgroup in $L_1^a$ and let

$$B(L_2, L_1) = \{(w_j - u^b_j \varphi(w_{j-1}))_{j \in \mathbb{Z}/s} | w_j \in F(L_1)\}$$

be a subgroup of $Z(L_2, L_1)$. Then there is a natural isomorphism of abelian groups $Z(L_2, L_1)/B(L_2, L_1) \simeq \text{Ext}L_{s^L}(L_2, L_1)$.

**Proposition 1.19.** Any $L \in \text{Ext}L_{s^L}(L_2, L_1)$ appears from a system of factors $(v(j))_{j \in \mathbb{Z}/s} \in Z(L_2, L_1)$ satisfying the following normalization condition

(C1) if $v_j = \sum_{i,t} \gamma_{ijt} u^{i_1}t_1$ with $\gamma_{ijt} \in k$, then $\gamma_{ijb_j} = 0$.

**Proof.** Choose a section $S$ of the projection $\varphi(F(L)) \rightarrow \varphi(F(L_2))$ with the minimal set $\gamma(S) = \{(i, j, \tilde{b}_j) | \gamma_{ijb_j} \neq 0\}$. Suppose $\gamma(S) \neq \emptyset$ (otherwise, the proposition is proved) and let $(v(j))_{j \in \mathbb{Z}/s}$ be the corresponding system of factors.

Suppose $(i_0, j_0, b_{j_0}) \in \gamma(S)$ and $\gamma = \gamma_{i_0j_0b_{j_0}}$. Replace $(v(j))_{j \in \mathbb{Z}/s}$ by an equivalent system $(v'(j))_{j \in \mathbb{Z}/s}$ via the elements $w_j \in F(L_1)$ such that $w_j = 0$ if $j \neq j_0 - 1$ and $w_{j+1} = \sigma^{-1}((\gamma) u_{i_0}^{a_{i_0} - 1}l_{i_0}^{(1)} - 1)$.

If $v'_j = \sum_{i,t} \gamma'_{ijt} u^{i'_{ijt}}$ then

- $\gamma'_{i_0j_0b_{j_0}} = 0$;
- $\gamma'_{i_0-1, j_0-1, a_{i_0} - 1} = \sigma^{-1}(\gamma) + \gamma_{i_0-1, j_0-1, a_{i_0} - 1}$;
- for all remaining indices $\gamma'_{ijt} = \gamma_{ijt}$.

Then $\gamma(S') \subset \gamma(S) \setminus \{(i_0, j_0, \tilde{b}_{j_0})\}$ and the minimality condition for $S$ implies $\gamma(i_0 - 1, j_0 - 1, a_{i_0} - 1) = \gamma(S') \setminus \gamma(S)$. Therefore, $\tilde{a}_{i_0} - 1 = \tilde{b}_{j_0} - 1$, $\gamma_{i_0-1, j_0-1, a_{i_0} - 1} = 0$, $\gamma'_{i_0-1, j_0-1, b_{j_0} - 1} = \sigma^{-1}(\gamma)$, and the new section $S'$ again satisfies the minimality condition.

Repeating the above procedure we obtain for all $n \in \mathbb{Z}/s$, that $\tilde{a}_{i_0} - n = \tilde{b}_{j_0} - n$, that is $\tilde{r}_1(i_0) = \tilde{r}_2(j_0)$.

Choose $\beta \in k$ such that $\sigma^s(\beta) - \beta = \gamma$ and consider $w_j \in F(L_1)$ such that for all $0 \leq n < s$, $w_{j+n} = \sigma^n(\beta) u_{i_0}^{a_{i_0} - 1}l_{i_0}^{(1)}$. Then for the corresponding new system of factors $(v'(j))_{j \in \mathbb{Z}/s}$, where

$$v'_j = v_j + w_j - u^b_j \varphi(w_{j-1}) = \sum_{i,t} \gamma'_{ijt} u^{i'_{ijt}}$$

one has $\gamma'_{i_0j_0b_{j_0}} = 0$, and $\gamma_{ijt} = \gamma'_{ijt}$ if $(i, j, t) \neq (i_0, j_0, b_{j_0})$. This contradicts to the minimality condition for $S$. \qed
Proposition 1.20. Any $\mathcal{L} \in \text{Ext}^\omega_{\mathcal{L}}(\mathcal{L}_2, \mathcal{L}_1)$ can be described via a system of factors $(v_j)_{j \in \mathbb{Z}/s}$, satisfying the above condition (C1) and the normalization condition

(C2) the coefficients $\gamma_{ijt} = 0$ if $t > \tilde{a}_i$.

Proof. Suppose $v^{(0)} = (v_j)_{j \in \mathbb{Z}/s}$ is such that $v_{j_0} = \gamma u^{t_0}l_0^{(1)}$ with $\gamma \in \mathcal{K}$, $t_0 > \tilde{a}_{i_0}$ and for $j \neq j_0$, $v_j = 0$. It will be sufficient to prove that any such system of factors is trivial.

Take the elements $w_j^{(0)}$, $j \in \mathbb{Z}/s$, such that $w_{j_0}^{(0)} = -\gamma u^{t_0}l_0^{(1)}$ and $w_j^{(0)} = 0$ if $j \neq j_0$. Then the corresponding equivalent system $(v_j^{(1)})_{j \in \mathbb{Z}/s}$ is such that $v_j^{(1)} = 0$ if $j \neq j_0 + 1$, and $v_j^{(1)} = \gamma p u^{t_1}l_{j_0+1}$, where $t_1 = \tilde{b}_{j_0+1} + (t_0 - \tilde{a}_{i_0})p$. This implies that $t_1 > p > \tilde{a}_{i_0+1}, t_1 - \tilde{a}_{i_0+1} > t_0 - \tilde{a}_{i_0}$, and $t_1 - \tilde{a}_{i_0+1} > t_0 - \tilde{a}_{i_0}$ unless $\tilde{b}_{j_0+1} = 0, t_1 = p$ and $\tilde{a}_{i_0+1} = p - 1$.

Repeat this procedure by using for all $n \geq 0$, the appropriate elements $w_j^{(n)}$, $j \in \mathbb{Z}/s$, to obtain the equivalent systems of factors $(v_j^{(n)})_{j \in \mathbb{Z}/s}$ such that $v_j^{(n)} = 0$ if $j \neq j_0 + n$, and $v_j^{(n)} = \gamma p^n u^{t_n}l_{j_0+n}^{(1)}$.

If $(\tilde{r}_2, \tilde{r}_1, t_0) \neq (0, 1, p)$ then $t_n \to \infty$ and we can use the elements $w_j = \sum_{n \geq 0} w_j^{(n)}$, $j \in \mathbb{Z}/s$, to trivialize the original system of factors $v^{(0)}$.

If $(\tilde{r}_2, \tilde{r}_1, t_0) = (0, 1, p)$, we can trivialize $v^{(0)}$ via the elements $w_j$, $j \in \mathbb{Z}/s$, for which $0 \leq n < s$, $w_{j_0+n} = \kappa p^n u^{t_n}l_{j_0+n}^{(1)}$ and $\kappa \in \mathcal{K}$ such that $\sigma^s(\kappa) - \kappa = \gamma$.

\[ N(u^\tilde{b}_j l_j + v_j) \equiv \sum_{i, t} \gamma_{ijt}(\tilde{b}_j - t)u^{t_i^{(1)}} + u^\tilde{b}_j N(l_j) \text{mod} F(L). \]

The condition $\mathcal{L} \in \mathcal{L}_r^r$ implies that $N(u^\tilde{b}_j l_j + v_j) \in F(L) \text{mod} u^{2p}L$ and $N(l_j) \in u^p L \text{mod} u^{2p}L \subset F(L) \text{mod} u^{2p}L$. This means that all $(\tilde{b}_j - t)\gamma_{ijt} u^{t_i^{(1)}} \in F(L_1)$.

Therefore, for $t \neq \tilde{b}_j$, $\gamma_{ijt} u^{t_i^{(1)}} \in F(L_1))$, and $v_j \in F(L_1)$. The proposition is proved.
implies that \( \tilde{\kappa} \). Let

Now we can set for all indices \( \tilde{\kappa} \) and this implies that \( \tilde{\kappa} \). Proof. Suppose \( \tilde{\kappa} \) and \( \tilde{\kappa} \) is trivial, cf. the proof of Proposition 1.22. This

Let \( \tilde{\kappa} \) and suppose \( \tilde{\kappa} \) if \( \tilde{\kappa} \). Proposition 1.22. Any element \( \tilde{\kappa} \) can be obtained as a sum of \( \tilde{\kappa} \) and \( \tilde{\kappa} \) runs over the set of \( \tilde{\kappa} \)-admissible pairs and all coefficients \( \tilde{\kappa} \) in \( \tilde{\kappa} \).

The above proposition describes the subgroup \( \tilde{\kappa} \) of \( \tilde{\kappa} \). In particular, working modulo this subgroup we can describe the extensions in the whole category \( \tilde{\kappa} \) via the systems of factors \( \tilde{\kappa} \) such that \( \tilde{\kappa} \) satisfy the normalization conditions (C1) and

\[(C3) \text{ if } t \geq \tilde{\kappa} \text{ then } \tilde{\kappa} = 0. \]

Proposition 1.23. Suppose the system of factors \( \tilde{\kappa} \) satisfies the conditions (C1) and (C3). If it determines \( \tilde{\kappa} \) then:

a) \( \tilde{\kappa} = 0 \) if \( t < \tilde{\kappa} - 1 \);

b) \( \tilde{\kappa} = \tilde{\kappa} + 1 \) and there is an \( m_0 \in \mathbb{N} \) such that for all \( 1 \leq m < m_0 \), \( \tilde{\kappa} + m = \tilde{\kappa} + m_0 \) but \( \tilde{\kappa} + m_0 = \tilde{\kappa} + m_0 \), then \( \tilde{\kappa} = 0 \);

c) \( \tilde{\kappa} = \tilde{\kappa} + 1 \) and for all \( m \in \mathbb{N} / \mathbb{S} \), \( \tilde{\kappa} + m = \tilde{\kappa} + m_0 \) then \( \tilde{\kappa} = 0. \)

Proof. Suppose \( \tilde{\kappa} \) and \( \tilde{\kappa} \) describes the image of \( \tilde{\kappa} \) in \( \tilde{\kappa} \). By the definition of \( N \), \( uN(u(b_i + v_j)) \in \tilde{\kappa} \), and this implies that \( \tilde{\kappa} = 0 \) if \( t < \tilde{\kappa} - 1 \), \( t \neq \tilde{\kappa} \) (use congruence (1.5)). This proves a).

Now we can set for all indices \( i \) and \( j \), \( \tilde{\kappa} := \tilde{\kappa} \). Let \( \kappa \) be such that \( N(\kappa) = \sum \kappa j \mod p \) and suppose \( \kappa \neq 0 \) (this implies that \( \tilde{\kappa} \neq \tilde{\kappa} \)). For \( m > 0 \), consider the relations

\[(1.6) \quad N(\kappa j) = \varphi(uN(u(b_i + v_j) + \tilde{\kappa} + m)). \]
Varieties with Bad Reduction at 3 Only

If $m = 0$ then (1.6) implies $\kappa_{i+1,j+1} = \gamma_i^p (\tilde{b}_j - \tilde{a}_i + 1)$. Suppose that there is an $m_0 \geq 0$ such that for all $1 \leq m < m_0$, $\tilde{a}_{i+m} - 1 = \tilde{b}_j + m$ but $\tilde{a}_{i+m} - 1 \neq \tilde{b}_j + m_0$. Then (1.6) together with (1.5) (where $j$ is replaced by $j + m$) imply that for $1 \leq m < m_0$,

$$\kappa_{i+m+1,j+m+1} = r_{i+m,j+m} = \gamma_i^{p+1} (\tilde{b}_j - \tilde{a}_i + 1).$$

In particular, $N(l_{j+m_0}) \mod u^pL$ contains $l_{j+m_0}^{(1)}$ with the coefficient $\gamma_i^{p+1} (\tilde{b}_j - \tilde{a}_i + 1)$. Therefore, $uN(\tilde{b}_j^{i+m_0}l_{j+m_0} + v_{j+m_0}) \mod u^pL$ contains $l_{j+m_0}^{(1)}$ with the coefficient $u^{\tilde{b}_j^{i+m_0}+1} \gamma_i^{p+1} (\tilde{b}_j - \tilde{a}_i + 1)$. But this monomial must belong to $F(L_i)$.

This proves that if $\gamma_{ij} \neq 0$ then $\tilde{b}_{j+m_0} + 1 > \tilde{a}_{i+m_0}$.

Finally, suppose that for all $m \geq 1$, $\tilde{a}_{i+m} - 1 = \tilde{b}_j + m$. Then $\tilde{a}_i - 1 = \tilde{a}_{i+s} - 1 = \tilde{b}_j + s = \tilde{b}_j$ and $\gamma_{ij} = \gamma_i, \tilde{a}_s - 1 = \tilde{a}_{ij}, \tilde{b}_j = 0$. □

**Remark.** With notation from the proof of above proposition the elements $v_j = \sum_{1} \gamma_{ij} u \tilde{a}_i - 1 l_{i+1}^{(1)}$ determine a system of factors from $Z(L_2, L_1)$ if $\gamma_{ij} = 0$ when either $\tilde{a}_i = 0$ or $\tilde{b}_j = p - 1$ (in this case $v_j$ should belong to $F(L_i)$).

**Definition.** A pair $(i_0, j_0) \in (\mathbb{Z}/s)^2$ is $(r_1, r_2)_s$-admissible if:

- $\tilde{b}_{j_0} \neq p - 1$ and $\tilde{a}_{i_0} \neq 0$; cf. above remark;
- $\tilde{b}_{j_0} - 1 \neq \tilde{b}_{j_0};$
- there is an $m_0 = m_{st}(i_0, j_0) \in \mathbb{N}$ such that for $1 \leq m < m_0$, $\tilde{a}_{i_0+m} - 1 = \tilde{b}_{j_0+m}$ but $\tilde{a}_{i_0+m} - 1 < \tilde{b}_{j_0+m}.$

**Definition.** A pair $(i_0, j_0) \in (\mathbb{Z}/s)^2$ is $(r_1, r_2)_s^p$-admissible if $i_0 = 0$ and for all $m \in \mathbb{Z}/s$, $\tilde{a}_m - 1 = \tilde{b}_{j_0+m}.$

**Proposition 1.24.** a) If $(i_0, j_0)$ is an $(r_1, r_2)_s$-admissible pair then $r_1(i_0) + 1/(p - 1) > r_2(j_0);$ b) if $(0, j_0)$ is an $(r_1, r_2)_s^p$-admissible pair then $r_1 + 1/(p - 1) = r_2(j_0).$

**Proof.** a) Here for $1 \leq m < m_0$, $a_{i_0+m} + 1 = b_{j_0+m}$ and $a_{i_0+m} \geq b_{j_0+m}.$ Therefore,

$$r_1(i_0) + 1/(p - 1) > \sum_{1 \leq m \leq m_0} (a_{i_0+m} + 1)p^{-m} > \sum_{1 \leq m \leq m_0} b_{j_0+m}p^{-m} + \sum_{m > m_0} (p - 1)p^{-m} \geq r_2(j_0).$$

The part b) can be obtained similarly. □

Using the calculations from the proof of Proposition 1.23 we obtain the following two statements.

**Proposition 1.25.** Suppose $(i_0, j_0) \in (\mathbb{Z}/s)^2$ is $(r_1, r_2)_s$-admissible and $\gamma \in k$. Then there is a unique $E_s(i_0, j_0, \gamma) \in \text{Ext}_{\mathbb{Z}/s}(L_2, L_1)$ given by the system of factors $(v_j)_{j \in \mathbb{Z}/s}$ such that $v_{j_0} = \gamma a_{i_0} - 1 l_{i_0}^{(1)}$ and $v_j = 0$ if $j \neq j_0,$ and the map $N$, which is uniquely determined by the condition:
The following proposition gives the uniqueness property of the decomposition of any element $L \in \text{Ext}^{1}(\mathcal{L}_2, \mathcal{L}_1)$ into a sum of standard extensions.

**Proposition 1.27.** Any element $L \in \text{Ext}^{1}(\mathcal{L}_2, \mathcal{L}_1)$ appears as a unique sum of the standard extensions $E_{cr}(i, j, \gamma_{ij})$, $E_{st}(i, j, \gamma_{ij}^{st})$ and $E_{sp}(j, \gamma_{0j}^{sp})$, where all $\gamma_{ij}^{cr}$, $\gamma_{ij}^{st}$ are equal to 0 but $\gamma_{0j}^{sp} \neq 0$, resp. $\gamma_{ij}^{st} = 0$, $\gamma_{0j}^{sp} = 0$, if the corresponding pair of lower indices is not $(r_1, r_2)_{cr}$-admissible, resp. $(r_1, r_2)_{st}$-admissible, $(r_1, r_2)_{sp}$-admissible.

**Proof.** By Proposition 1.23, any $L \in \text{Ext}^{1}(\mathcal{L}_2, \mathcal{L}_1)$ can be decomposed as a sum of the above special extensions. To prove the uniqueness of such decomposition, assume that $L$ represents a trivial element of $\text{Ext}^{1}(\mathcal{L}_2, \mathcal{L}_1)$ and prove that all involved coefficients $\gamma_{ij}^{cr}$, $\gamma_{ij}^{st}$ and $\gamma_{0j}^{sp}$ are equal to 0.

The image of $L$ in $\text{Ext}^{1}(\mathcal{L}_2, \mathcal{L}_1)$ is given by the system of factors $(v_{ij}^{cr} + v_{ij}^{st})_{j \in \mathbb{Z}/s}$ such that

- $v_{ij}^{cr} = \sum_{i} \gamma_{ij}^{cr} u_{i} l_{i}^{(1)}$.
- $v_{ij}^{st} = \sum_{i} \gamma_{ij}^{st} u_{i} l_{i}^{(1)}$.

Let $w_{j} \in F(L_{1})$ be such that for all $j$, $v_{j} = w_{j} - \tilde{b}_{j} \varphi(w_{j-1})$. If $w_{j} \equiv \sum_{i} \kappa_{ij} u_{i} l_{i}^{(1)} \mod u F(L)$ with $\kappa_{ij} \in k$, then for all $i$ and $j$, $j > 1$,

$$
(1.7) \quad \gamma_{ij}^{cr} u_{i}^{\tilde{a}_{i}} + \gamma_{ij}^{st} u_{i}^{\tilde{a}_{i}-1} \equiv \kappa_{ij} u_{i}^{\tilde{a}_{i}} - \kappa_{i-1,j-1} u_{i}^{\tilde{b}_{j}} \mod u^{\tilde{a}_{i}+1}.
$$

Suppose $(i_0, j_0)$ is $(r_1, r_2)_{cr}$-admissible. Then $\tilde{a}_{i_0} - 1 \neq \tilde{b}_{j_0}$ and comparing the coefficients for $u^{\tilde{a}_{i_0}+1}$ in (1.7) we deduce that $\gamma_{i_0,j_0}^{cr} = 0$. Therefore, all $\gamma_{ij}^{cr} = 0$.

Suppose $(i_0, j_0)$ is $(r_1, r_2)_{st}$-admissible. Then for $m_0 = m_{cr}(i_0, j_0)$, $\tilde{a}_{i_0} \neq \tilde{b}_{j_0}$, $\tilde{a}_{i_0} + m = \tilde{b}_{j_0} + m$ if $1 \leq m < m_0$, and $\tilde{a}_{i_0} + m_0 > \tilde{b}_{j_0} + m_0$. Then (1.7) implies that $\gamma_{i_0,j_0}^{cr} = \kappa_{i_0,j_0}$, $\kappa_{i_0,m,j_0} + m = \kappa_{i_0,m-1,j_0} + m - 1$ for $1 \leq m < m_0$, and $\kappa_{i_0+m-1,j_0+m-1} = 0$. Therefore, $\gamma_{i_0,j_0}^{st} = 0$.

Finally, $L$ is the trivial element of the group $\text{Ext}^{1}(\mathcal{L}_2, \mathcal{L}_1)$ and, therefore, for all $j$, $N(l_{j}) \in u^{2} L$. Then from the description of standard extensions $E_{sp}(j, \gamma_{0j}^{sp})$ in Proposition 1.26 it follows that all $\gamma_{0j}^{sp} = 0$. \hfill $\Box$
2. The functor $CV^* : \mathcal{L}^* \to \text{CMF}$

2.1. The object $R_{st}^0 \in \mathcal{L}^*$. Let $R = \varprojlim (\bar{O}/p)_n$ be Fontaine's ring; it has a natural structure of $k$-algebra via the map $k \to R$ given by $\alpha \mapsto \varprojlim (\sigma^{-n}\alpha \mod p)$, where for any $\gamma \in k$, $[\gamma] \in W(k) \subset \bar{O}$ is the Teichmüller representative of $\gamma$. Let $m_R$ be the maximal ideal of $R$.

Choose $x_0 = (x_0^{(n)} \mod p)_{n \geq 0} \in R$ and $\varepsilon = (\varepsilon^{(n)} \mod p)_{n \geq 0}$ such that for all $n \geq 0$, $x_0^{(n+1)p} = x_0^{(n)}$ and $\varepsilon^{(n+1)p} = \varepsilon^{(n)}$ with $x_0^{(0)} = -p$, $\varepsilon^{(0)} = 1$ but $\varepsilon^{(1)} \neq 1$. We shall denote by $v_R$ the valuation on $R$ such that $v_R(x_0) = 1$.

Let $Y$ be an indeterminate.

Consider the divided power envelope $R(Y)$ of $R[Y]$ with respect to the ideal $(Y)$. If for $j \geq 0$, $\gamma_j(Y)$ is the $j$-th divided power of $Y$ then $R(Y) = \bigoplus_{j \geq 0} R\gamma_j(Y)$. Denote by $R_{st}$ the completion $\prod_{j \geq 0} R\gamma_j(Y)$ of $R(Y)$ and set, $\text{Fil}^p R_{st} = \prod_{i \geq 1} R\gamma_i(Y)$. Define the $\sigma$-linear morphism of the $R$-algebra $R_{st}$ by the correspondence $Y \mapsto x_0^{2p}Y$; it will be denoted below by the same symbol $\sigma$.

Introduce a $W_1$-module structure on $R_{st}$ by the $k$-algebra morphism $W_1 \to R_{st}$ such that $u \mapsto \iota(u) := x_0 \exp(-Y) = x_0 \sum_{j \geq 0} (-1)^j \gamma_j(Y)$. Set $F(R_{st}) = \sum_{0 \leq i < p} x_0^{p-1-i} R\gamma_i(Y) + \text{Fil}^p R_{st}$. Define the continuous $\sigma$-linear morphism of $R$-modules $\varphi : F(R_{st}) \to R_{st}$ by setting for $0 \leq i < p$, $\varphi(x_0^{p-1-i} \gamma_i(Y)) = \gamma_i(Y)(1 - (i/2)x_0^{2p}Y)$, and for $i \geq p$, $\varphi(\gamma_i(Y)) = 0$. Let $N$ be a unique $R$-differential of $R_{st}$ such that $N(Y) = 1$.

**Proposition 2.1.** a) If $a \in R_{st}$ and $b \in F(R_{st})$ then

$$\varphi(ab) = \sigma(a) \varphi(b) \mod x_0^{2p} R_{st};$$

b) $\varphi \mod x_0^{2p} R_{st}$ is a $\sigma$-linear morphism of $W_1$-modules;

c) for any $b \in R_{st}$ and $w \in W_1$, $N(\varphi(w)) = N(\varphi(w)) \mod x_0^{2p} R_{st}$;

d) for any $l \in F(R_{st})$, $uN(l) \in F(R_{st})$ and

$$N(\varphi(l)) = \varphi(uN(l)) \mod x_0^{2p} R_{st}.$$

**Proof.** a) It is sufficient to verify it for $a = Y$ and $b = x_0^{p-1-i} \gamma_i(Y)$, $0 \leq i < p$.

b) Use that the multiplication by $\sigma(u) = u^p$ comes as the multiplication by $\iota(u)^p = x_0^{p-1-i} \gamma_i(Y)$, $0 \leq i < p$.

c) Use that $N(\iota(u)) = -\iota(u)$.

d) It will be enough to check the identity for $l = x_0^{p-1-i} \gamma_i(Y)$ with $1 \leq i < p$.

Then $N(\varphi(l)) = \gamma_i(Y)(1 - (1/2)(i + 1)x_0^{2p}Y)$. On the other hand, $uN(l) = x_0^{p-1-i} \gamma_i(Y) \exp(-Y)$ and $\varphi(uN(l))$ is equal to

$$\gamma_i(Y) \left(1 - \frac{i - 1}{2} x_0^{2p}Y\right) \mod x_0^{2p}$$

\[\Box\]

Introduce a $\Gamma_F$-action on $R_{st} \mod x_0^{2p}/(p-1) R_{st}$ as follows.
For any $\tau \in \Gamma_F$, let $k(\tau) \in \mathbb{Z}$ be such that $\tau(x_0) = e^{k(\tau)}x_0$ and let $\log(1 + X) = X - X^2/2 + \cdots - X^{p-1}/(p-1)$ be the truncated logarithm. For any $\tau \in \Gamma_F$, define a linear map $\tau : R_{st} \to R_{st}$ by extending the natural action of $\tau$ on $R$ and setting for $\tau \in \Gamma_F$ and $j \geq 0$,

$$\tau(\gamma_j(Y)) := \sum_{0 \leq i \leq \min(j, p-1)} \gamma_{j-i}(Y)\gamma_i(\log \varepsilon).$$

Note that the cocycle relation $\varepsilon^{k(\tau_1)(\gamma_1\varepsilon)}^{k(\tau)} = \varepsilon^{k(\tau_1\tau)}$, where $\tau_1, \tau \in \Gamma_F$, implies the cocycle relation

$$k(\tau_1)\log \varepsilon + k(\tau)\log(\varepsilon) \equiv k(\tau_1\tau)\log \varepsilon \bmod x_0^{p^2/(p-1)}.$$ 

(Use that $\log(1 + X)^k \equiv k\log(1 + X) \bmod (X^p)$ and $\varepsilon \equiv 1 \bmod x_0^{p/(p-1)}$.) In addition, for any $k \in \mathbb{Z}$, the obvious congruence

$$(1 + X)^k = \exp(k\log(1 + X)) \equiv \exp(k\log(1 + X)) \bmod (X^p)$$

implies that for any $\tau \in \Gamma_F$, $\tau(x_0 \exp(-Y)) \equiv x_0 \exp(-Y) \bmod x_0^{p^2/(p-1)}$.

Therefore, the correspondences $\gamma_j(Y) \mapsto \tau(\gamma_j(Y))$ induce a $\Gamma_F$-action on $\mathcal{W}_1$-algebra $R_{st}$ mod $x_0^{p^2/(p-1)}$, which extends the natural $\Gamma_F$-action on $R$.

**Proposition 2.2.** For any $\tau \in \Gamma_F$,

a) $\tau(F(R_{st})) = F(R_{st})$;

b) for any $a \in F(R_{st})$, $\tau(\varphi(a)) \equiv \varphi(\tau(a)) \bmod x_0^{p^2/(p-1)}$;

c) for any $b \in R_{st}$, $\tau(N(b)) = N(\tau(b))$.

*Proof.* The proof is straightforward in cases a) and c). Part b) follows by direct calculation from the following Lemma. \qed

**Lemma 2.3.** $\sigma(\log \varepsilon)/x_0^{p^2/(p-1)} \equiv \log \varepsilon \bmod x_0^{p^2/(p-1)} R$.

*Proof.* Consider Fontaine’s element

$$t^+ = \log[\varepsilon] = \sum_{n \geq 1} (-1)^{n-1} \frac{([\varepsilon] - 1)^n}{n} = \sum_{m \in \mathbb{Z}} p^m[\eta_m] \in A_{cr}$$

where all $\eta_m \in R$. Then $t^+ \in \text{Fil}^1 A_{cr}$ and $\sigma t^+ = pt^+$. This implies for all $m \in \mathbb{Z}$, that $\eta_m = \sigma^{-m}\eta_0$.

Consider $\mathcal{H} \subset A_{cr}$ consisting of the elements of the form $\sum_{m \in \mathbb{Z}} p^m[r_m]$ such that for $m \leq 0$, $v_R(r_m) \geq p^2/(p-1)$ (this is automatic for $m \leq -2$), $v_R(r_1) \geq p^2/(p-1) - 1$ and $v_R(r_2) \geq p^2/(p-1) - 2$. Then $\mathcal{H}$ is an additive subgroup in $A_{cr}$.

Verify that

- for all $n \geq p$, $([\varepsilon] - 1)^n/n \in \mathcal{H}$.

*Documenta Mathematica* 18 (2013) 547–619
Indeed, the congruence \([ε] ≡ 1 + [a_0] \mod pW(R)\) (where \(a_0 = ε - 1\)) implies that \([ε] = \lim_{m→∞}(1 + [σ^{-m}a_0])p^m\). Therefore,

\[
[ε] - 1 = \sum_{m \geq 0} [a_m]p^m = [a_0] \left(1 + \sum_{m \geq 1} p^m[a_m]\right),
\]

where \(v_R(a_m) = p^{1-m}/(p - 1)\), if \(v_R(1) = -1\) and \(v_R(2) = -1 - 1/p\). If \(n \not\equiv 0 \mod p\) then \(([ε] - 1)^n ≡ [a_0n] + p[a_1n] + p^2[a_2n] \mod p^3W(R)\) with \(v_R(a_0n) = v_R(a_1n) + 1 = v_R(a_2n) + 2 =pn/(p - 1)\). This proves that \(([ε] - 1)^n/n \in \mathcal{H}\) for all \(n \not\equiv 0 \mod p, n > p\).

As for all remaining \(n \geq p\), just note that for all \(M \geq 1,\)

\[
([ε] - 1)^{p^M} ≡ [a_0]^{p^M}(1 + p^{M+1}[b_1] + p^{M+2}[b_{2M}]) \mod p^{M+3}W(R),
\]

where \(v_R(b_{2M}) = -2\).

The above calculations mean that \(t^+ ≡ \log[ε] \mod \mathcal{H}\). Therefore, if

\[
\log[ε] = [ω_0] + p[ω_1] + p^2[ω_2] \mod p^3W(R),
\]

then \(<ω_0 = \log[ε] ≡ ω_0 \mod x_0^{p^2/(p - 1)}R, ω_1 ≡ ω_1 ≡ ω_0 ≡ ω_1 \mod x_0^{p^2/(p - 1)}R, ω_2 ≡ ω_2 \mod x_0^{p^2/(p - 1)}R,\>

and \(\log[ε] \mod x_0^{p^2/(p - 1)}R, t^+ = \log[ε] \mod x_0^{p^2/(p - 1)}R,\) the lemma is proved.

By above results we can introduce \(\mathcal{R}_*^{\mathcal{L}} = (\mathcal{R}_0^\mathcal{L}, F(\mathcal{R}_0^\mathcal{L}), \varphi, N) \in \mathcal{L}_*\), where \(\mathcal{R}_0^\mathcal{L} = R_0 \otimes \mathcal{M}_R \otimes \mathcal{F}(\mathcal{R}_0^\mathcal{L})\) and \(F(\mathcal{R}_0^\mathcal{L}) = F(\mathcal{R}_0^\mathcal{L}) \mod x_0^{p^2}N\) with induced \(\sigma\)-linear map \(\varphi\) and \(\mathcal{W}_1\)-differentiation \(N\). The above defined \(\Gamma_F\)-action on \(R_0 \otimes \mathcal{M}_R \otimes \mathcal{F}(\mathcal{R}_0^\mathcal{L})\) respects the structure of \(\mathcal{R}_0^\mathcal{L}\) as an object of the category \(\mathcal{L}_*\). In our setting the filtered Galois module \(\mathcal{R}_0^\mathcal{L}\) plays a role of Fontaine’s ring \(\mathcal{L}_1\).

2.2. The functor \(\mathcal{V}^*\). If \(\mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}_*\) then the triple \((L, F(L), \varphi)\) is an object of \(\mathcal{L}_*\) which will be denoted below by the same symbol \(\mathcal{L}\).

**Definition.** Let \(\mathcal{R}^\mathcal{L} = (R_0^\mathcal{L}, F(R_0^\mathcal{L}), \varphi) \in \mathcal{L}_*\), where \(R_0^\mathcal{L} = R_0 \otimes \mathcal{M}_R \otimes \mathcal{F}(\mathcal{R}_0^\mathcal{L})\) and \(F(R_0^\mathcal{L}) = F(\mathcal{R}_0^\mathcal{L}) \mod x_0^{p^2}N\) with induced \(\sigma\)-linear map \(\varphi\) and \(\mathcal{W}_1\)-differentiation \(N\). If \(f \in \mathcal{V}^*(\mathcal{L})\) and \(i \geq 0\), introduce the \(k\)-linear morphisms \(f_i : \mathcal{L} \rightarrow R_i^\mathcal{L}\) such that for any \(l \in L, f(l) = \sum_{i \geq 0} f_i(l)\gamma_i(Y)\). The correspondence \(f \mapsto f_0\) gives the homomorphism of abelian groups \(pr_0 : \mathcal{V}^*(\mathcal{L}) \rightarrow \mathcal{V}^0(\mathcal{L}) := \text{Hom}_{\mathcal{L}_0^*}(\mathcal{L}, R_0^\mathcal{L})\).
PROPOSITION 2.4. pr⁰ is isomorphism of abelian groups.

Proof. Clearly, pr⁰ is additive. Suppose f ∈ Ker pr⁰. Then for all i ⩾ 0 and
l ∈ L, fᵢ(l) = fᵢ(Nᵢ(l))) = 0, i.e. f = 0.
Suppose g ∈ Hom₂₋₁(L, R⁰). This means that g : L → R⁰ is a σ-linear
morphism of W₁-modules, g(F(L)) ⊂ F(R⁰) and for any l ∈ F(L), g(ϕ(l)) =
((g(l)/x₀⁻¹)ᵢ)ₚ.
Set for any l ∈ L, f(l) = g(l) + g(Nl)γ₁(Y) + · · · + g(Nᵢl)γᵢ(Y) + · · · Then for
any l ∈ L, f(N(l)) = N(f(l)) and our Proposition is implied by the following
Lemma.

LEMMA 2.5. a) For any l ∈ L, f(u[l]) = x₀ exp(−Y)f(l);
b) for any l ∈ F(L), ϕ(f(l)) = f(ϕ(l)).

Proof of Lemma. a) For any l ∈ L, f(u[l]) = ∑₀,₀ g(Nᵢ(u[l]))γᵢ(Y) =
x₀ ∑₀,₀ g(Nᵢ(u[l]))γᵢ(Y) =
x₀ ∑₀,₀ (−1)ᵢ)γᵢ(Y)γᵢ(Y) = x₀ exp(−Y)f(l).
b) Let l ∈ L. Prove by induction on i ⩾ 1 that
Nᵢ(ϕ(l)) = ϕ(u[Nᵢ](l)) = −i(i − 1)₂uᵢ(ϕ(uᵢ−¹Nᵢ−¹(l)) + ϕ(uᵢNᵢ(l)).
Then
\[ g(Nᵢ(ϕ(l))) = \frac{−i(i − 1)}{2} x₀ \left( \frac{g(uᵢ−¹Nᵢ−¹(l))}{x₀⁻¹} \right) + \left( \frac{g(uᵢNᵢ(l))}{x₀⁻¹} \right) \]
and f(ϕ(l)) is equal to ∑₀,₀ g(Nᵢ(ϕ(l)))γᵢ(Y) =
\[ \sum₀,₀ \left( \frac{g(Nᵢl)}{x₀⁻¹} \right) \left( γᵢ(Y) - \frac{i(i + 1)}{2} x₀²γᵢ+₁(Y) \right) = ϕ(f(l)). \]

COROLLARY 2.6. a) If rkวล L = s then |วล(ς)| = pᵢ;
b) the correspondence /us(ς) induces an exact functor υ from ς to the
category of F_p[ς]-modules.

Proof. a) Proceed as in [1, 3]. Suppose the structure of the filtered ϕ-module
ς is given by a choice of a W₁-basis m₁, ..., mᵢ of F(L) and non-degenerate
matrix A ∈ Ms(W₁) such that (m₁, ..., mᵢ) = (ϕ(m₁), ..., ϕ(mᵢ)). Let
X = (X₁, ..., Xᵢ) be a vector with s independent variables and let R₀ =
Frac R. Consider the quotient Aς of the polynomial ring R₀[X] by the ideal
generated by the coordinates of the vector (X A)ₚ − x₀ᵖ⁻¹X. (For a matrix
C the matrix Cₚ is obtained by raising all elements of C to p-power.)
Then Aς is etale R₀-algebra of rank pᵢ (use that (uᵢ−¹Iₖ)A⁻¹ ∈ Ms(W₁))
and all its $\hat{R}_0$-points give rise to elements of the group $\text{Hom}_{\mathcal{E}_0}(\mathcal{L}, \mathcal{R})$, where $\mathcal{R} = (R, x_0^{p-1}R, \varphi) \in \mathcal{E}_0$ is such that for any $r \in R$, $\varphi(x_0^{p-1}r) = r^p$. It remains to note that $\varphi |_{x_0^{p-1}m_R}$ is nilpotent, by Lemma 1.5, the natural projection $\mathcal{R} \to \mathcal{R}_0$ induces bijection from $\text{Hom}_{\mathcal{E}_0}(\mathcal{L}, \mathcal{R})$ to $\text{Hom}_{\mathcal{E}_0}(\mathcal{L}, \mathcal{L}) \subseteq \mathcal{L}$. Then it is strict if and only if $f$ is monomorphism ($\text{Ker} f$) is nilpotent. For $\mathcal{R}_0 = (R_0, x_0^{p-1}R_0, \varphi)$ of $R_0$, there is a natural projection $\mathcal{R}_0 \to \hat{\mathcal{R}}_0$ in $\mathcal{E}_0$ and for any $\mathcal{L} \in \mathcal{E}_0$, $\text{Hom}_{\mathcal{E}_0}(\mathcal{L}, \mathcal{R}_0) = \text{Hom}_{\mathcal{E}_0}(\mathcal{L}, \hat{\mathcal{R}}_0)$. This implies the following description of the $\hat{\mathcal{L}}_F$-modules $V^*(\mathcal{L})$ where $\mathcal{L} \in \mathcal{E}_0$ (use the identification $\text{pr}_0$ of Proposition 2.4). 

**Corollary 2.7.**

$$V^*(\mathcal{L}) = \left\{ \sum_{0 \leq i \leq p} N^*(f_0) \gamma_i(Y) \mod \hat{J} \mid f_0 \in \text{Hom}_{\mathcal{E}_0}(\mathcal{L}, \mathcal{R}_0) \right\}$$

**Remark.** a) In the above description of $V^*(\mathcal{L})$, for any $l \in \mathcal{L}$, $N^*(f_0)(l) = f_0(N(l))$. In addition, all $N^*(f_0) \gamma_i(Y)$ depend just on $N_1 = N \mod u^p L$. 
b) If $\mathcal{L} \in \mathcal{E}_0$ then in the above Corollary we can replace $\mathcal{R}_0$ and $\hat{J}$ by $\mathcal{R}_0 = (R/x_0^0R, x_0^0R, \varphi) \in \mathcal{E}_0$ and the ideal $\hat{J}_1 = \sum_{0 \leq i \leq p} R x_0^0 \gamma_i(Y) + \text{Fil} R_0^p$. In particular, for unipotent modules the whole theory can be developed in the context of $k[u]/u^p$-modules.

### 2.3. THE CATEGORY $\text{CMGF}_\Gamma$ AND THE FUNCTOR $\text{CV}^*$.

**Definition.** The objects of the category $\text{CMGF}_\Gamma$ are the triples $\mathcal{H} = (H, H^0, j)$, where $H, H^0$ are finite $\mathbb{Z}_p[\Gamma_F]$-modules, $\Gamma_F$ acts trivially on $H^0$ and $j : H \to H^0$ is an epimorphic map of $\mathbb{Z}_p[\Gamma_F]$-modules. If $\mathcal{H}_1 = (H_1, H_0^1, j_1) \in \text{CMGF}_\Gamma$, then $\text{Hom}_{\text{CMGF}_\Gamma}(\mathcal{H}_1, \mathcal{H})$ consists of the couples $(f, f^0)$, where $f : H_1 \to H$ and $f^0 : H_0^1 \to H^0$ are morphisms of $\Gamma_F$-modules such that $jf = f^0 j_1$.

The category $\text{CMGF}_\Gamma$ is pre-abelian, cf. Appendix A, and its objects have a natural group structure. In particular, with above notation, $\text{Ker}(f, f^0) = (\text{Ker} f, j_1(\text{Ker} f))$ together with the natural embedding to $\mathcal{H}_1$. Similarly, $\text{Coker}(f, f^0) = (H/f(H_1), H^0/j(f(H_1)))$. For example, the map $(\text{id}, 0) : (H, H) \to (H, 0)$ has the trivial kernel and cokernel. In addition, the monomorphism $(f_1, f_1^0) : \mathcal{H}_1 \to \mathcal{H}$ is strict if and only if $f_1 \circ \text{Ker}j_1 = f_1(\text{Ker} j) \cap \text{Ker} j$. Suppose $\mathcal{H}_2 = (H_2, H_0^2, j_2)$ and $(f_2, f_2^0) : \mathcal{H} \to \mathcal{H}_2$ is an epimorphism. Then it is strict if and only if $f_2^0$ induces epimorphic map from $\text{Ker} j$ to $\text{Ker} j_2$. In $\text{CMGF}_\Gamma$ we can use formalism of short exact sequences and the corresponding 6-terms $\text{Hom}_{\text{CMGF}_\Gamma} = \text{Ext}_{\text{CMGF}_\Gamma}$ exact sequences, cf. Appendix A.
2.4. A criterion. Suppose \( \mathcal{L} \in \mathcal{L}^* \) and \( i^{ct} : \mathcal{L}^{ct} \rightarrow \mathcal{L} \) is the maximal etale subobject. Then \( CV^* : \mathcal{L}^* \rightarrow CM\Gamma_F \) is the functor such that \( CV^*(\mathcal{L}) = (V^*(L), V^*(L^{ct}), V^*(i^{ct})) \).

The simple objects in \( CM\Gamma_F \) are of the form either \( (H, 0, 0) \), where \( H \) is a simple \( \mathbb{Z}_p[\Gamma_F] \)-module, or \( (\mathbb{F}_p, \mathbb{F}_p, id) \), where \( \mathbb{F}_p \) is provided with the trivial \( \Gamma_F \)-action. In this context it will be very convenient to use the following formalism.

For \( s \in \mathbb{N} \), consider Serre’s fundamental characters \( \chi_s : \Gamma_F \rightarrow k^* \). Here for \( \tau \in \Gamma_F \), \( \chi_s(\tau) = (\tau x_s) / x_s \mod x_s^m \), where \( x_s \in R \) is such that \( x_s^{p^s-1} = x_0 \). If \( \chi \) is any continuous (1-dimensional) character of \( \Gamma_F \) then there are \( s, m \in \mathbb{N} \) such that \( 0 < m \leq p^s - 1 \) and \( \chi = \chi_s^m \). Set \( r(\chi) = m / (p^s - 1) \). Then \( r(\chi) \) depends only on \( \chi \) and the correspondence \( \chi \mapsto r(\chi) \) gives a bijection from the set of all continuous (1-dimensional) characters of \( \Gamma_F \) with values in \( k^* \) to the set \( [0, 1]_p \cup \{0\} \).

For \( r \in [0, 1]_p \), \( r \neq 0 \), introduce the \( \Gamma_F \)-module \( \mathbb{F}(r) \) such that \( \mathbb{F}(r) = \mathbb{F}_{p(r)} \), where \( p(r) \) is the period of the \( p \)-digit expansion of \( r \), cf. Subsection 1.2, with the \( \Gamma_F \)-action given by the character \( \chi \) such that \( r(\chi) = r \). We have:

- \( \mathbb{F}(r) \) are simple \( \mathbb{Z}_p[\Gamma_F] \)-modules;
- \( \Gamma_F \)-modules \( \mathbb{F}(r_1) \) and \( \mathbb{F}(r_2) \) are isomorphic if and only if there is an \( n \in \mathbb{Z} \) such that \( r_1 = r_2(n) \);
- any simple \( \mathbb{Z}_p[\Gamma_F] \)-module is isomorphic to some \( \mathbb{F}(r) \).

It will be natural to set \( \mathcal{F}(r) := (\mathbb{F}(r), 0, 0) \) for all \( r \in [0, 1]_p \), and to set separately \( \mathcal{F}(0) := (\mathbb{F}_p, \mathbb{F}_p, id) \).

With above notation we have the following property, where the objects \( \mathcal{L}(r) \) were introduced in Subsection 1.3.

**Proposition 2.8.** For any \( r \in [0, 1]_p \), \( CV^*(\mathcal{L}(r)) = \mathcal{F}(r) \).

**Proof.** The proof goes along the lines of Subsection 4.2 of [1], cf. also the beginning of Subsection 2.4 below. \( \square \)

### 2.4. A criterion.

Suppose \( \mathcal{L}_1, \mathcal{L}_2 \) are given in notation of Subsection 1.4 and \( q = p^s \). Then for \( i = 1, 2 \), \( V^*(\mathcal{L}_i) = V_i \) are 1-dimensional vector spaces over \( \mathbb{F}_q \) with \( \Gamma_F \)-action given by the character \( \chi_i : \Gamma_F \rightarrow k^* \) such that \( r(\chi_i) = r_i \).

(Note that \( (q-1)r_i \in \mathbb{Z} \) and, therefore, \( \chi_i(\Gamma_F) \subset \mathbb{F}_q^* \)). Choose \( \pi_s \in \mathcal{F}_s \) such that \( \pi_s^{q-1} = -p \).

Then \( F_s = \mathcal{F}(\pi_s) \) is a tamely ramified extension of \( F \) of degree \( q-1 \) and all points of \( V_i \) are defined over \( F_s \). We can identify \( V_i \) with the \( \mathbb{F}_s[\Gamma_F] \)-module \( \mathbb{F}_s[\pi_s^{(q-1)r_i}] \subset \mathcal{O}/p\mathcal{O} \), where \( \pi_s = \pi_s \mod p \). These identifications allow us to fix the points \( h_{\alpha}^{0} := \pi_s^{(q-1)r_i} \in V_i \) and to identify \( V_i \) with the \( \Gamma_F \)-module \( \{\alpha h_{\alpha}^{0} | \alpha \in \mathbb{F}_q^* \} \).

Suppose \( h_1 \in V_1 \). Define the homomorphism

\[
F_{h_1} : \text{Ext}_{\mathbb{F}_s[\Gamma_F]}(V_1, V_2) \rightarrow Z^1(\Gamma_F, \mathbb{F}_q) = \text{Hom}(\Gamma_F, \mathbb{F}_q),
\]

where \( \Gamma_F = \text{Gal}(\bar{F}/F_s) \), as follows. If \( V \in \text{Ext}_{\mathbb{F}_s[\Gamma_F]}(V_1, V_2) \) and \( h \in V \) is a lift of \( h_1 \) then for any \( \tau \in \Gamma_F \), \( F_{h_1}(V)(\tau) = a_\tau \in \mathbb{F}_q \), where \( \tau h - h = a_\tau h_{\alpha}^{0} \).
Clearly, $F_{h_1}(V)$ does not depend on a choice of $h$ and it is the zero function if and only if the projection $V \to V_1$ admits a $\Gamma_F$-equivariant section. In other words, we have the following criterion.

**Proposition 2.9.** $V$ is the trivial extension if and only if for all $h_1 \in V_1$, one has $F_{h_1}(V) = 0$.

2.5. Galois Modules $V^*(E_{cr}(i_0, j_0, \gamma))$. Suppose we have an object $L = (L, F(L), \varphi, N)$ of the category $\mathcal{L}_{\Gamma}$. Then there is a special $\sigma(W_1)$-basis $l_1, \ldots, l_s$ of $\varphi(F(L))$ such that for some integers $0 \leq c_1, \ldots, c_s < p$ and a matrix $A \in GL_r(k)$, the elements $\varphi^{c_i}l_1, \ldots, \varphi^{c_s}l_s$ form a $W_1$-basis of $F(L)$ and $(\varphi(\varphi^{c_1}l_1), \ldots, \varphi(\varphi^{c_s}l_s)) = (l_1, \ldots, l_s)A$.

For $1 \leq i \leq s$, set $c_i = (p - 1) - c_i$. The following Proposition is a special case of Corollary 2.7 (remind that $R^0 = R/x_0^p\mathcal{M}_R$).

**Proposition 2.10.** With above notation, $V^*(L)$ is the $\mathbb{F}_p[\Gamma_F]$-module of all $(\theta_1, \ldots, \theta_s)$ mod $x_0^p\mathcal{M}_R$ in $(R^0)^s$ such that

$$(\theta_1/x_0^{c_1}, \ldots, \theta_s/x_0^{c_s}) = (\theta_1, \ldots, \theta_s)A.$$  

**Remark.** In [1, 2] it was proved (in the context of the Fontaine-Laffaille theory) that the family of $\mathbb{F}_p[\Gamma_F]$-modules $V^*(L)$, where $L \in \mathcal{L}_{\Gamma}$, coincides with the family of all killed by $p$ subquotients of crystalline representations of $\Gamma_F$ with weights from $[0, p)$. This result can be also extracted from Subsection 4, where we establish that the family of $\mathbb{F}_p[\Gamma_F]$-modules $V^*(L)$, where $L \in \mathcal{L}_{\Gamma}$, coincides with the family of all killed by $p$ subquotients of semi-stable representations of $\Gamma_F$ with weights from $[0, p)$.

For an $(r_1, r_2)_{\gamma,r}$-admissible pair $(i_0, j_0) \in (\mathbb{Z}/s)^2$ and $\gamma \in k$, use the description of $E_{cr}(i_0, j_0, \gamma)$ from Subsection 1.4. Then by Corollary 2.7, $V = V^*(E_{cr}(i_0, j_0, \gamma))$ is identified with the additive group of all taken modulo $x_0^p\mathcal{M}_R$ solutions in $R$ of the following system of equations

$$X^{(1)}_i/x_0^{c_i} = X^{(1)}_{i+1}, \quad \text{for all } i \in \mathbb{Z}/s;$$

$$X_{j+1}^p/x_0^{p \delta_j} = X_{j+1} - \delta_{j,0} \gamma^p X^{(1)}_{i+1}, \quad \text{for all } j \in \mathbb{Z}/s.$$  

Note that the first group of equations describes $V_1 = V^*(L_1)$ and the correspondences $X^{(1)}_i \mapsto 0$ and $X_j \mapsto X^{(2)}_j$ with $i, j \in \mathbb{Z}/s$, define the map $V \to V_2$, where $V_2 = V^*(L_2)$ is associated with all taken modulo $x_0^p\mathcal{M}_R$ solutions in $R$ of the equations $X^{(2)}_{j+1}/x_0^{p \delta_j} = X^{(2)}_{j+1}, j \in \mathbb{Z}/s$. As it was noted in Subsection 2.2, the corresponding $\Gamma_F$-action on $V, V_1$ and $V_2$ comes from the natural $\Gamma_F$-action on $R^0$.

Take $x_s \in R$ such that $x_2^{-1} = x_0$ and $x_s \mapsto \pi_s \text{mod } p$ under the natural identification $R/x_0^p\mathcal{M}_R \simeq O/pO$. (This identification is given by the correspondence $r = \lim_{n \to \infty} (r_n \text{mod } p) \mapsto X^{(1)}_n$.) For $i, j \in \mathbb{Z}/s$, set $x^{(i)}_{r_1} := x^{(q^{-1} r_1)}_r$ and $x^{(i)}_{r_2} := x^{(q^{-1} r_2)}_r$, and introduce the variables $Z^{(1)}_i = x^{(1)}_{r_1} X^{(1)}_i$, $Z^{(2)}_j = x^{(1)}_{r_2} X^{(2)}_j$. Then the elements of $V$ appear as
the taken modulo \( m_R \) solutions in \( R_0 := \text{Frac}(R) \) of the following system of equations
\[
\begin{align*}
Z_i^{(1)} p &= Z_{i+1}^{(1)}, & \text{for all } i \in \mathbb{Z}/s; \\
Z_j^{(2)} &= Z_{j+1}, & \text{for all } j \neq j_0 + 1;
\end{align*}
\]
\[
Z_{j_0+1} - Z_{j_0+1}^{(2)} = \gamma^p Z_{i+1}^{(1)} p^r \pi_0 \quad (r_1(i_0) - r_2(j_0))
\]
Note that for the points \( h_1 \in V_1 \) and \( h_2 \in V_2 \) chosen in Subsection 2.4, one has \( Z_i^{(1)}(h_1^0) = Z_i^{(2)}(h_2^0) = 1 \), where \( i \in \mathbb{Z}/s \).
Suppose \( \alpha \in \mathbb{F}_q \) and \( h_1 = ah_0^0 \in V_1 \).
Let \( F_s = k((x_s)) \subset R_0 = \text{Frac} R \). The field-of-norms functor gives a natural embedding of the absolute Galois group \( \Gamma_{F_s} \) of \( F_s \) into \( \Gamma_{F_r} \), where \( F_s = F(\pi_s) \).
Then the restriction \( F_{h_1}(V)|_{\Gamma_{F_r}} \) of the cocycle
\[
\{ F_{h_1}(V)(\tau) = A_{\tau,\alpha}(i_0, j_0, \gamma) \in \mathbb{F}_q \mid \tau \in \Gamma_{F_r} \}
\]
from Subsection 2.4 can be described as follows.
Let \( U \in R_0 \) be such that \( U - U^q = \gamma x_{0}^{(i_0) - r_2(j_0)} \). Then for any \( \tau \in \Gamma_{F_r} \),
\[
\sigma^{\tau}(A_{\tau,\alpha}(i_0, j_0, \gamma)) = \sigma^{\tau}(\alpha)(\tau(U) - U)
\]
and therefore
\[
A_{\tau,\alpha}(i_0, j_0, \gamma) = \sigma^{\tau}(\alpha)\sigma^{-\gamma}(\tau U - U).
\]
The following Lemma is an immediate consequence of the definition of \((r_1, r_2, \alpha)\)-admissible pairs.

**Lemma 2.11.** With above notation let \( C = -(q-1)(r_1(i_0) - r_2(j_0)) \). Then \( C \)
is a prime to \( p \) integer and \( 1 \leq C \leq q - 1 \).

2.6. **Galois Modules** \( V^* (E_{st}(i_0, j_0, \gamma)) \). For an \((r_1, r_2, \alpha)\)-admissible pair \((i_0, j_0) \in (\mathbb{Z}/s)^2\) and \( \gamma \in k \), use the description of \( E_{st}(i_0, j_0, \gamma) \) from Subsection 1.5.

By Subsection 2.2, \( V = V^* (E_{st}(i_0, j_0, \gamma)) \) is identified (as an abelian group)
with the solutions \( \{(X_i^{(1)} \mid i \in \mathbb{Z}/s), \{X_j \mid j \in \mathbb{Z}/s\} \} \in R^{2s} \) of the following system of equations
\[
\begin{align*}
X_i^{(1)} p / x_0^{p a_i} &= X_{i+1}^{(1)}, & \text{for all } i \in \mathbb{Z}/s; \\
X_j^{p} / x_0^{p b_j} + \delta_{j_0} \gamma^p X_0^{(1)} p / x_0^{p a_{i_0} + p} &= X_{j+1}, & \text{for all } j \in \mathbb{Z}/s
\end{align*}
\]
The structure of \( V \) as an element of \( \text{Ext} F_s(\Gamma_{F_s})(V_1, V_2) \) can be described along the lines of Subsection 2.5. The action of \( \Gamma_{F_s} \) on \( V \) comes from the natural \( \Gamma_{F_s} \)-action on \( \mathbb{R}^{2s}_{st} \), and the embedding of \( V \) into \( (R^{0}_{st})^{2s} \) given by the following correspondences:
- if \( i \in \mathbb{Z}/s \) then \( X_i^{(1)} \mapsto X_i^{(1)} \bmod x_0^{p a_i} R; \)
- if \( j \notin \{j_0 + 1, \ldots, j_0 + m_0\} \) then \( X_j \mapsto X_j \bmod x_0^{p a_i} R; \)
- for \( 1 \leq m \leq m_0 \), \( X_{j_0 + m} \mapsto X_{j_0 + m} + \gamma^p(b_{j_0} - a_{i_0} + 1)X_{i_0 + m}^{(1)} Y \mod x_0^{p a_i} R. \)

Similarly to Subsection 2.5, introduce new variables by the relations \( Z_i^{(1)} = x_0^{p r_1(i)} X_i^{(1)}, \ Z_i = x_0^{-r_2(i)} X_i \) and \( Z_i^{(2)} = x_0^{-r_2(i)} X_i^{(2)} \), \( i \in \mathbb{Z}/s \), and rewrite
system of equations (2.1) in the following form:

\[
\begin{align*}
Z_i^{(1)p} &= Z_{i+1} \quad &\text{for all } i \in \mathbb{Z}/s; \\
Z_j^q &= Z_{j+1} \quad &\text{for all } j \notin j_0 + 1; \\
Z_{j_0+1} - Z_{j_0+1}^q &= \gamma^p Z_i^{(1)} \beta_0^q (r_1(i_0) - r_2(j_0) - 1)
\end{align*}
\]

If \( \alpha \in \mathbb{F}_q \) and \( h_1 = \alpha h_0^q \in V_1 \), then the restriction to \( \Gamma_{F} \) of the cocycle \( \{ F_{h_1}(V)(\tau) = A_{\tau,\alpha}(i_0, j_0, \gamma) \mid \tau \in \Gamma_{F_1} \} \) can be described as follows. Let \( U \in R_0 \) be such that

\[
U - U^q = \gamma^p Z_i^{(1)} \beta_0^q (r_1(i_0) - r_2(j_0) - 1).
\]

Then for any \( \tau \in \Gamma_{F_1} \), \( \sigma^j_{h_1}(A_{\tau,\alpha}(i_0, j_0, \gamma)) = \sigma^j_{h_0}(\alpha)(\tau U - U) \).

Thus

\[
A_{\tau,\alpha}(i_0, j_0, \gamma) = \sigma^j_{h_0}(\alpha) \sigma^j_{h_0}(\tau U - U).
\]

The following Lemma is a direct consequence of the definition of \((r_1, r_2)_m\)-admissible pairs, cf. also Proposition 1.24.

**Lemma 2.12.** Let \( C = -(q - 1)(r_1(i_0) - r_2(j_0) - 1). \) Then \( C \) is a prime to \( p \) integer such that \( 1 \leq C < (q - 1)(1 + 1/(p - 1)) \).

### 2.7. Galois Modules \( E_{sp}(j_0, \gamma) \)

In this subsection \((0, j_0)\) is some \((r_1, r_2)_m\)-admissible pair (i.e. \( r_1 + 1/(p - 1) = r_0(j_0) \)) and \( \gamma \in \mathbb{F}_q \). Then \( V = \mathcal{V}^s(E_{sp}(j_0, \gamma)) \) is identified as an abelian group with the solutions

\[
\{(X_i^{(1)} \mid i \in \mathbb{Z}/s), (X_j^{(2)} \mid j \in \mathbb{Z}/s)\} \in R^{2s}
\]

of the following system of equations

\[
\begin{align*}
X_i^{(1)p} / X_i^{(1)} &= X_{i+1}^{(1)} \quad &\text{for all } i \in \mathbb{Z}/s, \\
X_j^{(2)p} / X_j^{(2)} &= X_{j+1}^{(2)} \quad &\text{for all } j \in \mathbb{Z}/s.
\end{align*}
\]

The corresponding \( \Gamma_{F}\)-action comes from the natural \( \Gamma_{F}\)-action on \( R_0^{1} \), and the embedding of \( V \) into \((R_0^{1})^{2s}\) given by the following correspondences:

- if \( i \in \mathbb{Z}/s \) then \( X_i^{(1)} \mapsto X_i^{(1)} \mod x_0^{p_1} m_R \);
- if \( m \in \mathbb{Z}/s \) then \( X_{j_0 + m}^{(2)} \mapsto X_{j_0 + m}^{(2)} + \gamma^m X_i^{(1)} Y \mod x_0^{p_2} m_R \).

If \( \alpha \in \mathbb{F}_q \) and \( h_1 = \alpha h_0^q \in V_1 \) then the cocycle

\[
\{ F_{h_1}(V)(\tau) = A_{\tau,\alpha}^{sp}(j_0, \gamma) \mid \tau \in \Gamma_{F_1} \}
\]

can be described as follows. Note that the point \( h_1 \) corresponds to the collection

\[
\{(\sigma^i(\alpha)x_0^{pr_1(i)} \mid i \in \mathbb{Z}/s), \{\sigma^i_{h_0}(\alpha) x_0^{pr_1(i-j_0)} Y | i \in \mathbb{Z}/s\} \}
\]

Then for \( \tau \in \Gamma_{F_1}, \) \( \tau(h_1) \) corresponds to the collection

\[
\{(\sigma^i(\alpha)x_0^{pr_1(i)} \mid i \in \mathbb{Z}/s), \{\sigma^i_{h_0}(\alpha) x_0^{pr_1(i-j_0)} Y + k(\tau) \log \varepsilon | i \in \mathbb{Z}/s\} \}
\]

Therefore, \( \tau(h_1) - h_1 \) corresponds to the collection

\[
\{0 \mid i \in \mathbb{Z}/s\}, \{\sigma^i h_0(\alpha) x_0^{pr_2(i)} k(\tau) \mid i \in \mathbb{Z}/s\},
\]

which corresponds to \( \sigma^i_{h_0}(\alpha) h_0^{pr_2} \). Therefore, \( A_{\tau,\alpha}^{sp}(j_0, \gamma) = \sigma^i_{h_0}(\alpha) k(\tau) \).

Notice that for any \( \tau \in \Gamma_{F_1} \subset \Gamma_{F_1} \), \( A_{\tau,\alpha}^{sp}(j_0, \gamma) = 0 \).

**Documenta Mathematica 18 (2013) 547–619**
2.8. Fully faithfulness of $CV^*$.

In this subsection we prove the following important property.

**Proposition 2.13.** The functor $CV^*$ is fully faithful.

**Proof.** We must prove that for all $L_1, L_2 \in L^*$, the functor $CV^*$ induces a bijective map

$$\Pi(L_1, L_2) : \text{Hom}_{L^*}(L_2, L_1) \to \text{Hom}_{CM^*}(CV^*(L_1), CV^*(L_2)).$$

By induction on lengths of composition series for $L_1$ and $L_2$ it will be sufficient to verify that for any two simple objects $L_1$ and $L_2$:

- $\Pi(L_1, L_2)$ is bijective;
- the functor $CV^*$ induces injective map

$$\text{EII}(L_1, L_2) : \text{Ext}_{L^*}(L_2, L_1) \to \text{Ext}_{CM^*}(CV^*(L_1), CV^*(L_2)).$$

The first fact has been already checked in Subsection 2.3.

In order to verify the second property, notice that for any two objects $L_1, L_2 \in L^*$, the natural map

$$\text{Ext}_{CM^*}(CV^*(L_1), CV^*(L_2)) \to \text{Ext}_{CM^*}(V^*(L_1), V^*(L_2))$$

is injective. Therefore, we can prove injectivity of $\text{EII}(L_1, L_2)$ on the level of functor $V^*$. In addition, for $n_1, n_2 \in \mathbb{N}$, $\text{Ext}_{L^*}(L_2^{n_1}, L_1^{n_2}) = \text{Ext}_{L^*}(L_2, L_1)^{n_1/n_2}$ (the formation of $\text{Ext}$ is compatible with direct sums). So, by Lemma 1.17, we can replace $L_1$ and $L_2$ by the objects introduced in Subsection 1.5 (where they are denoted also by $L_1$ and $L_2$).

By Proposition 1.27, any element of $\text{Ext}_{L^*}(L_2, L_1)$ appears as a sum of standard extensions of the form $E_{cr}(i, j, \gamma_{ij})$, $E_{st}(i, j, \gamma_{ij})$ and $E_{sp}(j, \gamma_{ij}^{sp})$. Here:

a) $(i, j) \in (\mathbb{Z}/s)^2$ is either $(r_1, r_2)_{cr}$-admissible or $(r_1, r_2)_{sp}$-admissible and all $\gamma_{ij} \in k$; b) $j \in \mathbb{Z}/s$ is such that $(0, j)$ is $(r_1, r_2)_{sp}$-admissible and $\gamma_{ij}^{sp} \in \mathbb{F}_q$.

**Remark.** A couple $(i, j)$ can’t be both $(r_1, r_2)_{cr}$-admissible and $(r_1, r_2)_{st}$-admissible, but it can be $(r_1, r_2)_{cr}$-admissible and $(r_1, r_2)_{sp}$-admissible at the same time.

By Subsections 2.5-2.7, we can attach to these standard extensions the 1-cocycles $A_{r, \alpha}(i, j, \gamma_{ij})$ and $A_{r, \alpha}^{sp}(j, \gamma_{ij}^{sp})$, where $\tau \in \Gamma_{F^*}$. It remains to prove that the sum of these cocycles is trivial only if all corresponding coefficients $\gamma_{ij}$ and $\gamma_{ij}^{sp}$ are equal to 0.

First, we need the following lemma.

**Lemma 2.14.** Suppose for all $(i, j) \in (\mathbb{Z}/s)^2$, the elements $U_{ij} \in R_0 = \text{Frac } R$ are such that $U_{ij} = U_{ij}^{q} = \gamma_{ij} x_i\overline{C}_i$, where all $\gamma_{ij} \in k$ and all $C_i$ are prime to $p$ natural numbers. For $\tau \in \Gamma_{F^*}$, let $B_{\tau}(i, j, \gamma_{ij}) = \tau(U_{ij}) - U_{ij} \in \mathbb{F}_q$. If for all $\alpha \in \mathbb{F}_q$ and all $\tau \in \Gamma_{F^*}$,

$$\sum_{i,j \in \mathbb{Z}/s} \sigma^{i-j}(\alpha)\sigma^{-j} B_{\tau}(i, j, \gamma_{ij}) = 0$$

(2.2)
then all $\gamma_{ij} = 0$.

Proof of Lemma. For different prime to $p$ natural numbers $C_{ij}$ the extensions $\mathcal{F}_s(U_{ij})$ behave independently. Therefore, we can assume that all $C_{ij} = C$ are the same.

Let $j_0 = j_0(j)$ be such that $0 \leq j_0 < s$ and $j_0 \equiv -j \mod s$. Then (2.2) means that for any $\alpha \in \mathbb{F}_q$,

$$B_\alpha := \sum_{i,j \in \mathbb{Z}/s} \sigma^{i-j}(\alpha)\sigma^{j_0}(U_{ij}) \in \mathbb{F}_s.$$ 

Then

$$B_\alpha - B^0_\alpha = \sum_{j \in \mathbb{Z}/s} \left( \sum_{i \in \mathbb{Z}/s} \sigma^{i-j}(\alpha)\gamma_{ij}^{-p} \right) x^{-p\alpha}.$$ 

Looking at the Laurent series of $B_\alpha \in \mathbb{F}_s$ we conclude that all $B_\alpha \in \mathbb{F}_q$. This means that for all $j \in \mathbb{Z}/s$ and $\alpha \in \mathbb{F}_q$, $\sum_{i \in \mathbb{Z}/s} \sigma^i(\alpha)\gamma_{ij} = 0$ and, therefore, all $\gamma_{ij} = 0$. The lemma is proved.

Now suppose that for all $\alpha \in \mathbb{F}_q$ and $\tau \in \Gamma_F$, the sum of cocycles $A_{\tau,\alpha}(i,j,\gamma_{ij})$ and $A^p_{\tau,\alpha}(i,j,\gamma_{ij}^p)$ is zero. Restrict this sum to the subgroup $\Gamma_F$. Then all $\gamma_{ij}$ will disappear and by above Lemma 2.14 all $\gamma_{ij} = 0$. So, for all $\tau \in \Gamma_F$ and $\alpha \in \mathbb{F}_q$, $\sum_{j \in \mathbb{Z}/s} \sigma^{-j}(\alpha)\gamma_{ij}^p = 0$, and this implies that all $\gamma_{ij}^p = 0$.

Corollary 2.15. The functor $\mathcal{V}^*$ is fully faithful on the subcategories of unipotent objects $\mathcal{L}^u$ and of connected objects $\mathcal{L}^c$.

Proof. Indeed, on both categories the map $\Pi(\mathcal{L}_1, \mathcal{L}_2)$ is already bijective on the level of functor $\mathcal{V}^*$.

2.9. Ramification estimates. Suppose $\mathcal{L} \in \mathcal{L}^u$ and $H = \mathcal{V}^*(\mathcal{L})$. For any rational number $v \geq 0$, denote by $\Gamma_F^{(v)}$ the ramification subgroup of $\Gamma_F$ in upper numbering, [22].

Proposition 2.16. If $v > 2 - 1/p$ then $\Gamma_F^{(v)}$ acts trivially on $H$.

A proof can be obtained along the lines of the paper [17] (which adjusts Fontaine’s approach from [14]). Alternatively, one can apply author’s method from [3]: if $\tau \in \Gamma_F^{(v)}$ with $v > 2 - 1/p$ then there is an automorphism $\psi$ of $R$ such that $\psi(x_0) = \tau(x_0)$ and $\psi$ induces the trivial action on $H$; therefore we can assume that $\tau$ comes from the absolute Galois group of $k((x_0))$ and the characteristic $p$ approach from [3] gives the ramification estimate which coincides with the required by the theory of field-of-norms.

Corollary 2.17. If $\tilde{F}$ is the common field-of-definition of points of $\mathbb{F}_p[\Gamma_F]$-modules $\mathcal{V}(\mathcal{L})$ for all $\mathcal{L} \in \mathcal{L}^u$, then $v_p(D(\tilde{F}/F)) < 3 - \frac{1}{p}$, where $D(\tilde{F}/F)$ is the different of the field extension $\tilde{F}/F$.
3. The ring $S$. Let $v = a + p \in W$ and let $S$ be the $p$-adic closure of the divided power envelope of $W$ with respect to the ideal generated by $v$. Use the same symbols $\sigma$ and $N$ for natural continuous extensions of $\sigma$ and $N$ from $W$ to $S$. For $i \geq 0$, denote by $\text{Fil}^i S$ the $i$-th divided power of the ideal $(v)$ in $S$. Then for $0 \leq i < p$, there are $\sigma$-linear morphisms $\phi_i = \sigma/p^i : \text{Fil}^i S \rightarrow S$.\footnote{Note that $\phi_0 = \sigma$ and agree to use the notation $\phi$ for $\phi_{p - 1}$. One can see also that $S$ is the $p$-adic closure of $W(k)[v_1, v_2, \ldots, v_n, \ldots]$, where $v_0 = v$ and for all $n \geq 0$, $v_{n+1}/p = v_n$.}

Consider the ideals $m_S = (p, v, v_1, \ldots, v_n, \ldots)$, $I = (p, v_1, v_2, \ldots)$ and $J = (p, v_1 v, v_2, \ldots, v_n, \ldots)$ of $S$. Then

- $m_S$ is the maximal ideal in $S$;
- $I = \text{Fil}^p S + pS \supset J$;
- $\varphi(I) \subset S$ and $\varphi(J) \subset pS$;
- $\varphi(v_i - 1) = -v_1 (\text{mod } J)$ and $\varphi(v_i) \equiv 1 (\text{mod } J)$.

3.2. The ring of semi-stable periods $\hat{A}_{st}$. Let $R$ be Fontaine’s ring and let $x_0, \varepsilon \in R$ be the elements chosen in Subsection 2.1. Denote by $A_{cr}$ the Fontaine crystalline ring. It is the $p$-adic closure of the divided power envelope of $W(R)$ with respect to the ideal $([x_0] + p)$ of $W(R)$, where $[x_0] \in W(R)$ is the Teichmüller representative of $x_0$. Then for $i \geq 0$, $\text{Fil}^i A_{cr}$ is the $i$-th divided power of the ideal $([x_0] + p)$ in $A_{cr}$. Denote by $\sigma : A_{cr} \rightarrow A_{cr}$ the natural morphism induced by the $p$-th power on $R$. Then for $0 \leq i < p$, there are $\sigma$-linear maps $\phi_i = \sigma/p^i : \text{Fil}^i A_{cr} \rightarrow A_{cr}$. We shall often use the simpler notation $\varphi = \phi_{p - 1}$ and $F(A_{cr}) = \text{Fil}^{p - 1} A_{cr}$.

Notice that $A_{st}$ is provided with the natural continuous $\Gamma_F$-action.

Let $X$ be an indeterminate. Then $\hat{A}_{st}$ is the $p$-adic closure of the ring $A_{cr}[\gamma_i(X) \mid i \geq 0] \subset A_{cr}[X] \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$, where for all $i \geq 0$, $\gamma_i(X) = X^{i!}$.\footnote{The ring $\hat{A}_{st}$ has the following additional structures:
- the $S$-module structure given by the natural $W(k)$-algebra structure and the correspondence $u \mapsto [x_0]/(1 + X)$;
- the ring endomorphism $\sigma$, which is the extension of the above defined endomorphism $\sigma$ of $A_{cr}$ via the condition $\sigma(X) = (1 + X)^p - 1$;
- the continuous $A_{cr}$-derivation $N : A_{st} \rightarrow A_{st}$ such that $N(X) = X + 1$;
- for any $i \geq 0$, the ideal $\text{Fil}^i \hat{A}_{st}$, which is the closure of the ideal $\sum_{i_1 + i_2 \geq i} (\text{Fil}^{i_1} A_{cr}) \gamma_{i_2}(X)$;
- the action of $\Gamma_F$, which is the extension of the $\Gamma_F$-action on $A_{cr}$ such that for all $\tau \in \Gamma_F$, $\tau(X) = [\varepsilon]^{k(\tau)}(X + 1) - 1$. Here all $k(\tau) \in \mathbb{Z}_p$ are such that $\tau(x_0) = \varepsilon^{k(\tau)} x_0$.
}

Note that for $0 \leq m < p$, $\sigma(\text{Fil}^m \hat{A}_{st}) \subset p^m \hat{A}_{st}$ and, as earlier, we can set $\phi_m = p^{-m} \sigma|_{\text{Fil}^m \hat{A}_{st}}$ and introduce the simpler notation $\varphi = \phi_{p - 1}$ and $F(\hat{A}_{st}) = \text{Fil}^{p - 1} \hat{A}_{st}$.

\footnote{Documenta Mathematica 18 (2013) 547–619}
3.3. Construction of semi-stable representations of $\Gamma_F$ with weights from $[0, p]$. For $0 \leq m < p$, consider the category $\mathcal{S}_m$ of quadruples $\mathcal{M} = (M, \text{Fil}^m M, \phi_m, N)$, where $\text{Fil}^m M \subset M$ are $S$-modules, $\phi_m : \text{Fil}^m M \to M$ is a $\sigma$-linear map and $N : M \to M$ is a $W(k)$-linear endomorphism such that for any $s \in S$ and $m \in M$, $N(sx) = N(s)x + sN(x)$.

The morphisms of the category $\mathcal{S}_m$ are $S$-linear morphisms commuting with the corresponding morphisms $\phi_m$ and $N$. Notice that for $0 \leq m < p$, $\mathcal{S}_m$ has a natural structure of the object of the category $\mathcal{S}_m$. As earlier, we shall use the simpler notation $\phi = \phi_{p-1}$ and $F(M) = \text{Fil}^{p-1} M$.

For $0 \leq m < p$, the Breuil category $\mathcal{S}_m$ of strongly divisible $S$-modules of weight $\leq m$ is a full subcategory of $\mathcal{S}_m$ consisting of the objects $\mathcal{M} = (M, \text{Fil}^m M, \phi_m, N)$ such that

1. $M$ is a free $S$-module of finite rank;
2. $(\text{Fil}^m S) M \subset \text{Fil}^m M$;
3. $(\text{Fil}^m M) \cap p M = p\text{Fil}^m M$;
4. $\phi_m(\text{Fil}^m M)$ spans $M$ over $S$;
5. $N\phi_m = p\phi_m N$;
6. $(\text{Fil}^1 S) N(\text{Fil}^m M) \subset \text{Fil}^m M$.

For $\mathcal{M} \in \mathcal{S}_m$, let $T_{st}^* (\mathcal{M})$ be the $\Gamma_F$-module of all $S$-linear and commuting with $\phi_m$ and $N$, maps $f : M \to \mathcal{A}_{st}$ such that $f(\text{Fil}^m M) \subset \text{Fil}^m \mathcal{A}_{st}$. Then one has the following two basic facts:

- $T_{st}^* (\mathcal{M})$ is a continuous $\mathbb{Z}_p [\Gamma_F]$-module without $p$-torsion, its $\mathbb{Z}_p$-rank equals $\text{rk}_S M$, and $V_{st}^* (\mathcal{M}) = T_{st}^* (\mathcal{M}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is semi-stable $\Gamma_F$-module with Hodge-Tate weights from $[0, m]$;

- any semi-stable representation of $\Gamma_F$ with Hodge-Tate weights from $[0, m]$, $0 \leq m < p$, appears in the form $V_{st}^* (\mathcal{M})$ for a suitable $\mathcal{M} \in \mathcal{S}_m$.

By Theorem 1.3 [6] these facts follow from the existence of strongly divisible lattices in $S \otimes W F$-modules associated with weakly admissible $(\phi_0, \hat{\mathbb{N}})$-modules with filtration of length $m$. Breuil proved this for all $m \leq p - 2$ but his method can be easily extended to cover the case $m = p - 1$ as well, cf. also [7].

3.4. The category $\mathcal{L}_f$. In this section we introduce $W$-analogues of Breuil’s $S$-modules from the category $\mathcal{S}_{p-1}$ and prove that they can be also used to construct semi-stable representations of $\Gamma_F$ with Hodge-Tate weights from $[0, p)$.

**Definition.** Let $\mathcal{L}$ be the category of $\mathcal{L} = (L, F(L), \varphi, N_S)$, where $L \supset F(L)$ are $W$-modules, $\varphi : F(L) \to L$ is a $\sigma$-linear morphism of $W$-modules and $N_S : L \to L_S := L \otimes_W S$ is such that for all $w \in W$ and $l \in L$, $N_S(wl) = N(w)l + (w \otimes 1)N_S(l)$. For $L_1 = (L_1, F(L_1), \varphi, N_S) \in \mathcal{L}$, the morphisms $\text{Hom}_W (\mathcal{L}, \mathcal{L}_1)$ are $W$-linear $f : L \to L_1$ such that $f(F(L)) \subset F(L_1)$, $f\varphi = \varphi f$ and $fN_S = N_S(f \otimes 1)$.
Let $\mathcal{A}_s = (\hat{A}_s, F(\hat{A}_s), \varphi, N_S)$, where $N_S = N \otimes 1$. Then $\mathcal{A}_s$ is an object of the category $\hat{\mathcal{L}}$.

Suppose $\mathcal{L} = (L, F(L), \varphi, N_S) \in \hat{\mathcal{L}}$.

Set $L_S := L \otimes_{W} S$, $F(L_S) = (F(L) \otimes 1)S + (L \otimes 1)\text{Fil}^P S$, and $\varphi_S : F(L_S) \longrightarrow F(L_S)$ is a unique $\sigma$-linear map such that $\varphi_S|_{F(L) \otimes 1} = \varphi \otimes 1$ and for any $s \in \text{Fil}^P S$ and $l \in L$, $\varphi_S(l \otimes s) = (\varphi(v^{p-1}l) \otimes 1)\varphi(s)/\varphi(v^{p-1})$.

**Definition.** Denote by $\mathcal{L}^f$ the full subcategory in $\hat{\mathcal{L}}$ consisting of the quadruples $\mathcal{L} = (L, F(L), \varphi, N_S)$ such that

- $L$ is a free $W$-module of finite rank;
- $v^{p-1}L \subset F(L)$, $F(L) \cap pL = pF(L)$ and $L = \varphi(F(L)) \otimes_{\sigma W} W$;
- for any $l \in F(L)$, $vN_S(l) \in F(L_S)$ and $\varphi_S(vN(l)) = cN_S(\varphi(l))$, where $c = 1 + v^p/p$.

It can be easily seen that for $\mathcal{L} = (L, F(L), \varphi, N_S) \in \mathcal{L}^f$ and the map $N = N_S \otimes 1 : L_S \longrightarrow L_S$, the quadruple $\mathcal{L}_S = (L_S, F(L_S), \varphi, N)$ is the object of the category $S_{p-1}$.

The main result of this Subsection is the following statement.

**Proposition 3.1.** For any $\mathcal{M} = (M, F(M), \varphi, N) \in S_{p-1}$, there is an $\mathcal{L} = (L, F(L), \varphi, N_S) \in \mathcal{L}^f$ such that $\mathcal{M} = \mathcal{L}_S$.

**Corollary 3.2.** a) If $\mathcal{L} \in \mathcal{L}^f$ and $T^{st}_{\mathcal{L}}(\mathcal{L}) = \text{Hom}_{\mathcal{L}}(\hat{\mathcal{L}}, \hat{\mathcal{A}}_{st})$ with the induced structure of $\mathbb{Z}_p[\Gamma_F]$-module then $V^{st}_{\mathcal{L}}(\mathcal{L}) = T^{st}_{\mathcal{L}}(\mathcal{L}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a semi-stable $\mathbb{Q}_p[\Gamma_F]$-module with Hodge-Tate weights from $[0, p)$ and $\dim_{\mathbb{Q}_p} V^{st}_{\mathcal{L}}(\mathcal{L}) = \text{rk}_{W} L$.

b) For any semi-stable $\mathbb{Q}_p[\Gamma_F]$-module $V^{st}_{\mathcal{L}}$ with Hodge-Tate weights from $[0, p)$, there is an $\mathcal{L} \in \mathcal{L}^f$ such that $V^{st}_{\mathcal{L}} \simeq V^{st}_{\mathcal{L}}(\mathcal{L})$.

**Proof of Proposition 3.1.** Let $d$ be a rank of $M$ over $S$. If $L \subset M$ is a free $W$-submodule of rank $d$ and $M$ is generated by the elements of $L$ over $S$ we say that $L$ is $W$-structural (with respect to $M$).

Let $F(L) = F(M) \cap L$.

**Lemma 3.3.** If $L$ is $W$-structural for $M$ then

a) $F(L) \supset v^{p-1}L$;

b) $F(L) \cap pL = pF(L)$;

c) $F(L)$ is a free $W$-module of rank $d$.

**Proof.** a) $v^{p-1}L \subset (\text{Fil}^{p-1}S)M \cap L \subset F(M) \cap L = F(L)$.

b) $F(L) \cap pL = L \cap F(M) \cap pL = F(M) \cap pL = F(M) \cap pM \cap pL = pF(M) \cap pL = pF(L)$.

c) $F(L)$ has no $p$-torsion. Therefore, it will be sufficient to prove that $F(L)/pF(L)$ is a free $k[[\nu]]$-module of rank $d$. Consider the following natural embeddings of $k[[\nu]]$-modules

$L/pL \supset F(L)/pF(L) \supset v^{p-1}L/pv^{p-1}L \simeq L/pL$.
(Use b) and that $pl \cap v^{p-1}L = pv^{p-1}L$.) It remains to note that $L/pL$ is free of rank $d$ over $k[[v]]$.

The Lemma is proved. \hfill \Box

Suppose $L$ is $W$-structural for $M$.

**Lemma 3.4.** If $L$ is $W$-structural then $\varphi(F(L))$ spans $M$ over $S$.

**Proof.** The equality $S = W + \text{Fil}_p S$ implies that $M = L + (\text{Fil}_p S)L = L + (\text{Fil}_p S)M$. Therefore,

$$F(M) = F(M) \cap L + (\text{Fil}_p S)M = F(L) + (\text{Fil}_p S)L$$

(use that $F(M) \supset (\text{Fil}_p S)M$) and in notation of Subsection 3.1 one has

$$F(M) = F(L) + v(L) + JM.$$  

This implies that $\varphi(F(L)), \varphi(v(L))$ and $\varphi(JM)$ span $M$ over $S$. But for any $l \in L, \varphi(v(L)) = \varphi(v_1)/(\varphi(v^{p-1}) = (1 - v_1)^{-1}(v^{p-1}) \equiv \varphi(v^{p-1}) \mod mS.$ For similar reasons, $\varphi(JM) \subset pM \subset mS.$ This means that $\varphi(F(L))$ spans $M$ modulo $mS.$ The lemma is proved. \hfill \Box

By above lemma it remains to prove the existence of a $W$-structural $L$ for $M$ such that $\varphi(F(L)) \subset L$.

Let $\phi_0$ be a $\sigma$-linear endomorphism of the $S$-module $M \in S_{p-1}$ such that for all $m \in M, \phi_0(m) = \varphi(v^{p-1})m/\varphi(v^{p-1})$. Clearly, $\phi_0(mS) \subset mS$ and, therefore, it induces a $\sigma$-linear endomorphism $\sigma_0$ of the $k$-vector space $M_k = M/mS$.  

**Lemma 3.5.** Suppose $n \in \mathbb{Z}_{\geq 0}, L$ is $W$-structural and $\varphi(F(L)) \subset L + p^nM$. Then there is a $W$-structural $L'$ for $M$ such that $\varphi(F(L')) \subset L' + p^nJM$.  

**Proof.** Denote by $F(L)_k$ the image of $F(L)$ in the $k$-vector space $M/mS = L/(mS \cap W)L = L_k$. Let $s = \dim_k F(L)_k$, then $s \leq d = \dim_k L_k$. Choose a $W$-basis $e^{(1)}, \ldots, e^{(d)}$ of $L$ and a $W$-basis $f^{(1)}, \ldots, f^{(d)}$ of $F(L)$ such that

- $1 \leq i \leq s, f^{(i)} = e^{(i)}$ and for $s < i \leq d, f^{(i)} \in vL$.

It will be convenient to use the following vector notation; $\bar{e} = (\bar{e}_1, \bar{e}_2)$, where $\bar{e}_1 = (e^{(1)}, \ldots, e^{(s)})$ and $\bar{e}_2 = (e^{(s+1)}, \ldots, e^{(d)})$, and $\bar{f} = (f_1, f_2)$, where $f_1 = \bar{e}_1$ and $f_2 = (f^{(s+1)}, \ldots, f^{(d)})$.

Then in obvious notation one has $(\bar{e}_1, \bar{f}_2) = (\bar{e}_1, \bar{e}_2)C$, where $C \in \text{GL}_d(S)$. Clearly, $C \equiv C_0 + p^n v_1 C_1 \mod p^nJ$ with $C_0 \in \text{GL}_d(W)$ and $C_1 \in M_d(W)$. Clearly, $\varphi(F(L)) \subset L + p^nJM$ iff $C_1 \equiv 0 \mod mS$. Choose $\bar{g} = (\bar{g}_1, \bar{g}_2) \in L^d$ and set

$$\bar{e}_1' = (e^{(1)}, \ldots, e^{(s)}) = e_1 + p^n(v_1 - v^{p-1})\bar{g}_1$$

$$\overline{e}_2' = (e^{(s+1)}, \ldots, e^{(d)}) = e_2 + p^n(v_1 - v^{p-1})\bar{g}_2$$

Clearly, the coordinates of $\bar{e}' = (\bar{e}_1', \bar{e}_2')$ give an $S$-basis of $M$ and we can introduce the structural $W$-module $L' = \sum e^{(i)}$ for $M$.  

*Documenta Mathematica* 18 (2013) 547–619
Prove that the elements $e^{(i)}$, $1 \leq i \leq s$, and $f^{(i)}$, $s < i \leq d$, generate $F(L') \mod p^n JM$. Indeed, we have

$$L + p^n JM = L' + p^n IM$$

and this implies that the image $F(L)_k$ of $F(L)$ in $L_k$ coincides with its analogue $F(L')_k$. In addition, for $1 \leq i \leq s$,

$$e^{(i)} \in L' \cap (F(L) + p^n IM) \subset L' \cap F(M) = F(L').$$

Therefore, it would be sufficient to prove that $(vL) \cap F(L') \mod p^n JM$ is generated by the images of $ve^{(i)}$, $1 \leq i \leq s$, and $f^{(s+1)}, \ldots, f^{(d)}$. But relation (3.1) implies that $vL + p^n JM = vL' + p^n JM$ and

$$(vL') \cap F(L') \mod p^n JM = (vL) \cap F(L) \mod p^n JM.$$ 

It remains to note that for $1 \leq i \leq s$, $ve^{(i)} \equiv ve^{(i)} \mod p^n JM$.

Therefore, we can define special bases for $L'$ and $F(L')$ by the relations $f_1 = e_1$ and $f'_2 = f_2$ and obtain that

$$(\varphi(f'_1), \varphi(f'_2)) = (\varphi(f_1), \varphi(f_2)) + p^n v_1(\sigma_0 \bar{g}_1, 0) \mod p^n JM$$

and

$$(\varphi(f'_1), \varphi(f'_2)) \equiv (e'_1, e'_2)C_0 + p^n v_1(\bar{g}_1, \bar{g}_2)C_0 + p^n v_1((e_1, e_2)C_1 - (\bar{g}_1, \bar{g}_2)C_0 + (\sigma \bar{g}_1, 0)) \mod p^n JM$$

So, $\varphi(F(L')) \subset L' + p^n JM$ if and only if there is an $\bar{g} = (\bar{g}_1, \bar{g}_2) \in L^d$ such that $(\sigma_0 \bar{g}_1, 0) \equiv (\bar{g}_1, \bar{g}_2)C_0 + \bar{h} \mod (m_S \cap W)L$, where $\bar{h} = (e_1, e_2)C_1 \in L$ and $C_0 \mod m_S \in GL_d(k)$. The existence of such vector $\bar{g}$ is implied by Lemma 3.6 below.

**Lemma 3.6.** Suppose $V$ is a d-dimensional vector space over $k$ with a $\sigma$-linear endomorphism $\sigma_0 : V \rightarrow V$ and $\bar{a} = (\bar{a}_1, \bar{a}_2) \in V^d$, where $\bar{a}_1 \in V^s$ and $\bar{a}_2 \in V^{d-s}$. Then for any $C \in GL_d(k)$ there is an $\bar{g} = (\bar{g}_1, \bar{g}_2) \in V^d$ with $\bar{g}_1 \in V^s$ and $\bar{g}_2 \in V^{d-s}$ such that

$$\begin{align*}
(\sigma_0 \bar{g}_1, 0) & = \bar{g}C + \bar{a}.
\end{align*}$$

**Proof.** Let $C^{-1} = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$ with the block matrices of sizes $s \times s$, $(d-s) \times s$, $s \times (d-s)$ and $(d-s) \times (d-s)$. Then the equality (3.2) can be rewritten as

$$\begin{align*}
(\sigma_0 \bar{g}_1)D_{11} &= \bar{g}_1 + \bar{a}_1' \\
(\sigma_0 \bar{g}_1)D_{21} &= \bar{g}_2 + \bar{a}_2'
\end{align*}$$

where $(\bar{a}_1', \bar{a}_2') = \bar{a}C^{-1}$. Clearly, it will be sufficient to solve the first equation in $\bar{g}_1$, but this is a special case of Lemma 1.1.

**Lemma 3.7.** Suppose $n \geq 0$ and $L$ is $W$-structural for $M$ such that $\varphi(F(L')) \subset L + p^n JM$. Then there is a $W$-structural $L'$ for $M$ such that $\varphi(F(L')) \subset L' + p^{n+1} M$.
Varieties with Bad Reduction at 3 Only

585

Definition. Denote by $\mathcal{W}$-module in $L \otimes W$ is such that the coordinates of $f = \phi D$ form a $\mathcal{W}$-basis of $F(L)$. Then $\phi(f) = \epsilon + p^n \bar{h}$, where $\bar{h} \equiv 0 \mod JM$. Let $e' = \epsilon + p^n \bar{h}$ and let $L'$ be a $\mathcal{W}$-submodule in $M$ spanned by the coordinates of $e'$. Clearly, $L'$ is $\mathcal{W}$-structural. Prove that $F(L')$ is spanned by the coordinates of $e'$. Indeed, suppose $\phi$ and $\phi'$ have the coordinates $\epsilon^{(i)}$ and, resp., $\epsilon'^{(i)}$, $1 \leq i \leq s$. Then for all $i$, $e^{(i)} = e^{(i)} + p^n h^{(i)}$, where $h^{(i)} \in JM \subset (\text{Fil}^p S)M$. This means that a $\mathcal{W}$-linear combination of $e^{(i)}$ belongs to $F(M)$ if and only if the same linear combination of $e^{(i)}$ belongs to $F(M)$. This implies that $\phi' D$ spans $F(L')$ over $\mathcal{W}$ because $\epsilon D$ spans $F(L)$ over $\mathcal{W}$. Then $\phi(F(L')) \subset L' + p^{n+1}M$ because $\varphi(\bar{h}) \in \varphi(M)$ (use that $\varphi(J) \subset pS$) and

$$\varphi(e') = \varphi(\epsilon D + p^n \bar{h} D) = \epsilon + p^n \bar{h} + p^n \varphi(\bar{h}) \sigma(D) \equiv \epsilon' \mod p^{n+1} M$$

It remains to notice that applying above Lemmas 3.6 and 3.7 one after another we shall obtain a sequence of $\mathcal{W}$-structural modules $L_n$ such that for all $n \geq 0$, $L_n + p^{n+1}M = L_{n+1} + p^{n+1} M$, where $L_0 \otimes \mathcal{W} S = M$. Therefore, $L = \lim_{\to} L_n/p^n$ is $\mathcal{W}$-structural and $\varphi(L) \subset L$.

The proposition is completely proved.

3.5. The categories $\mathcal{L}^1$ and $\mathcal{L}^1'$

Definition. $\mathcal{W}$-module $L$ is $p$-strict if it is isomorphic to $\otimes_{1 \leq i \leq s} \mathcal{W}/p^{n_i}$, where $n_1, \ldots, n_s \in \mathbb{N}$.

In particular, if $L$ is $p$-strict and $pL = 0$ then $L$ is a free $\mathcal{W}_L$-module. The $p$-strict modules can be efficiently studied via devissage due to the following property.

Lemma 3.8. $L$ is $p$-strict if and only if $pL$ and $L/pL$ are $p$-strict.

Proof. Specify Breuil’s proof of a similar statement but for more complicated ring $S = \mathcal{W}_{dp}$ from [6].

Definition. Denote by $\mathcal{L}'$ the full subcategory in $\mathcal{L}$ consisting of the quadruples $L = (L, F(L), \varphi, N_S)$ such that

- $L$ is $p$-strict;
- $L \subset F(L)$, $F(L) \cap pL = \varphi(F(L)) \otimes_{\mathcal{W}} \mathcal{W}$;
- for any $l \in F(L)$, $vN_S(l) \in F(L_S)$ and $\varphi_S(vN_S(l)) = cN_S(\varphi(l))$, where $c = 1 + u^p/p$.

Definition. Denote by $\mathcal{L}'[1]$ the full subcategory in $\mathcal{L}'$, which consists of objects killed by $p$.

The category $\mathcal{L}'[1]$ is not very far from the category $\mathcal{L}^1$ introduced in Section 1. Indeed, suppose $L = (L, F(L), \varphi, N_S) \in \mathcal{L}'[1]$. Note that $N_S(L) \subset L_{S_1} := L \otimes_{\mathcal{W}_1} S_1 = L/u^p L \oplus (L \otimes \text{Fil}^p S_1)$. (Remind that $S_1 = S/pS = \mathcal{W}_1/u^p \mathcal{W}_1 \oplus \text{Fil}^p S_1$.) With this notation we have the following property.

Documenta Mathematica 18 (2013) 547–619
Proposition 3.9. There is a unique $N : L \rightarrow L/u^{2p}$ such that

a) for any $l \in L$, $N(l) \otimes 1 = cN_S(l)$ in $L_{S_1}$, where $c = 1 + u^p/p \in S^*$;

b) $(L, F(L), \varphi, N) \in \mathcal{L}^*$.

Proof. Let $N_1 := cN_S : L \rightarrow L_{S_1}$. Then for any $w \in W_l$ and $l \in L$, one has

\[ N_1(wl) = N(w)l + wN_1(l) \] (use that $N(c) = 0$ in $S_1$) and there is a commutative diagram (use that $\sigma(c) = 1$ in $S_1$)

\[
\begin{array}{ccc}
F(L) & \xrightarrow{\varphi} & L \\
\downarrow{uN_1} & & \downarrow{N_1} \\
F(L)_{S_1} & \xrightarrow{\varphi} & L_{S_1}
\end{array}
\]

Prove that $N_1(\varphi(F(L)) \subset L/u^pL$ and, therefore, $N_1(L) \subset L/u^pL$.

Indeed, $(uN_1)(F(L)) \subset uN_1(L) = F(L)_S \subset (uL/u^pL \oplus (uL)\operatorname{Fil}^pS_1)
\cap (F(L)/u^pL \oplus (uL)\operatorname{Fil}^pS_1) \subset F(L)/u^pL \oplus (uL)\operatorname{Fil}^pS_1$. This implies that $N_1(\varphi(F(L)) \subset \varphi_S(uN_1(F(L))) \subset L/u^pL$ because $\varphi_S(u\operatorname{Fil}^pS_1) = 0$. So, by Proposition 1.3 there is a unique $N : L \rightarrow L/u^{3p}$ such that $N \operatorname{mod} u^p = N_1$ and $(L, F(L), \varphi, N) \in \mathcal{L}^*$.

\[ \square \]

Corollary 3.10. With above notation the correspondence

\[ (L, F(L), \varphi, N_S) \mapsto (L, F(L), \varphi, N) \]

induces the equivalence of categories $\Pi : \mathcal{L}'[1] \rightarrow \mathcal{L}'^*$.

Proof. We must verify that our correspondence is surjective on objects and bijective on morphisms. The first holds because $N_S = c^{-1}N \operatorname{mod} u^p$ and the second — because a $W_l$-linear map $f$ commutes with $N$ iff it commutes with $N \operatorname{mod} u^p$ (use Proposition 1.2) iff $f \otimes_{W_l} S_1$ commutes with $N_S$.

\[ \square \]

Corollary 3.11. The category $\mathcal{L}'[1]$ is pre-abelian.

Proof. Corollary 3.10 and Proposition 1.3 imply that $\mathcal{L}'[1]$ is pre-abelian. This can be extended then to the whole category $\mathcal{L}'$ by Breuil’s method from [6] via above Lemma 3.8.

\[ \square \]
Varieties with Bad Reduction at 3 Only

According to Appendix A, we can use the concept of $p$-divisible group \( \{ L(n), i_n \}_{n \geq 0} \) in \( L^e \). In this case \( L(n) = (L_n, F(L_n), \varphi, N_S) \), where all \( L_n \) are free \( W/p^n \)-modules of the same rank equal to the height of this \( p \)-divisible group. We have obvious equivalence of the category \( L^e \) and the category of \( p \)-divisible groups of finite height in \( L^e \).

**Definition.** Denote by \( L^{f, e} \) the full subcategory in \( L^{e} \), which consists of strict subobjects of \( p \)-divisible groups in \( L^{e} \). By \( L^{f, e}[1] \) we denote the full subcategory in \( L^{f, e} \) consisting of all objects killed by \( p \).

It is easy to see that \( L^{f, e} \) contains all strict subquotients of the corresponding \( p \)-divisible groups. Contrary to the case of filtered modules coming from crystalline representations, the categories \( L^{f, e} \) and \( L^{e} \) do not coincide but they have the same simple objects.

Note that the functor \( \Pi \) from Corollary 3.10 identifies simple objects of the categories \( L^{f, e} \) and \( L^{e} \) and for any two objects \( L_1, L_2 \in L^{e}[1] \), we have a natural isomorphism \( \text{Ext}_{L^{e}[1]}(L_1, L_2) = \text{Ext}_{L^{e}}(\Pi(L_1), \Pi(L_2)) \). One can use the methods of Subsection 1.2 to extend the concepts of etale, connected, unipotent and multiplicative objects to the whole category \( L^{f, e} \). The starting point for this extension is the case of \( W(k) \)-modules, which is well-known from the classical Dieudonne theory [10]. Then we obtain the following standard properties:

- for any \( L \in L^{e} \), there are a unique maximal etale subobject \( (L^{e, t}, i^{e, t}) \) and a unique maximal connected quotient object \( (L^{e, c}, j^{e, c}) \) in \( L^{e} \) such that the sequence \( 0 \to L^{e, t} \xrightarrow{i^{e, t}} L \xrightarrow{j^{e, c}} L^{e, c} \to 0 \) is exact and the correspondences \( L \mapsto L^{e, t} \) and \( L \mapsto L^{e, c} \) are also objects of \( L^{f, e} \);

- for any \( L \in L^{e} \), there are a unique maximal unipotent subobject \( (L^{u, t}, i^{u, t}) \) and a unique maximal multiplicative quotient object \( (L^{m, c}, j^{m, c}) \) in \( L^{e} \) such that the sequence \( 0 \to L^{u, t} \xrightarrow{i^{u, t}} L \xrightarrow{j^{m, c}} L^{m, c} \to 0 \) is exact and the correspondences \( L \mapsto L^{u, t} \) and \( L \mapsto L^{m, c} \) are also objects of \( L^{f, e} \).

Denote by \( L^{e, t, f, e}, L^{e, u, t, f, e} \) and \( L^{m, t, f, e} \) the full subcategories in \( L^{e} \) consisting of, resp., etale, connected, unipotent and multiplicative objects. We have also the corresponding full subcategories \( L^{e, t, f, e}, L^{u, t, f, e} \) and \( L^{m, t, f, e} \) in \( L^{f, e} \).

The results of Subsection 1.5 and Appendix A imply that in the category \( L^{f, e} \):

- there is a unique etale \( p \)-divisible group \( L^\infty(0) := \{ L(n)(0), i_n \}_{n \geq 0} \) of height 1 such that \( L^{(1)}(0) = L(0) \);

- there is a unique multiplicative \( p \)-divisible group of height 1, \( L^\infty(1) := \{ L(n)(1), i_n \}_{n \geq 0} \) such that \( \Lambda^{(1)}(1) = \Lambda(1) \);

- for any \( p \)-divisible group \( L^\infty \) there are functorial exact sequences of \( p \)-divisible groups

\[
0 \to L^\infty, e \to L^\infty \to L^\infty, c \to 0
\]
Here $\mathcal{L}^{c,ft}$ and $\mathcal{L}^{\infty,\ast}$ are products of several copies of $\mathcal{L}^{\infty}(0)$ and, resp., $\mathcal{L}^{\infty}(1)$, and $\mathcal{L}^{\infty,c}$ and $\mathcal{L}^{\infty,\ast}$ are $p$-divisible groups in the categories $\mathcal{L}^{c,ft}$ and, resp., $\mathcal{L}^{u,ft}$.

4. Semistable modular representations with weights $[0, p)$

In this section we prove that all killed by $p$ subquotients of Galois invariant lattices of semistable $\mathbb{Q}_p[\Gamma_F]$-modules with Hodge-Tate weights $[0, p)$ can be obtained via the functor $\mathcal{V}^t$ from Section 2.

4.1. The functor $\mathcal{V}^t : \mathcal{L}^t \to \mathcal{M}_{\Gamma_F}^t$. For $n \geq 1$, introduce the objects $\mathcal{A}_{st,n} = (\hat{A}_{st,n}, F(\hat{A}_{st,n}), \varphi, N_S)$ of the category $\mathcal{L}_n^t$ with $\hat{A}_{st,n} = \hat{A}_{st}/p^n\hat{A}_{st}$, $F(\hat{A}_{st,n}) = F(\hat{A}_{st})/p^n F(\hat{A}_{st})$ and induced $\varphi$ and $N_S$. Let $\mathcal{A}_{st,\infty} = (\hat{A}_{st,\infty}, F(\hat{A}_{st,\infty}), \varphi, N_S)$ be the inductive limit of all $\mathcal{A}_{st,n}$. For $\mathcal{L} \in \mathcal{L}_n^t$, set $\mathcal{V}^t(\mathcal{L}) = \text{Hom}_\pi(\mathcal{L}, \mathcal{A}_{st,\infty})$ with the induced structure of $\Gamma_F$-module. This gives the functor $\mathcal{V}^t : \mathcal{L}^t \to \mathcal{M}_{\Gamma_F}^t$. We shall use the same notation for its restriction to the category $\mathcal{L}^{c,ft}$.

**Proposition 4.1.** Suppose $\mathcal{L} = (L, F(L), \varphi, N_S) \in \mathcal{L}^t$. Then $N_S|_{\varphi(F(L))}$ is nilpotent.

By deissage and Corollary 3.10 this is implied by the following statement for the objects of the category $\mathcal{L}_n^t$.

**Lemma 4.2.** If $\mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}^t$ then $N^p(\varphi(F(L))) \subseteq u^p L$.

**Proof.** For any $l \in F(L), N(\varphi(l)) = \varphi(u^N(l))$. Use induction to prove that for $1 \leq m \leq p, N^m(\varphi(l)) \equiv \varphi(u^m N^m(l)) \mod u^p L$ and use then that $\varphi(u^p N^p(l)) \in \varphi(u F(L)) \subseteq u^p L$. \hfill $\square$

**Proposition 4.3.** For $n \geq 1$, $\oplus_{j \geq 0} A_{cr,n} \gamma_j(\log(1 + X))$ is the maximal $W(k)$-submodule of $\hat{A}_{st,n}$ where $N$ is nilpotent.

**Proof.** For any $j \geq 1$, one has $N(\gamma_j(\log(1 + X))) = \gamma_{j-1}(\log(1 + X))$ and $N$ is nilpotent on $\oplus_{j \geq 0} A_{cr,n} \gamma_j(\log(1 + X))$. Therefore, it will be sufficient to prove that

$$\text{Ker} \left(N^p|_{\hat{A}_{st,1}}\right) = \oplus_{0 \leq j < p} A_{cr,1} \gamma_j(\log(1 + X)).$$

Let $C = F_p\langle X \rangle$ be the divided power envelope of $F_p[X]$ with respect to the ideal $(X)$. Then $C = F_p[X_0, X_1, \ldots, X_n, \ldots]_{<p}$ is the ring of polynomials in $X_i := \gamma_i(X)$, where for all $i \geq 0, X_i^p = 0$.

Let $m_C$ be the maximal ideal of $C$ and $Y = \log(1 + X) \in C$. Then $Y \equiv X_0 - X_1 \mod m_C^2$, and for all $j \geq 0, \gamma_j(Y) \equiv X_j - X_{j+1} \mod m_C^2$. This implies that with $Y_j = \gamma_j(Y)$ for all $j \geq 0$,

$$C = F_p[X_0, Y_0, \ldots, Y_n, \ldots]_{<p} = F_p\langle Y \rangle[X]_{<p} = \oplus_{0 \leq i < p} F_p(Y) \gamma_i(X).$$

**Documenta Mathematica 18 (2013) 547–619**
Indeed, take $i$ and $N^\prime$. If $N^\prime$ is an $A_{cr,1}$-derivation, $N^\prime(X) = X + 1$ and $N^\prime(\gamma_j(Y)) = \gamma_j(Y)$, we obtain that for any $P = \sum_{i,j} \alpha_{ij}X^i\gamma_j(Y) \in F_p\{Y\}[X]_{<p}$ with $\alpha_{ij} \in F_p$,

$$N^p(P) = \sum_{i,j} \alpha_{ij}X^i\gamma_j(Y) + \sum_{i,j} (i+1)\alpha_{i+1,j}\gamma_j(Y)X^i + \sum_{i,j} \alpha_{i,j+p}X^i\gamma_j(Y).$$

If $P \in \text{Ker}N^p$ then for all involved indices $i, j$,

$$i\alpha_{ij} + (i+1)\alpha_{i+1,j} + \alpha_{i,j+p} = 0.$$

This implies that $\alpha_{ij} = 0$ if either $i \neq 0$ or $j \geq p$.

Indeed, take $i = p-1$. Then $-\alpha_{p-1,j} + \alpha_{p-1,j+p} = 0$. Because for $j \geq 0$, $\alpha_{p-1,j} = 0$ it implies that all $\alpha_{p-1,j} = 0$. Then proceed similarly with $i = p-2$ and so on. This proves that all $\alpha_{ij} = 0$ if $i \neq 0$. It remains to note that for $i = 0$, our relations give $\alpha_{0,j+p} = 0$ for all $j \geq p$.

As earlier, consider the category $\tilde{\mathcal{L}}$. Remind that its objects are the triples $(L, F(L), \varphi)$, where $L \supset F(L)$ are $W$-modules and $\varphi : F(L) \rightarrow L$ is a $A$-linear morphism. For any object $\mathcal{L} = (L, F(L), \varphi) \in \tilde{\mathcal{L}}$, agree to use the same notation $\mathcal{L}$ for the corresponding object $(L, F(L), \varphi) \in \tilde{\mathcal{L}}$.

For all $n \geq 0$, set $A_{cr,n} = (A_{cr}, F(A_{cr,n}), \varphi) \in \tilde{\mathcal{L}}$ with $A_{cr,n} = A_{cr}/p^nA_{cr}$, $F(A_{cr,n}) = F(A_{cr})/p^nF(A_{cr})$ and induced $\varphi$. Here the $W$-module structure on $A_{cr,n}$ is defined by the morphism of $W(k)$-algebras $W \rightarrow A_{cr,n}$ such that $u \mapsto [x_0]_{\mathcal{L}}$. Denote by $A_{cr,\infty}$ the inductive limit of all $A_{cr,n}$.

Suppose $\mathcal{L} \in \mathcal{L}^\prime$ and $f \in \text{Hom}_\mathcal{L}(\mathcal{L}, A_{cr,n})$. Then by Propositions 4.1 and 4.3,

$$f(\varphi(F(L))) \subset \oplus_{j \geq 0} A_{cr,n}\gamma_j(\log(1 + X)).$$

Consider the formal embedding of the algebra $A_{st,n}$ into the completion $\prod_{j \geq 0} A_{cr,n}\gamma_j(\log(1 + X))$ of $\oplus_{j \geq 0} A_{cr,n}\gamma_j(\log(1 + X))$ such that $X \mapsto \sum_{j \geq 1} \gamma_j(\log(1 + X))$. Then any element of $A_{st,n}$ can be uniquely written in the form $\sum_{j \geq 0} a_j \gamma_j(\log(1 + X))$, where all $a_j \in A_{cr,n}$. Note that the $W$-module structure on $A_{st,n}$ is given via the map

$$u \mapsto [x_0](1 + X) = [x_0] \sum_{j \geq 0} (-1)^j \gamma_j(\log(1 + X)).$$

For $j \geq 0$, introduce the $W(k)$-linear maps $f_j \in \text{Hom}(L, A_{cr,n})$ such that for any $l \in L$, one has $f(l) = \sum_{j \geq 0} f_j(l)\gamma_j(\log(1 + X))$. Then using methods from [6] obtain the following property.

**Proposition 4.4.** a) The correspondence $f \mapsto f_0$ induces isomorphism of abelian groups $\mathcal{V}^\prime(\mathcal{L}) = \text{Hom}_\mathcal{L}(\mathcal{L}, A_{cr,n})$;
b) for any $j \geq 0$ and $l \in L$, $f_j(l) = f_0(N^j(l))$.

**Corollary 4.5.** The functor $\mathcal{V}^\prime$ is exact.
Proof. Let $L_1'$ be the full subcategory of $L_1$, consisting of the triples $(L, F(L), \varphi)$ coming from all $L = (L, F(L), \varphi, N) \in L_1'$. By Proposition 4.4 it will be sufficient to prove that the functor $\mathcal{V}_0^0 : L_1' \to (Ab)$, such that $\mathcal{V}_0^0(L) = \text{Hom}_{\mathcal{C}_0}(L, \mathcal{A}_{cr,\infty})$, is exact. The verification can be done by devisage along the lines of paper [13].

Remark. One can simplify the verification of above corollary by replacing $\mathcal{A}_{cr,1}$ by the corresponding object $\tilde{\mathcal{A}}_{cr,1}$ related to the module $\tilde{A}_{cr,1} = (R/x_0^pT_1, \ldots) \subset \mathcal{C}_0$. Let $\tilde{\mathcal{A}}_{cr,1}$ be related to the module $\tilde{A}_{cr,1} = (R/x_0^pT_1, \ldots) \subset \mathcal{C}_0$. Then in notation of Corollary 3.2, $T_{st}(L) = \lim_{\to_n}^{\text{Hom}}(L_1^0(L))$.

4.2. The functor $\mathcal{V}[1]^*$. Note the following case of Proposition 4.4.

Proposition 4.7. Suppose $L = (L, F(L), \varphi, N) \in L_1'[1]$. Then there is an isomorphism of abelian groups $\mathcal{V}^1(L) \cong \text{Hom}_{\mathcal{C}_0}(L, \mathcal{A}_{cr,1})$. In addition, $\Gamma_F$ acts on $\mathcal{V}^1(L)$ via its natural action on $\mathcal{A}_{st,1}$ and the identification $\mathcal{V}^1(L) \cong \text{Hom}_{\mathcal{C}_0}(L, \mathcal{A}_{cr,1})$ such that if $f_0 \in \text{Hom}_{\mathcal{C}_0}(L, \mathcal{A}_{cr,1})$ then for any $l \in L$,

$$
\iota_L(f_0)(l) = \sum_{j \geq 0} f_0(N^j(l))\gamma_j(\log(1 + X))
$$

Introduce the functor $\mathcal{V}[1]^* := \mathcal{V}[1]^{\text{C}_0} \circ \Pi^{-1} : \mathcal{L}^* \to \mathcal{M}_{\Gamma_F}$, where $\Pi : \mathcal{L}'[1] \to \mathcal{L}^*$ is the equivalence of categories from Corollary 3.10.

Proposition 4.8. On the subcategory of unipotent objects $\mathcal{L}^u$ of $\mathcal{L}^*$ the functors $\mathcal{V}[1]^*$ and $\mathcal{V}^*$ coincide.

Proof. The definition of $\mathcal{A}_{cr}$ implies that $\mathcal{A}_{cr,1} = (R/x_0^pT_1, T_2, \ldots) \subset \mathcal{C}_0$, where for all indices $i \geq 1$, $T_i$ comes from $\gamma_i^p([x_0^p] + p)$ and $T_i^p = 0$. Set $F(\mathcal{A}_{cr,1}) = \text{Fil}^{-1} \mathcal{A}_{cr,1} = (x_0^pR/x_0^pR) \oplus (R/x_0^pT_1)$, where the ideal $T_1$ is generated by all $T_i$. Then the corresponding map $\varphi : F(\mathcal{A}_{cr,1}) \to \mathcal{A}_{cr,1}$ is uniquely determined by the conditions $\varphi(x_0^p) = 1 - T_1$, $\varphi(T_1) = 1$ and $\varphi(T_i) = 0$ if $i \geq 2$. In particular, $\varphi(\mathcal{A}_{cr,1}) \subset (R/x_0^pT_1) \oplus (R/x_0^p)$.

Let $\tilde{\mathcal{A}}_{cr,1} = \mathcal{A}_{cr,1}/J_1$ with the induced structure of filtered $\varphi$-module $\tilde{\mathcal{A}}_{cr,1}$, where the ideal $J_1$ of $\mathcal{A}_{cr,1}$ is generated by the elements $T_1x_0^p$ and $T_i$ with $i \geq 2$. Then the projection $\mathcal{A}_{cr,1} \to \tilde{\mathcal{A}}_{cr,1}$ induces for any object $L = (L, F(L), \varphi, N)$ of the category $\mathcal{L}^u$, the identification (use that $\varphi|_J = 0$)

$$
\text{Hom}_{\mathcal{C}_0}(L, \mathcal{A}_{cr,1}) = \text{Hom}_{\mathcal{C}_0}(L, \tilde{\mathcal{A}}_{cr,1}).
$$

Introduce $a_0, a_{-1} \in \text{Hom}(L, R/x_0^p)$ such that for any $m \in L$, $f_0(m) = a_{-1}(m)T_1 + a_0(m)$. Note that $a_0$ and $a_{-1}$ are $\mathcal{W}_1$-linear, where the multiplication by $u$ on $L$ corresponds to the multiplication by $x_0$ in $R/x_0^p$.

Then for any $m \in F(L)$, the requirement $f_0(\varphi(m)) = \varphi(f_0(m))$ is equivalent to the conditions
\begin{align*}
a_0(\varphi(m)) &= a_{-1}(m)p + \frac{a_0(m)p}{x_0^{p(p-1)}} \\
a_{-1}(\varphi(m)) &= -\frac{a_0(m)p}{x_0^{p(p-1)}}
\end{align*}

Note that these conditions depend only on \( m = m \mod u^pL \).

Consider the operator \( V : L \to L \) from Subsection 1.5. Clearly, \( V(u^pL) \subset uF(L) \) and for \( \bar{L} := L/u^pL \), we obtain the induced operator \( V : \bar{L} \to \bar{L} \) (use that \( F(L)/uF(L) \subset L/u^pL \)).

For any \( m \in L \), relations (4.1) can be rewritten as follows:
\begin{align*}
a_0(m) &= \frac{a_0(\bar{V}m)p}{x_0^{p(p-1)}} + a_{-1}(\bar{V}m)p \\
a_{-1}(m) &= -\frac{a_0(\bar{V}m)p}{x_0^{p(p-1)}}
\end{align*}

Therefore, if \( \mathcal{L} \) is unipotent then for any \( m \in \bar{L} \),
\begin{align*}
a_{-1}(m) &= -a_0(m) + a_{-1}(\bar{V}m)p = -a_0(m) + a_{-1}(V^2m)p = \cdots = -a_0(m).
\end{align*}

This implies that for any \( m \in F(\bar{L}) \), \( a_0(\varphi(m)) = a_0(m)p/x_0^{p(p-1)} \). In other words, we have a natural identification
\[
\Hom_{\mathcal{L}}(\mathcal{L}, \tilde{\mathcal{R}}^u) = \Hom_{\mathcal{L}}(\mathcal{L}, \tilde{\mathcal{A}}_{cr,1})
\]

coming from the map of filtered \( \varphi \)-modules \( \tilde{\mathcal{R}}^u \to \tilde{\mathcal{A}}_{cr,1} \) given by the \( R \)-linear map \( R/x_0^{p} \to \tilde{\mathcal{A}}_{cr,1} = (R/x_0^p)T_1 \oplus (R/x_0^p) \) such that for any \( r \in R/x_0^p \),
\[
r \mapsto (-\bar{r}T_1, r).
\]
(For the definition of \( \tilde{\mathcal{R}} \in \mathcal{L}^{u*} \), cf. Subsection 2.2.)

This implies that for all unipotent \( \mathcal{L} \in \mathcal{L}^{u*} \), there is a natural identification of \( \Gamma_F \)-modules \( V[1]^*\mathcal{L} = \mathcal{V}^*\mathcal{L} \). Indeed, the above embedding \( R/x_0^p \to \tilde{\mathcal{A}}_{cr,1} \) can be extended to the embedding of \( R_{st}/x_0^pR_{st} \) to
\[
\tilde{\mathcal{A}}_{st,1} = \prod_{j \geq 0} \tilde{\mathcal{A}}_{cr,1} \gamma_j(\log(1 + X)),
\]

which induces the above identification. \( \square \)

4.3. Splittings \( \Theta \) and \( \tilde{\Theta} \). Suppose \( \mathcal{L} = (L, F(L), \varphi, N) \in \mathcal{L}^{*} \). Then there is a standard short exact sequence
\[
(4.2) \quad 0 \to \mathcal{L}_u \overset{i}{\to} \mathcal{L} \overset{j}{\to} \mathcal{L}^m \to 0,
\]
where \( (\mathcal{L}_u, i) \) is the maximal unipotent subobject and \( (\mathcal{L}^m, j) \) is the maximal multiplicative quotient of \( \mathcal{L} \).

If \( \mathcal{L}^m = (L^m, F(L^m), \varphi, N) \) then \( F(L^m) = L^m \otimes_{\mathbb{F}_p} \mathcal{W}_1 \), where \( L_0 = \{ l \in L^m \mid \varphi(l) = l \} \). Suppose \( S : L^m \to F(L) \subset L \) is a \( \mathcal{W}_1 \)-linear section. Then for any \( l_0 \in L_0 \), \( S(l_0) = \varphi(S(l_0)) + g(l_0) \), where \( g \in \Hom(L_0, L^m) \). If \( S' : L^m \to \mathcal{X} \), then \( S'(l_0) = \varphi(S'(l_0)) + g'(l_0) \), where \( g' \in \Hom(L_0, L^m) \).
\[ F(L) \] is another \( W \)-linear section then for any \( l_0 \in L_0 \), \( S'(l_0) = \varphi(S(l_0)) + g'(l_0) \). Here \( g' \in \text{Hom}(L_0, L^u) \) is such that for some \( h \in \text{Hom}(L_0, L^u) \), one has
\[
(g' - g)(l_0) = h(l_0) - \varphi(h(l_0)).
\]

**Proposition 4.9.** a) There is a section \( S \) such that \( g(L_0) \subset uL^u \).
b) If \( g(L_0), g'(L_0) \subset uL^u \) then \( h(L_0) \subset uF(L^u) \).

**Proof.** a) It will be sufficient to prove that for any \( l \in L^u \), there is an \( h \in F(L^u) \) such that \( l \equiv h - \varphi(h) \mod uL^u \).

Suppose \( n_0 \geq 1 \) is such that \( V^{n_0}(L^u) \subset uF(L^u) \). Then for all \( n \geq n_0 \), \( V^n(L^u) \subset uF(L^u) \). Let \( h = -(Vl + V^2l + \cdots + V^{n_0+l}l) \). By the definition of the operator \( V \) for all \( 1 \leq i \leq n_0 + 1 \), \( V^i l \in F(L^u) \) and \( \varphi(V^i l) \equiv V^{i-1}l \mod uL^u \). Therefore, \( h \in F(L^u) \) and \( \varphi(h) = -(l + Vl + \cdots + V^{n_0+l}l) \equiv -l - h \mod uL^u \).

b) We must prove that if \( h \in F(L^u) \) and \( h - \varphi(h) \in uL^u \) then \( h \in uF(L^u) \).

Indeed, we have \( V(h) - h \in V(uL^u) \subset uF(L^u) \) and for all \( n \geq 1 \), \( V^n(h) \equiv h \mod uF(L^u) \) implies that \( h \in uL^u \). Therefore, \( \varphi(h) \in uL^u \) and \( h \in uF(L^u) \).

\[ \square \]

**Proposition 4.10.** With above notation the short exact sequence
\[ 0 \longrightarrow V[1]^*(L^u) \longrightarrow V[1]^*(L) \longrightarrow V[1]^*(L^c) \longrightarrow 0 \]

obtained from (4.2) by applying \( V[1]^* \), has a canonical functorial splittings \( \Theta : V[1]^*(L^u) \longrightarrow V[1]^*(L) \) and \( \tilde{\Theta} : V[1]^*(L) \longrightarrow V[1]^*(L^c) \) in the category \( \mathcal{M}_F^w \).

**Proof.** It will be sufficient to prove the existence of a functorial splitting
\[ \Theta : \text{Hom}_F(L^u, \tilde{A}_{cr,1}) \longrightarrow \text{Hom}_F(L, \tilde{A}_{cr,1}) \]
of the epimorphism \( \text{Hom}_F(L^u, \tilde{A}_{cr,1}) \rightarrow \text{Hom}_F(L, \tilde{A}_{cr,1}) \), obtained from exact sequence (4.2).

Suppose \( f_0 = (a_{-1}, a_0) : L^u \longrightarrow (R/x_0^p)T_1 \oplus (R/x_0^p) \) belongs to \( \text{Hom}_F(L^u, \tilde{A}_{cr,1}) \). Here \( a_{-1}, a_0 \in \text{Hom}_{\mathbb{W}_L}(L^u, R/x_0^p) \) and for any \( l \in L^u \), \( a_{-1}(l) = -a_0(l) \), cf. Subsection 4.2.

Let \( S : L^m \rightarrow L \) be a \( W \)-linear section such that for any \( l \in L_0 \), \( S(l_0) = \varphi(S(l_0)) + g(l_0) \), where \( g \in \text{Hom}(L_0, uL^u) \).

Extend \( f_0 \) to \( \Theta(f_0) = (a_{-1}, a_0) : L \longrightarrow (R/x_0^p)T_1 \oplus (R/x_0^p) \) by setting \( a_{-1}(S(l_0)) = -a_0(S(l_0)) = X \), where \( X \) is a unique element of \( R/x_0^p \) such that \( X = X^p/x_0^{p(p-1)} = a_0(g(l_0)) \). One can prove that \( \Theta(f_0) \in \text{Hom}_F(L, \tilde{A}_{cr,1}) \) by verifying relations (4.1) with \( m = S(l_0) \).

\[ \square \]

4.4. A modification of Breuil’s functor. Remind that Breuil’s functor \( V^t : L^t \longrightarrow \mathcal{M}_F^w \) attaches to any \( L \in L^t \), the \( \Gamma_F \)-module \( V(L) = \text{Hom}_F(L, \tilde{A}_{cr,\infty}) \).

**Proposition 4.11.** The functor \( V^t \) is fully faithful on the subcategory of unipotent objects \( L^{t,u} \).
Proposition 4.10 implies that $\mathcal{V}^t$ is very far from to be fully faithful on the whole $\mathcal{L}^{ft}$: if $\mathcal{L} \in \mathcal{L}^{ft}[1]$ and $0 \rightarrow \mathcal{L}^u \rightarrow \mathcal{L} \rightarrow \mathcal{L}^m \rightarrow 0$ is the standard exact sequence then the corresponding sequence of $\Gamma_{\mathcal{F}}$-modules admits a functorial splitting.

Introduce a modification $\widetilde{\mathcal{V}}^{ft}: \mathcal{L}^{ft} \rightarrow \text{M}_{\mathcal{F}}$ of Breuil’s functor. Suppose $\mathcal{L} \in \mathcal{L}^{ft}$. From the definition of the category $\mathcal{L}^{ft}$ in Subsection 3 it follows the existence of $\mathcal{L}' \in \mathcal{L}^{ft}$ such that $p\mathcal{L}' = \mathcal{L}$. More precisely, there are a strict monomorphism $i_{\mathcal{L}'}: \mathcal{L} \rightarrow \mathcal{L}'$ and a strict epimorphism $j_{\mathcal{L}'}: \mathcal{L}' \rightarrow \mathcal{L}$ such that $p \circ i_{\mathcal{L}'} = i_{\mathcal{L}'} \circ j_{\mathcal{L}'}$. (Note that $j_{\mathcal{L}'} \circ i_{\mathcal{L}'} = p \circ i_{\mathcal{L}'}$.)

Consider the following short exact sequences

$$0 \rightarrow \mathcal{L} \xrightarrow{i_{\mathcal{L}'} \circ} \mathcal{L}' \xrightarrow{C_p} \mathcal{L}' \rightarrow 0$$

and the corresponding sequence of $\Gamma_{\mathcal{F}}$-modules and their morphisms

$$\mathcal{V}^t(\mathcal{L}^u) \xrightarrow{\Theta} \mathcal{V}^t(\mathcal{L}') \xrightarrow{\mathcal{V}^t(C_p)} \mathcal{V}^t(\mathcal{L}') \xrightarrow{\mathcal{V}^t(K_{p})} \mathcal{V}^t(\mathcal{L}_p^m) \xrightarrow{\Theta} \mathcal{V}^t(\mathcal{L}_p^m)$$

As earlier, for any $\mathcal{L} \in \mathcal{L}^{ft}$, $\mathcal{L}^u$ is the maximal unipotent subobject and $\mathcal{L}^m$ is the maximal multiplicative quotient object for $\mathcal{L}$.

**Lemma 4.12.** $\text{Ker} (\widetilde{\Theta} \circ \mathcal{V}^t(K_p)) \supset \text{Im} (\mathcal{V}^t(C_p) \circ \Theta)$.

**Proof.** The section $\Theta$ depends functorially on objects of the category $\mathcal{L}^{ft}[1] \supset \mathcal{L}^{ft}[1]$. Therefore, we have the following commutative diagram

$$\begin{array}{ccc}
\mathcal{V}^t(\mathcal{L}^u) & \xrightarrow{\mathcal{V}^t(C_p \circ K_p)} & \mathcal{V}^t(\mathcal{L}_p^m) \\
\widetilde{\Theta} \circ \mathcal{V}^t(K_p) \circ \mathcal{V}^t(C_p) & \Theta & \mathcal{V}^t(C_p \circ K_p)
\end{array}$$

and $\widetilde{\Theta} \circ \mathcal{V}^t(K_p) \circ \mathcal{V}^t(C_p) \circ \Theta = (\widetilde{\Theta} \circ \Theta) \circ \mathcal{V}^t(C_p \circ K_p) = 0$. □

**Definition.** Set $\mathcal{V}_{\mathcal{L},\mathcal{L}'}(\mathcal{L}) = \text{Ker} (\widetilde{\Theta} \circ \mathcal{V}^t(K_p))/\text{Im} (\mathcal{V}^t(C_p) \circ \Theta)$.

**Proposition 4.13.** With above notation one has:

a) $\mathcal{V}_{\mathcal{L},\mathcal{L}'}(\mathcal{L}) = \text{Coker} \mathcal{V}^t(C_p) = \mathcal{V}^t(\mathcal{L})$ if $\mathcal{L} \in \mathcal{L}^{u,ft}$;

b) $\mathcal{V}_{\mathcal{L},\mathcal{L}'}(\mathcal{L}) = \text{Ker} \mathcal{V}^t(K_p) = \mathcal{V}^t(\mathcal{L})$ if $\mathcal{L} \in \mathcal{L}^{m,ft}$;

c) for any $\mathcal{L} \in \mathcal{L}^{ft}$, we have the induced exact sequence of $\Gamma_{\mathcal{F}}$-modules

$$0 \rightarrow \mathcal{V}^t(\mathcal{L}_p^m) \rightarrow \mathcal{V}^t(\mathcal{L}_p^m) \rightarrow \mathcal{V}^t(\mathcal{L}_p^u) \rightarrow 0.$$ This sequence depends functorially on the pair $(\mathcal{L}, \mathcal{L}')$. 

---

*Documenta Mathematica 18 (2013) 547–619*
The correspondence $L \rightarrow n$ is sufficient to verify that for maps on $V$ from functoriality and diagram chasing implies that it induces the identity map $p$. This gives a functorial sequence

$$0 \rightarrow V_{\mathcal{L}}^n(L^m) \rightarrow V_{\mathcal{L}}^n(L) \rightarrow V_{\mathcal{L}}^n(L^n) \rightarrow 0.$$ 

Then standard diagram chasing proves that this sequence is exact.

**Proposition 4.14.** Suppose for a given $L \in \mathcal{L}^{ft}$, the objects $L', L'' \in \mathcal{L}^{ft}$ are such that $pL' = pL'' = L$. Then there is a natural isomorphism $f(L', L'')$ of $\Gamma_F$-modules such that the following diagram is commutative

$$
\begin{array}{ccc}
0 & \rightarrow & V^1(L^m) \\
\downarrow{id} & & \downarrow{id} \\
0 & \rightarrow & V^1(L') \\
\end{array}
\begin{array}{ccc}
V^1(L^m) & \rightarrow & V^1(L'') \\
\downarrow{f(L', L'')} & & \downarrow{\text{id}} \\
V^1(L') & \rightarrow & V^1(L'') \\
\end{array}
\begin{array}{ccc}
0 & \rightarrow & V^1(L^m) \\
\end{array}
\begin{array}{ccc}
0 & \rightarrow & V^1(L') \\
\end{array}
\begin{array}{ccc}
0 & \rightarrow & V^1(L'') \\
\end{array}
(\text{The lines of this diagram are given by Prop 4.13)}

**Proof.** By replacing $L''$ by $\prod_{\mathcal{L}} L''$ with respect to strict epimorphisms $j_{\mathcal{L}}$ and $j_{\mathcal{L}', \mathcal{L}''}$, we can assume that there is a map $f : L'' \rightarrow L'$ which induces the identity map $pL'' = L \rightarrow pL' = L$. Then the existence of $f(L', L'')$ follows from functoriality and diagram chasing implies that it induces the identity maps on $V^1(L^n)$ and $V^1(L^m)$.

**Definition.** For $L, L' \in \mathcal{L}^{ft}$ such that $pL' = L$, set $\hat{V}^{ft}(L) = V_{\mathcal{L}}^n(L)$.

The correspondence $L \rightarrow \hat{V}^{ft}(L)$ induces the additive exact functor $\hat{V}^{ft} : \mathcal{L}^{ft} \rightarrow \text{MF}_F$.

4.5. $\varphi$-filtered module $\tilde{A}_{cr, 2} \subseteq \tilde{B}_0$. Let $\xi = [x_0] + p \in W(R) \subset A_{cr}$, and for $n \in \mathbb{N}$, $\gamma_n(\xi) = \xi^n/n!$

**Lemma 4.15.** If $n \geq 2p$ then $\varphi(\gamma_n(\xi)) \in p^2A_{cr}$.

**Proof.** We have $\varphi(\gamma_n(\xi)) = (p^{n-p+1}/n!)([x_0]^p/p + 1)^n$. Therefore, it will be sufficient to verify that for $n \geq 2p$, $v_p(n!) + p + 1 \leq n$. Using the estimate $v_p(n!) < n/(p - 1)$ we obtain that the required inequality holds for $p \geq 5$ if $n \geq p + 3$ and for $p = 3$ if $n \geq 8$. It remains to check that our inequality holds for $p = 3$ and $n \in \{6, 7\}$.

Let $J_2$ be the closed ideal in $A_{cr}$ generated by $[x_0]^p\xi^n/p$ and all $\xi^n/n!$ with $n \geq 2p$. Then $J_2 \subset F(A_{cr})$ and $\varphi(J_2) \subset p^2A_{cr}$. Introduce $\tilde{A}_{cr, 2} = A_{cr}/(J_2 + p^2A_{cr})$ and consider the corresponding induced filtered $\varphi$-module $\tilde{A}_{cr, 2} = (\tilde{A}_{cr, 2}, F(\tilde{A}_{cr, 2}), \varphi) \subseteq \tilde{B}_0$. Clearly, for any $L \in \mathcal{L}$, the natural projection $A_{cr, 2} \rightarrow \tilde{A}_{cr, 2}$ induces the identification $\text{Hom}_{\mathcal{L}}(L, A_{cr, 2}) = \text{Hom}_{\mathcal{L}}(L, \tilde{A}_{cr, 2})$.

Consider the structure of $\tilde{A}_{cr, 2}$ more closely.
Let $T_1 = \xi^p/p$. With obvious notation the elements of $\tilde{A}_{cr,2}$ can be written uniquely modulo the subgroup $[x_0^pR]T_1 + [x_0^pR] + p[x_0^pR] + p^2W(R)$ in the form $[r_1]T_1 + [r_0] + p[r_1]$, where $r_1, r_0, r_1 \in R$. In formally, we shall use that $r_1, r_0 \in R/x_0^p$ and $r_0 \in R/x_0^2$. The $W(R)$-module structure on $\tilde{A}_{cr,2}$ is induced by usual operations on Teichmuller’s representatives and the relation $pT_1 \equiv [x_0]^p \mod p^2W(R)$. (Use that $T_1 \equiv [x_0]^p/p + p[x_0]^{p-1} \mod p^2W(R)$.)

The $S$-module structure on $\tilde{A}_{cr,2}$ is induced by the $W(k)$-algebra morphism $S \rightarrow W(R)$ such that $u \mapsto [x_0]$. Then $F(\tilde{A}_{cr,2})$ is generated over $W(R)$ by the images of $T_1$ and $\xi^{p-1}$. Note that $\xi^{p-1} \equiv [x_0]^{p-1} - p[x_0]^{p-2} \mod p^2W(R)$. The map $\varphi : F(\tilde{A}_{cr,2}) \rightarrow \tilde{A}_{cr,2}$ is uniquely determined by the knowledge of $\varphi(T_1)$ and $\varphi(\xi^{p-1})$. Note that

$$
\varphi(T_1) = \left(\frac{1 + [x_0]^p}{p}\right)^p \equiv 1 + [x_0]^p \mod (J + p^2A_{cr,2} + p[m_R])
$$

$$
\varphi(\xi^{p-1}) = \left(\frac{1 + [x_0]^p}{p}\right)^{p-1} \equiv 1 - T_1 \mod (J + p^2A_{cr,2} + p[m_R])
$$

Suppose $L \in \mathcal{L}^{f^1[1]}$ and $L' \in \mathcal{L}^{f^2}$ is such that $pL' = L$. Consider short exact sequences (4.3) and (4.4). Then the points $f \in \mathcal{V}^t(pL')$ and $\mathcal{V}^t(C_p)(f) \in \mathcal{V}^t(L')$ are related via the commutative diagram

\[
\begin{array}{ccc}
\mathcal{L}' & \xrightarrow{\mathcal{V}^t(C_p)(f)} & \tilde{A}_{cr,2} \\
\downarrow{\mathcal{V}^t(C_p)(f)} & & \downarrow{\mathcal{V}^t(C_p)(f)} \\
pL' & \xrightarrow{f} & \tilde{A}_{cr,1}
\end{array}
\]

where the right vertical arrow is induced by the correspondence $[r_1]T_1 + [r_0] + p[r_1] \mapsto [r_1]T_1 + [r_0] \mod x_0^p]$. Similarly, the points $g \in \mathcal{V}^t(L')$ and $\mathcal{V}^t(K_p)(g) \in \mathcal{V}^t(L'_p)$ are related via the commutative diagram

\[
\begin{array}{ccc}
\mathcal{L}' & \xrightarrow{g} & \tilde{A}_{cr,2} \\
\downarrow{K_p} & & \downarrow{K_p} \\
L'_p & \xrightarrow{\mathcal{V}^t(K_p)(g)} & \tilde{A}_{cr,1}
\end{array}
\]

where the right vertical arrow is induced by the correspondence $[r_1]T_1 + [r_0] \mapsto [r_1]x_0^p] + p[r_0]$. 

*Documenta Mathematica* 18 (2013) 547–619
4.6. Filtered $\varphi$-modules $A_{cr,1}^0$ and $A_{cr,2}^0$. Let $A_{cr,2}^0$ be the $W(R)$-submodule of $\tilde{A}_{cr,2}$ consisting of elements $[r_{-1}]T_1 + [r_0] + p[r_1]$ such that $r_{-1} = -r_0 \mod x_0^p$. Then $F(A_{cr,2}^0) = F(\tilde{A}_{cr,2}) \cap A_{cr,2}^0$ is generated over $W(R)$ by $[x_0^{p-1}]T_1 + x_0^{p-1}$ and the congruence
\[
\varphi([x_0^{p-1}]T_1 + x_0^{p-1}) \equiv -T_1 + 1 \mod (J_2 + p^2 A_{cr,2} + p|m_R])
\]
implies that $\varphi(F(A_{cr,2}^0)) \subset A_{cr,2}^0$ and $A_{cr,2}^0 = (A_{cr,2}^0, F(A_{cr,2}^0), \varphi) \in \tilde{L}_0$.

Note that $pA_{cr,2}^0 = (pA_{cr,2}^0, pF(A_{cr,2}^0), \varphi) \in \tilde{L}_0$. Then in notation from Subsection 4.4, one has:

- $\text{Im } \Theta = \text{Hom}_{\tilde{L}_0}(p\mathcal{L}', pA_{cr,2}^0)$;
- $\text{Ker } \Theta = \text{Hom}_{\tilde{L}_0}((\mathcal{L}', pA_{cr,2}^0)$;
- $\text{Ker } (\Theta \circ \mathcal{V}^t(K_p)) = \text{Hom}_{\tilde{L}_0}((\mathcal{L}', A_{cr,2}^0)$.

Therefore, $\tilde{\mathcal{V}}^t(\mathcal{L}) = \mathcal{V}^t(\mathcal{L}) = \text{Hom}_{\tilde{L}_0}((\mathcal{L}', A_{cr,2}^0/pA_{cr,2}^0)$.

4.7. The functor $\tilde{\mathcal{V}}^t$. Let $\mathcal{L} \in L_{\mathcal{L}}^t$ and let $i^t : \mathcal{L}^t \to \mathcal{L}$ be the maximal etale subobject of $\mathcal{L}$.

DEFINITION. $\tilde{\mathcal{V}}^t : L_{\mathcal{L}}^t \to \text{CMF}_R$ is the functor induced by the correspondence $\mathcal{L} \mapsto \tilde{\mathcal{V}}^t(\mathcal{L}) = (\tilde{\mathcal{V}}^t(\mathcal{L}), \tilde{\mathcal{V}}^t(\mathcal{L}^t), \tilde{\mathcal{V}}^t(i^t))$.

The functor $\tilde{\mathcal{V}}^t$ is not very far from Breuil’s functor $\mathcal{V}^t$ but it satisfies the following important property.

PROPOSITION 4.16. The functor $\tilde{\mathcal{V}}^t$ is fully faithful.

Proof. By standard devissage it will be sufficient to verify this property for the restriction $\tilde{\mathcal{V}}^t|_{\mathcal{L}^t[X]}$. Due to Proposition 2.13 it will be sufficient to verify that the functor $\tilde{\mathcal{V}}^t|_{\mathcal{L}^t[X]} \circ \Pi^{-1}$ coincides with the functor $\mathcal{V}^t$ from Subsection 2.2. This can be proved similarly to the proof of the corresponding fact for unipotent objects in Subsection 4.2 as follows.

Let
\[
A_{st,2}^0 = \prod_{j \geq 0} A_{cr,2}^0 \gamma_j(\log(1 + X)) \subset \tilde{A}_{st,2} = \prod_{j \geq 0} \tilde{A}_{cr,2} \gamma_j(\log(1 + X))
\]
with induced structures of the objects $A_{st,2}^0$ and $\tilde{A}_{st,2}$ of the category $\tilde{L}$. Then from Subsection 4.6 it follows that
\[
\mathcal{V}^t(\mathcal{L}) = \text{Hom}_{\tilde{L}_0}(\mathcal{L}, A_{st,2}^0/p A_{st,2}^0).
\]
One can see easily that the correspondence
\[
[r_0 \mod x_0^p]T_1 + [r_0] + p[r_1] \mapsto (r_0 + x_0^p r_1) \mod x_0^p m_R
\]
induces the morphism $A_{cr,2}^0/pA_{cr,2}^0 \to R^0$ in the category $\tilde{\mathcal{L}}_0$. This morphism induces a unique identification of the abelian groups $V^t(L)$ and $\text{Hom}(L, R^0) = V^t(L)$. Now going to a suitable factor of the object $A_{st,2}^0/pA_{st,2}^0$ we obtain that this identification is compatible with the $\Gamma_F$-actions on both abelian groups. □

Now we can describe all Galois invariant lattices of semi-stable $\mathbb{Q}_p[\Gamma_F]$-modules with weights from $[0, p)$.

**Corollary 4.17.** Suppose $V$ is a semi-stable representation of $\Gamma_F$ with weights from $[0, p)$, $\dim_{\mathbb{Q}_p} V = s$ and $T$ is a $\Gamma_F$-invariant lattice in $V$. Then there is a $p$-divisible group $\{L^{(n)}, i_n\}_{n \geq 0}$ of height $s$ in $\mathcal{L}_f$ such that $\lim_{\leftarrow} \tilde{C}V^{(n)}(L^{(n)}) = (T, T^{et}, i^{et}) \in \text{CM}_{\Gamma_F}$.

5. Proof of Theorem 0.1.

As earlier, $p$ is a fixed prime number, $p \neq 2$. Starting Subsection 5.2 we assume $p = 3$.

5.1. For all prime numbers $l$, choose embeddings of algebraic closures $\overline{\mathbb{Q}} \subset \overline{\mathbb{Q}}_l$ and use them to identify the inertia groups $I_l = \text{Gal}(\overline{\mathbb{Q}}_l/\mathbb{Q}_l_{ur})$, where $\mathbb{Q}_l_{ur}$ is the maximal unramified extension of $\mathbb{Q}_l$, with the appropriate subgroups in $\Gamma_\mathbb{Q} = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

Introduce the category $\mathcal{M}_\Gamma^{\mathbb{Q}}$. Its objects are the pairs $H_\mathbb{Q} = (H, \tilde{H}_{st})$, where $H$ is a finite $\mathbb{Z}_p[\Gamma_\mathbb{Q}]$-module unramified outside $p$ and $\tilde{H}_{st} = (H_{st}, H^0_{st}, i) \in \text{CM}_{\Gamma_F}$, where $H|_{I_p} = H_{st}$, $F = W(\overline{\mathbb{F}}_p)[1/p]$ and $\text{CM}_{\Gamma_F}$ is the image of the functor $\tilde{C}V^{f}$ from Subsection 4.7. The morphisms in $\mathcal{M}_\Gamma^{\mathbb{Q}}$ are compatible morphisms of Galois modules. Clearly, the category $\mathcal{M}_\Gamma^{\mathbb{Q}}$ is special pre-abelian, cf. Appendix A.

Let $\mathcal{M}_\Gamma^{\mathbb{Q}}[1]$ be the full subcategory of killed by $p$ objects in $\mathcal{M}_\Gamma^{\mathbb{Q}}$. Denote by $\mathcal{K}(p)$ an algebraic extension of $\mathbb{Q}$ such that for any $H_\mathbb{Q} = (H, \tilde{H}_{st}) \in \mathcal{M}_\Gamma^{\mathbb{Q}}[1]$, $\Gamma_{\mathcal{K}(p)} = \text{Gal}(\overline{\mathbb{Q}}/\mathcal{K}(p))$ acts trivially on $H$. In other words, $\mathcal{K}(p)$ can be taken as a common field-of-definition of points of all such $\Gamma_\mathbb{Q}$-modules $H$.

Now assume that

(C) $\mathcal{K}(p)$ is totally ramified at $p$.

Under this assumption we have a natural identification $\text{Gal}(\mathcal{K}(p)/\mathbb{Q}) = \text{Gal}(\mathcal{K}(p)/F/F)$, that is the Galois group of the global extension $\mathcal{K}(p)/\mathbb{Q}$ comes as the Galois group of its completion over $F$. Therefore, we can identify $\mathcal{M}_\Gamma^{\mathbb{Q}}[1]$ with the full subcategory of $\text{CM}_{\Gamma_F}$, consisting of $(H_{st}, H^0_{st}, i)$ such that $pH_{st} = 0$ and all points of $H_{st}$ are defined over $\mathcal{K}(p)F$. In other words, the objects of $\mathcal{M}_\Gamma^{\mathbb{Q}}[1]$ can be described via our local results about killed by $p$ subquotients of semistable representations of $\Gamma_F$.
Denote by $\mathbf{M}_\ell^{f \ell}[1]$ a full subcategory in $\mathbf{M}_\ell^{f}[1]$ which consists of killed by $p$ subquotients of $p$-divisible groups in the category $\mathbf{M}_\ell^{f}$. Let $F'$ be the maximal tamely ramified extension of $F$ in $\mathcal{K}(p)F$. Then $\text{Gal}(F'/F)$ is abelian group of order prime to $p$ (use that the residue field of $F'$ is algebraically closed) and $\text{Gal}(\mathcal{K}(p)F/F')$ is a $p$-group. This gives an abelian extension $\mathcal{K}'$ of $\mathbb{Q}$ in $\mathcal{K}(p)$ of prime-to-$p$ degree and such that $\mathcal{K}(p)/\mathcal{K}'$ is a $p$-extension. This extension is unramified outside $p$ and, therefore, it coincides (use class field theory) with $\mathbb{Q}(\zeta_p)$. In particular, all simple objects in $\mathbf{M}_\ell^{f}[1]$ are of the form $\mathcal{F}(j) = (\mathbb{F}_p(j), 0, 0)$ if $1 \leq j < p$ and $\mathcal{F}(0) = (\mathbb{F}_p(0), \mathbb{F}_p(0), \text{id})$ if $j = 0$.

Let $\mathbf{L}_\ell^{f}[1]$ and $\mathbf{L}_\ell'[1]$ be the full subcategories of $\mathbf{L}^{f}[1]$ mapped by the functor $\widetilde{\mathcal{V}}^f$ to the objects of $\mathbf{M}_\ell^{f}[1]$ and, resp., $\mathbf{M}_\ell^{f}[1]$. Clearly, $\mathbf{L}_\ell^{f}[1]$ is a full subcategory in $\mathbf{L}^{f}[1]$ and the only simple objects in these categories are $\mathcal{L}(r)$, where $r \in \{ j/(p - 1) \mid j = 0, 1, \ldots, p - 1 \}$.

Suppose $H^\infty = \{ H^{(n)}_\mathbb{Q}, n \} \_{n \geq 0}$ is a $p$-divisible group in the category $\mathbf{M}_\ell^{f}$. Here all $H^{(n)}_\mathbb{Q} = (H^{(n)}_\mathbb{Q}, \hat{H}^{(n)}_\text{st})$ are objects of the category $\mathbf{M}_\ell^{f}$. Let $\mathcal{L} \in \mathbf{L}_\ell^{f}[1]$ be such that $\widetilde{\mathcal{V}}^f(\mathcal{L}) = \hat{H}^{(1)}_\text{et}$. Note that the maximal etale subobject $\mathcal{L}^\text{et}$ of $\mathcal{L}$ is isomorphic to $\mathcal{L}(0)^{n_{\text{et}}}$, where $n_{\text{et}} = n_{\text{et}}(\mathcal{L}) \in \mathbb{Z}_{\geq 0}$, and $\mathcal{L}/\mathcal{L}^\text{et}$ has no simple subquotients isomorphic to $\mathcal{L}(0)$. Similarly, the corresponding maximal multiplicative quotient $\mathcal{L}^\text{m}$ is isomorphic to $\mathcal{L}(1)^{n_{\text{m}}}$, where $n_{\text{m}} = n_{\text{m}}(\mathcal{L}) \in \mathbb{Z}_{\geq 0}$, and the kernel of the canonical projection $\mathcal{L} \rightarrow \mathcal{L}^\text{m}$ has no simple subquotients isomorphic to $\mathcal{L}(1)$. Therefore, for any $\mathcal{M} \in \mathbf{L}_\ell^{f}[1]$,

$$\text{Ext}_{\mathbf{L}_\ell^{f}[1]}(\mathcal{L}(0), \mathcal{M}) = \text{Ext}_{\mathbf{L}_\ell^{f}[1]}(\mathcal{M}, \mathcal{L}(1)) = 0.$$ 

This implies that for any $H \in \mathbf{M}_\ell^{f}[1]$,

$$\text{Ext}_{\mathbf{M}_\ell^{f}[1]}(H, \mathcal{F}(0)) = \text{Ext}_{\mathbf{M}_\ell^{f}[1]}(\mathcal{F}(1), H) = 0.$$ 

Therefore, by Theorem A.5 of Appendix A there is an embedding of $p$-divisible groups $H^\infty,^m \subset H^\infty$, where $H^{(1)}^m = \mathcal{F}(1)^{n_{\text{et}}}$, and there is a projection of $p$-divisible groups $H^\infty \twoheadrightarrow H^\infty,^c\text{t}$, where $H^{(1)}^c\text{t} = \mathcal{F}(0)^{n_{\text{m}}}$. For similar reasons,

$$\text{Ext}_{\mathbf{M}_\ell^{f}[1]}(\mathcal{F}(0), \mathcal{F}(0)) = \text{Ext}_{\mathbf{M}_\ell^{f}[1]}(\mathcal{F}(1), \mathcal{F}(1)) = 0$$

and by Theorem A.4 of Appendix A, the corresponding $p$-divisible groups $H^\infty,^m_\mathbb{Q}$ and $H^\infty,^c\text{t}_\mathbb{Q}$ are unique. Therefore they coincide with the products of trivial $p$-divisible groups $(\mathbb{Q}_p/\mathbb{Z}_p)(p - 1)$ and, resp., $(\mathbb{Q}_p/\mathbb{Z}_p)(0)$.

We state this result in the following form.

**Proposition 5.1.** Under assumption (C), for any $p$-divisible group $H^\infty$ in the category $\mathbf{M}_\ell^{f}$ there is a filtration of $p$-divisible groups

$$H^\infty \supset H^\infty_1 \supset H^\infty_0$$

---

**Documenta Mathematica** 18 (2013) 547–619
such that $H_0^\infty = (\mathbb{Q}/\mathbb{Z}_p)(p-1)^n$, $H^\infty/H_1^\infty = (\mathbb{Q}/\mathbb{Z}_p)(0)^n$, and all simple subquotients of $H_1^\infty/H_0^\infty$ belong to $\{F_p(j)\}_{1 \leq j \leq p-2}$.

5.2. Assume that $p = 3$.

**Lemma 5.2.** $\mathcal{K}(3) = \mathbb{Q}(\sqrt[3]{3}, \zeta_3)$, where $\zeta_3$ is the 9-th primitive root of 1.

This Lemma will be proved in Subsection 5.3 below.

In particular, $\mathcal{K}(3)$ satisfies the assumption (C).

**Proposition 5.3.** If $H^\infty$ is a 3-divisible group in $\text{MF}^3_\mathcal{Q}$ then in its filtration from Proposition 5.1 the 3-divisible group $H^\infty = H_1^\infty/H_0^\infty$ is a product of finitely many trivial 3-divisible groups $(\mathbb{Q}_3/\mathbb{Z}_3)(1)$.

**Proof.** Let $\hat{\mathcal{L}}_3$ be the full subcategory of $\mathcal{L}_3[1]$ consisting of objects $\mathcal{L}$ such that $L^3 = L^e = 0$. This category has only one simple object $\mathcal{L}(1/2)$. Let $\hat{\text{MF}}_3$ be the full subcategory in $\text{MF}^3_\mathcal{Q}[1]$ consisting of the objects $\hat{\text{CV}}^\dagger(\mathcal{L})$, where $\mathcal{L} \in \hat{\mathcal{L}}_3$. Then $\hat{\mathcal{L}}_3$ and $\hat{\text{MF}}_3$ are antiequivalent categories and $\hat{H}^{(1)}(1) \in \hat{\text{MF}}_3$. By Theorems A.4 and A.5 our Proposition is implied by the following result.

**Proposition 5.4.** $\text{Ext}_{\mathcal{L}_3}(\mathcal{L}(1/2), \mathcal{L}(1/2)) = 0$.

**Proof.** Consider the equivalence of the categories $\Pi : \mathcal{L}^\dagger \rightarrow \mathcal{L}^*$ from Corollary 3.10. This equivalence transforms the functor $\hat{\text{CV}}^\dagger$ into the functor $\text{CV}^*$ from Section 2, cf. the proof of Proposition 4.16. Therefore, the objects $\mathcal{L}$ of the category $\Pi(\hat{\mathcal{L}}_3) := \mathcal{L}_3^*$ are characterised by the condition that all points of $V^*(\mathcal{L})$ are defined over the field $K(3)$. The objects $\mathcal{L}$ of the category $\Pi(\mathcal{L}_3)(1)$ are characterised by the additional properties: they are all obtained by subsequent extensions via $\mathcal{L}(1/2)$ and $V^*(\mathcal{L})$ appears as a subquotient of $\Gamma_\mathcal{L}$ with Hodge-Tate weights from $[0,2]$.

Introduce the object $\mathcal{L}(1/2, 1/2) = (L, F(L), \varphi, N)$ of the category $\mathcal{L}_3^*$ as follows:

- $L = W_1l + W_1l$;
- $F(L)$ is spanned by $u_1$ and $u_1 + l$;
- $\varphi(u_1) = l$, $\varphi(u_1 + l) = l$;
- $N(l_1) \equiv 0 \mod u^3L$, $N(l_1) \equiv l_1 \mod u^3L$.

Clearly, $\mathcal{L}(1/2, 1/2)$ has a natural structure of an element of the group $\text{Ext}_{\mathcal{L}_3}(\mathcal{L}(1/2), \mathcal{L}(1/2))$.

**Lemma 5.5.** a) $\mathcal{L}(1/2, 1/2) \in \mathcal{L}_3^*$; b) $\text{Ext}_{\mathcal{L}_3}(\mathcal{L}(1/2), \mathcal{L}(1/2)) \cong \mathbb{Z}/3$ and is generated by the class of $\mathcal{L}(1/2, 1/2)$; c) $\text{Ext}_{\mathcal{L}_3}(\mathcal{L}(1/2), \mathcal{L}(1/2, 1/2)) = \text{Ext}_{\mathcal{L}_3}(\mathcal{L}(1/2, 1/2), \mathcal{L}(1/2)) = 0$.

This Lemma will be proved in Subsection 5.4 below.

Lemma 5.5 implies that $\text{Ext}_{\mathcal{L}_3}(\mathcal{L}(1/2, 1/2), \mathcal{L}(1/2, 1/2)) = 0$ and, therefore, any object $\mathcal{L}$ of $\mathcal{L}_3^*$ is the product of several copies of $\mathcal{L}(1/2)$ and $\mathcal{L}(1/2, 1/2)$.
Suppose $\mathcal{L} = \mathcal{L}_1 \times L(1/2,1/2) \in \hat{\mathcal{L}}_Q^*$. Then there is a 3-divisible group $\tilde{H}^\infty$ in $\check{\text{M}^{\Gamma}_2}$ such that $\tilde{H}^{(1)} = H' \times H(1/2,1/2)$, where $H'$ and $H(1/2,1/2) = CV^*(L(1/2,1/2))$ belong to $\check{\text{M}^{\Gamma}_2}$. Clearly, we have $\text{Ext}_{\check{\text{M}^{\Gamma}_2}(1)}(H',H(1/2,1/2)) = 0$ and applying Theorem A.5 we obtain a 3-divisible group $H^\infty$ in $\check{\text{M}^{\Gamma}_2}$ such that $H^{(1)} = H(1/2,1/2)$. This implies the existence of 2-dimensional semi-stable (and non-crystalline) representation of $\Gamma_F$ with the only simple subquotient $\mathbb{F}_3(1)$, that is for any Galois invariant lattice $T$ of such representation, the $\Gamma_F$-module $T/3T$ has semi-simple envelope $\mathbb{F}_3(1) \times \mathbb{F}_3(1)$. This situation appears as a very special case of Breuil’s description of 2-dimensional semi-stable (and non-crystalline) representations. According to Theorem 6.1.1.2 of [5] the corresponding semi-simple envelope is either $\mathbb{F}_3(0) \times \mathbb{F}_3(1)$ or $\mathbb{F}_3(1) \times \mathbb{F}_3(2)$. The proposition is proved. □

Now our main Theorem appears as the following Corollary.

**Corollary 5.6.** If $Y$ is a projective variety with semi-stable reduction modulo 3 and good reduction modulo all primes $l \neq 3$ then $h^2(Y_\mathbb{C}) = h^{1,1}(Y_\mathbb{C})$.

**Proof.** Indeed, let $V$ be the $\mathbb{Q}_3[\Gamma_F]$-module of 2-dimensional etale cohomology of $Y$. Then it is a semi-stable representation of $F$ and its $\Gamma_F$-invariant lattice determines a 3-divisible group in the category $\check{\text{M}^{\Gamma}_2}$. By Proposition 5.3 this 3-divisible group can be built from the Tate twists $(\mathbb{Q}_3/\mathbb{Z}_3)(i)$, $i = 0,1,2$. Equivalently, all $\Gamma_F$-equivariant subquotients of $V$ are $\mathbb{Q}_3(i)$ with $i = 0,1,2$. Applying the Riemann Conjecture (proved by Deligne) to the reductions $Y \text{ mod } l$ with $l \neq 3$, we obtain that $\mathbb{Q}(0)$ and $\mathbb{Q}(2)$ do not appear. Therefore, $V$ is the product of finitely many $\mathbb{Q}_3(1)$ and $h^2(Y_\mathbb{C}) = h^{1,1}(Y_\mathbb{C})$. □

**5.3. Proof of Lemma 5.2.** Use the ramification estimate from Subsection 2.9 to deduce that the normalized discriminant of $\mathcal{K}(3)$ over $\mathbb{Q}$ satisfies the inequality $|D(\mathcal{K}(3)/\mathbb{Q})|^{[\mathcal{K}(3):\mathbb{Q}]^{-1} < 3^{3-1/3} = 18.72075}$. Then Odlyzko estimates imply that $|\mathcal{K}(3) : \mathbb{Q}| < 230 [11]$.

Let $K_0 = \mathbb{Q}(\zeta_3)$ and $K_1 = \mathbb{Q}(\sqrt{3}, \zeta_3)$. Then $K_0$ is the maximal abelian extension of $\mathbb{Q}$ in $\mathcal{K}(3)$ and $K_1 \subset \mathcal{K}(3)$. We have also the inequality $[\mathcal{K}(3) : K_1] < 60$ and, therefore, $\text{Gal}(\mathcal{K}(3)/\mathbb{Q})$ is soluble. Prove that $K_1 = \mathcal{K}(3)$.

Suppose the field $K_2$ is the maximal abelian extension of $K_1$ in $\mathcal{K}(3)$. One can apply the computer package SAGE to prove that the group of classes of $K_1$ is trivial. Therefore, $K_2$ is totally ramified at 3 and $\text{Gal}(K_2/\mathbb{Q})$ coincides with the Galois group of the corresponding 3-completions. In particular, the maximal tamely ramified subextension of these completions comes from $\mathbb{Q}(\zeta_3)$ and, therefore, $K_2/K_1$ is 3-extension. Therefore, there is an $\eta \in \mathcal{O}_{K_1}^*$ such that $K_1(\sqrt[3]{\eta}) \subset K_2$. Then a routine computation shows that the normalized discriminant for $K_1(\sqrt[3]{\eta})$ over $\mathbb{Q}$ is less than $3^{3-1/3}$ if and only if $\eta \equiv 1 \text{ mod } \mathcal{O}_{K_1}^*(1 + 3 \mathcal{O}_{K_1})$. The Lemma will be proved if we show that such $\eta \in \mathcal{O}_{K_1}^*$. (This is equivalent to the Leopoldt Conjecture for the field $K_1$.)
was proved via a SAGE computer program written by R.Henderson (Summer-2009 Project at Durham University supported by Nuffield Foundation). This program, cf. Appendix B, constructed a basis $\varepsilon_i \mod O_K^{x^i}$, $1 \leq i \leq 9$, of $O_K^{x^9}/O_K^{x^3}$ such that $18v_3(\varepsilon_i - 1)$ takes values in the set $\{1, 2, 4, 5, 7, 8, 10, 13, 16\}$. In other words, $v_3(\eta - 1) \geq 1 > 16/18$ implies that $\eta \in O_K^{x^3}$.

Lemma 5.2 is proved.

5.4. Proof of Lemma 5.5. a) Use the notation from the definition of the functor $\mathcal{V}^f$ in Subsection 4. If $f_0 \in \mathcal{V}^f(\mathcal{L}(1/2, 1/2))$ then the correspondence $f_0 \mapsto (f_0(l_1), f_0(l_2))$ identifies $\mathcal{V}^f(\mathcal{L}(1/2, 1/2))$ with the $\mathbb{F}_2$-module of couples $(X_{10}, X_0) \in (R/x_0)^2$ such that $X_{10}^3/x_0^3 = X_{10}$ and $(X_0^3 + X_{10})/x_0^3 = X_0$. Then the $\mathbb{F}_2[\Gamma_F]$-module $\mathcal{V}^f(\mathcal{L}(1/2, 1/2))$ is identified with the module formed by the images of all $(X_{10}, X_0 + X_{10}Y) \in (R/\mathfrak{m})^2$ in $R/\mathfrak{m} = R_0^x/(x_0^3m_R + x_0^3m_RY + x_0m_RY^2)$. In particular, the corresponding $\Gamma_F$-action on $\mathcal{V}^f(\mathcal{L}(1/2, 1/2))$ comes from the natural $\Gamma_F$-equivariant identification

$$\iota : m_R/(x_0^3m_R) \to \mathfrak{m}/3\mathfrak{m},$$

where $\mathfrak{m}$ is the maximal ideal of the valuation ring of $\mathbb{Q}_3$. This isomorphism $\iota$ comes from the map $r \mapsto r^{(1)}$, where for $r = \lim_{n \to \infty} (r_n \mod p)$, $r^{(1)} := \lim_{n \to \infty} r_n^{3n+1}$.

Then Hensel’s Lemma implies the existence of unique $Z_{10}, Z_0 \in \mathfrak{m}$ such that the following equalities hold $\iota(X_{10} \mod x_0^3m_R) = Z_{10} \mod 3\mathfrak{m}$, $\iota(X_0 \mod x_0^3m_R) = Z_0 \mod 3\mathfrak{m}$, $Z_{10} + 3Z_0 = 0$ and $Z_0^3 + 3Z_0 = -Z_{10}$.

Clearly, $F(Z_{10}, Z_0) = F(\zeta_3)$. Therefore, if $\tau \in \Gamma_F$ is such that $\tau(\zeta_3) = \zeta_3$ then $\tau(X_{10}) = X_0$ and $\tau(X_0) = X_{10}$. Finally, it follows directly from definitions that if $\tau(\sqrt{3}) = \sqrt{3}$ then $\tau$ acts as identity on the image of $Y$ in $R_0^x$. The part a) of the Lemma is proved.

b) Suppose $\mathcal{L} = (L, F(L), \varphi, N) \in \text{Ext}_{\mathcal{L}_3}(\mathcal{L}(1), \mathcal{L}(1))$. Then $L = W_1l \oplus W_1l_1$, there is an $w \in W_1$ such that $F(L)$ is spanned by $ul_1$ and $ul_1 + w_1l_1$ over $W_1$, and one has $\varphi(w_1l_1) = l_1, \varphi(ul + w_1l_1) = l, N(l_1) \in w^3L$ and $N(l_1) \equiv w^3l_1 \mod w^3L$. Notice that $\mathcal{L}$ splits in $\mathcal{L}_3$ if $w \in uW_1$. Therefore, we can assume that $w = \alpha \in k$.

Then the field-definition of all points of $\mathcal{V}^f(\mathcal{L})$ contains the field-definition of all solutions $(X_1, X) \mod x_0^3m_R \in (R/x_0^3m_R)^2$ of the following congruences: $X_1^3/x_0^3 \equiv X_1 \mod x_0^3m_R$ and $(X^3 + \alpha^3X_1)/x_0^3 \equiv X \mod x_0^3m_R$. Let $x_1 \in R$ be such that $x_1^3 = x_0$. Then we can take $X_1 = x_1^3$ and for $T = X/x_1^3$ one has the following Artin-Schreier-type congruence:

$$T^3 - T \equiv -\alpha^3/x^6 \mod m_R.$$

Using calculations from above part a) we can conclude that $\mathcal{L} \in \mathcal{L}_3^0$ if and only if the field-definition of $T \mod m_R$ over $k((x_1))$ belongs to the field-definition of $T_0 \mod m_R$ over $k((x_1))$, where $T_3^3 - T_0 \equiv -x_1^{-6} \mod m_R$. By Artin-Schreier theory this happens if and only if $\alpha \in \mathbb{F}_3$ and, therefore, $\mathcal{L} \simeq \mathcal{L}(1/2, 1/2)$. 

---

**Documenta Mathematica 18 (2013) 547–619**
c) Suppose \( L = (L, F(L), \varphi, N) \in \text{Ext}_{\mathcal{L}}(L(1/2), L(1/2, 1/2)). \)

Then we can assume that:

- \( L = W_1 l \oplus W_1 l_1 \oplus W_1 m; \)
- \( F(L) \) is spanned over \( W_1 \) by \( ul, ul_1 \) and \( um + wl + w_1 l_1 \) with \( w, w_1 \in W_1; \)
- \( \varphi(ul) = l, \varphi(ul_1) = l \) and \( \varphi(um + wl + w_1 l_1) = m. \)

Then the condition \( u^2 m \in F(L) \) implies that \( w_1l \in F(L), \) or \( w \in uW_1 \) and we can assume that \( w = 0. \) Then the submodule \( W_1 m + W_1 l_1 \) determines a subobject \( L' \) of \( L, L' \in \mathcal{L} \) and using calculations from b) we conclude that \( w_1 \in \mathbb{F}_3 \) mod \( uW_1. \) Therefore, we can assume that \( w_1 = \alpha \in \mathbb{F}_3 \) and for \( m' = m - \alpha l \) we have \( m' \in F(L) \) and \( \varphi(m') = m', \) i.e. \( L \) is a trivial extension. Now suppose \( L = (L, F(L), \varphi, N) \in \text{Ext}_{\mathcal{L}}(L(1/2, 1/2), L(1/2)). \)

Then we can assume that:

- \( L = W_1 m \oplus W_1 m_1 \oplus W_1 l; \)
- \( F(L) \) is spanned over \( W_1 \) by \( ul, um_1 + wt \) and \( um + m_1 + w_1 l \) with \( w, w_1 \in W_1; \)
- \( \varphi(ul) = l, \varphi(um_1 + wt) = m_1 \) and \( \varphi(um + m_1 + w_1 l) = m. \)

Again, the condition \( u^2 m \in F(L) \) implies that \( w \in uW_1 \) and, therefore, we can assume that \( w = 0. \) Then the quotient module \( L/W_1 m_1 \) is the quotient of \( L \) in the category \( \mathcal{L}. \) This quotient must belong to the subcategory \( \mathcal{L}/. \) This implies that \( w_1 \in \mathbb{F}_3 \) mod \( uW_1, \) and, as earlier, \( L \) becomes a trivial extension. The Lemma is completely proved.

Appendix A. \( p \)-divisible groups in pre-abelian categories

A.1. Short exact sequences in pre-abelian categories.

A.1.1. Pre-abelian categories. Introduce the concept of special pre-abelian category following mainly [28], cf. also [25, 26, 29]. Remind that a category \( S \) is pre-abelian if it is additive and for any morphism \( u \in \text{Hom}_S(A, B), \) there exist \( \text{Ker} u = (A_1, i) \) and \( \text{Coker} u = (B_1, j), \) where \( i \in \text{Hom}_S(A_1, A) \) and \( j \in \text{Hom}_S(B_1, B). \) For any objects \( A, B \in S, \) let \( A \coprod B \) and \( A \coprod B \) be their product and coproduct, respectively. There is a canonical isomorphism \( A \coprod B \simeq A \coprod B \) in \( S. \) More generally, for given morphisms:

- \( \alpha \in \text{Hom}_S(C, A), \beta \in \text{Hom}_S(C, B), \) there is a fibered coproduct \( (A \coprod_C B, i_A, i_B), \) with \( i_A \in \text{Hom}_S(A, A \coprod_C B), i_B \in \text{Hom}_S(B, A \coprod_C B) \) which completes the diagram \( A \xleftarrow{\alpha} C \xrightarrow{\beta} B \) to a cocartesian square;

- \( f \in \text{Hom}_S(A, C) \) and \( g \in \text{Hom}_S(B, C), \) there is a fibered product \( (A \coprod_C B, p_A, p_B), \) with \( p_A \in \text{Hom}_S(A, A \coprod_C B), p_B \in \text{Hom}_S(A, A \coprod_C B), \) which completes the diagram \( A \xrightarrow{f} C \xleftarrow{g} B \) to a cartesian square.

Suppose \( i \in \text{Hom}_S(A_1, A), f \in \text{Hom}_S(A_1, B) \) and \( (B \coprod_A A, i_A, i_B) \) is their fibered coproduct. If \( (A_2, j) = \text{Coker} i \) then there is a morphism \( j_B : \)

\[ \text{Documenta Mathematica 18 (2013) 547–619} \]
Varieties with Bad Reduction at 3 Only

$B \coprod_{A_1} A \to A_2$ such that the following diagram

\[
\begin{array}{ccccccccc}
A_1 & \xrightarrow{i} & A & \xrightarrow{j} & A_2 \\
| & & \downarrow f & & \downarrow i_A & & \downarrow id \\
B & \xrightarrow{i_B} & B \coprod_{A_1} A & \xrightarrow{j_B} & A_2 \\
\end{array}
\]

is commutative (use the zero morphism from $B$ to $A_2$). A formal verification shows that $(A_2, j_B) = \text{Coker} i_B$.

Suppose $j \in \text{Hom}_S(A, A_2), g \in \text{Hom}_S(B, A_2)$ and $(B \coprod_{A_2} A, p_B, p_A)$ is their fibered product. If $(A_1, i) = \text{Ker} j$ then there is an $i_B : A_1 \to B \coprod_{A_1} A$ (use the zero map from $A_1$ to $B$) such that the following diagram

\[
\begin{array}{ccccccccc}
A_1 & \xrightarrow{i} & A & \xrightarrow{j} & A_2 \\
| & & \downarrow id & & \downarrow p_A & & \downarrow p_B & & \downarrow g \\
A_1 & \xrightarrow{i_B} & B \coprod_{A_1} A & \xrightarrow{j_B} & A_2 & \xrightarrow{p_B} & B \\
\end{array}
\]

is commutative and $(A_1, i_B) = \text{Ker} p_B$.

A.1.2. **Strict morphisms.** A morphism $u \in \text{Hom}_S(A, B)$ is strict if the canonical morphism $\text{Coim} u = \text{Coker}(\text{Ker} u) \to \text{Im} u = \text{Ker}(\text{Coker} u)$ is isomorphism. One can verify that always $\text{Ker} u = (A_1, i)$ is a strict monomorphism and $\text{Coker} u = (B_1, j)$ is a strict epimorphism. By definition, a sequence of objects and morphisms

\[(A.1) \quad 0 \to A_1 \xrightarrow{i} A \xrightarrow{j} A_2 \to 0\]

in $S$ is short exact if $(A_1, i) = \text{Ker} j$ and $(A_2, j) = \text{Coker} i$. In particular, any strict monomorphism (resp. strict epimorphism) can be included in a short exact sequence.

**Definition.** A pre-abelian category is special if it satisfies the following two axioms:

- **SP1.** if $\alpha : C \to A$ is strict monomorphism then $i_B : B \to A \coprod_C B$ is also strict monomorphism;
- **SP2.** if $f : A \to C$ is strict epimorphism then $p_B : A \coprod_C B \to B$ is also strict epimorphism.

**Remark.** A typical example of pre-abelian special category is the category of modules with filtration.
Consider short exact sequence (A.1) in $S$. If $f \in \text{Hom}_S(A_1, B)$ then we have the following commutative diagram

\[
\begin{array}{ccccccccc}
0 & \longrightarrow & A_1 & \longrightarrow & A & \longrightarrow & A_2 & \longrightarrow & 0 \\
& & \downarrow f & & \downarrow i & & \downarrow j & & \\
0 & \longrightarrow & B & \longrightarrow & A \prod_i A_i B & \longrightarrow & A_2 & \longrightarrow & 0 \\
\end{array}
\]

Then $j_B = \text{Coker}i_B$ is strict epimorphism and by axiom SP1, $i_B$ is strict monomorphism. Then $\text{Ker}j_B = \text{Ker}(\text{Coker}i_B) = \text{Im}i_B = (B, i_B)$ and, therefore, the lower row of the above diagram is exact.

Dually, for any $g \in \text{Hom}_S(B, A_2)$ there is a commutative diagram

\[
\begin{array}{ccccccccc}
0 & \longrightarrow & A_1 & \longrightarrow & A & \longrightarrow & A_2 & \longrightarrow & 0 \\
& & \uparrow id & & \uparrow p_A & & \uparrow g & & \\
0 & \longrightarrow & A_1 & \longrightarrow & A \prod_i A_i B & \longrightarrow & B & \longrightarrow & 0 \\
\end{array}
\]

where $i_B = \text{Ker}j_B$ is strict monomorphism, by Axiom SP2, $p_B$ is strict epimorphism and the lower row of this diagram is exact.

With relation to above diagram (A.2) we have the following properties.

\textbf{Lemma A.1.} a) The natural map $\delta : \text{Ker}f \longrightarrow \text{Ker}i_A$ is isomorphism;

b) if $f$ is strict epimorphism then $\text{Ker}i_A$ is also strict epimorphism.

\textbf{Proof.} a) Suppose that $\text{Ker}f = (K_1, \alpha_1)$ and $\text{Ker}i_A = (K, \alpha)$. Then $\delta : \text{Ker}f \longrightarrow \text{Ker}i_A$ appears from the universal property of $(K, \alpha)$ because $i_A \circ i \circ \alpha_1 = i_B \circ f \circ \alpha_1 = 0$. The relation $j \circ \alpha = j_B \circ i_A \circ \alpha = 0$ implies the existence of $\tilde{\alpha} : K \longrightarrow A_1$ such that $i \circ \tilde{\alpha} = \alpha$. Then $i_B \circ f \circ \tilde{\alpha} = i_A \circ \alpha = 0$ and $f \circ \tilde{\alpha} = 0$ (use that $i_B$ is monomorphism). By the universal property of $(K_1, \alpha_1)$ this gives the map $\delta_1 : K \longrightarrow K_1$ such that $\alpha_1 \circ \delta_1 = \tilde{\alpha}$ and this map is inverse to $\delta$.

b) Suppose $f$ is a strict epimorphism, then $(B, f) = \text{Coker} \alpha_1$. Let $(\tilde{C}, \tilde{j}) = \text{Coker} \alpha$. By functoriality, there is $\varepsilon : B \longrightarrow \tilde{C}$ such that $\varepsilon \circ f = \tilde{j} \circ i$. Then $\tilde{j}$ and $\varepsilon$ define a unique $\omega : A \prod_i A_i B \longrightarrow \tilde{C}$ such that $\omega \circ i_B = \varepsilon$ and $\omega \circ i_A = \tilde{j}$. But $i_A \circ \alpha = 0$ implies by the universal property of $(\tilde{C}, \tilde{j})$ the map $\omega_1 : \tilde{C} \longrightarrow A \prod_i A_i B$ and one can verify that it is inverse to $\omega$. \hfill \square

\textbf{Remark.} If $f$ is strict monomorphism then $i_A$ is also strict monomorphism by axiom SP2.

With relation to diagram (A.3) we have the following Lemma which is dual to above Lemma A.1.

\textbf{Lemma A.2.} a) The natural map $\text{Coker}p_A \longrightarrow \text{Coker}g$ is isomorphism; b) if $g$ is strict epimorphism then $p_A$ is also strict epimorphism.

\textbf{Proof.} The proof is dual to the proof of Lemma A.1. \hfill \square
Lemma A.3. A composition of two strict monomorphisms (resp., epimorphisms) is again strict monomorphism (resp., epimorphism).

Proof. It will be sufficient to consider only the case of monomorphisms. Suppose $i \in \text{Hom}_S(A_1, A)$ and $i_1 \in \text{Hom}_S(A, B)$ are two strict monomorphisms. Construct the following commutative diagram:

\[
\begin{array}{c}
0 & \rightarrow & A_1 & \xrightarrow{i} & A & \xrightarrow{j} & A_2 & \rightarrow & 0 \\
& & \downarrow{\beta} & & \downarrow{i_1} & & \downarrow{i_{A_2}} & & \\
0 & \rightarrow & K & \xrightarrow{\alpha} & B & \xrightarrow{i_B} & A_2 \coprod_A B & \rightarrow & 0 \\
& & \downarrow{j_1} & & \downarrow{j} & & \downarrow{\gamma} & & \\
& & C_1 & \xrightarrow{\gamma} & C & & & & \\
\end{array}
\]

Here $j = \text{Coker} i$ is strict epimorphism and the upper right square is co-cartesian. Therefore, $i_B$ is strict epimorphism and we obtain the second line which is short exact. The morphisms $i_1$ and $i_{A_2}$ are strict monomorphisms and we can complete our diagram by $(C_1, j_1) = \text{Coker} i_1$ and $(C, j) = \text{Coker} i_{A_2}$. The maps $\beta$ and $\gamma$ are obtained by functoriality and we have proved that they are isomorphisms. Therefore, $i_1 \circ i = \beta \circ \alpha$ is strict (use that $\alpha = \text{Ker} i_B$ is strict). \qed

A.1.3. Bifunctor $\text{Ext}_S$. If $S$ is pre-abelian category then in the following commutative diagram with exact rows

\[
\begin{array}{c}
0 & \rightarrow & A_1 & \xrightarrow{id} & A & \xrightarrow{f} & A_2 & \rightarrow & 0 \\
0 & \rightarrow & A_1 & \xrightarrow{id} & A' & \xrightarrow{id} & A_2 & \rightarrow & 0 \\
\end{array}
\]

the morphism $f$ is isomorphism. Therefore, we can introduce the set of equivalence classes of short exact sequences $\text{Ext}_S(A_2, A_1)$. This set is functorial in both arguments due to axioms SP1 and SP2.

Suppose the objects of $S$ are provided with commutative group structure respected by morphisms of $S$. Then for any $A, B \in S$, $\text{Ext}_S(A, B)$ has a natural group structure, where the class of split short exact sequences plays a role of neutral element. Remind that the sum $\varepsilon_1 + \varepsilon_2$ of two extensions $\varepsilon_1 : 0 \rightarrow A_1 \xrightarrow{\iota'} A' \xrightarrow{\iota''} A_2 \rightarrow 0$ and $\varepsilon_2 : 0 \rightarrow A_1 \xrightarrow{\iota'''} A'' \xrightarrow{\iota'''} A_2 \rightarrow 0$ is the lower line of the following commutative diagram relating the rows
l = ε₁ ⊕ ε₂, ∇⁺(l) and (+), ∇⁺(l),

\[ l : 0 \rightarrow A_1 \prod_{A_1} A_2 \rightarrow 0 \]
\[ \nabla⁺(l) : 0 \rightarrow A_1 \prod_{A_1} A_2 \rightarrow 0 \]
\[ (+), \nabla⁺(l) : 0 \rightarrow A_1 \rightarrow A_2 \rightarrow 0 \]

Here ∇ is the diagonal morphism, + is the morphism of the group structure on S. For any \( f \in \text{Hom}_S(A_1, B) \) and \( g \in \text{Hom}_S(B, A_2) \) the corresponding morphisms \( f_* : \text{Ext}_S(A_2, A_1) \rightarrow \text{Ext}_S(A_2, B) \) and \( g^* : \text{Ext}_S(A_2, A_1) \rightarrow \text{Ext}_S(B, A_1) \) are homomorphisms of abelian groups. The proof is completely formal and goes along the lines of [27].

Suppose \( \varepsilon \in \text{Ext}_S(A_2, A_1) \), then the extension \( \varepsilon + (-id)^* \varepsilon \) splits. We shall need below the following explicit description of this splitting.

Let \( \varepsilon : 0 \rightarrow A_1 \rightarrow A \rightarrow A_2 \rightarrow 0 \). Then \( \varepsilon + (-id)^* \varepsilon \) is the lower row in the following diagram

\[ 0 \rightarrow A_1 \prod_{A_1} A_2 \rightarrow A_2 \rightarrow 0 \]
\[ 0 \rightarrow A_1 \rightarrow A_0 \rightarrow A_2 \rightarrow 0 \]

where the left vertical arrow is the cokernel of the diagonal embedding \( \nabla : A_1 \rightarrow A_1 \prod_{A_1} A_1. \) One can see that the epimorphic map \( A_0 \rightarrow A_1 \), which splits the lower exact sequence, is induced by the morphism \( p_1 - p_2 : A \prod_{A_2} A \rightarrow A. \)

Finally, one can apply Serre’s arguments [30] to obtain for any short exact sequence \( 0 \rightarrow A_1 \rightarrow A \rightarrow A_2 \rightarrow 0 \) and any \( B \in S \), the following standard 6-terms exact sequences of abelian groups

\[ 0 \rightarrow \text{Hom}_S(B, A_1) \xrightarrow{i_*} \text{Hom}_S(B, A) \xrightarrow{j_*} \text{Hom}_S(B, A_2) \]
\[ \xrightarrow{\delta} \text{Ext}_S(B, A_1) \xrightarrow{i_*} \text{Ext}_S(B, A) \xrightarrow{j_*} \text{Ext}_S(B, A_2) \]
\[ 0 \rightarrow \text{Hom}_S(A_2, B) \xrightarrow{j^*} \text{Hom}_S(A, B) \xrightarrow{i^*} \text{Hom}_S(A_1, B) \]
\[ \xrightarrow{\delta} \text{Ext}_S(A_2, B) \xrightarrow{j^*} \text{Ext}_S(A, B) \xrightarrow{i^*} \text{Ext}_S(A_1, B) \]

A.2. \( p \)-DIVISIBLE GROUPS. In this section \( S \) is a special pre-abelian category consisting of group objects. Denote by \( S_1 \) the full subcategory of objects killed by \( p \) in \( S \), where \( p \) is a fixed prime number. Clearly, \( S_1 \) is again special pre-abelian.
A.2.1. Basic definitions. Consider an inductive system \((C^{(n)}, i^{(n)})_{n \geq 0}\) of objects of \(S\), where \(C^{(0)} = 0\) and all \(i^{(n)} : C^{(n)} \to C^{(n+1)}\) are strict monomorphisms. Let for all \(n \geq m \geq 0\), \(i_{mn} = i^{(n-1)} \circ \cdots \circ i^{(m+1)} \circ i^{(m)} \in \text{Hom}_S(C^{(m)}, C^{(n)})\). Then all \(i_{mn}\) are strict monomorphisms. Follow Tate’s paper [Ta] to define a \(p\)-divisible group in \(S\) as an inductive system \((C^{(n)}, i^{(n)})_{n \geq 0}\) such that for all \(0 \leq m \leq n\),

a) \(\text{Coker } i_{mn} = (C^{(n-m)}, j_{n,n-m})\), i.e. there are short exact sequences:

\[
0 \to C^{(m)} \xrightarrow{i_{mn}} C^{(n)} \xrightarrow{j_{n,m-n}} C^{(n-m)} \to 0
\]

b) there are commutative diagrams

\[
\begin{array}{ccc}
C^{(n)} & \xrightarrow{p^m \text{id}_{C^{(n)}}} & C^{(n)} \\
\downarrow j_{n,n-m} & & \downarrow \text{id} \\
C^{(n-m)} & & C^{(n-m)}
\end{array}
\]

The above definition implies the existence of the following commutative diagrams with exact rows (where \(m \leq n \leq n_1\)):

\[
\begin{array}{cccc}
0 & \to & C^{(m)} & \xrightarrow{i_{mn}} C^{(n)} & \xrightarrow{j_{n,m-n}} C^{(n-m)} & \to & 0 \\
0 & \to & C^{(m)} & \xrightarrow{i_{m1}} C^{(n_1)} & \xrightarrow{j_{n_1,m-n_1}} C^{(n_1-m)} & \to & 0 \\
0 & \to & C^{(n)} & \xrightarrow{i_{n1}} C^{(n_1)} & \xrightarrow{j_{n_1,m-n_1}} C^{(n_1-m)} & \to & 0 \\
0 & \to & C^{(m)} & \xrightarrow{i_{m,n+m-n_1}} C^{(m+n_1-n)} & \xrightarrow{j_{m+n_1-n,m-n_1}} C^{(m+n_1-n,m-n_1)} & \to & 0 \\
\end{array}
\]

Also, for all \(n \geq m \geq 0\), one has

• \((C^{(m)}, i_{mn}) = \text{Ker}(p^m \text{id}_{C^{(m)}}), (C^{(m)}, j_{mn}) = \text{Coker}(p^m \text{id}_{C^{(m)}})\);

• \(i_{mn} = i_{n-1,n} \circ \cdots \circ i_{m,m+1} \circ j_{m,n} = j_{m+1,n} \circ \cdots \circ j_{n,1}\).

The set of \(p\)-divisible groups in \(S\) has a natural structure of category. This category is pre-abelian. In particular,

\[
0 \to (C^{(n)}, i^{(n)})_{n \geq 0} \xrightarrow{(\gamma_n)_{n \geq 0}} (C^{(n)}, i^{(n)})_{n \geq 0} \xrightarrow{(\delta_n)_{n \geq 0}} (C^{(n)}, i^{(n)})_{n \geq 0} \to 0
\]

is a short exact sequence of \(p\)-divisible groups iff for all \(n \geq 1\), there are the following commutative diagrams with short exact rows in \(S\)

\[
\begin{array}{cccc}
0 & \to & C^{(n)} & \xrightarrow{\gamma_n} C^{(n)} & \xrightarrow{\delta_n} C^{(n)} & \xrightarrow{\gamma_{n+1}} C^{(n+1)} & \to & 0 \\
0 & \to & C^{(n+1)} & \xrightarrow{\gamma_{n+1}} C^{(n+1)} & \xrightarrow{\delta_{n+1}} C^{(n+1)} & \to & 0 \\
\end{array}
\]
A.2.2. A property of uniqueness of $p$-divisible groups.

**Theorem A.4.** Let $D$ be an object of $S_1$ such that $\text{Ext}_{S_1}(D,D) = 0$. If $(C^{(n)}, i^{(n)})_{n \geq 0}$ and $(C^{(n)}, i^{(n)})_{n \geq 0}$ are $p$-divisible groups in $S$ such that $C^{(1)} \simeq C^{(1)}_1 \simeq D$ then these $p$-divisible groups are isomorphic.

**Proof.** We must prove that for all $n \geq 1$, there are isomorphisms $f_n : C^{(n)} \to C^{(n)}_1$ such that $i^{(n)}_1 \circ f_n = f_{n+1} \circ i^{(n)}$. Suppose $n_0 \geq 1$ and all such isomorphisms have been constructed for $1 \leq n \leq n_0$. Therefore, we can assume that $C^{(n)} = C^{(1)}_1$ for $1 \leq n \leq n_0$. Consider the following commutative diagrams with exact rows:

(A.5) \[
\begin{array}{ccccccccc}
0 & \xrightarrow{i_1} & C^{(n_0+1)} & \xrightarrow{j_1} & C^{(n_0)} & \xrightarrow{j^{(n_0-1)}} & 0 \\
\quad \downarrow{\text{id}} & & \quad \downarrow{i^{(n_0)}} & & \quad \downarrow{j^{(n_0-1)}} & & \\
0 & \xrightarrow{i} & C^{(n)} & \xrightarrow{j} & C^{(n-1)} & \xrightarrow{j^{(n-1)}} & 0 \\
\end{array}
\]

(A.6) \[
\begin{array}{ccccccccc}
0 & \xrightarrow{i'_1} & C^{(n_0+1)}_1 & \xrightarrow{j'_1} & C^{(n_0)} & \xrightarrow{j'_{(n_0-1)}} & 0 \\
\quad \downarrow{\text{id}} & & \quad \downarrow{i^{(n_0)}} & & \quad \downarrow{j^{(n_0-1)}} & & \\
0 & \xrightarrow{i} & C^{(n)} & \xrightarrow{j} & C^{(n-1)} & \xrightarrow{j^{(n-1)}} & 0 \\
\end{array}
\]

Here in standard notation of Subsection A.2.1, $i_1 = i_{1,n_0+1}$, $i'_1 = i'_{1,n_0+1}$, $i = i_{1,n_0}$, $j = j_{n_0,n_0-1}$, $j_1 = j_{n_0+1,n_0}$ and $j'_1 = j'_{n_0+1,n_0}$ (the dash means that the corresponding morphism is related to the second $p$-divisible group). We must construct an isomorphism $f_{n_0+1} : C^{(n_0+1)} \to C^{(n_0+1)}_1$ such that $f_{n_0+1} \circ i^{(n_0)} = i^{(n_0)}_1$. Consider the following commutative diagram obtained from above two diagrams

(A.7) \[
\begin{array}{ccccccccc}
0 & \xrightarrow{i} & C^{(1)} & \xrightarrow{j} & C^{(n)} & \xrightarrow{j^{(n-1)}} & 0 \\
\downarrow{\text{id}} & & \quad \downarrow{i^{(n_0)}} & & \quad \downarrow{j^{(n_0-1)}} & & \\
0 & \xrightarrow{i} & C^{(n_0)} & \xrightarrow{j} & C^{(n_0-1)} & \xrightarrow{j^{(n_0-1)}} & 0 \\
\end{array}
\]

Notice that the morphisms of multiplication by $p$ in $C^{(n_0+1)}$ and $C^{(n_0+1)}_1$ can be factored as follows

\[
\begin{array}{ccccccccc}
C^{(n_0+1)} & \xrightarrow{p} & C^{(n_0+1)} & \xrightarrow{j_1} & C^{(n_0)} & \xrightarrow{i^{(n_0)}} & 0 \\
\downarrow{j_1} & & \quad \downarrow{i^{(n_0)}} & & \quad \downarrow{j^{(n_0)}} & & \\
C^{(n_0)} & \xrightarrow{p} & C^{(n_0+1)}_1 & \xrightarrow{j'_1} & C^{(n_0)} & \xrightarrow{i^{(n_0)}} & 0 \\
\end{array}
\]
Therefore, we obtain the following commutative diagram

\[
\begin{array}{c}
\begin{array}{c}
\xymatrix{
C_{(n_0)} \times \prod C_{(n_0)} & C_{(n_0+1)} \\
\downarrow & \downarrow \\
C_{(n_0)} \times \prod C_{(n_0)} & C_{(n_0+1)} \\
\end{array}
\end{array}
\]

\[
\begin{array}{c}
\begin{array}{c}
\xymatrix{
& \ar[r]^{p} & \\
\downarrow & & \\
& \ar[r]^{i^{(n_0)}_{1} \times i^{(n_0)}} & \\
\downarrow & & \\
\n& \ar[r]^{\nabla} & \\
\downarrow & \downarrow & \downarrow \\
C_{(n_0)} \times \prod C_{(n_0)} & C_{(n_0+1)} & C^{(n_0+1)} \\
\end{array}
\end{array}
\]

(here $\nabla$ is the diagonal morphism). Let $\alpha : C_{(1)} \prod C_{(1)} \to C_{(1)}$ be the cokernel of the diagonal morphism $\nabla : C_{(1)} \to C_{(1)} \prod C_{(1)}$. Clearly, $\nabla$ and $\alpha$ are, resp., strict monomorphism and strict epimorphism. Set $(D_{n_0+1}, \alpha_1) = \operatorname{Coker}((i_1 \prod i'_1) \circ \nabla)$ and $(D_{n_0}, \alpha_0) = \operatorname{Coker}((i \prod i) \circ \nabla)$. Applying $\alpha_\ast$ to diagram (A.7) obtain the two lower rows of the following diagram

\[
\begin{array}{c}
\begin{array}{c}
\xymatrix{
0 & C^{(1)} & D_0 & C^{(1)} & 0 \\
\downarrow & \downarrow \downarrow & \downarrow s & \downarrow f_{n_0+1} & \downarrow \downarrow \\
0 & C^{(1)} & D_{n_0+1} & C^{(n_0)} & 0 \\
\downarrow & \downarrow \downarrow & \downarrow u & \downarrow i_{n_0-1,n_0} & \downarrow \downarrow \\
0 & C^{(1)} & D_{n_0} & C^{(n_0-1)} & 0 \\
\end{array}
\end{array}
\]

Note that the middle line of this diagram equals $\varepsilon_{n_0+1} - \varepsilon'_{n_0+1} \in \operatorname{Ext}(C_{(n_0)}, C^{(1)})$, and at the third row we have a trivial extension. This implies the existence of the first row of our diagram. As it was pointed out earlier, a splitting of the third line can be done via the morphism $f$ from the commutative diagram

\[
\begin{array}{c}
\begin{array}{c}
\xymatrix{
C_{(n_0)} \times \prod C_{(n_0-1)} & C^{(n_0)} \\
\downarrow & \downarrow \\
C^{(1)} & D_{n_0} \\
\end{array}
\end{array}
\]

(Notice that the morphism $s : D_{n_0+1} \to D_0$ is the cokernel of the composition $\ker f \to D_{n_0} \to D_{n_0+1}$.)

Above diagram (A.8) means that the morphism of multiplication by $p$ on $C_{(n_0+1)} \prod C_{(n_0)}$ factors through the diagonal embedding of $C_{(n_0)}$ into $C^{(n_0)} \prod C_{(n_0-1)}$. From diagram (A.10) it follows that $p \id_{D_{n_0+1}}$ factors through the embedding $\ker f \to D_{n_0} \to D_{n_0+1}$. Therefore, $pD_0 = 0$ i.e. the first line in diagram (A.9) is an element of the trivial group $\operatorname{Ext}_S(C^{(1)}, C^{(1)}) = 0$. So, the second row in (A.9) is a trivial extension, i.e. the extensions $\varepsilon_{n_0+1}$ and $\varepsilon'_{n_0+1}$ from diagrams (A.5) and (A.6) are equivalent. This implies the existence of isomorphism $f_{n_0+1}$.

\[\square\]

A.2.3. Splitting of extensions of $p$-divisible groups.
Theorem A.5. Suppose \((C^{(n)}, i^{(n)})_{n \geq 0}\) is a \(p\)-divisible group in the category \(\mathcal{S}\) and there are \(D_1, D_2 \in \mathcal{S}_1\) such that \(C^{(1)} \in \text{Ext}_{\mathcal{S}_1}(D_2, D_1)\) and \(\text{Ext}_{\mathcal{S}_1}(D_1, D_2) = 0\). Then there is an exact sequence of \(p\)-divisible groups

\[
0 \rightarrow (C^{(n)}, i^{(n)})_{n \geq 0} \rightarrow (C^{(n)}, i^{(n)})_{n \geq 0} \rightarrow (C^{(2)}, i^{(2)})_{n \geq 0} \rightarrow 0
\]

in \(\mathcal{S}\) such that \(C^{(1)} = D_1\) and \(C^{(1)}(2) = D_2\).

Proof. We have the exact sequence \(0 \rightarrow D_1 \xrightarrow{i} C^{(1)} \xrightarrow{j} D_2 \rightarrow 0\). We must show for all \(n \geq 1\), the existence of objects \(C^{(n)}\), strict monomorphisms \(\gamma_n : C^{(n)}(n) \rightarrow C^{(n)}\) and \(i^{(n)} : C^{(n)} \rightarrow C^{(n+1)}\) such that \((C^{(n)}, i^{(n)})_{n \geq 0}\) is a \(p\)-divisible group, the system \((\gamma_n)_{n \geq 0}\) defines an embedding of this \(p\)-divisible group into the original \(p\)-divisible group \((C^{(n)}, i^{(n)})_{n \geq 0}\, C^{(1)} = D_1\) and \(\gamma_1 = i\). Agree to use for all \(0 \leq m \leq n\), the notation \(i_{mn}\) and \(j_{nm}\) from Subsection A.2.1 for the original \(p\)-divisible group and set \(C^{(n)} = C_{n0}\).

Illustrate the idea of proof by considering the case \(n = 2\). Set \(C_{11} = D_1\) and consider the following commutative diagram with exact rows \(\varepsilon_2\) and \(\varepsilon^{(1)}_2 = i^*\varepsilon_2\):

\[
\begin{array}{c}
\varepsilon_2 : 0 \rightarrow C_{10} \xrightarrow{i^{(2)}} C_{20} \xrightarrow{j^{(2)}} C_{10} = 0 \\
\varepsilon^{(1)}_2 : 0 \rightarrow C_{10} \xrightarrow{i^{(1)}} C_{21} \xrightarrow{j^{(1)}} C_{11} = 0 \\
\end{array}
\]

Then \(\gamma^{(1)}_2 \circ p \circ \text{id}_{C^{21}} = p \circ \text{id}_{C^{20}} \circ \gamma^{(1)}_2 = i_{12} \circ j_{21} \circ \gamma^{(1)}_2 = i_{12} \circ i_{21} \circ j_{21}\). By Lemma A.2, \(\gamma^{(1)}_2\) is (strict) monomorphism. Therefore, \(p \circ \text{id}_{C^{21}} = i_{21} \circ i_{21}\).

Then the morphism \(j : \text{Ext}_{\mathcal{S}}(C_{11}, C_{10}) \rightarrow \text{Ext}_{\mathcal{S}}(C_{11}, D_2)\) induces the following commutative diagram

\[
\begin{array}{c}
0 \rightarrow C_{10} \xrightarrow{i^{(2)}} C_{21} \xrightarrow{j^{(2)}} C_{11} = 0 \\
0 \rightarrow D_2 \xrightarrow{\alpha} D_{21} \xrightarrow{\text{id}} C_{11} = 0 \\
\end{array}
\]

Here \(p \circ \text{id}_{D^{21}} \circ f = f \circ p \circ \text{id}_{C^{21}} = f \circ i_{12} \circ j_{21} = \alpha \circ i_{21} = j_{21} = 0\). By Lemma A.1, \(f\) is (strictly) epimorphic. Therefore, \(p \circ \text{id}_{D^{21}} = 0\), i.e. \(D_{21} \in \text{Ext}_{\mathcal{S}_1}(C_{11}, D_2) = 0\). Then the exact sequence \(\text{Hom}_{\mathcal{S}} - \text{Ext}_{\mathcal{S}}\) implies the commutative diagram

\[
\begin{array}{c}
0 \rightarrow C_{10} \xrightarrow{i^{(2)}} C_{21} \xrightarrow{j^{(2)}} C_{11} = 0 \\
0 \rightarrow C_{11} \xrightarrow{i^{(2)}} C_{22} \xrightarrow{j^{(2)}} C_{11} = 0 \\
\end{array}
\]

Verify that one can set \(C^{(2)}_{1} = C_{22}\) and \(i^{(2)}_{1} = i^{(2)}_{12}\). Indeed,

\[
\gamma^{(2)}_2 \circ p \circ \text{id}_{C^{22}} = p \circ \text{id}_{C^{21}} \circ \gamma^{(2)}_2 = (i^{(1)}_{12} \circ i) \circ (j^{(1)}_{21} \circ \gamma^{(2)}_2) = \gamma^{(2)}_2 \circ i^{(2)}_{12} \circ j^{(2)}_{21}
\]
and because $\gamma_2^{(2)}$ is monomorphism (use axiom SP1), $p\text{id}_{C_{22}} = j_{22}^{(2)} \circ j_{21}$. Thus, we constructed a segment of length 2 of the $p$-divisible group $(C^{(n)}, j^{(n)})_{n \geq 0}$. Consider the general case.

**Lemma A.6.** In the category $S$ there are the following commutative diagrams with exact lines:

- For $k \geq 1$,

\[
\begin{array}{cccccccc}
E_k^1: & 0 & \to & C_{k-1,0} & \xrightarrow{i_{k-1,k}^{(0)}} & C_{k0} & \xrightarrow{j_{k1}^{(0)}} & C_{10} & \to & 0 \\
& & \downarrow{\text{id}} & & \downarrow{\gamma_k^{(1)}} & & \downarrow{i} & & \\
& 0 & \to & C_{k-1,0} & \xrightarrow{i_{k-1,k}^{(1)}} & C_{k1} & \xrightarrow{j_{k1}^{(1)}} & C_{11} & \to & 0 \\
\end{array}
\]

- For $2 \leq t \leq k$,

\[
\begin{array}{cccccccc}
E_t^1: & 0 & \to & C_{k-1,t-2} & \xrightarrow{i_{k-1,k-2}^{(t-1)}} & C_{k,t-1} & \xrightarrow{j_{k1}^{(t-1)}} & C_{11} & \to & 0 \\
& & \downarrow{\gamma_{k-1}^{(t-1)}} & & \downarrow{\gamma_k^{(t)}} & & \downarrow{\text{id}} & & \\
& 0 & \to & C_{k-1,t-1} & \xrightarrow{i_{k-1,k-1}^{(t)}} & C_{kt} & \xrightarrow{j_{k1}^{(t)}} & C_{11} & \to & 0 \\
\end{array}
\]

- For $1 \leq t < k$,

\[
\begin{array}{cccccccc}
\Delta_k^1: & 0 & \to & C_{k-1,t-1} & \xrightarrow{i_{k-1,k-1}^{(t)}} & C_{kt} & \xrightarrow{j_{k1}^{(t)}} & C_{11} & \to & 0 \\
& & \downarrow{j_{k-1,k-2}^{(t-1)}} & & \downarrow{j_{k1}^{(t)}} & & \downarrow{\text{id}} & & \\
& 0 & \to & C_{k-2,t-1} & \xrightarrow{i_{k-2,k-1}^{(t)}} & C_{k-1,t} & \xrightarrow{j_{k1}^{(t)}} & C_{11} & \to & 0 \\
\end{array}
\]

- For $1 \leq t < k$,

\[
\begin{array}{cccccccc}
\Omega_k^1: & 0 & \to & C_{kt} & \xrightarrow{\gamma_k^{(t)}} & C_{k,t-1} & \xrightarrow{f_{k}^{(t)}} & D_2 & \to & 0 \\
& & \downarrow{j_{k1}^{(0)}} & & \downarrow{j_{k1}^{(t-1)}} & & \downarrow{\text{id}} & & \\
& 0 & \to & C_{k-1,t} & \xrightarrow{\gamma_k^{(0)}} & C_{k-1,t-1} & \xrightarrow{f_{k-1}^{(0)}} & D_2 & \to & 0 \\
\end{array}
\]

where for all indices $k$, $C_{k0} = C^{(k)}$, $i_{k,k+1}^{(0)} = i^{(k)}$, $j_{k+1,k}^{(0)} = j_{k+1,1}^{(0)} = j_{k+1,k} = j_{k+1,k}$. 

**Proof.** Construct the diagram $E_k^1$ by setting $\gamma_1^{(1)} = i$, $j_{11}^{(0)} = \text{id}_{C_{10}}$, $j_{11}^{(1)} = \text{id}_{C_{11}}$. Then for any $k \geq 2$, the upper row of $E_k^1$ is the short exact sequence $\varepsilon_k \in \text{Ext}_S(C_{10}, C_{k-1,0})$ from the original $p$-divisible group $(C_{k0}, j^{(k)})_{k \geq 0}$. Then $E_k^1$ is just a standard diagram relating $\varepsilon_k$ and $i^*\varepsilon_k$. For any $k \geq 2$, we have $(j_{k,k-1} \circ \varepsilon_k = \varepsilon_{k-1}$, therefore, $(j_{k,k-1} \circ i^*\varepsilon_k = i^*\varepsilon_{k-1}$ and we obtain $\Delta_k^1$. The upper row of $\Omega_k^1$ is obtained from the middle column of $E_k^1$ because Coker $\gamma_k^{(1)} \simeq \text{Coker } i = (D_2, j)$. Similarly, the lower row of $\Omega_k^1$ is obtained from $E_{k-1}^1$. The
left square of \( \Omega^1_k \) is commutative by the definition of \( j^{(1)}_{k,k-1} \). The right square is commutative because \( \Omega^1_k \) relates diagrams \( E^1_k \) and \( E^1_{k-1} \).

Suppose now we are given integers \( k_0 \geq 2 \) and \( t_0 < k_0 \) such that the required diagrams \( E^1_k \), \( \Delta^1_k \) and \( \Omega^1_k \) have been already constructed for all \( k < k_0 \) with all relevant \( t \) and for \( k = k_0 \) with \( 1 \leq t \leq t_0 \). Clearly, all \( i^{(t)}_{k,k-1} \) and \( \gamma^{(t)}_k \) are strict monomorphisms and all \( j^{(t)}_{k,k-1} \), \( j^{(t)}_k \) and \( f^{(t)}_k \) are strict epimorphisms.

Constructing \( E^{t_0+1}_{k_0} \). Apply \((f^{(t_0)}_{k_0-1})_*\) to \( E^{t_0}_{k_0} \):

Then \( \text{Ker}(C_{k_0,t_0} \to D^*) = (C_{k_0-1,t_0}, j^{(t_0)}_{k_0-1,k_0} \circ \gamma^{(t_0)}_{k_0-1}) \). Consider the strict monomorphism \( \gamma_{k_0,t_0} := \gamma^{(1)}_0 \circ \ldots \circ \gamma^{(t_0)}_{k_0} : C_{k_0,t_0} \to C_{k_0,0} \) and its analogue \( \gamma_{k_0-1,t_0-1} : C_{k_0-1,t_0-1} \to C_{k_0-1,0} \). Because \( t_0 \neq 0 \), the diagrams \( \Omega^0_{k_0} \) and \( E^1_{k_0} \) give the commutative diagram

\[ \begin{array}{ccc}
C_{k_0,t_0} & \gamma^{(t_0)}_{k_0} & C_{k_0,0} \\
j^{(t_0)}_{k_0,k_0-1} & & \\
C_{k_0-1,t_0} & \gamma^{(t_0)}_{k_0-1} & C_{k_0-1,t_0-1} & \gamma^{(t_0)}_{k_0-1,t_0-1} & C_{k_0-1,0} \\
j^{(t_0)}_{k_0-1,k_0} & & \\
C_{k_0,t_0} & \gamma^{(t_0)}_{k_0} & C_{k_0,0} \\
\end{array} \]

Then \( p\text{id}C_{k_0,0} = i^{(0)}_{k_0-1,k_0} \circ j^{(0)}_{k_0,k_0-1} \) implies \( p\text{id}C_{k_0,t_0} = (i^{(t_0)}_{k_0-1,k_0} \circ \gamma^{(t_0)}_{k_0-1}) \circ j^{(t_0)}_{k_0,k_0-1} \), i.e. \( p\text{id}C_{k_0,t_0} \) factors through \( \text{Ker}(C_{k_0,t_0} \to D^*) \) and \( p\text{id}D^* = 0 \). Then \( \text{Ext}_S(C_{11}, D_2) = 0 \) implies \((f^{(t_0)}_{k_0-1}, \varepsilon_{k_0}) = 0 \), and 6-terms \( \text{Hom}_S - \text{Ext}_S \) exact sequence gives \( E^{t_0+1}_{k_0} \):

\[ \begin{array}{ccc}
0 & \rightarrow & C_{k_0-1,t_0-1} & \rightarrow & C_{k_0,t_0} & \rightarrow & C_{11} & \rightarrow & 0 \\
\gamma^{(t_0)}_{k_0-1} & & i^{(t_0)}_{k_0-1,k_0} & & j^{(t_0)}_{k_0,k_0-1} & & \text{id} & & \\
0 & \rightarrow & C_{k_0-1,t_0} & \rightarrow & C_{k_0,t_0+1} & \rightarrow & C_{11} & \rightarrow & 0 \\
\end{array} \]

We shall denote the rows of this diagram by \( \varepsilon^{(t_0)}_{k_0} \) and \( \varepsilon^{(t_0+1)}_{k_0} \).
Constructing $\Delta_{t_0 \to 1}^{t_0+1}$. Assume that $t_0 + 1 < k_0$. The above extension $\varepsilon_{k_0}^{(t_0+1)}$ is not uniquely defined by $\varepsilon_{k_0}^{(t_0)}$. Show that its choice can be done in such a way that the diagram $\Delta_{t_0 \to 1}^{t_0+1}$ commutes. Consider the short exact sequences from $\Omega_{k_0-1}^{t_0}$. They give rise to the following exact sequences of abelian groups, where $H := \text{Hom}(C_{t_0}, D_{t_0})$ and $E := \text{Ext}(C_{t_0}, D_{t_0})$

\[
\begin{array}{cccc}
H & \xrightarrow{\gamma_{k_0-1}^{(t_0)}} & \text{Ext}(C_{t_0}, C_{k_0-1},D_{t_0}) & \xrightarrow{id} & E \\
\downarrow & & \downarrow & & \downarrow \\
H & \xrightarrow{\gamma_{k_0-2}^{(t_0)}} & \text{Ext}(C_{t_0-1}, C_{k_0-2},D_{t_0-1}) & \xrightarrow{id} & E \\
\end{array}
\]

(A.12)

As we saw earlier, the commutativity of $E_{k_0}^{t_0+1}$ is equivalent to the following relation

\[(\gamma_{k_0-1}^{(t_0)})_* \varepsilon_{k_0}^{(t_0+1)} = \varepsilon_{k_0}^{(t_0)} \tag{A.13}\]

From $\Delta_{t_0}^{k_0}$ it follows that $\varepsilon_{k_0-1}^{(t_0)} = (\gamma_{k_0-1}^{(t_0-1)})_* \varepsilon_{k_0}^{(t_0)}$, and from $E_{k_0-1}^{t_0+1}$ it follows that $(\gamma_{k_0-2}^{(t_0)})_* \varepsilon_{k_0-1}^{(t_0+1)} = \varepsilon_{k_0-1}^{(t_0)}$. Then (A.12) implies that $\varepsilon_{k_0}^{(t_0+1)}$ from (A.13) can be chosen in such a way that $(\gamma_{k_0-1, k_0-2}^{(t_0)})_* \varepsilon_{k_0}^{(t_0+1)} = \varepsilon_{k_0-1}^{(t_0+1)}$. This gives the diagram $\Delta_{t_0}^{t_0+1}$.

Constructing $\Omega_{k_0}^{t_0+1}$. The above arguments imply that the left squares of diagrams $E_{k_0}^{t_0+1}$ and $E_{k_0-1}^{t_0}$ are related via the following commutative diagram

\[
\text{Documenta Mathematica 18 (2013) 547–619}
From diagrams $\Omega_{k0}^{t0}$, $E_{k0}^{t0+1}$ and $E_{k0}^{t0}$ it follows that the induced map $\text{Coker}\gamma_{k0}^{(t0+1)} : \text{Coker}\gamma_{k0}^{(t0)} \simeq D_2$ is isomorphism. This is equivalent to the existence of diagram $\Omega_{k0}^{t0+1}$. The lemma is proved. \hfill \Box

For any $k \geq 1$, set $C_{kk} = C_{1}^{(k)}, i_{k-1:k}^{(k)} = i_{1}^{(k)}$. Then use diagrams $E_{k}^{t}$ to define the inductive system $(C_{1}^{(k)}, i_{1}^{(k)})_{k \geq 0}$. Denote by $\gamma_{k}$ the strict monomorphism $\gamma_{k}^{(1)} \circ \ldots \circ \gamma_{k}^{(k)} : C_{1}^{(k)} \rightarrow C^{(k)}$. From diagrams $E_{k}^{t}$, $1 \leq t \leq k$, obtain the following commutative diagrams:

$$(A.14) \quad 0 \rightarrow C_{1}^{(k-1)} \overset{i_{k-1}^{(k)}}{\rightarrow} C_{1}^{(k)} \overset{\gamma_{k}}{\rightarrow} C_{1}^{(k)} \overset{\gamma_{k}}{\rightarrow} C_{1}^{(1)} \rightarrow 0$$

It remains only to prove that the inductive system $(C_{1}^{(n)}, i_{1}^{(n)})_{n \geq 0}$ is a $p$-divisible group in $S$. From diagrams $E_{k}^{t}$ and $\Delta_{k}^{k-1}$ obtain the following commutative
diagrams with exact rows

\begin{align*}
\text{(A.15)} & \quad 0 \rightarrow C_{k-1,k-1} \xrightarrow{j_{k-1,k}} C_{kk} \xrightarrow{j_{k1}} C_{11} \rightarrow 0 \\
& \quad 0 \rightarrow C_{k-2,k-2} \xrightarrow{j_{k-2,k-2}\gamma_{k-1}} C_{k-1,k-1} \xrightarrow{j_{k-1,k-1}\gamma_k} C_{11} \rightarrow 0 \\
& \quad \varepsilon_{k-2,k-2} \xrightarrow{j_{k-1,k-2}\gamma_{k-1}} C_{k-1,k-1} \xrightarrow{j_{k-1,k-1}\gamma_k} C_{11} \rightarrow 0
\end{align*}

If \( k = 3 \) then the left vertical morphism of this diagram is equal to \( j_{21}^{(1)} \circ \gamma_2^{(2)} = j_{21}^{(2)} \) and is a strict monomorphism. By induction all morphisms \( j_{k,k-1} := j_{k,k-1} \circ \gamma_k \) are strict epimorphisms and are included in the following commutative diagrams

\begin{align*}
\text{(A.16)} & \quad C^{(k)} \xrightarrow{j_{k,k-1}} C^{(k-1)} \\
& \quad C^{(k)} \xrightarrow{j_{k,k-1}} C^{(k-1)} \\
& \quad C^{(k-1)} \xrightarrow{j_{k,k-1}} C^{(k-1)}
\end{align*}

For \( 0 \leq m \leq n \), set \( \varepsilon_{nm} = \varepsilon_{m+1,m} \circ \cdots \circ \varepsilon_{n,n-1} \) and \( \varepsilon_{nm}' = \varepsilon_{n-1,n} \circ \cdots \circ \varepsilon_{m+1,m+1} \).

Composing diagrams (A.15) obtain the following commutative diagram with exact rows

\begin{align*}
0 & \rightarrow C_1^{(n-1)} \xrightarrow{\varepsilon_{n-1,n}} C_1^{(n)} \xrightarrow{j_{n1}} C_1^{(1)} \rightarrow 0 \\
& \quad 0 \rightarrow C_1^{(m-1)} \xrightarrow{\varepsilon_{m-1,m}} C_1^{(m)} \xrightarrow{j_{m1}} C_1^{(1)} \rightarrow 0 \\
& \quad 0 \rightarrow C_1^{(1)} \xrightarrow{id} C_1^{(1)} \rightarrow 0
\end{align*}

Thus, \( \varepsilon_{n-1,n} \) induces the isomorphism \( \text{Ker} \ j_{n-1,n-1} \cong \text{Ker} \ j_{nm} \). Therefore, \( \text{Ker} \ j_{nm} = (C_1^{(n-m)}, \varepsilon_{n-m,n}) \) if we prove that

\begin{align*}
\text{(A.17)} & \quad \text{Ker} \ j_{k1} = (C_1^{k-1}, \varepsilon_{k-1,k}).
\end{align*}

As we noticed earlier, \( j_{k1} = j_{21} \circ \cdots \circ j_{k,k-1} \). Therefore, diagrams (A.16) imply that \( j_{k1} \circ \gamma_k = \gamma_1 \circ j_{k1} \). Now diagram (A.14) implies that \( \gamma_1 \circ j_{k1} = \gamma_1 \circ j_{k1} \), and, therefore, \( j_{k1} = j_{k1}' \) because \( \gamma_1 \) is monomorphism. Hence equality (A.17) follows from diagram (A.14) and \((C_1^{(n)}, \varepsilon_{1}^{(n)})_{n \geq 0}\) satisfies the part a) of the definition of \( p \)-divisible groups.
From diagrams (A.14) and (A.16) one can easily obtain for all indices $0 \leq m \leq n$, the commutativity of the following diagrams:

\[
\begin{array}{ccc}
C(n) & \xrightarrow{j_{n,n-m}} & C(n-m) \\
\gamma_n & \downarrow & \downarrow \\
C_1(n) & \xrightarrow{j'_{n,n-m}} & C_1(n-m)
\end{array}
\]

\[
\begin{array}{ccc}
i_{n-m,n} & \xrightarrow{i_{n-m,n}} & i_{n-m,n} \\
\gamma_n & \downarrow & \downarrow \\
i_{n-m,n} & \xrightarrow{i'_{n-m,n}} & i_{n-m,n}
\end{array}
\]

Because $\gamma_n$ is monomorphism, the equality $i_{n-m,n} \circ j_{n,n-m} = p^m \text{id}_{C(n)}$ implies the equality $i'_{n-m,n} \circ j'_{n,n-m} = p^m \text{id}_{C_1(n)}$. This gives the part b) of the definition of $p$-divisible groups for $(C_1(n), i_{1(n)})_{n \geq 0}$. The proposition is proved. \hfill \Box

**Appendix B. SAGE program**

This program computes the class number of the field $\mathbb{Q}(\sqrt[3]{3}, \zeta_9)$ and finds the basis $f_1, f_2, \ldots, f_9$ of the 3-subgroup of units in this field such that for the normalized 3-adic valuation $v_3$ and all $1 \leq i \leq 9$, the natural numbers $a_i = 18v_3(f_i \pm 1)$ are prime to 3 and $1 \leq a_1 < a_2 < \cdots < a_9$. The result appears as the vector $af = (a_1, a_2, \ldots, a_9) = (1, 2, 4, 5, 7, 8, 10, 13, 16).

```sage
sage: L.<b>=NumberField(x^3-3);
sage: R.<t>=L[]
sage: M.<c>=L.extension(t^6+t^3+1);
sage: X.<d>=M.absolute_field();
sage: h=X.class_number();
sage: e=list(X.unit_group().gens())

sage: def p(x):
... for i in range(1,3):
...     if valuation(norm(X(x+2*i-3)),3)!=0:
...         break
... return valuation(norm(X(x+2*i-3)),3)

sage: a=[p(x) for x in e]
sage: f=[e.pop(a.index(min(a)))]
sage: while len(e)!=0:
...     a=[p(x) for x in e]
...     i0=a.index(min(a))
...     for j in range(len(f)):
...         for k in range(5):
...             s=0
...             if p(f[j]^(3^k))>p(e[i0]):
...                 break
```

Documenta Mathematica 18 (2013) 547–619
Varieties with Bad Reduction at 3 Only

if p(e[i0])==p(f[j]^(3^k)):
    s=1
    break
if s == 1:
    for i in range(1,3):
        if p(e[i0])<p(e[i0]/(f[j]^(i*3^k))):
            e[i0]=e[i0]/(f[j]^(i*3^k))
            break
    break
if j+1==len(f) and s==0:
    f.append(e.pop(i0))
sage: af=[p(x) for x in f];
sage: print h
sage: print af
1
[1, 2, 4, 5, 7, 8, 10, 13, 16]

Remark. First 4 lines introduce the field $X = \mathbb{Q}(\sqrt[3]{3}, \zeta_9)$; its elements appear as polynomials in variable $d$ of degree $\leq 17$. Then we find the class number of $X$ and form the array $e = (e[1], \ldots, e[9])$ of minimal generators of the group $U/U^3$, where $U$ is the group of units in $X$. Next block gives a standard procedure to determine for any $x \in U$ the maximal natural number $p(x)$ such that $x \pm 1$ is divisible precisely by $\pi^{p(x)}$, where $\pi \in X$, $\langle \pi^{18} \rangle = (3)$. The remaining part of the program is based on a standard technique from Linear algebra to rearrange the given system of generators $e$ into a new system $f$ with required properties. As a matter of fact, we use that the class number of $X$ is prime to 3 (it equals 1) by allowing $k < 5$ on line 21. (Any unit $x \equiv 1 \mod \pi^{2k}$ is a cube in the 3-completion of $X$ by Hensel’s Lemma and, therefore, is a cube in $X$.) The last two lines contain the values of the class number of $X$ and the exponents $(a(f[1]), \ldots, a(f[9]))$.

References


Documenta Mathematica 18 (2013) 547–619


[11] Diaz Y Diaz, Tables minorant la racine $n$-ième du discriminant d’un corps de degré $n$, Publications Mathématiques d’Orsay 80.06, Université de Paris-Sud, Département de Mathématique, Bat. 425, 91405, Orsay, France


Victor Abrashkin
Department of Mathematical Sciences
Durham University
Science Laboratories,
South Rd, Durham DH1 3LE
United Kingdom
victor.abrashkin@durham.ac.uk
and
Steklov Mathematical Institute
Gubkina str. 8,
119991 Moscow
Russia
NEKOVÁŘ DUALITY OVER \( p \)-ADIC LIE EXTENSIONS OF GLOBAL FIELDS

MENG FAI LIM AND ROMYAR T. SHARIFI

Received: July 12, 2012
Revised: March 11, 2013

Communicated by Peter Schneider

Abstract. Tate duality is a Pontryagin duality between the \( i \)th Galois cohomology group of the absolute Galois group of a local field with coefficients in a finite module and the \((2 - i)\)th cohomology group of the Tate twist of the Pontryagin dual of the module. Poitou-Tate duality has a similar formulation, but the duality now takes place between Galois cohomology groups of a global field with restricted ramification and compactly-supported cohomology groups. Nekovář proved analogues of these in which the module in question is a finitely generated module \( T \) over a complete commutative local Noetherian ring \( R \) with a commuting Galois action, or a bounded complex thereof, and the Pontryagin dual is replaced with the Grothendieck dual \( T^* \), which is a bounded complex of the same form. The cochain complexes computing the Galois cohomology groups of \( T \) and \( T^*(1) \) are then Grothendieck dual to each other in the derived category of finitely generated \( R \)-modules. Given a \( p \)-adic Lie extension of the ground field, we extend these to dualities between Galois cochain complexes of induced modules of \( T \) and \( T^*(1) \) in the derived category of finitely generated modules over the possibly noncommutative Iwasawa algebra with \( R \)-coefficients.

2010 Mathematics Subject Classification: 11R23, 11R34, 11S25, 16E35, 18E30
Keywords and Phrases: Galois cohomology, Tate duality, Poitou-Tate duality, Grothendieck duality
1 Introduction

1.1 Duality

In [Nek], Nekovár gave formulations of analogues of Tate and Poitou-Tate duality for finitely generated modules over a complete commutative local Noetherian ring $R$ with finite residue field of characteristic a fixed prime $p$. In the usual formulation of these dualities, one takes the Pontryagin dual, which does not in general preserve the property of finite generation. Nekovár takes the dual with respect to a dualizing complex of Grothendieck so as to have a duality between bounded complexes of $R$-modules with finitely generated cohomology groups. This paper is devoted to a generalization of this result to the setting of nonabelian $p$-adic Lie extensions.

Recall that a dualizing complex $\omega_R$ is a bounded complex of $R$-modules with cohomology finitely generated over $R$ that has the property that for every complex $M$ of finitely generated $R$-modules, the Grothendieck dual $R\text{Hom}_R(M, \omega_R)$ in the derived category of $R$-modules $D(\text{Mod}_R)$ has finitely generated cohomology, and moreover, the canonical morphism

$$M \longrightarrow R\text{Hom}_R(R\text{Hom}_R(M, \omega_R), \omega_R)$$

is an isomorphism in $D(\text{Mod}_R)$. Such a complex exists and is unique up to quasi-isomorphism and translation (see [Har1]).

One can choose $\omega_R$ to be a bounded complex of injectives, in which case the derived homomorphism complexes are represented by the complexes of homomorphisms themselves. If $R$ is regular, then $R$ itself, as a complex concentrated in degree 0, is a dualizing complex, but $R$ is not in general $R$-injective. If $R = \mathbb{Z}_p$, for instance, then the complex $[\mathbb{Q}_p \rightarrow \mathbb{Q}_p/\mathbb{Z}_p]$ concentrated in degrees 0 and 1 provides a complex of injective $\mathbb{Z}_p$-modules quasi-isomorphic to $\mathbb{Z}_p$.

Let us explain Nekovár’s theorem, and our generalization of it, in the setting of Poitou-Tate duality. Let $F$ be a global field with characteristic not equal to $p$. Let $S$ be a finite set of primes of $F$ that, if $F$ is a number field, contains all primes over $p$ and any real places, and let $G_{F,S}$ denote the Galois group of the maximal unramified outside $S$ extension of $F$. We remark that a finitely generated $R$-module has a canonical topology arising from its filtration by powers of the maximal ideal of $R$. Let $T$ be a bounded complex in the category of finitely generated (topological) $R$-modules with $R$-linear continuous $G_{F,S}$-actions. We use $R\Gamma(G_{F,S}, T)$ to denote the object in $D(\text{Mod}_R)$ corresponding to the complex of continuous $G_{F,S}$-cochains with coefficients in $T$, and we use $R\Gamma_{(c)}(G_{F,S}, T)$ to denote the derived object attached to the complex of continuous compactly supported cochains (using Tate groups for real places in its definition as a cone: see Section [4]).

There exists a bounded complex $T^*$ of finitely generated $R$-modules with $R$-linear continuous $G_{F,S}$-actions that represents $R\text{Hom}_R(T, \omega_R)$ in the derived category of such modules. Nekovár’s duality theorem is then as follows (see Documenta Mathematica 18 (2013) 621–678).
Theorem (Nekovář). We have an isomorphism
\[ R\Gamma(G_{F,S}, T) \to R\text{Hom}_R \left( R\Gamma_{(c)}(G_{F,S}, T^*(1)), \omega_R \right)[-3] \]
in the derived category of finitely generated \( R \)-modules.

In this paper, we consider a generalization of this to the setting of noncommutative Iwasawa theory. Suppose that \( F_\infty \) is a \( p \)-adic Lie extension of \( F \) contained in \( F_S \). We denote by \( \Gamma \) the Galois group of the extension \( F_\infty/F \), and we let \( \Lambda = R[\Gamma] \) denote the resulting Iwasawa algebra over \( R \). For a finitely generated \( R \)-module \( T \) with a continuous \( R \)-linear \( G_{F,S} \)-action, we define a finitely generated \( \Lambda \)-module \( F\Gamma(T) \) with a continuous \( \Lambda \)-linear \( G_{F,S} \)-action by
\[ F\Gamma(T) = \Lambda^\iota \otimes_R T, \]
where the superscript \( \iota \) denotes that an element of \( \Gamma \) in \( \Lambda \) acts here on \( \Lambda \) by right multiplication by its inverse, and where \( G_{F,S} \) acts on \( \Lambda \) through the quotient map \( G_{F,S} \to \Gamma \) by left multiplication and then diagonally on the tensor product.

The \( G_{F,S} \)-cohomology of \( F\Gamma(T) \) is interesting in that a version of Shapiro’s lemma provides natural isomorphisms of continuous cohomology groups
\[ H^n(G_{F,S}, F\Gamma(T)) \cong \varprojlim H^n(G_{F_\alpha,S}, T) \]
for every \( n \geq 0 \), where the limit is taken over \( \alpha \) indexing the finite Galois extensions \( F_\alpha \) of \( F \) that are contained in \( F_\infty \) (cf. [Lim2, Lemma 5.3.1]). That is, the cohomology groups of \( F\Gamma(T) \) are the Iwasawa cohomology groups of the module \( T \) itself for the extension \( F_\infty/F \), and this identification is one of \( \Lambda \)-modules. We also have the analogous statements for compactly supported cohomology, as seen in [Lim2, Proposition 5.3.3]. Therefore, we can reduce the question of finding dualities among Iwasawa cohomology groups of compact modules to that of obtaining dualities among cohomology groups for \( G_{F,S} \) itself. Note that, however, it is not even a priori clear in this setting that \( H^n(G_{F,S}, F\Gamma(T)) \) is a finitely generated \( \Lambda \)-module, let alone that we can find such a duality of \( \Lambda \)-modules.

The following is our main theorem (cf. Theorem 4.5.1).

Theorem. We have isomorphisms
\[ R\Gamma(G_{F,S}, F\Gamma(T)) \to R\text{Hom}_{\Lambda^\iota} \left( R\Gamma_{(c)}(G_{F,S}, F\Gamma(T^*(1)), \Lambda \otimes_R \omega_R \right)[-3] \]
\[ R\Gamma_{(c)}(G_{F,S}, F\Gamma(T)) \to R\text{Hom}_{\Lambda^\iota} \left( R\Gamma(G_{F,S}, F\Gamma(T^*(1)), \Lambda \otimes_R \omega_R \right)[-3] \]
in the derived category of finitely generated \( \Lambda \)-modules.

\footnote{Actually, Nekovář’s longer treatise is particularly concerned with a generalization of this duality that takes place between Selmer complexes, which we do not address in this article.}
Nekovář proves the above theorem in the case that \( \Gamma \) is abelian [Nek] Theorem 8.5.6. In fact, in that case, it is an almost immediate consequence of his above-mentioned theorem. That is, suppose for instance that \( \Gamma \cong \mathbb{Z}_p^r \) for some \( r \geq 1 \). Then \( \Lambda \) is a complete commutative Noetherian local ring with finite residue field of characteristic \( p \). Moreover, its dualizing complex is isomorphic to \( \Lambda \otimes_R \omega_R \) in the derived category of \( \Lambda \)-modules [Nek] Lemma 8.4.5.6. Therefore, the commutative theory described above applies to \( \Lambda \), and Nekovář is able to deduce the result from this. On the other hand, since we are working with nonabelian \( \Gamma \) and hence noncommutative \( \Lambda \), we do not know that there exists a (nice enough) dualizing complex, and so the proof of our main theorem takes a different route. The idea is a simple one, though the proof is rather involved: after reducing to the case that \( R \) is regular and \( \mathbb{Z}_p \)-flat, we perform an inductive argument, using the grading on \( \Lambda \) arising from the powers of its augmentation ideal, to deduce our result from Nekovář’s.

We remark that, in their manuscript on the noncommutative main conjecture, Fukaya and Kato stated an analogue of our main theorem, with \( \Lambda \) replacing \( \Lambda \otimes_R \omega_R \) [FK1, (1.6.12)], which in turn generalized a result of Burns and Flach [BF, Lemma 12(b)] for a narrower class of rings. The result of Fukaya-Kato applies to a more general class of (adic) rings \( \Lambda \) than ours and replaces \( \mathcal{F}_T(T) \) and \( \mathcal{F}_T(T^*) \) by a bounded complex of \( \Lambda[\mathbf{G}_F,\mathcal{S}] \)-modules \( X \) and its \( \Lambda \)-dual \( \text{Hom}_\Lambda(X, \Lambda) \). However, in order to be able to work with the \( \Lambda \)-dual, they assume that \( X \) consists of (finitely generated) projective \( \Lambda \)-modules. In the case of Iwasawa cohomology that we study, the complex \( T \) need not be quasi-isomorphic to a bounded complex of \( R[\mathbf{G}_F] \)-modules that are projective and finitely generated over \( R \). Moreover, if \( R \) is Gorenstein, then \( R \) serves as an \( R \)-dualizing complex, and our result reduces to a duality with respect to \( \Lambda \) itself, as in the result of Fukaya-Kato. We also note that Vauclair proved a noncommutative duality theorem for induced modules in the case that \( R = \mathbb{Z}_p \) and \( T \) is \( \mathbb{Z}_p \)-free, via a rather different method [Vau, Theorem 6.4].

1.2 An application

Since applications of our main result are not discussed in the body of this paper, we end the introduction with an indication of one setting in which our results naturally apply. Fix \( N \geq 1 \) not divisible by \( p \), and suppose that \( p \geq 5 \). Let \( \mathbb{Z}_{p,N} \) denote the inverse limit of the rings \( \mathbb{Z}/Np^r\mathbb{Z} \) over \( r \geq 1 \). Hida’s ordinary cuspidal \( \mathbb{Z}_p \)-Hecke algebra \( \mathfrak{h} \) of level \( Np^\infty \) is a direct product of local rings that is free of finite rank over the subalgebra \( \Omega = \mathbb{Z}_p[1 + p\mathbb{Z}_p] \) of the algebra \( \mathbb{Z}_p[[\mathbb{Z}_p,N]/\langle -1 \rangle] \) of diamond operators in \( \mathfrak{h} \) [Hid, Theorem 3.1]. Hida showed that the \( \mathfrak{h} \)-module \( \mathcal{S} \) of \( \Omega \)-adic cusp forms is \( \Omega \)-dual to \( \mathfrak{h} \) (see Ohi1 Theorem 2.5.3) for a proof), from which it follows that \( \mathcal{S} \) is a dualizing complex for \( \mathfrak{h} \).

The inverse limit

\[
\mathcal{H} = \lim_{\rightarrow} H^1_{\text{ét}}(X(Np^r)/\mathbb{Q}, \mathbb{Z}_p(1))^{\text{ord}}
\]
of ordinary parts of étale cohomology groups of modular curves is an \( \mathfrak{h}[G_{Q,S}] \)-module for the dual action of \( \mathfrak{h} \), where \( S \) is the set of primes dividing \( Np\infty \). As an \( \mathfrak{h} \)-module, Ohta showed in \([\text{Oht}2]\) that \( \mathcal{H} \) is an extension of \( S \) by \( \mathfrak{h} \). Since \( \mathfrak{h} \) is not always Gorenstein (in Eisenstein components: see \([\text{Oht}3]\) Corollary 4.2.13 for conditions), it is not at all clear that \( \mathcal{H} \) is quasi-isomorphic to a bounded complex of \( \mathfrak{h}[G_{Q,S}] \)-modules that are finitely generated and projective over \( \mathfrak{h} \), precluding the use of the duality result of Fukaya-Kato. On the other hand, there is a perfect, \( \mathfrak{h} \)-bilinear pairing \( \mathcal{H} \times \mathcal{H} \rightarrow S(1) \) which is \( G_{Q,S} \)-equivariant for the action of \( \sigma \in G_{Q,S} \) on \( S \) via the diamond operator \((\chi(\sigma))\), where \( \chi: G_{Q,S} \rightarrow \mathbb{Z}_{p,N}^\times \) is the cyclotomic character (see \([\text{FK}2]\) Section 1.6], where it is quickly derived from a pairing of Ohta’s \([\text{Oht}1]\) Definition 4.1.17]). Now fix a \( p \)-adic Lie extension \( F_\infty \) of \( \mathbb{Q} \) that is unramified outside \( S \) and contains \( \mathbb{Q}(\mu_{Np\infty}) \), and set \( \Gamma = \text{Gal}(F_\infty/\mathbb{Q}) \). The complex \( \text{R} \Gamma(\sigma;G_{Q,S},\mathcal{F}_T(\mathcal{H})) \) is the subject of the noncommutative Tamagawa number conjecture of Fukaya and Kato for \( \mathcal{H} \) (though to be precise, said conjecture is only formulated in the case that \( \mathcal{H} \) is \( \mathfrak{h} \)-projective), which is directly related to the noncommutative main conjecture for ordinary \( \Omega \)-adic cusp forms (see \([\text{FK}1]\) Sections 2.3 and 4.2)). It is also perhaps worth remarking that, in the commutative setting, the first Iwasawa cohomology group \( H^1(G_{Q,S},\mathcal{F}_T(\mathcal{H})) \) contains zeta elements constructed out of Kato’s Euler system (see \([\text{Kat}\] Section 12) or \([\text{FK}2]\) Section 3.2).

For a finitely generated \( \Lambda = \mathfrak{h}[\Gamma] \)-module \( A \), let us use \( A(\chi) \) to denote the \( \Lambda \)-module that is \( A \) as an \( \mathfrak{h} \)-module but for which the original \( \mathfrak{h} \)-linear action of \( \gamma \in \Gamma \) on \( A \) has been twisted by multiplication by the diamond operator \((\chi(\gamma))\), for any lift \( \tilde{\gamma} \in G_{Q,S} \) of \( \gamma \). Using the pairing on \( \mathcal{H} \), our main result can be seen to yield two interesting isomorphisms in the derived category of finitely generated modules over, including

\[
\text{R} \Gamma(\sigma;G_{Q,S},\mathcal{F}_T(\mathcal{H})) \longrightarrow \text{R} \text{Hom}_{\Lambda^+}(\text{R} \Gamma(\sigma;G_{Q,S},\mathcal{F}_T(\mathcal{H})),\Lambda \otimes^L \mathcal{S})(\chi)[-3].
\]

Perhaps more concretely, we have a spectral sequence

\[
\text{Ext}_{\Lambda^+}^i \left( H^{j-1}(G_{Q,S},\mathcal{F}_T(\mathcal{H})),\Lambda \otimes^L \mathcal{S} \right)(\chi) \Rightarrow H^{i+j}(\sigma;G_{Q,S},\mathcal{F}_T(\mathcal{H})).
\]

We remark that one has a similar result with \( X_1(Np^r) \) replaced by \( Y_1(Np^r) \); in this case, the pairing on \( \mathcal{H} \) to \( \mathcal{S} \) is replaced by a pairing between ordinary parts of cohomology and compactly supported cohomology groups to ordinary \( \Omega \)-adic modular forms. We also note that Fouquet has constructed an analogue of Ohta’s pairing for towers of Shimura curves attached to indefinite quaternion algebras over totally real fields \([\text{Fou}\] Proposition 2.8], providing a related setting for an application of our main result.

Acknowledgments. The material presented in this article forms a revised and extended version of the part of the Ph.D. thesis \([\text{Lim1}\] of the first author that was a collaboration between the two authors. The results of the rest of the thesis can be found in the article \([\text{Lim2}\]. The first author would like to
thank Manfred Kolster for his encouragement. The second author would like to thank Takako Fukaya, Uwe Jannsen, Jan Nekovář, Otmar Venjakob, and Amnon Yekutieli for helpful discussions. He was supported by NSF Grant DMS-0901526, and his work was partially conducted during visits to the Newton Institute in 2009 and IHES in 2010. The authors also thank the anonymous referee for pointing out a mistake in an earlier version of this work.

2 Preliminaries

In this section, we lay out some facts regarding modules over profinite rings that we shall require later in the paper. Elsewhere in the paper, we use some of the standard terminology and conventions regarding complexes, cones, shifts, total complexes of homomorphisms and tensor products, and derived categories, as can be found in [Nek], or in [Lim2] off of which this paper builds. In particular, all sign conventions will be as in the latter two papers. For further background on derived categories, we suggest the excellent book [Wei], as well as the more advanced text [KS].

2.1 Tensor products and homomorphism groups

In this subsection, $R$ is a commutative ring. By an $R$-algebra, we will mean a ring with a given homomorphism from $R$ to its center. In this section, we construct derived bifunctors of homomorphism groups and tensor products for bimodules over $R$-algebras and study a few isomorphisms that result from these constructions.

We use $\Lambda^\circ$ to denote the opposite ring to an $R$-algebra $\Lambda$. Given two $R$-algebras $\Lambda$ and $\Sigma$, we are interested in a subclass of the class of $\Lambda$-$\Sigma$-bimodules, namely, the class of $\Lambda$-$\Sigma$-bimodules with the extra property that the left $R$-action coincides with the right $R$-action. We can (and shall) identify the category of such $\Lambda$-$\Sigma$-bimodules with the category of $\Lambda \otimes_R \Sigma^\circ$-modules, and there are natural exact forgetful functors

$$\text{res}_\Lambda : \text{Mod}_{\Lambda \otimes_R \Sigma^\circ} \longrightarrow \text{Mod}_\Lambda \quad \text{and} \quad \text{res}_\Sigma^\circ : \text{Mod}_{\Lambda \otimes_R \Sigma^\circ} \longrightarrow \text{Mod}_{\Sigma^\circ},$$

which extend to exact functors on the derived categories. One observes that the categories $\text{Mod}_{\Lambda \otimes_R \Sigma^\circ}$ and $\text{Mod}_\Lambda$ are equivalent, as are $\text{Mod}_{R \otimes_R \Sigma^\circ}$ and $\text{Mod}_{\Sigma^\circ}$.

We have following lemma, which extends [Yek Lemma 2.1].

**Lemma 2.1.1.**

(a) If $\Sigma$ is a projective (resp., flat) $R$-algebra, then $\text{res}_\Lambda$ preserves projective (resp., flat) modules. Similarly, if $\Lambda$ is a projective (resp., flat) $R$-algebra, then $\text{res}_\Sigma^\circ$ preserves projective (resp., flat) modules.

(b) If $\Sigma$ is a flat $R$-algebra, then $\text{res}_\Lambda$ preserves injectives. Similarly, if $\Lambda$ is a flat $R$-algebra, then $\text{res}_\Sigma^\circ$ preserves injectives.
Proof. (a) We prove the statements of the first sentence, those of the second being a consequence. Suppose that $\Sigma$ is a projective $R$-algebra. Since projective modules are the summands of free modules, it suffices to show that $\Lambda \otimes_R \Sigma^\circ$ is a projective $\Lambda$-module. Since $\Sigma$ is a central $R$-algebra, we have $\Sigma \cong \Sigma^\circ$ as $R$-modules. Therefore, we have $\Sigma^\circ \oplus P \cong L$ for some projective $R$-module $P$ and free $R$-module $L$. Then $\Lambda \otimes_R \Sigma^\circ$ is a direct summand of $\Lambda \otimes_R L$, which is a free $\Lambda$-module. Hence, $\Lambda \otimes_R \Sigma^\circ$ is a projective $\Lambda$-module.

Now suppose that $\Sigma$ is flat over $R$. Since flat modules are direct limits of finitely generated free modules (see [Lam, Theorem 4.34]) and tensor products preserve direct limits, it suffices to show that $\Lambda \otimes_R \Sigma^\circ$ is a flat $\Lambda$-algebra. Since $\Sigma$ is flat over $R$, we have that $\Sigma$ is a direct limit of finitely generated free $R$-modules, which implies that $\Lambda \otimes_R \Sigma^\circ$ is a direct limit of finitely generated free $\Lambda$-modules.

(b) We shall prove this for $\mathrm{res}_\Lambda$, the case of $\mathrm{res}_\Sigma^\circ$ being a consequence. The functor from $\mathrm{Mod}_\Lambda$ to $\mathrm{Mod}_{\Lambda \otimes_R \Sigma^\circ}$ sending $M$ to $M \otimes_R \Sigma^\circ$ is exact by our assumption and is left adjoint to the functor $\mathrm{res}_\Lambda$. The conclusion then follows from [Wei, Prop. 2.3.10].

Let us briefly introduce the notions of q-projective and q-injective complexes of $\Lambda$-modules and several facts regarding them arising from the work of Spaltenstein, as can be found in [Spa, Kel, Lip, Section 2.3], and [KS, Chapter 14]. A complex $P$ of $\Lambda$-modules is called q-projective if for every map $g: P \to B$ and quasi-isomorphism $s: A \to B$ of complexes of $\Lambda$-modules, there exists a map $f: P \to A$ of complexes of $\Lambda$-modules such that $g$ and $s \circ f$ are homotopy equivalent. We remark that $P$ is q-projective if and only if it is homotopy equivalent to a direct limit of bounded above complexes $P_n$ of projective $\Lambda$-modules via maps $t_n: P_n \to P_{n+1}$ that are split injective in each degree with quotients $P_{n+1}/t_n(P_n)$ having zero differentials. In particular, bounded above complexes of projectives are q-projective. If $A$ is a complex of $\Lambda$-modules, then there exists a quasi-isomorphism $P \to A$ with $P$ q-projective.

We also have the dual notion of q-injective complexes. A complex $I$ of $\Lambda$-modules is q-injective if for every map $f: A \to I$ and quasi-isomorphism $s: A \to B$ of complexes of $\Lambda$-modules, there exists a map $g: B \to I$ such that $f$ and $g \circ s$ are homotopy equivalent. A complex $I$ of $\Lambda$-modules is q-injective if and only if it is homotopy equivalent to an inverse limit of bounded below complexes $I_n$ of injective $\Lambda$-modules via maps $\pi_n: I_{n+1} \to I_n$ that are split surjective in each degree with kernels $\ker \pi_n$ having zero differentials. In particular, bounded below complexes of injectives are q-injective. If $A$ is a complex of $\Lambda$-modules, then there exists a quasi-isomorphism $A \to I$ with $I$ q-injective.

In addition to $\Lambda$ and $\Sigma$, we now let $\Omega$ be an $R$-algebra. If $A$ is a $\Lambda \otimes_R \Omega^\circ$-module and $B$ is a $\Lambda \otimes_R \Sigma^\circ$-module, we give $\mathrm{Hom}_\Lambda(A, B)$ the structure of an $\Omega \otimes_R \Sigma^\circ$-module via the left $\Omega$ and right $\Sigma$-actions

$$(\omega \cdot f)(a) = f(\omega a) \quad \text{and} \quad (f \cdot \sigma)(a) = f(a)\sigma$$
for $f \in \text{Hom}_A(A, B)$, $a \in A$, $\omega \in \Omega$, and $\sigma \in \Sigma$.

Moreover, if $A$ is a complex of $\Lambda \otimes_R \Omega^\circ$-modules and $B$ is a complex of $\Lambda \otimes_R \Sigma^\circ$-modules, we define a complex $\text{Hom}_A(A, B)$ of $\Omega \otimes_R \Sigma^\circ$-modules by

$$\text{Hom}_A^\Lambda(A, B) = \prod_{i \in \mathbb{Z}} \text{Hom}_\Lambda(A^i, B^{i+n}),$$

with the usual differentials, as in [Lim2, Section 2].

**Proposition 2.1.2.** Let $A$ be a complex of $\Lambda \otimes_R \Omega^\circ$-modules, and let $B$ be a complex of $\Lambda \otimes_R \Sigma^\circ$-modules. If $\Omega$ is a projective $R$-algebra or $\Sigma$ is a flat $R$-algebra, then we have a derived bifunctor

$$\mathbf{R}\text{Hom}_\Lambda(-, -) : \mathbf{D}(\text{Mod}_{\Lambda \otimes_R \Omega^\circ}) \to \mathbf{D}(\text{Mod}_{\Omega \otimes_R \Sigma^\circ}),$$

and $\text{Hom}_A(A, B)$ represents $\mathbf{R}\text{Hom}_\Lambda(A, B)$ if $A$ is $\pi$-projective as a complex of $\Lambda$-modules and $\Omega$ is $R$-projective or $B$ is $\pi$-injective as a complex of $\Lambda$-modules and $\Sigma$ is $R$-flat.

**Proof.** Let us first assume that $\Omega$ is a projective $R$-algebra. Since every complex of $\Lambda \otimes_R \Omega^\circ$-modules is quasi-isomorphic to a $\pi$-projective complex of $\Lambda \otimes_R \Omega^\circ$-modules, we have a derived functor

$$\mathbf{R}\text{Hom}_\Lambda(-, B) : \mathbf{D}(\text{Mod}_{\Lambda \otimes_R \Omega^\circ}) \to \mathbf{D}(\text{Mod}_{\Omega \otimes_R \Sigma^\circ}).$$

(see [Lim, Corollary 2.3.2.3]). Suppose that $f : B \to B'$ is a quasi-isomorphism of complexes of $\Lambda \otimes_R \Omega^\circ$-modules. Let $\varepsilon : P \to A$ be a quasi-isomorphism of complexes of $\Lambda \otimes_R \Omega^\circ$-modules, where $P$ is $\pi$-projective. Then $\mathbf{R}\text{Hom}_\Lambda(P, B)$ (resp., $\mathbf{R}\text{Hom}_\Lambda(P, B')$) is represented by $\text{Hom}_\Lambda(P, B)$ (resp., $\text{Hom}_\Lambda(P, B')$).

Applying Lemma 2.1.1(a), we see that $P$ is also a $\pi$-projective complex of $\Lambda$-modules, so by [Lim, Proposition 2.3.8], the induced map

$$f_* : \text{Hom}_\Lambda(P, B) \to \text{Hom}_\Lambda(P, B')$$

is a quasi-isomorphism of complexes of abelian groups. Since $f_*$ is a morphism of complexes of $\Omega \otimes_R \Sigma^\circ$-modules, it is a quasi-isomorphism of such complexes. Hence, $f$ induces isomorphisms

$$\mathbf{R}\text{Hom}_\Lambda(A, B) \to \mathbf{R}\text{Hom}_\Lambda(A, B'),$$

proving the existence of the derived bifunctor. Moreover, if $A$ is $\pi$-projective as a complex of $\Lambda$-modules, then

$$\varepsilon_* : \text{Hom}_\Lambda(A, B) \to \text{Hom}_\Lambda(P, B)$$

is a quasi-isomorphism of complexes of $\Sigma^\circ$-modules, hence of $\Omega \otimes_R \Sigma^\circ$-modules as well. Thus, $\text{Hom}_\Lambda(A, B)$ represents $\mathbf{R}\text{Hom}_\Lambda(A, B)$, as desired.

If $\Sigma$ is $R$-flat, the argument for the existence of the derived bifunctor and its computation by $\text{Hom}_\Lambda(A, B)$ in the case that $B$ is $\pi$-injective as a complex of
\(\Lambda\)-modules is the direct analogue of the above argument, employing Lemma 2.1.1(b). To see that the resulting derived functor agrees with that constructed above in the case that both \(\Omega\) is \(R\)-projective and \(\Sigma\) is \(R\)-flat, consider a quasi-isomorphism \(\varepsilon: P \to A\) as above and a quasi-isomorphism \(\delta: B \to I\) with \(I\) a \(q\)-injective complex of \(\Lambda \otimes_R \Sigma^o\)-modules. As we have quasi-isomorphisms

\[
\hom\Lambda(P, B) \xrightarrow{\delta_*} \hom\Lambda(P, I) \xleftarrow{\varepsilon_*} \hom\Lambda(A, I)
\]

of complexes of \(\Omega \otimes_R \Sigma^o\)-modules, the derived functors coincide.

Note that one always has a map \(\hom\Lambda(A, B) \to R\hom\Lambda(A, B)\) (canonical up to isomorphism in the derived category), induced either by a quasi-isomorphism \(P \to A\) with \(P\) \(q\)-projective or a quasi-isomorphism \(B \to I\) with \(I\) \(q\)-injective.

**Corollary 2.1.3.** If \(\Omega\) is a projective \(R\)-algebra, then we have a commutative diagram

\[
\begin{array}{ccc}
D(\text{Mod}_{\Lambda \otimes R \Omega})^o \times D(\text{Mod}_{\Lambda \otimes R \Sigma^o}) & \xrightarrow{R\hom\Lambda(-,-)} & D(\text{Mod}_{\Omega \otimes R \Sigma^o}) \\
\text{res_A} \times \text{id} & & \text{res_{\Sigma^o}}
\end{array}
\]

In the case that \(\Sigma\) is a flat \(R\)-algebra, we have a commutative diagram

\[
\begin{array}{ccc}
D(\text{Mod}_{\Lambda \otimes R \Omega})^o \times D(\text{Mod}_{\Lambda \otimes R \Sigma^o}) & \xrightarrow{R\hom\Lambda(-,-)} & D(\text{Mod}_{\Omega \otimes R \Sigma^o}) \\
\text{id} \times \text{res}_A & & \text{res}_I
\end{array}
\]

If \(A\) is a complex of \(\Omega \otimes_R \Lambda^o\)-modules and \(B\) is a complex of \(\Lambda \otimes_R \Sigma^o\)-modules, we define a complex \(A \otimes_{\Lambda} B\) of \(\Omega \otimes_R \Sigma^o\)-modules by

\[
(A \otimes_{\Lambda} B)^n = \bigoplus_{i \in \mathbb{Z}} A^i \otimes_{\Lambda} B^{n-i},
\]

again with the usual differentials, as in [Lim2, Section 2].

We also have a notion of a \(q\)-flat complex of \(\Lambda^o\)-modules (see [Lip, Section 2.5] and [Spa, Section 5] in the case of commutative rings, the proofs being identical); that is, a complex of \(\Lambda^o\)-modules \(A\) is said to be \(q\)-flat if for every quasi-isomorphism \(B \to C\) of complexes of \(\Lambda\)-modules, the resulting map \(A \otimes_{\Lambda} B \to C\) is also a quasi-isomorphism. In particular, \(q\)-projective complexes of \(\Lambda^o\)-modules are \(q\)-flat, any bounded above complex of flat \(\Lambda^o\)-modules is \(q\)-flat, and any filtered direct limit of \(q\)-flat complexes is \(q\)-flat. As in the case of homomorphism complexes, the total tensor product induces derived bifunctors as follows. We omit the analogous proof.
Proposition 2.1.4. Let \( A \) and \( B \) be complexes of \( \Omega \otimes_R \Lambda^\circ \)-modules and \( \Lambda \otimes_R \Sigma^\circ \)-modules, respectively. If either \( \Omega \) or \( \Sigma \) is a flat \( R \)-algebra, then we have a derived bifunctor

\[
\bigodot_k : D(\text{Mod}_{\Omega \otimes_R \Lambda^\circ}) \times D(\text{Mod}_{\Lambda \otimes_R \Sigma^\circ}) \to D(\text{Mod}_{\Omega \otimes_R \Sigma^\circ}),
\]

and \( A \otimes \Lambda B \) represents \( A \otimes_k \Lambda B \) if \( A \) is \( q \)-flat as a complex of \( \Lambda^\circ \)-modules and \( \Omega \) is \( R \)-flat or if \( B \) is \( q \)-flat as a complex of \( \Lambda \)-modules and \( \Sigma \) is \( R \)-flat.

Proposition 2.1.4 has the following direct corollary.

Corollary 2.1.5. If \( \Omega \) is a flat \( R \)-algebra, then we have a commutative diagram

\[
\begin{array}{ccc}
D(\text{Mod}_{\Omega \otimes_R \Lambda^\circ}) \times D(\text{Mod}_{\Lambda \otimes_R \Sigma^\circ}) & \to & D(\text{Mod}_{\Omega \otimes_R \Sigma^\circ}) \\
\text{id} \times \text{res}_{\Lambda^\circ} & & \text{res}_{\Sigma^\circ} \\
\text{res}_{\Lambda^\circ} \times \text{id} & & \\
D(\text{Mod}_{\Lambda^\circ}) \times D(\text{Mod}_{\Lambda \otimes_R \Sigma^\circ}) & \to & D(\text{Mod}_{\Sigma^\circ}),
\end{array}
\]

and if \( \Sigma \) is a flat \( R \)-algebra, we have a commutative diagram

\[
\begin{array}{ccc}
D(\text{Mod}_{\Omega \otimes_R \Lambda^\circ}) \times D(\text{Mod}_{\Lambda \otimes_R \Sigma^\circ}) & \to & D(\text{Mod}_{\Omega \otimes_R \Sigma^\circ}) \\
\text{id} \times \text{res}_{\Lambda} & & \text{res}_{\Omega} \\
\text{res}_{\Lambda} \times \text{id} & & \\
D(\text{Mod}_{\Omega \otimes_R \Lambda^\circ}) \times D(\text{Mod}_{\Lambda}) & \to & D(\text{Mod}_{\Omega}).
\end{array}
\]

We end this section with some general lemmas regarding the passing of tensor products through homomorphism groups and the resulting isomorphisms in the derived categories. Let us use \( \bar{m} \) to denote the degree of an element \( m \) of a term of a complex \( M \) of modules over a ring. We fix a fourth \( R \)-algebra \( \Xi \).

Lemma 2.1.6. Suppose that \( \Xi \) is a flat \( R \)-algebra and \( \Sigma \) is a projective \( R \)-algebra. Let \( A \) be a complex of \( \Omega \otimes_R \Lambda^\circ \)-modules, let \( B \) be a complex of \( \Lambda \otimes_R \Sigma^\circ \)-modules, and let \( C \) be a complex of \( \Omega \otimes_R \Xi^\circ \)-modules. Fix a quasi-isomorphism \( Q \to A \) of complexes of \( \Omega \otimes_R \Lambda^\circ \)-modules with \( Q \) \( q \)-flat over \( \Lambda^\circ \), a quasi-isomorphism \( P \to B \) of complexes of \( \Lambda \otimes_R \Sigma^\circ \)-modules with \( P \) \( q \)-projective over \( \Lambda \), and a quasi-isomorphism \( C \to I \) of complexes of \( \Omega \otimes_R \Xi^\circ \)-modules with \( I \) \( q \)-injective over \( \Omega \). Then the adjunction isomorphism

\[
\text{Hom}_\Omega(A \otimes \Lambda P, I) \to \text{Hom}_\Lambda(P, \text{Hom}_\Omega(A, I))
\]

\[
f \mapsto (p \mapsto (-1)^{\bar{a} \bar{p}}f(a \otimes p))
\]

induces an isomorphism

\[
\text{RHom}_\Omega(A \otimes_k^\Lambda B, C) \to \text{RHom}_\Lambda(B, \text{RHom}_\Omega(A, C))
\]

in \( D(\text{Mod}_{\Sigma \otimes_R \Xi^\circ}) \), as does the adjunction isomorphism

\[
\text{Hom}_\Omega(Q \otimes \Lambda B, I) \to \text{Hom}_\Lambda(B, \text{Hom}_\Omega(Q, I)).
\]
Proof. By Propositions 2.1.2 and 2.1.4, the complex \( \text{Hom}_\Omega(A \otimes \Lambda P, I) \) represents \( R \text{Hom}_\Omega(A \otimes L B, C) \), and the complex \( \text{Hom}_A(P, \text{Hom}_\Omega(A, I)) \) represents \( R \text{Hom}_A(B, R \text{Hom}_\Omega(A, C)) \). Therefore, we are reduced for the first part to showing that the adjunction map is an isomorphism of complexes, and this is standard (see [Lim2, Lemma 2.2]). The second part follows easily from the first part and the commutative diagram

\[
\begin{align*}
\text{Hom}_\Omega(Q \otimes \Lambda B, I) & \xrightarrow{\sim} \text{Hom}_A(B, \text{Hom}_\Omega(Q, I)) \\
\text{Hom}_\Omega(Q \otimes \Lambda P, I) & \xrightarrow{\sim} \text{Hom}_A(P, \text{Hom}_\Omega(Q, I)),
\end{align*}
\]
in that the left-hand vertical map is a quasi-isomorphism.

We consider the next isomorphism first on the level of complexes.

Lemma 2.1.7. Let \( A \) be a bounded complex of \( \Omega \otimes R \Sigma^\circ \)-modules that are flat as \( \Sigma^\circ \)-modules, let \( B \) be a complex of \( \Xi \otimes R \Lambda^\circ \)-modules that are finitely presented as \( \Lambda^\circ \)-modules, and let \( C \) be a complex of \( \Sigma \otimes R \Lambda^\circ \)-modules. Suppose also that either the terms of \( A \) are finitely presented as \( \Sigma^\circ \)-modules or at least one of \( B \) and \( C \) is bounded above and at least one is bounded below. Then the map

\[
A \otimes \Sigma \text{Hom}_{\Lambda^\circ}(B, C) \rightarrow \text{Hom}_{\Lambda^\circ}(A, A \otimes \Sigma C)
\]
\[
a \otimes f \mapsto (b \mapsto a \otimes f(b))
\]
is an isomorphism of complexes of \( \Omega \otimes R \Xi^\circ \)-modules.

Proof. That the stated map is a map of complexes is an easy check of the actions and signs. To see that it is an isomorphism, consider first the case that \( A, B, \) and \( C \) are concentrated in degree 0 and take a presentation

\[
F_1 \rightarrow F_2 \rightarrow B \rightarrow 0
\]

with \( F_1 \) and \( F_2 \) finitely generated free \( \Lambda^\circ \)-modules. Then the two rightmost vertical arrows in the commutative diagram with exact (noting the \( \Sigma^\circ \)-flatness of \( A \)) rows

\[
\begin{align*}
0 & \rightarrow A \otimes \Sigma \text{Hom}_{\Lambda^\circ}(B, C) \rightarrow A \otimes \Sigma \text{Hom}_{\Lambda^\circ}(F_2, C) \rightarrow A \otimes \Sigma \text{Hom}_{\Lambda^\circ}(F_1, C) \\
0 & \rightarrow \text{Hom}_{\Lambda^\circ}(B, A \otimes \Sigma C) \rightarrow \text{Hom}_{\Lambda^\circ}(F_2, A \otimes \Sigma C) \rightarrow \text{Hom}_{\Lambda^\circ}(F_1, A \otimes \Sigma C)
\end{align*}
\]

are isomorphisms of \( \Omega \)-modules by universal property of the direct sum and the commutativity of direct sums and tensor products, so the other is as well. Since the latter map is a morphism of complexes of \( \Omega \otimes R \Xi^\circ \)-modules, we have the result in the case of modules.
In the general setting, we have that

$$\operatorname{Hom}_{\Lambda^\circ}(B, A \otimes \Sigma C) = \prod_{j \in \mathbb{Z}} \operatorname{Hom}_{\Lambda^\circ}(B^j, \bigoplus_{i \in \mathbb{Z}} A^i \otimes \Sigma C^{n-i+j}).$$

Since $B^j$ is finitely presented over $\Lambda^0$, the argument given above implies that the latter group is naturally isomorphic to

$$\prod_{j \in \mathbb{Z}} \bigoplus_{i \in \mathbb{Z}} A^i \otimes \operatorname{Hom}_{\Lambda^\circ}(B^j, C^{n-i+j}).$$

On the other hand, we have

$$(A \otimes \Sigma \operatorname{Hom}_{\Lambda^\circ}(B, C))^n = \bigoplus_{i \in \mathbb{Z}} A^i \otimes \left( \prod_{j \in \mathbb{Z}} \operatorname{Hom}_{\Lambda^\circ}(B^j, C^{n-i+j}) \right).$$

If the terms of $A$ are finitely presented over $\Sigma^\circ$, then we can use a finite presentation of $A^i$ to see that the latter term is naturally isomorphic to

$$\bigoplus_{i \in \mathbb{Z}} \prod_{j \in \mathbb{Z}} A^i \otimes \operatorname{Hom}_{\Lambda^\circ}(B^j, C^{n-i+j}),$$

and the boundedness of $A$ allows us to commute the direct product and direct sum. Finally, if at least one of $B$ and $C$ is bounded above and at least one is bounded below, then the products over $j \in \mathbb{Z}$ involve only finitely many nonzero terms (again using that $A$ is bounded), and the result follows as direct sums commute with tensor products and each other.

In the derived category, we have the following generalization of [Ven, Proposition 6.1].

**Lemma 2.1.8.** Suppose that $\Omega$ is a flat $R$-algebra and $\Xi$ is a projective $R$-algebra. Let $A$ be a bounded above complex of $\Omega \otimes R \Sigma^\circ$-modules, let $B$ be a complex of $\Xi \otimes R \Lambda^\circ$-modules, and let $C$ be a bounded below complex of $\Sigma \otimes R \Lambda^\circ$-modules. Suppose that there exists a quasi-isomorphism $Q \to A$ of $\Omega \otimes R \Sigma^\circ$-modules, where $Q$ is bounded and has terms that are flat as $\Sigma^\circ$-modules. Suppose also that there is a quasi-isomorphism $P \to B$ of complexes of $\Xi \otimes R \Lambda^\circ$-modules, where $P$ is $q$-projective as a complex of $\Lambda^\circ$-modules with terms that are finitely presented over $\Lambda^\circ$. We assume also that either $C$ is bounded, $P$ is bounded above, or the terms of $Q$ are finitely presented as $\Sigma^\circ$-modules. Then the map

$$Q \otimes \Sigma \operatorname{Hom}_{\Lambda^\circ}(P, C) \to \operatorname{Hom}_{\Lambda^\circ}(P, Q \otimes \Sigma C)$$

gives rise to an isomorphism

$$A \otimes \Sigma\mathcal{R}\operatorname{Hom}_{\Lambda^\circ}(B, C) \xrightarrow{\sim} \mathcal{R}\operatorname{Hom}_{\Lambda^\circ}(B, A \otimes \Sigma C)$$

in $D(\text{Mod}_{\Omega \otimes R \Xi^\circ})$. 

Documenta Mathematica 18 (2013) 621–678
Proof. By Propositions 2.1.2 and 2.1.4, the complex \( Q \otimes \Sigma \text{Hom}_\Lambda^-(P, C) \) represents \( A \otimes R \text{Hom}_\Lambda^-(B, C) \) and the complex \( \text{Hom}_\Lambda^-(P, Q \otimes \Sigma) \) represents \( R \text{Hom}_\Lambda^-(B, A \otimes \Sigma) \). That the stated map is an isomorphism follows from Lemma 2.1.7.

Remark 2.1.9. In this article, we need only compute derived tensor products in the case that at least one of the complexes is bounded above and derived homomorphism complexes in the case that the target complex is bounded below. However, in Section 4.5, we will nevertheless be forced to represent derived homomorphism complexes using \( q \)-projective resolutions in an unbounded first variable. Hence, we have given a general treatment.

2.2 Modules over group rings

In this subsection, we preserve the notation of Section 2.1, and in addition, let \( G \) be a group. We begin by extending the notions of derived bifunctors of homorphism groups and tensor products to incorporate an additional action of \( G \).

If \( A \) is a complex of \( \Lambda[G] \)-modules and \( B \) is a complex of \( (\Lambda \otimes R \Sigma^\circ)[G] \)-modules, we give \( \text{Hom}_\Lambda^-(A, B) \) the structure of a complex of \( \Sigma^\circ[G] \)-modules via the standard \( G \)-action

\[
(g \cdot f)(a) = g \cdot f(g^{-1}a)
\]

for \( g \in G \), \( a \in A \), and \( f \in \text{Hom}_\Lambda^-(A, B) \). The following is then a slight weakening of the natural analogue of Proposition 2.1.2, written in condensed form.

Proposition 2.2.1. There is a derived bifunctor

\[
R \text{Hom}_\Lambda^-(\cdot, \cdot) : \text{D}(\text{Mod}_{\Lambda[G]}^\circ) \times \text{D}(\text{Mod}_{\Lambda \otimes R \Sigma^\circ[G]}) \rightarrow \text{D}(\text{Mod}_{\Sigma^\circ[G]}).
\]

Moreover, \( R \text{Hom}_\Lambda^-(A, B) \) can be represented by \( \text{Hom}_\Lambda^-(A, B) \) if \( A \) is \( q \)-projective as a complex of \( \Lambda \)-modules or, if \( \Sigma \) is \( R \)-flat, \( B \) is \( q \)-injective as a complex of \( \Lambda \)-modules.

Proof. Proposition 2.1.2 (with \( \Omega \) replaced by \( R[G]^\circ \) and \( \Sigma \) replaced by \( \Sigma^\circ[G] \)) implies the existence of derived bifunctors with the desired properties that take values in the category \( \text{D}(\text{Mod}_{R[G]^\circ \otimes R \Sigma^\circ[G]}) \). There is a natural exact functor

\[
\text{Mod}_{R[G]^\circ \otimes R \Sigma^\circ[G]} \rightarrow \text{Mod}_{\Sigma^\circ[G]}
\]

taking \( M \) to the same \( \Sigma^\circ \)-module with new \( G \)-action

\[
g \cdot m = (g^{-1} \otimes g) \cdot m
\]

for \( g \in G \) and \( m \in M \). This in turn induces a functor on derived categories, and composition of the above derived functor with this functor yields the result.
Next, if \( A \) is a complex of \((\Omega \otimes_R \Lambda^\circ)[G]\)-modules and \( B \) is a complex of \((\Lambda \otimes_R \Sigma^\circ)[G]\)-modules, we give \( A \otimes_A B \) the structure of a complex of \((\Omega \otimes_R \Sigma^\circ)[G]\)-modules via the \( G \)-action defined by \( g(a \otimes b) = ga \otimes gb \) for \( a \in A, b \in B \), and \( g \in G \). We state a very slight weakening (for brevity) of the analogue of Proposition 2.1.4.

**Proposition 2.2.2.** If \( \Omega \) or \( \Sigma \) is a flat \( R \)-algebra, then we have a derived bifunctor

\[
- \otimes^L_A : \text{D}(\text{Mod}_{\Omega \otimes_R \Lambda^\circ}[G]) \times \text{D}(\text{Mod}_{\Lambda \otimes_R \Sigma^\circ}[G]) \to \text{D}(\text{Mod}_{\Omega \otimes_R \Lambda^\circ}[G]).
\]

Moreover, \( A \otimes_A B \) represents \( \Lambda \otimes^L_A \mathbf{B} \) if either \( \Omega \) is \( R \)-flat and \( A \) is a \( q \)-flat as a complex of \( \Lambda^\circ \)-modules or \( \Sigma \) is \( R \)-flat and \( B \) is \( q \)-flat as a complex of \( \Lambda \)-modules.

**Proof.** Proposition 2.1.4 with \( \Omega \) replaced by \( \Omega[G] \) and \( \Sigma^\circ \) replaced by \( \Sigma^\circ[G] \) yields the existence of a derived bifunctor

\[
- \otimes^L_A : \text{D}(\text{Mod}_{\Omega \otimes_R \Lambda^\circ}[G]) \times \text{D}(\text{Mod}_{\Lambda \otimes_R \Sigma^\circ}[G]) \to \text{D}(\text{Mod}_{\Omega[G] \otimes_R \Sigma^\circ}[G])
\]

which has the desired properties. There is a natural exact functor

\[
\text{Mod}_{\Omega[G] \otimes_R \Sigma^\circ}[G] \to \text{Mod}_{(\Omega \otimes_R \Sigma^\circ)[G]}
\]

that takes an module \( M \) to the same \( \Omega \otimes_R \Sigma^\circ \)-module with new \( G \)-action

\[
g \cdot m = (g \otimes g) \cdot m
\]

for \( g \in G \) and \( m \in M \), and the desired bifunctor is the resulting composition of derived functors.

Suppose now that \( \Lambda \) is an \( R \)-algebra and that we are given a homomorphism \( \chi : G \to \text{Aut}_{\Lambda^\circ}(\Lambda) \), allowing us to endow \( \Lambda \) with the structure of a \( \Lambda^\circ[G] \)-module. We denote the resulting module by \( \chi \Lambda \). (If \( \chi \) is trivial, then we continue to write \( \Lambda \) for \( \chi \Lambda \).) We also have a map \( \chi^{-1} : G \to \text{Aut}_\Lambda(\Lambda) \) defined by \( \chi^{-1}(g)(\lambda) = \lambda \cdot \chi(g^{-1})(1) \) for \( g \in G \) and \( \lambda \in \Lambda \), so the resulting object \( \Lambda \chi \) is a \( \Lambda[G] \)-module. The relationship between \( \chi \Lambda \) and \( \Lambda \chi \) can be expressed by the evaluation-at-1 maps

\[
\text{Hom}_\Lambda(\chi \Lambda, \Lambda) \to \chi \Lambda \text{ and } \text{Hom}_{\Lambda^\circ}(\chi \Lambda, \Lambda) \to \Lambda \chi,
\]

which are isomorphisms of \( \Lambda^\circ[G] \) and \( \Lambda[G] \)-modules, respectively.

**Lemma 2.2.3.** Suppose that \( \Lambda \) is a flat \( R \)-algebra, and let \( \chi : G \to \text{Aut}_{\Lambda^\circ}(\Lambda) \) be a homomorphism. For any bounded above complex \( A \) of \( R[G] \)-modules that are finitely presented over \( R \) and any bounded below complex \( B \) of \( R \)-modules, the two maps

\[
\theta : \chi \Lambda \otimes_R \text{Hom}_R(A, B) \to \text{Hom}_\Lambda(\chi \Lambda \otimes_R A, \Lambda \otimes_R B)
\]

\[
\theta' : \Lambda \otimes_R \text{Hom}_R(A, B) \to \text{Hom}_{\Lambda^\circ}(\chi \Lambda \otimes_R A, \Lambda \otimes_R B)
\]

are isomorphisms.

---

**Meng Fai Lim and Romyar T. Sharifi**

**Documenta Mathematica 18 (2013) 621–678**
defined by \( \theta(\lambda \otimes f)(\mu \otimes x) = \mu \lambda \otimes f(x) \) and \( \theta'(\lambda \otimes f)(\mu \otimes x) = \lambda \mu \otimes f(x) \) are isomorphisms of complexes of \( \Lambda^\circ[G] \)-modules and of complexes of \( \Lambda[G] \)-modules, respectively.

Proof. We focus on the case of \( \theta \), as the other case follows from it. Since \( \chi \Lambda \) is \( R \)-flat, Lemma \[2.1.7\] (with \( \Omega = \Lambda^\circ[G] \) and \( \Xi = R[G] \)) implies that the natural map

\[ \chi \Lambda \otimes_R \text{Hom}_R(A, B) \xrightarrow{\sim} \text{Hom}_R(A, \chi \Lambda \otimes_R B) \]

is an isomorphism of complexes of \( \Lambda^\circ[G] \otimes_R R[G] \)-modules if we take the actions of \( G \) independently, and hence of \( \Lambda^\circ[G] \)-modules if we take the \( G \)-actions prescribed earlier in this subsection. The inverse of the isomorphism provided by the evaluation map provides the first of the isomorphisms of complexes of \( \Lambda^\circ[G] \)-modules

\[ \chi \Lambda \otimes_R B \xrightarrow{\sim} \text{Hom}_\Lambda(A, \Lambda \otimes_R B) \xrightarrow{\sim} \text{Hom}_\Lambda(A, \chi \Lambda \otimes_R B), \]

the canonical second map being an isomorphism as \( \Lambda \chi \) is free of rank 1 over \( \Lambda \). We therefore have an isomorphism

\[ \text{Hom}_R(A, \chi \Lambda \otimes_R B) \xrightarrow{\sim} \text{Hom}_R(A, \text{Hom}_\Lambda(A, \chi \Lambda \otimes_R B)). \]

Finally, the inverse of the adjoint morphism is an isomorphism

\[ \text{Hom}_R(A, \text{Hom}_\Lambda(A, \Lambda \otimes_R B)) \xrightarrow{\sim} \text{Hom}_\Lambda(A \otimes_R \Lambda, \Lambda \otimes_R B) \]

of complexes of \( \Lambda^\circ[G] \)-modules, and the resulting composite of three isomorphisms is easily computed to be \( \theta \).

2.3 Projective modules over a profinite ring

Let \( \Lambda \) be a profinite ring, and fix a directed fundamental system \( I \) of open neighborhoods of zero consisting of two-sided ideals of \( \Lambda \). We say that a topological \( \Lambda \)-module \( M \) is endowed with the \( I \)-adic topology if the collection \( \{\mathcal{M}\}_{\mathcal{M} \in I} \) forms a fundamental system of neighborhoods of zero. It was shown in \[Lim2, \text{Section 3.1}\] that any finitely generated compact (Hausdorff) \( \Lambda \)-module necessarily has the \( I \)-adic topology, and, moreover, any homomorphism between such modules is necessarily continuous.

In this subsection, we recall several facts about projective \( \Lambda \)-modules that will be of use to us. We denote the abelian category of compact \( \Lambda \)-modules by \( \mathcal{C}_\Lambda \). The free profinite \( \Lambda \)-module on a set \( X \) is canonically isomorphic to the topological direct product of one copy of \( \Lambda \) for each element of \( X \) \[Wil\ Proposition 7.4.1\], and a profinite \( \Lambda \)-module \( P \) is projective if and only if it is continuously isomorphic to a direct summand of the free profinite module on a set of generators of \( P \) \[Wil\ Proposition 7.4.7\]. In particular, the category \( \mathcal{C}_\Lambda \) has enough projectives. Any projective object in \( \mathcal{C}_\Lambda \) that is finitely generated over \( \Lambda \) is a projective \( \Lambda \)-module, and conversely, any finitely generated projective
Λ-module endowed with the \( \mathcal{I} \)-adic topology is a projective object in \( C_{\Lambda} \) [Lim2 Proposition 3.1.8].

Recall that the projective dimension of an abstract Λ-module \( M \) is the minimum integer \( n \) (if it exists) such that there is a resolution of \( M \) by projective Λ-modules

\[
0 \rightarrow P^{-n} \rightarrow \cdots \rightarrow P^{-1} \rightarrow P^0 \rightarrow M \rightarrow 0.
\]

The topological projective dimension of a compact Λ-module is defined similarly, replacing projective Λ-modules by projective objects in \( C_{\Lambda} \).

For the remainder of the subsection, we suppose that Λ is left Noetherian. Note that any projective resolution \( P \) of a finitely generated Λ-module \( M \) is quasi-isomorphic to a projective resolution \( Q \) of \( M \) by finitely generated Λ-modules via a map \( Q \rightarrow P \) compatible with the augmentations to \( M \). In particular, the projective dimension of \( M \) is the length of its shortest resolution by projectives in the category of finitely generated Λ-modules. It follows that the notions of projective dimension and topological projective dimension coincide on finitely generated (compact) Λ-modules.

In general, if \( \mathcal{C} \) is a category with objects that are Λ-modules, we will use \( \mathcal{C}_{\Lambda}^{\text{ft}} \) to denote the full subcategory of objects that are finitely generated (i.e., of finite type) over Λ. Since \( \text{Mod}_{\Lambda}^{\text{ft}} \) has enough projectives, the equivalent category \( \mathcal{C}_{\Lambda}^{\text{ft}} \) also has enough projectives. Denote by \( \mathbf{D}_{\Lambda}^{\text{ft}}(\text{Mod}_{\Lambda}) \) the full subcategory of the bounded above derived category \( \mathbf{D}^-(\text{Mod}_{\Lambda}) \) which has as its objects those bounded above complexes \( X \) of Λ-modules for which all of the \( H^i(X) \) are finitely generated Λ-modules. The following standard lemma tells us that any such complex is quasi-isomorphic to a complex of finitely generated modules. (See [Nek] Proposition 3.2.6 for an analogous statement, which has a similar proof.)

**Lemma 2.3.1.** Let \( \Omega \) be a left Noetherian ring. Every bounded above complex \( X \) of \( \Omega \)-modules for which every \( H^i(X) \) is a finitely generated \( \Omega \)-module has a quasi-isomorphic subcomplex of finitely generated \( \Omega \)-modules.

In sum, we have equivalences of categories

\[
\mathbf{D}^-(\mathcal{C}_{\Lambda}^{\text{ft}}) \cong \mathbf{D}^-(\text{Mod}_{\Lambda}^{\text{ft}}) \cong \mathbf{D}_{\Lambda}^{\text{ft}}(\text{Mod}_{\Lambda}),
\]

the first being induced by the forgetful functor and the second by the inclusion of categories \( \text{Mod}_{\Lambda}^{\text{ft}} \hookrightarrow \text{Mod}_{\Lambda} \). We use these equivalences to identify these categories with each other.

### 2.4 Completed tensor products

In this subsection, we review some basic facts about completed tensor products and briefly study their derived functors. Let \( R \) be a commutative profinite ring, and let \( \Lambda, \Omega, \) and \( \Sigma \) be profinite \( R \)-algebras, by which we shall mean that they are profinite and the maps from \( R \) to their centers are continuous.
Let $\mathcal{I}$ (resp., $\mathcal{J}$) denote a directed fundamental system of open neighborhoods of zero consisting of two-sided ideals of $\Lambda$ (resp., $\Omega$). We then define a completed tensor product algebra by
\[
\Omega \hat{\otimes}_R \Lambda^\circ = \lim_{\leftarrow A \in \mathcal{I}, B \in \mathcal{J}} \Omega / \mathfrak{A} \otimes_R (\Lambda / \mathfrak{A})^\circ.
\]
This is clearly a profinite $R$-algebra.

We shall denote the category of compact $\Omega \hat{\otimes}_R \Lambda^\circ$-modules by $\mathcal{C}_{\Omega^\circ \Lambda^\circ}$. Let $M$ be an object of $\mathcal{C}_{\Omega^\circ \Lambda^\circ}$, and let $N$ be an object of $\mathcal{C}_{\Lambda^\circ \Sigma^\circ}$. By [RZ, Lemma 5.1.1(a)], we have that the set of open, finite index $\Omega \hat{\otimes}_R \Lambda^\circ$-submodules (resp., $\Lambda \hat{\otimes}_R \Sigma^\circ$-submodules) of $M$ (resp., $N$) forms a basis of neighborhoods of zero on $M$ (resp., $N$). We define the completed tensor product to be the compact $\Omega \hat{\otimes}_R \Sigma^\circ$-module
\[
M \hat{\otimes}_\Lambda N = \lim_{\leftarrow M', N'} M / M' \otimes_\Lambda N / N',
\]
where $M'$ and $N'$ run through the respective bases for $M$ and $N$, with the topology defined by the inverse limit. The completed tensor product is associative and commutative (i.e., $M \hat{\otimes}_\Lambda N \cong N \hat{\otimes}_\Lambda M$) in the same sense as the usual tensor product.

Note that the canonical $\Lambda$-balanced map
\[
t: M \times N \to M \hat{\otimes}_\Lambda N
\]
induces a homomorphism
\[
M \otimes_\Lambda N \to M \hat{\otimes}_\Lambda N
\]
of $\Omega \hat{\otimes}_R \Sigma^\circ$-modules that has dense image. The completed tensor product of $M$ and $N$ then satisfies the following universal property (see [RZ, Section 5.5] in the case that $\Omega = \Sigma = R$).

**Lemma 2.4.1.** For any compact $\Omega \hat{\otimes}_\Lambda \Sigma^\circ$-module $L$ and any continuous, $\Lambda$-balanced, left $\Omega$-linear and right $\Sigma$-linear map $f: M \times N \to L$, there is a unique continuous map $\hat{f}: M \hat{\otimes}_\Lambda N \to L$ of $\Omega \hat{\otimes}_R \Sigma^\circ$-modules such that $\hat{f} \circ t = f$.

It follows that, in defining the completed tensor product, it suffices to run through a basis of neighborhoods of zero consisting of open $\Omega \hat{\otimes}_R \Lambda^\circ$-submodules of $M$ and a basis of neighborhoods of zero consisting of open $\Lambda \hat{\otimes}_R \Sigma^\circ$-submodules of $N$.

The following is also standard.

**Lemma 2.4.2.** Let $M$ and $N$ be objects of $\mathcal{C}_{\Omega^\circ \Lambda^\circ}$ and $\mathcal{C}_{\Lambda^\circ \Sigma^\circ}$, respectively.

(a) Suppose that $M = \lim_{\alpha} M_\alpha$ and $N = \lim_{\beta} N_\beta$, where each $M_\alpha$ (resp., $N_\beta$) is a compact $\Omega \hat{\otimes}_R \Lambda^\circ$-module (resp., compact $\Lambda \hat{\otimes}_R \Sigma^\circ$-module). Then there is an isomorphism
\[
M \hat{\otimes}_\Lambda N \cong \lim_{\alpha, \beta} M_\alpha \hat{\otimes}_\Lambda N_\beta
\]
of compact $\Omega \hat{\otimes}_R \Sigma^\circ$-modules.
Meng Fai Lim and Romyar T. Sharifi

The canonical map $M \otimes \Lambda N \to M \hat{\otimes} \Lambda N$ is an isomorphism if either $M$ is finitely generated as a $\Lambda^\circ$-module or $N$ is finitely generated as a $\Lambda$-module.

The functor $M \hat{\otimes} \Lambda - : C_{\Lambda^\circ} \to C_{\Omega^\circ}$ is right exact.

The next lemma describes connections between projective objects in different categories of compact modules and flat objects in categories of abstract modules. A ring is said to be left (resp., right) coherent if all its finitely generated left (resp., right) ideals are finitely presented.

**Lemma 2.4.3.**

(a) If $\Lambda$ is left (resp., right) coherent, then every projective object in $C_{\Lambda^\circ}$ (resp., $C_\Lambda$) is flat with respect to the usual left (resp., right) tensor product over $\Lambda$.

(b) If $\Omega$ (resp., $\Lambda$) is a projective object of $C_R$, then every projective object in $C_{\Omega^\circ}$ is a projective object in $C_{\Lambda^\circ}$ (resp., $C_\Omega$).

**Proof.** (a) Since every projective object in $C_{\Lambda^\circ}$ is continuously isomorphic to a direct summand of a direct product of copies of $\Lambda^\circ$, it suffices to show that any direct product of copies of $\Lambda$ is flat as an abstract $\Lambda^\circ$-module. By a theorem of Chase (cf. [Lam, Theorem 4.47]), this is equivalent to the fact that $\Lambda$ is left coherent.

(b) Suppose that $\Omega$ is projective in $C_R$, so $\Omega$ is continuously isomorphic to a direct summand of a direct product of copies of $R$. Lemma 2.4.2(a) then tells us that $\Omega \hat{\otimes} R \Lambda^\circ$ is topologically a direct summand of a direct product of copies of $\Lambda^\circ$, hence is projective.

Let $G$ be a profinite group. We use $C_{\Lambda,G}$ to denote the category of compact $\Lambda$-modules with a continuous commuting action of $G$, the morphisms being continuous homomorphisms of $\Lambda[G]$-modules. Note the following.

**Remark 2.4.4.** The category $C_{\Lambda,G}$ is equivalent to the category $C_{\Lambda[G]}$, where $\Lambda[G]$ is given the profinite topology defined by $\Lambda[G] \cong \lim_{\Delta \in \mathcal{I}} \lim_{N} (\Lambda/N)[G/N]$, where $N$ runs over the open normal subgroups of $G$.

Consequently, $C_{\Lambda,G}$ is an abelian category with enough projectives, and every element of $C_{\Lambda,G}$ is an inverse limit of finite $\Lambda[G]$-quotients (see also [Lim2, Section 3.2]). Moreover, $\Lambda[G]$ is a projective object of $C_\Lambda$ by [RZ, Lemma 1.4.1].
5.3.5(d)], so the forgetful functor $\mathcal{C}_{\Lambda,G} \to \mathcal{C}_{\Lambda}$ takes projective objects to projective objects. To shorten notation, we use $\mathcal{C}_{\Lambda-\Sigma,G}$ to denote the category $\mathcal{C}_{\Lambda \otimes R \Lambda^\circ, G}$.

We remark that if $M$ and $N$ are, respectively, objects of $\mathcal{C}_{\Omega-\Lambda,G}$ and $\mathcal{C}_{\Lambda-\Sigma,G}$, possibly with trivial $G$-actions, we may give $\hat{M} \otimes \Lambda N$ the structure of an object of $\mathcal{C}_{\Omega-\Sigma,G}$ via the diagonal action of $G$. That is, the $G$-action is defined by choosing bases of open $(\Omega \otimes R \Lambda^\circ)[G]$-submodules of $M$ (resp., $(\Lambda \otimes R \Sigma^\circ)[G]$-submodules of $N$) and taking the inverse limits of the tensor products of the finite quotients.

**Remark 2.4.5.** The analogous result to Lemma 2.4.2 holds, as a consequence of said lemma, if we take $M$ and $N$ to be objects of $\mathcal{C}_{\Omega-\Lambda,G}$ and $\mathcal{C}_{\Lambda-\Sigma,G}$, respectively. (That is, in part (a), one must take $M_\alpha$ and $N_\beta$ to be objects of these categories to attain an isomorphism in $\mathcal{C}_{\Omega-\Sigma,G}$, and in part (c), the functor is now a functor from $\mathcal{C}_{\Lambda-\Sigma,G}$ to $\mathcal{C}_{\Omega-\Sigma,G}$.)

Note that we may form completed tensor products of bounded above complexes, or of a bounded complex with any complex, as with the usual tensor products.

**Proposition 2.4.6.** If $\Omega$ is projective in $\mathcal{C}_R$, then the completed tensor product induces the following derived bifunctor

$$- \hat{\otimes}^L_{\Lambda} : \mathcal{D}^{-}(\mathcal{C}_{\Omega-\Lambda,G}) \times \mathcal{D}^{-}(\mathcal{C}_{\Lambda-\Sigma,G}) \to \mathcal{D}^{-}(\mathcal{C}_{\Omega-\Sigma,G}),$$

where $A \hat{\otimes} \Lambda B$ represents $A \hat{\otimes}^L_{\Lambda} B$ if the terms of $A$ are projective as objects in $\mathcal{C}_{\Lambda^\circ}$. Furthermore, there is a commutative diagram

$$\begin{array}{c}
\mathcal{D}^{-}(\mathcal{C}_{\Omega-\Lambda,G}) \times \mathcal{D}^{-}(\mathcal{C}_{\Lambda-\Sigma,G}) \\
\downarrow \\
\mathcal{D}^{-}(\mathcal{C}_{\Omega-\Sigma,G}) \end{array} \quad \rightarrow \quad \begin{array}{c}
\mathcal{D}^{-}(\mathcal{C}_{\Lambda-\Sigma,G}) \\
\downarrow \\
\mathcal{D}^{-}(\mathcal{C}_{\Omega-\Sigma,G}) \end{array}$$

in which the vertical arrows are induced by forgetful functors, and if $\Lambda$ is left Noetherian, there is a commutative diagram

$$\begin{array}{c}
\mathcal{D}^{-}(\mathcal{C}_{\Omega-\Lambda,G}) \times \mathcal{D}^{-}(\mathcal{C}_{\Lambda-\Sigma,G}) \\
\downarrow \\
\mathcal{D}^{-}(\mathcal{C}_{\Omega-\Sigma,G}) \end{array} \quad \rightarrow \quad \begin{array}{c}
\mathcal{D}^{-}(\mathcal{C}_{\Lambda-\Sigma,G}) \\
\downarrow \\
\mathcal{D}^{-}(\mathcal{C}_{\Omega-\Sigma,G}) \end{array}$$

in which the vertical arrows are induced by forgetful functors and embeddings of categories.
Proof. The proof of the first part and the commutativity of the first set of diagrams follow by similar arguments to those of Propositions 2.1.2 and 2.2.2, making use of Lemma 2.4.3(b) (using bounded above complexes of projective objects in place of \( q \)-projective complexes). The commutativity of the final diagram then follows from Lemmas 2.4.2(b) and 2.4.3.

Let \( \text{Ch}^b_{\Lambda^+} \) denote the category of bounded complexes in \( \text{C}_{\Omega^-\Lambda,G} \) that are quasi-isomorphic to bounded complexes in \( \text{C}_{\Omega^-\Lambda,G} \) of objects that are projective in \( \text{C}_{\Lambda^+} \), and let \( \text{D}^b_{\Lambda^+} \) denote its derived category.

**Proposition 2.4.7.** If \( \Omega \) is projective in \( \text{C}_R \), then we have a derived bifunctor

\[
- \hat{\otimes}_\Lambda : \text{D}^b_{\Lambda^+} \times \text{D}(\text{C}_{\Lambda^-\Sigma,G}) \to \text{D}(\text{C}_{\Omega^-\Sigma,G}),
\]

where \( A \hat{\otimes}_\Lambda B \) represents \( A \otimes^L \Lambda B \) if \( A \) is a bounded complex with terms that are projectives in \( \text{C}_{\Lambda^+} \).

Proof. Let \( B \) be a complex of objects of \( \text{C}_{\Omega^-\Lambda,G} \). Since every object of \( \text{Ch}^b_{\Lambda^+} \) is by definition quasi-isomorphic to a bounded complex of objects that are acyclic for the functor \( - \hat{\otimes}_\Lambda B \), we have a derived functor \( - \hat{\otimes}_\Lambda B \). Let \( A \) be a bounded complex in \( \text{C}_{\Omega^-\Lambda,G} \) of projectives in \( \text{C}_{\Lambda^+} \) and \( f : B \to B' \) be a quasi-isomorphism. Then \( f \) induces isomorphisms between the \( E_2 \)-terms of the convergent spectral sequence

\[
E_2^{i,j}(B) = H^i(A \hat{\otimes}_\Lambda B^j) \Rightarrow H^{i+j}(A \hat{\otimes}_\Lambda B)
\]

and its analogue for \( B' \), and therefore it induces isomorphisms on the abutments.

2.5 Ind-admissible modules

The notion of an ind-admissible \( R[G] \)-module was introduced in [Nek, Section 3.3] for a complete commutative Noetherian local ring \( R \) with finite residue field and a profinite group \( G \). An \( R[G] \)-module is ind-admissible if it can be written as a union of \( R[G] \)-submodules that are finitely generated over \( R \) and on which \( G \) acts continuously with respect to the topology defined by the maximal ideal of \( R \). In this section, we discuss an analogous construction of ind-admissible modules over noncommutative profinite rings.

As in Nekovar’s treatment, we do not consider the seemingly delicate issue of placing topologies on ind-admissible modules, as it proves unnecessary. In particular, it is still possible to define the continuous cochain complex of an ind-admissible module as a direct limit.

We maintain the notation of Section 2.4. Moreover, we suppose that \( \Omega \) is left Noetherian and \( \Lambda \) is right Noetherian. For an \( (\Omega \otimes_R \Lambda^\omega)[G] \)-module \( M \), we denote by \( S(M) \) the set of \( (\Omega \otimes_R \Lambda^\omega)[G] \)-submodules of \( M \) that are finitely generated as \( \Lambda^\omega \)-modules and on which \( \Omega \) and \( G \) act continuously with respect to the \( I \)-adic topology. The following is a straightforward generalization of [Nek, Lemma 3.3.2].
Lemma 2.5.1. Let $M$ be an $(\Omega \otimes_R \Lambda^\circ)[G]$-module.

(a) If $M' \in S(M)$, then $N \in S(M)$ for every $(\Omega \otimes_R \Lambda^\circ)[G]$-submodule $N$ of $M'$.

(b) If $f : M \to N$ is a homomorphism of $(\Omega \otimes_R \Lambda^\circ)[G]$-modules and $M' \in S(M)$, then $f(M') \in S(N)$.

(c) If $M', M'' \in S(M)$, then $M' + M'' \in S(M)$.

Proof. For part (a), Corollary 3.1.6 and Proposition 3.1.7 of [Lim2] imply that the subspace topology on $N$ from the $I$-adic topology on $M'$ agrees with the $I$-adic topology on $N$, which implies that $\Omega$ and $G$ act continuously on $N$. In (b), the continuity of the $\Omega$ and $G$-actions on the finitely generated $\Lambda^\circ$-module $f(M')$ is a consequence of the fact that the map $M' \to f(M')$ is a continuous quotient map with respect to the $I$-adic topology (as follows from Corollary 3.1.5 and Proposition 3.1.7 of [Lim2]). Part (c) follows from (b), using the addition map $M \times M \to M$.

Note that Lemma 2.5.1(c) implies that $S(M)$ is a directed set with respect to inclusion. We say that an $(\Omega \otimes_R \Lambda^\circ)[G]$-module $M$ is (right) ind-admissible if

$$M = \bigcup_{N \in S(M)} N.$$  

We list some basic properties of the full subcategory $\mathcal{F}_{\Omega-\Lambda,G}$ of Mod$_{(\Omega \otimes_R \Lambda^\circ)[G]}$ with objects the ind-admissible $(\Omega \otimes_R \Lambda^\circ)[G]$-modules.

Lemma 2.5.2.

(a) The category $\mathcal{F}_{\Omega-\Lambda,G}$ is abelian and stable under subobjects, quotients and colimits.

(b) The embedding functor

$$i : \mathcal{F}_{\Omega-\Lambda,G} \hookrightarrow \text{Mod}_{(\Omega \otimes_R \Lambda^\circ)[G]}$$

is exact and is left adjoint to the functor

$$j : \text{Mod}_{(\Omega \otimes_R \Lambda^\circ)[G]} \to \mathcal{F}_{\Omega-\Lambda,G}$$

that takes a module $M$ to the union of the elements of $S(M)$.

(c) The category $\mathcal{F}_{\Omega-\Lambda,G}$ has enough injectives.

(d) Let $M$ be an ind-admissible $(\Omega \otimes_R \Lambda^\circ)[G]$-module, and let $N$ be a finitely generated $\Lambda^\circ$-submodule of $M$. Then $(\Omega \otimes_R \Lambda^\circ)[G] \cdot N$ is an ind-admissible $(\Omega \otimes_R \Lambda^\circ)[G]$-module which is a finitely generated $\Lambda^\circ$-module.

Documenta Mathematica 18 (2013) 621–678
(e) Let \( M \) be an \((\Omega \otimes_R \Lambda^\circ)[G]\)-module. Then \( M \in S(M) \) if and only if \( M \) is an ind-admissible \((\Omega \otimes_R \Lambda^\circ)[G]\)-module which is finitely generated as a \( \Lambda^\circ \)-module.

Proof. For parts (a), (b) and (c), similar arguments to those of [Nek] Proposition 3.3.5] apply. The “only if” direction of (e) is obvious. For (d), since \( N \) is \( \Lambda^\circ \)-finitely generated, we can find a finite subset \( \{M_1, \ldots, M_n\} \) of \( S(M) \) such that

\[
N \subseteq M_1 + \cdots + M_n.
\]

The assertion then follows from Lemma 2.5.3(c) and the “only if” direction of (e). The “if” direction of (e) follows from (d), since \( M = (\Omega \otimes_R \Lambda^\circ)[G] \cdot M \).

If \( G \) is trivial, we write \( \mathcal{I}_{\Omega-\Lambda} \) for \( \mathcal{I}_{\Omega-\Lambda,G} \). We leave to the reader the proof of the following.

**Lemma 2.5.3.**

(a) Let \( M \) be an \((\Omega \otimes_R \Lambda^\circ)[G]\)-module. If \( N \) is an \((\Omega \otimes_R \Lambda^\circ)[G]\)-submodule of \( M \), then \( N \in S(M) \) if and only if \( N \) is an object of \( C^\Lambda_{\Omega-\Lambda,G} \).

(b) The category \( \mathcal{I}_{\Omega-\Lambda,G} \) is equivalent to the ind-category of \( C^\Lambda_{\Omega-\Lambda,G} \).

(c) The category \( \mathcal{I}_{\Omega-\Lambda} \) is equivalent to the full subcategory of \( \mathcal{I}_{\Omega-\Lambda,G} \) with objects the modules on which \( G \) acts trivially.

**Remark 2.5.4.** In the case that \( \Omega = R \), we denote \( \mathcal{I}_{R-\Lambda,G} \) by \( \mathcal{I}_{\Lambda,G} \), and the latter category is equivalent to the ind-category of \( C^\Lambda_{\Lambda,G} \). The subcategory \( \mathcal{I}_{\Lambda^\circ} \) is simply Mod_{\Lambda^\circ}.

**Lemma 2.5.5.** Let \( A \) be an object of \( C^\Lambda_{\Omega,G} \), and let \( B \) be an object of \( C^\Lambda_{\Omega-\Lambda,G} \). Then \( \hom_{\Omega}(A, B) \) with the \( I \)-adic topology is an object of \( C^\Lambda_{\Lambda,G} \).

Proof. Since \( A \) is finitely generated over \( \Omega \) and \( B \) is a compact \( \Omega \)-module, we have by [Lim2] Lemma 3.1.4[3] that \( \hom_{\Omega}(A, B) = \hom_{\Omega,cts}(A, B) \), where the latter group is the group of continuous homomorphisms of \( \Omega \)-modules. Note that \( \Lambda^\circ \) acts continuously on \( \hom_{\Omega}(A, B) \) with respect to the compact-open topology by [Flo] Proposition 3[3], and similarly \( G \) acts continuously on it as a consequence of [Flo] Lemma 2. We have a continuous isomorphism of \( \Lambda^\circ[G]\)-modules

\[
\hom_{\Omega}(A, B) \xrightarrow{\sim} \varprojlim_{\beta} \hom_{\Omega}(A, B_{\beta}),
\]

where \( B_{\beta} \) runs over the finite Hausdorff \((\Omega \otimes_R \Lambda^\circ)[G]\)-quotients of \( B \). As \( B_{\beta} \) is finite and therefore discrete, the group \( \hom_{\Omega}(A, B_{\beta}) \) is finite and discrete as well, so \( \hom_{\Omega}(A, B) \) is compact.

Note that \( \hom_{\Omega}(A, B) \) injects into \( \hom_{\Omega}(\Omega^r, B) \cong B^r \) for some \( r \geq 0 \), since \( A \) is \( \Omega \)-finitely generated. We therefore have that \( \hom_{\Omega}(A, B) \) is \( \Lambda^\circ \)-finitely generated. Finally, [Lim2] Proposition 3.1.7 then implies that the compact-open topology on \( \hom_{\Omega}(A, B) \) agrees with the \( I \)-adic topology.
We define
\[ \text{Hom}_{\Omega,\text{cts}}(-, -): C_{\Omega,G} \times \mathcal{F}_{\Omega-\Lambda,G} \rightarrow \mathcal{F}_{\Lambda^*,G}. \]
by
\[ \text{Hom}_{\Omega,\text{cts}}(A, B) = \lim_{\alpha} \lim_{\beta} \text{Hom}_{\Omega}(A_\alpha, B_\beta), \]
upon making the identifications
\[ A_\alpha = \lim_{\leftarrow} A_\alpha, \]
where \( A_\alpha \) runs over the quotients of \( A \) in \( C_{\Omega,G}^{-ft} \), and
\[ B_\beta = \lim_{\rightarrow} B_\beta, \]
where \( B_\beta \) runs over the elements of \( S(B) \), i.e.,
by Lemma 2.5.3 the subobjects of \( B \) that lie in \( C_{\Omega^{-ft},G}^{\Lambda^*} \). Note that it actually suffices to let \( A_\alpha \) and \( B_\beta \) run over cofinal subsets.

**Remark 2.5.6.** We note that
\[ \text{Hom}_{\Omega,\text{cts}}(A, B) \cong \lim_{\beta} \text{Hom}_{\Omega,\text{cts}}(A, B_\beta), \]
where \( \text{Hom}_{\Omega,\text{cts}}(A, B_\beta) \) here is the \((\Omega \otimes R^{\Lambda^*})[G]\)-module of continuous \( \Omega \)-module homomorphisms in the usual sense.

We also note the following.

**Lemma 2.5.7.** Let \( A \) be an object of \( C_{\Omega,G}^{-ft} \), and let \( B \) be an object of \( \mathcal{F}_{\Omega-\Lambda,G} \). Then we have
\[ \text{Hom}_{\Omega,\text{cts}}(A, B) = \text{Hom}_{\Omega}(A, B). \]

**Proof.** This follows from the computation
\[ \text{Hom}_{\Omega,\text{cts}}(A, B) \cong \lim_{\beta} \text{Hom}_{\Omega,\text{cts}}(A, B_\beta) = \lim_{\beta} \text{Hom}_{\Omega}(A, B_\beta) \cong \text{Hom}_{\Omega}(A, B), \]
the latter isomorphism following from the case in which \( A \) is free of finite rank, since \( A \) is finitely presented over \( \Omega \).

As usual, we may extend our definition to consider the complex of continuous homomorphisms from a complex of compact modules to a complex of ind-admissible modules, supposing that at least one of these complexes is bounded above and at least one is bounded below.

**Proposition 2.5.8.** There is a derived bifunctor
\[ \mathbf{R}\text{Hom}_{\Omega,\text{cts}}(-, -): D^- (C_{\Omega,G})^\circ \times D^+ (\mathcal{F}_{\Omega-\Lambda,G}) \rightarrow D^+ (\mathcal{F}_{\Lambda^*,G}), \]
and \( \mathbf{R}\text{Hom}_{\Omega,\text{cts}}(A, B) \) can be represented by \( \text{Hom}_{\Omega,\text{cts}}(A, B) \) if the terms of \( A \) are projective as objects of \( C_{\Omega} \). Moreover, we have a commutative diagram
\[
\begin{array}{ccc}
D^-(C_{\Omega,G})^\circ \times D^+ (\mathcal{F}_{\Omega-\Lambda,G}) & \xrightarrow{\mathbf{R}\text{Hom}_{\Omega,\text{cts}}(-, -)} & D^+ (\mathcal{F}_{\Lambda^*,G}) \\
\downarrow & & \downarrow \\
D^-(C_{\Omega,G}^{-ft})^\circ \times D^+ (\mathcal{F}_{\Omega-\Lambda,G}) & \xrightarrow{\mathbf{R}\text{Hom}_{\Omega}(-, -)} & D^+ (\text{Mod}_{\Lambda^*[G]}) \\
\downarrow & & \downarrow \\
D^-(\text{Mod}_{\Omega[G]})^\circ \times D^+ (\text{Mod}_{(\Omega \otimes R^{\Lambda^*})[G]}) & \xrightarrow{\mathbf{R}\text{Hom}_{\Omega}(-, -)} & D^+ (\text{Mod}_{\Lambda^*[G]})
\end{array}
\]
Proof. The existence of the derived bifunctor in question follows from an argument similar to that of Proposition 2.1.2, and the second statement is a consequence of the fact that a projective object in $C_{\Omega,G}$ is a projective object of $C_{\Omega}$, noting Remark 2.5.6.

Let $A$ be a bounded above complex of objects of $C_{\Omega-G}^{\Omega-ft}$, and let $B$ be a bounded below complex of objects of $\mathscr{A}_{\Omega-\Lambda,G}$. Choose bounded above complexes $L$, $P$ and $Q$, consisting of projective objects in $C_{\Omega,G}$, $\text{Mod}_{\Omega[G]}$ and $\text{Mod}_{\Omega-G}^{\Omega-ft}$ (i.e., $C_{\Omega-G}^{\Omega-ft}$) respectively, with quasi-isomorphisms to $A$ of complexes in the respective categories. Projectivity yields maps $Q \to L$, $Q \to P$ and $P \to L$ of complexes in $C_{\Omega}$, $\text{Mod}_{\Omega}$ and $\text{Mod}_{\Omega[G]}$, respectively. We then have a commutative diagram

$$
\begin{array}{ccc}
\text{Hom}_{\Omega,\text{cts}}(L,B) & \longrightarrow & \text{Hom}_{\Omega,\text{cts}}(Q,B) \\
\downarrow & & \downarrow \\
\text{Hom}_{\Omega}(P,B) & \longrightarrow & \text{Hom}_{\Omega}(Q,B).
\end{array}
$$

Here, the right-hand vertical arrow is the identity map by Lemma 2.5.7 (viewing $Q$ as a complex of objects in $C_{\Omega-G}^{\Omega-ft}$ with trivial $G$-action), and the lower horizontal map is a quasi-isomorphism as a consequence of Proposition 2.2.1, as $\Omega[G]$ is $\Omega$-free. We want to show that the left-hand vertical arrow is a quasi-isomorphism of complexes of $\Lambda^0[G]$-modules, and we will be done if we can show the upper horizontal arrow is a quasi-isomorphism of complexes of $\Lambda^0$-modules.

In the case that $B$ is a module, exactness of direct limit reduces us to the case that $B$ is an object of $C_{\Omega-G}^{\Lambda^0-ft}$, hence of $C_{\Omega}$. In this case, since both $L$ and $Q$ are complexes of projectives in $C_{\Omega}$, we obtain that the map is a quasi-isomorphism. In the general setting, since $L$ and $Q$ are bounded above and $B$ is bounded below, we have

$$
\text{Hom}_{\Omega,\text{cts}}^n(X,B) = \bigoplus_{j \in \mathbb{Z}} \text{Hom}_{\Omega,\text{cts}}(X^j,B^{j+n})
$$

for $X = L$ and $X = Q$, and the direct sum commutes with direct limits. As $B$ is the direct limit of its truncations $\tau_{\leq i}B$ (with $i$th terms $\ker d^j_B$), we may therefore assume that $B$ is bounded with $B^j = 0$ for $j > i$. In this case, brutal truncations (i.e., $\sigma_{\leq i}B$ having $i$th term $B^j$) provide an exact triangle

$$(\sigma_{\leq i-1}B)[i-1] \to B^i \to (\sigma_{\leq i}B)[i] \to (\sigma_{\leq i-1}B)[i],$$

and we are reduced recursively to the above-proven case of a module.

The following analogue of Lemma 2.2.3 will be of later use.
Lemma 2.5.9. Suppose that $\Lambda$ is a flat, Noetherian $R$-algebra, and let $\chi: G \to \text{Aut}_{\Lambda^e}(\Lambda)$ be a homomorphism. Let $A$ be a bounded above complex of objects in $C_{R,G}$, and let $B$ be a bounded below complex of $R$-modules. Then there are isomorphisms

$$
\theta: \chi \Lambda \otimes_R \text{Hom}_{R,cts}(A, B) \to \text{Hom}_{\Lambda^e,cts}(\chi \Lambda \Lambda \otimes_R A, \Lambda \otimes_R B)
$$

$$
\theta': \chi \Lambda \otimes_R \text{Hom}_{R,cts}(A, B) \to \text{Hom}_{\Lambda^e,cts}(\chi \Lambda \Lambda \otimes_R A, \Lambda \otimes_R B)
$$

that arise as direct limits of the maps in Lemma 2.2.3.

Proof. Write $A = \lim_{\leftarrow} A_\alpha$ with $A_\alpha$ an object of $C_{R,ft}_{R,G}$ and $B = \lim_{\to} B_\beta$ with $B_\beta$ an object of $C_{R,ft}$. Note that

$$
\chi \Lambda \otimes_R A_\alpha \cong \lim_{\leftarrow} \chi \Lambda \otimes_R A_\alpha,
$$

so, applying Lemma 2.2.3, we have

$$
\text{Hom}_{\Lambda,cts}(\chi \Lambda \Lambda \otimes_R A, \Lambda \otimes_R B) = \lim_{\alpha,\beta} \chi \Lambda \otimes_R \text{Hom}_R(A_\alpha, B_\beta)
$$

$$
\cong \chi \Lambda \otimes_R \text{Hom}_{R,cts}(A_\alpha, B_\beta)
$$

$$
\cong \chi \Lambda \otimes_R \text{Hom}_{R,cts}(A, B).
$$

This yields the first isomorphism, and the argument for the second is similar.

Let $M$ be an ind-admissible $(\Omega \otimes_R \Lambda^e)[G]$-module. The continuous cochain complex of $G$ with values in $M$ is the complex of $\Omega \otimes_R \Lambda^e$-modules that is the direct limit of complexes

$$
C(G, M) = \lim_{N \in S(M)} C(G, N),
$$

where $C(G, N)$ is the usual complex of (inhomogeneous) continuous cochains, with $N$ given the $I$-adic topology.

We remark that, in this definition, it clearly suffices to take the direct limit over a cofinal subset of $S(M)$. In particular, if $M$ itself is finitely generated over $\Lambda^e$, then by Lemma 2.2.3(e) the above definition of $C(G, M)$ as a direct limit agrees with its definition as continuous cochains, considering $M$ as an object of $C_{\Omega,\Lambda[G]}$. Finally, we can extend the above definition to consider the total cochain complex of a complex $M$ of ind-admissible modules, which has $k$th term

$$
C^k(G, M) = \bigoplus_{i+j=k} C^i(G, M^j)
$$

and differentials as in [Nek (3.4.1.3)], and the cochain functor induces a functor

$$
\text{RI}^*(G, -): D^+(\mathcal{A}_{\Omega,\Lambda[G]}) \to D^+(\text{Mod}_R \otimes_R \Lambda^e)
$$

between derived categories.
In this paper, we are interested in the case that our profinite ring $\Lambda$ is the Iwasawa algebra of a compact $p$-adic Lie group. With this in mind, we introduce some notation that will be used from this point forward. Fix a prime $p$. Let $R$ be a commutative complete Noetherian local ring with maximal ideal $m$ and residue field $k$, where $k$ is finite of characteristic $p$. We let $\Gamma$ denote a compact $p$-adic Lie group. We are interested in the completed group ring $\Lambda = R[[\Gamma]]$.

We note that $\Lambda$ is a profinite ring, endowed with the topology given by the canonical isomorphism

$$\Lambda \cong \lim_{\leftarrow} \bigoplus_{u \in \mathcal{U}} \lim_{n \to \infty} R/m^n[\Gamma/U],$$

where $\mathcal{U}$ denotes the set of open normal subgroups of $\Gamma$. In fact, it is a projective object in $\mathcal{C}_R$ (cf. [RZ, Lemma 5.3.5(d)]). Moreover, we have the following (cf. [Wil, Theorem 8.7.8]).

**Proposition 3.0.1.** The ring $\Lambda$ is Noetherian.

**Proof.** As $R$ is a complete local Noetherian ring with finite residue field $k$, the Cohen structure theorem ([Coh, Theorem 12]) implies that it is isomorphic to a quotient of a power series ring $S$ in $n = \dim k/m/m^2$ variables over the ring of Witt vectors $O$ of $k$. Therefore, $R[[\Gamma]]$ is a quotient of $S[[\Gamma]]$, so it suffices to prove that $S[[\Gamma]]$ is Noetherian. Since $S[[\Gamma]] \cong O[[\mathbb{Z}_p \times \Gamma]]$ and $\mathbb{Z}_p \times \Gamma$ is a compact $p$-adic Lie group, we have that $S[[\Gamma]]$ is Noetherian by a mild extension of a classical theorem of Lazard’s (cf. [Ven2, Corollary 2.4] and [Laz, Proposition V.2.2.4]).

Since Noetherian rings are necessarily coherent, we may apply Lemma 2.4.3(a) to conclude from Proposition 3.0.1 that $\Lambda$ is a flat $R$-algebra.

### 3.1 Induced modules and descent

In this subsection, we will exhibit an interesting spectral sequence relating the cohomology of an induced module over an Iwasawa algebra to the cohomology of the module itself. We start work in a more general setting.

Let $G$ be a profinite group. Let $\Sigma$ be a left coherent and right Noetherian profinite $R$-algebra. For a complex $M$ of objects in $\mathcal{C}_\Sigma, G$, we let $C(G, M)$ denote its total direct sum complex of continuous $G$-cochains (cf. the end of Section 2.3), we let $R\Gamma(G, M)$ denote the corresponding object in $D(\text{Mod}_\Sigma)$, and we let $H^i(G, M)$ denote its $i$th continuous $G$-cohomology (or, more precisely, hypercohomology) group. As in [Lim2, Proposition 3.2.11], the functor

$$R\Gamma(G, -) : D^+(\mathcal{C}_\Sigma, G) \to D^+(\text{Mod}_\Sigma)$$

is well-defined and exact. (While $\mathcal{C}_\Sigma, G$ may not have enough injectives, the treatment of [KS] asserts the existence of $D^+(\mathcal{C}_\Sigma, G)$ and $D(\mathcal{C}_\Sigma, G)$ under certain set-theoretic assumptions, which we make here.)
In the case that $G$ has finite $p$-cohomological dimension, we can do better. For a double complex $X$ (or, by abuse of notation, its total complex) and $n \in \mathbb{Z}$, we let $\tau_{\leq n}(X)$ (resp., $\tau_{\geq n}(X)$) be the total complex of the quotient complex (resp., subcomplex) of $X$ with $j$th row equal to $\tau_{\geq n}(X^j)$ (resp., $\tau_{\leq n}(X^j)$). For lack of a sufficiently precise reference, we provide a short proof of the following.

**Lemma 3.1.1.** Suppose that $G$ has finite $p$-cohomological dimension. Then we have a convergent hypercohomology spectral sequence

$$E_2^{i,j} = H^i(G, H^j(M)) \Rightarrow H^{i+j}(G, M)$$

for any complex $M$ of objects in $C_{\Sigma, G}$.

**Proof.** Since $G$ has finite cohomological dimension, for sufficiently large $k$ we have a convergent hypercohomology spectral sequence

$$E_2^{i,j} = H^i(G, H^j(M)) \Rightarrow H^{i+j}(\tau_{\leq k}C(G, M))$$

arising from the filtration on rows of the indicated truncation of the double $G$-cochain complex of $M$ (cf. [NSW, §II.2]). As $C(G, M)$ is the direct limit of the complexes $\tau_{\leq k}C(G, M)$ under maps that are quasi-isomorphisms by the above convergence, the natural map

$$H^{i+j}(\tau_{\leq k}C(G, M)) \to H^{i+j}(G, M)$$

is an isomorphism by exactness of the direct limit. 

As quasi-isomorphisms of complexes induce isomorphisms on the truncated double complexes defining their hypercohomology spectral sequences, we have the following corollary.

**Corollary 3.1.2.** Suppose that $G$ has finite $p$-cohomological dimension. The functor $C(G, -)$ preserves quasi-isomorphisms of chain complexes and induces an exact functor

$$R\Gamma(G, -) \colon D(C_{\Sigma, G}) \to D(\text{Mod}_\Sigma).$$

Fix a profinite $R$-algebra $\Omega$ that is a projective object of $C_R$ for the remainder of the section. We now derive the following spectral sequence.

**Proposition 3.1.3.** Let $Y$ be a complex of objects in $C_{\Sigma, G}$, and let $N$ be a bounded above complex of objects in $C_{\Omega^-\Sigma}$. Consider the conditions

(1) $G$ has finite $p$-cohomological dimension,

(2) $N$ is bounded with terms of finite projective dimension over $\Sigma^\circ$.

If (1) holds and $Y$ is bounded above, (2) holds and $Y$ is bounded below, or both (1) and (2) hold, then we have an isomorphism

$$N \otimes_{\Sigma^\circ} R\Gamma(G, Y) \simto R\Gamma(G, N \otimes_{\Sigma^\circ} Y)$$

in $D(\text{Mod}_\Omega)$. 

**Documenta Mathematica** 18 (2013) 621–678
Proof. Let $L$ be a bounded above complex of projective objects in $\mathcal{C}_{\Omega-\Sigma}$ mapping quasi-isomorphically to $N$. If (1) holds and $Y$ is bounded above, we set $P = L$, and we have by Proposition 2.4.6 that $P \otimes_{\Sigma} Y$ represents $N \otimes_{\Sigma} Y$ in $D^-(\mathcal{C}_{\Omega,G})$. Otherwise, there exists a quasi-isomorphic bounded quotient complex $P = \tau_{\geq n} L$ of $L$ of objects in $\mathcal{C}_{\Omega-\Sigma}$, through which the quasi-isomorphism $L \to N$ factors, such that the terms of $P$ are projective in $\mathcal{C}_{\Sigma^\omega}$. Proposition 2.4.7 then tells us that $P \otimes_{\Sigma} Y$ represents $N \otimes_{\Sigma} Y$ in $D^-(\mathcal{C}_{\Omega,G})$. In all cases, the terms of $P$ are flat $\Sigma^\omega$-modules by Lemma 2.4.3, so Proposition 2.1.4 tells us that $P \otimes_{\Sigma} C(G,Y)$ represents $N \otimes_{\Sigma} R\Gamma(G,Y)$ in $D(\text{Mod}_{\Omega})$.

We have a map of complexes of $\Omega$-modules

$$P \otimes_{\Sigma} C(G,Y) \to C(G, P \otimes_{\Sigma} Y)$$

with sign conventions as in [Nek, Proposition 3.4.4]. (The continuity of the cochains in the image is insured, for instance, by the fact that any term of $P$ is a topological direct summand of a direct product of copies of $\Sigma$.) It suffices to show that this map is a quasi-isomorphism.

Let $Q$ be a bounded above complex of finitely generated projective $\Sigma^\omega$-modules mapping quasi-isomorphically to $N$, which we take to be bounded if (2) holds. We then have a map of complexes of $R$-modules

$$Q \otimes_{\Sigma} C(G,Y) \to C(G, Q \otimes_{\Sigma} Y)$$

that, much as in [Nek, Proposition 3.4.4], is an isomorphism. (To see that it is an isomorphism, note that it is immediate if $Q$ is a finitely generated free $\Sigma^\omega$-module, and therefore also for direct summands of such, i.e., the finitely generated projective $\Sigma^\omega$-modules. The case of a complex is then immediate.) By the projectivity of $Q$ in both the categories of abstract and compact $\Sigma^\omega$-modules, we have a commutative diagram

$$
\begin{array}{ccc}
Q \otimes_{\Sigma} C(G,Y) & \longrightarrow & C(G, Q \otimes_{\Sigma} Y) \\
\downarrow & & \downarrow \\
P \otimes_{\Sigma} C(G,Y) & \longrightarrow & C(G, P \otimes_{\Sigma} Y)
\end{array}
$$

in which the vertical arrows are quasi-isomorphisms of complexes of $R$-modules, noting Lemma 2.4.2(b) and, when (1) holds, Corollary 3.1.2. It follows that the lower horizontal arrow is a quasi-isomorphism as well.

We make a couple of remarks.

Remark 3.1.4. Suppose that $G$ is a finite group. The proof of Proposition 3.1.3 under condition (2) then goes through word-for-word for bounded $Y$, with Tate cochains and the resulting derived complexes replacing the usual cochains of $G$ and its derived complexes.
Remark 3.1.5. In [FK1, Proposition 1.6.5(3)], an identical isomorphism to that of Proposition 3.1.3 is proved for a very general class of “adic” rings Ω and Λ, with stronger conditions on G, N, and Y.

From now on, we focus on case that Σ is the Iwasawa algebra Λ = \( R[\Gamma] \). Fix a continuous homomorphism \( \chi : G \to \Gamma \) of profinite groups. Since Γ may be viewed as a subgroup of \( \text{Aut}_A(\Lambda) \) by left multiplication, we may define the \( \Lambda \otimes R[\Gamma] \)-module \( \chi A \) as in Section 2.2. If \( A \) is a \( \Lambda \otimes R \)-module, we then set \( \chi A = \chi A \otimes \Lambda A \) and \( A \chi = A \otimes \Lambda \chi \), which are \( \Lambda \otimes [G] \) and \( [G] \)-modules, respectively.

Let \( M \) be a \( R[G] \)-module. As in [Lim2, Section 5.1], we define a \( \Lambda_\otimes [G] \)-module \( F_\Gamma(M) \) by

\[
F_\Gamma(M) = \lim_{U \in \mathcal{U}} (R[\Gamma/U] \otimes_R M)
\]

with \( G \) acting diagonally and \( \Lambda \) acting on the left on the terms in the inverse limit. This construction applied to compact modules provides a functor

\[
F_\Gamma : C_{R,G} \longrightarrow C_{\Lambda,G},
\]

which is exact as \( R[\Gamma/U] \) is \( R \)-flat and the inverse limit is exact on inverse systems in \( C_{\Lambda,G} \). We extend this to an exact functor on complexes in the obvious fashion.

Given an complex \( T \) of objects in \( C_{R,G} \), we have, noting Lemma 2.4.2(a), a natural continuous isomorphism

\[
\mathcal{F}_T(T) \cong A \hat{\otimes}_R T
\]

of complexes in \( C_{\Lambda,G} \). We may therefore use Lemma 2.4.2(b) to note that \( \mathcal{F}_T \) takes a complex \( T \) of objects in \( C_{R,G}^{\text{ft}} \) to the complex \( \Lambda \hat{\otimes}_R T \) of objects in \( C_{\Lambda,G}^{\text{ft}} \). We make the latter identification freely.

We now supply a key ingredient for descent.

Lemma 3.1.6. Let \( T \) be a complex in \( C_{R,G} \). Let \( \Gamma' \) be a quotient of \( \Gamma \) by a closed normal subgroup, and set \( \Lambda' = R[\Gamma'] \). Let \( N \) be a bounded above complex of objects of \( C_{\Omega - \Lambda'} \). Then \( N \) can be viewed as a complex of objects in \( C_{\Omega - \Lambda} \) via the quotient map \( \pi : \Lambda \to \Lambda' \). Suppose either that \( T \) is bounded above or that \( N \) is bounded and its terms have finite projective dimension in \( C_{\Lambda'} \) and in \( C_{(\Lambda')}^{\text{ft}} \). Then \( \pi \) induces an isomorphism

\[
N \hat{\otimes}_\Lambda \mathcal{F}_T(T) \simto N \hat{\otimes}_{\Lambda'} \mathcal{F}_{T'}(T)
\]

in \( D^- (C_{\Omega,G}) \).

Proof. Let \( Q \to N \) be a quasi-isomorphism, where \( Q \) is a bounded above complex of \( C_{\Omega - \Lambda'} \)-projectives, and note that its terms are projective objects of \( C_{(\Lambda')}^{\text{ft}} \) by Lemma 2.4.2(b). If \( N \) is bounded with terms of finite projective dimension
in \( C(\Lambda') \), then for sufficiently small \( n \), the truncation \( \tau_n Q \) also consists of projectives in \( C(\Lambda') \), so we may assume in that case that \( Q \) is bounded. Next, and similarly, let \( P \to Q \) be a quasi-isomorphism in \( C_\Omega \) with \( P \) a bounded above complex in \( C_\Omega \) of projectives in \( C_{\Lambda'} \), and suppose that \( P \) is bounded if \( N \) is bounded with terms having finite projective dimension in \( C_{\Lambda'} \). The derived completed tensor products in question are then represented by \( P \hat{\otimes} \Lambda \chi \hat{\otimes} R T \) and \( Q \hat{\otimes} \Lambda' \chi \hat{\otimes} R T \), respectively, by Propositions 2.4.6 and 2.4.7. Since the induced map \( P \hat{\otimes} \Lambda \chi \to Q \hat{\otimes} \Lambda' \chi \) is clearly a quasi-isomorphism and the terms of these complexes are projective in \( C_{\Lambda'} \), we have by the same propositions the desired quasi-isomorphism between the two complexes.

We are particularly interested in the case that the complex \( N \) above is the Iwasawa algebra of a quotient of \( \Gamma \), for which the following lemma is crucial.

**Lemma 3.1.7.** Let \( \Gamma_0 \) be a closed normal subgroup of \( \Gamma \) with no elements of order \( p \). Set \( \Gamma' = \Gamma / \Gamma_0 \) and \( \Lambda' = R[\Gamma'] \). Then \( \Lambda' \) has finite projective dimension over \( \Lambda \).

**Proof.** Let \( c\text{-Tor}^\Xi_i (\_ , \_) \) denote the \( i \)th derived bifunctor of the completed tensor product over a profinite \( R \)-algebra \( \Xi \). A standard argument yields a convergent spectral sequence

\[
H_i(\Gamma, c\text{-Tor}^R_j(\Lambda', Z)) \Rightarrow c\text{-Tor}_i^{\Lambda_\chi}(\Lambda', Z)
\]

for any compact \( \Lambda \)-module \( Z \), as in [NSW, §V.2, Exercise 3], where elements of \( \Gamma \) act inversely on \( \Lambda_\chi \)-modules and diagonally on completed \( R \)-tensor products. Since \( \Lambda' \) is a projective object in \( C_R \), the spectral sequence degenerates to yield isomorphisms

\[
H_n(\Gamma, \Lambda' \hat{\otimes} R Z) \cong c\text{-Tor}^\Lambda_n(\Lambda', Z).
\]

Moreover, by a standard extension of Shapiro’s lemma [RZ Theorem 6.10.9], we have

\[
H_n(\Gamma_0, Z) \cong H_n(\Gamma, \Lambda' \hat{\otimes} R Z).
\]

Since \( \Gamma_0 \) has finite \( p \)-cohomological dimension by [Ser, Corollaire 1], there then exists \( n_0 \geq 0 \) independent of \( Z \) such that \( c\text{-Tor}_n^\Lambda(\Lambda', Z) = 0 \) for every \( n > n_0 \). It follows from [NSW, Proposition 5.2.11] that \( \Lambda' \) has finite topological projective dimension as a compact \( \Lambda \)-module, so it has finite projective dimension over \( \Lambda \) by the discussion of Section 2.3.

The following descent spectral sequence is reduced by the above results to a special case of Proposition 3.1.8.

**Theorem 3.1.8.** Let \( T \) be a complex in \( C_{R,G} \). Let \( \Gamma' \) be a quotient of \( \Gamma \) by a closed normal subgroup \( \Gamma_0 \), and set \( \Lambda' = R[\Gamma'] \). Consider the conditions

1. \( G \) has finite cohomological dimension,
2. \( \Gamma_0 \) contains no elements of order \( p \).
Suppose that (1) holds and \( T \) is bounded above, (2) holds and \( T \) is bounded below, or both (1) and (2) hold. Then we have an isomorphism
\[
\Lambda' \otimes_{\Lambda} R\Gamma(G, \mathcal{F}_T(T)) \cong R\Gamma(G, \mathcal{F}_{\Lambda'}(T))
\]
in \( \mathbf{D}(\text{Mod}_{\Lambda'}) \).

**Proof.** This is simply Proposition 3.1.3 for \( N = \Lambda' \) and \( Y = \mathcal{F}_T(T) \), noting Lemma 3.1.7 and applying the isomorphism of Lemma 3.1.6.

3.2 Finite generation of cohomology groups

In this subsection, we shall show that the cohomology groups of induced modules are finitely generated under a certain assumption on the group. We maintain the notation of the previous subsection. In particular, \( \Gamma \) is a compact \( p \)-adic Lie group, \( R \) is a complete commutative local Noetherian ring with finite residue field of characteristic \( p \), and \( \Lambda = R[\Gamma] \). Moreover, we are given a continuous homomorphism \( \chi : G \to \Gamma \) from a profinite group \( G \).

**Lemma 3.2.1.** Suppose that \( \Gamma \) is pro-\( p \). Let \( M \) be a compact \( \Lambda \)-module. Then \( M \) is finitely generated over \( \Lambda \) if and only if \( R \otimes_{\Lambda} M \) is finitely generated over \( R \).

**Proof.** Since \( \Gamma \) is pro-\( p \), the Jacobson radical \( \mathfrak{m} \) of \( \Lambda \) is \( \mathfrak{m}\Lambda + I \), where \( I \) is the augmentation ideal of \( \Lambda \) (see [NSW, Proposition 5.2.16(iii)]). This implies that
\[
M/\mathfrak{m}M \cong M_{\Gamma}/\mathfrak{m}M_{\Gamma},
\]
where \( M_{\Gamma} = M/I \). Therefore, Nakayama’s lemma tells us that \( M_{\Gamma} \) is finitely generated over \( R \) if and only if \( M/\mathfrak{m}M \) is finite. On the other hand, Nakayama’s lemma for compact \( \Lambda \)-modules (cf. [NSW Lemma 5.2.18(ii)]) tells us that \( M \) is finitely generated over \( \Lambda \) if and only if \( M/\mathfrak{m}M \) is finite.

Lemma 3.2.1 requires a compact \( \Lambda \)-module. Therefore, we give a sufficient condition for an abstract \( \Lambda \)-module to be a compact \( \Lambda \)-module under an appropriate topology.

**Lemma 3.2.2.** Suppose that \( M \) is an abstract \( \Lambda \)-module which is the inverse limit of an inverse system of finite quotient modules. Then \( M \) is a compact \( \Lambda \)-module with respect to the resulting profinite topology.

**Proof.** We need only to show that the \( \Lambda \)-action
\[
\theta : \Lambda \times M \to M
\]
is continuous with respect to the topology given by the inverse limit. Suppose first that \( M \) is finite. Let \( \mathfrak{m} \) denote the Jacobson radical of \( \Lambda \). Then \( \mathfrak{m}^n M \) stabilizes, and it follows from Nakayama’s lemma that we have \( \mathfrak{m}^n M = 0 \) for large enough \( n \). By [NSW, Corollary 5.2.19], there exist \( r \geq 0 \) and \( U \in \mathcal{U} \)
such that \( m^r \Lambda + I(U) \subseteq M^n \), where \( I(U) \) denotes the ideal of \( \Lambda \) generated by the augmentation ideal in \( R[U] \). Hence, \( \theta \) is continuous with respect to the discrete topology on \( M \).

In general, we can write \( M \cong \varprojlim M/M_\alpha \), where \( \{ M_\alpha \} \) is a directed system of \( \Lambda \)-submodules of finite index. Let \( (\lambda, x) \in \theta^{-1}(y + M_\alpha) \) for \( \lambda \in \Lambda \) and \( x, y \in M \).

Since \( M/M_\alpha \) is finite, it follows from the above discussion that there exist \( r \) and \( U \) such that

\[
(\lambda + m^r \Lambda + I(U)) \cdot (x + M_\alpha) \subseteq y + M_\alpha,
\]
as desired. \( \square \)

We shall also require the following lemma.

**Lemma 3.2.3.** If \( M \) is a finitely generated \( \Lambda \)-module, then \( \text{Tor}_i^\Lambda(R, M) \) is finitely generated over \( R \) for every \( i \geq 0 \).

**Proof.** To see this, we first choose a resolution \( P \) of \( M \) consisting of finitely generated projective \( \Lambda \)-modules. Then \( R \otimes_\Lambda P \) is a complex of finitely generated \( R \)-modules. Therefore, the homology groups \( \text{Tor}_i^\Lambda(R, M) \) of the latter complex are finitely generated over \( R \). \( \square \)

We recall from [Nek] Proposition 4.2.3 that if \( G \) is a profinite group such that \( H^i(G, M) \) is finite for every finite \( G \)-module \( M \) of \( p \)-power order and \( i \geq 0 \), then \( H^i(G, T) \) is a finitely generated \( R \)-module for every \( T \in C_{R,G}^{R-\text{ft}} \) and \( i \geq 0 \). We will refer to a profinite group \( G \) as \( p \)-cohomologically finite if there exists an integer \( n \) such that, for every finite \( G \)-module \( M \) of \( p \)-power order, \( H^i(G, M) = 0 \) for all \( i > n \) and \( H^i(G, M) \) is finite for all \( i \). We remark that if \( G \) is \( p \)-cohomologically finite, then so is every open subgroup of \( G \), as follows directly from Shapiro’s lemma.

**Theorem 3.2.4.** Suppose that \( G \) is \( p \)-cohomologically finite. Let \( T \) be a complex of objects in \( C_{R,G}^{R-\text{ft}} \). Then the cohomology groups \( H^i(G, \mathcal{F}_T(T)) \) are finitely generated over \( \Lambda \) for all \( i \). That is, \( R^i(G, \mathcal{F}_T(T)) \) is an object of \( D_{\Lambda-\text{pt}}(\text{Mod}_\Lambda) \).

**Proof.** Suppose for now that \( T \) is concentrated in degree 0. Let \( \Gamma_0 \) be an open uniform normal pro-\( p \)-subgroup of \( \Gamma \), and set \( H = \chi^{-1}(\Gamma_0) \). Fix a set of double coset representatives \( \gamma_1, \ldots, \gamma_t \) of \( \Gamma_0 \backslash \Gamma / \chi(G) \). For each \( i \), define \( \chi_i : G \to \Gamma \) by \( \chi_i(g) = \gamma_i g \gamma_i^{-1} \). The reader may check that

\[
\bigoplus_{i=1}^t \left( \mathbb{Z}[G] \otimes_{\mathbb{Z}[H]} R[\Gamma_0]_{\chi_i} \right) \cong \Lambda_X
\]

is an isomorphism of \( R[\Gamma_0][G] \)-modules. Given this, Shapiro’s lemma induces isomorphisms

\[
H^i(G, \mathcal{F}_T(T)) \cong \bigoplus_{i=1}^t H^i(H, R[\Gamma_0]_{\chi_i} \otimes_R T)
\]

**Documenta Mathematica** 18 (2013) 621–678
Nekovář Duality

of \( R[[\Gamma_0]]\)-modules. The finite generation of \( H^1(G, \mathcal{F}_T(T)) \) over \( \Lambda \) then reduces to the finite generation of each \( H^1(H, R[[\Gamma_0]]_I \otimes_R T) \) over \( R[[\Gamma_0]] \). Thus, we can and do assume that \( \Gamma \) has no elements of finite order.

Since \( R \) has finite projective dimension over \( \Lambda \) by Lemma 3.1.7, Theorem 3.1.8 provides the convergent spectral sequence

\[
E^{r,s}_2 = \text{Tor}^{\Lambda}_{r-s}(R, H^s(G, \mathcal{F}_T(T))) \Rightarrow H^{r+s}(G, T).
\]

Since \( G \) has finite \( p \)-cohomological dimension, say \( n \), we may suppose induc- 

tively that \( H^i(G, \mathcal{F}_T(T)) \) is a finitely generated \( \Lambda \)-module for all \( i \) greater than some \( j \leq n \). Note that \( E^{0,j}_\infty \) is a quotient of \( E^{0,j}_2 \) and that \( H^j(G, T) \) is a finitely generated \( R \)-module by [Nek, Proposition 4.2.3]. Since \( E^{0,j}_\infty \) is a quotient of \( H^j(G, T) \), it is also a finitely generated \( R \)-module. On the other hand, it follows from the definition of \( E^{0,j}_2 \) that the kernel of the surjective map \( E^{0,j}_2 \to E^{0,j}_\infty \) is isomorphic to a subquotient of the finite direct sum

\[
\bigoplus_{i=j+1}^n E^{j-i, i}_2 = \bigoplus_{i=j+1}^n \text{Tor}^{\Lambda}_{i-1}[R, H^i(G, \mathcal{F}_T(T))].
\]

By our induction hypothesis and Lemma 3.2.3, the above module is a finitely generated \( R \)-module. It follows that

\[
E^{0,j}_2 \cong R \otimes_{\Lambda} H^j(G, \mathcal{F}_T(T))
\]

is a finitely generated \( R \)-module. As \( H^j(G, \mathcal{F}_T(T)) \) is an inverse limit of finite \( \Lambda \)-modules by [Lim2, Proposition 5.2.4], Lemma 3.2.2 gives it the structure of a compact \( \Lambda \)-module, and we may therefore apply Lemma 3.2.1 to conclude that \( H^j(G, \mathcal{F}_T(T)) \) is finitely generated over \( \Lambda \).

Now let us work without condition on \( \Gamma \) or the complex \( T \). By Lemma 3.1.4 and the exactness of \( \mathcal{F}_T(-) \), we have the convergent spectral sequence of \( \Lambda \)-modules

\[
H^r(G, \mathcal{F}_T(H^s(T))) \Rightarrow H^{r+s}(G, \mathcal{F}_T(T)).
\]

Thus, \( H^i(G, \mathcal{F}_T(T)) \) has a filtration consisting of subquotients of the finitely generated \( \Lambda \)-modules \( H^i(G, \mathcal{F}_T(H^{i-j}(T))) \) with \( 0 \leq j \leq n \). Consequently, \( H^i(G, \mathcal{F}_T(T)) \) is also a finitely generated \( \Lambda \)-module.

**Remark 3.2.5.** In the case that \( \Gamma \) is abelian, Theorem 3.2.4 for bounded below \( T \) is essentially [Nek, Proposition 4.2.3], as \( R[[\Gamma_0]] \) for an open pro-\( p \), torsion-free subgroup \( \Gamma_0 \) is itself a commutative complete local Noetherian ring with finite residue field. In the case that \( T \) is a bounded complex of \( R \)-projectives, it is a corollary of [FK1, Proposition 1.6.5(2)].

### 3.3 Duals of induced modules

In this subsection, we describe the pairings and the resulting cup products that will be used in proving a duality theorem for the cohomology of induced
modules. Key to our discussion is the following lemma, which we record for convenience. We let \( \Omega \) and \( \Sigma \) denote auxiliary Iwasawa algebras: e.g., \( \Omega = R[\Phi] \) for some compact \( p \)-adic Lie group \( \Phi \).

**Lemma 3.3.1.** Suppose that \( M \) is a bounded above complex of objects of \( C_{\Omega-\Lambda,G} \), let \( N \) be a bounded above complex of objects of \( C_{\Lambda-\Sigma,G} \), and let \( L \) be a bounded above complex of objects of \( C_{\Omega-\Sigma,G} \). Then any map

\[
\phi: M \otimes_{\Lambda} N \to L
\]

of complexes in \( C_{\Omega-\Sigma,G} \) gives rise to a cup product morphism

\[
\cup: C(G,M) \otimes_{\Lambda} C(G,N) \to C(G,L)
\]

of complexes of \( \Omega \otimes_{\Lambda} \Sigma^\vee \)-modules.

**Proof.** Briefly, the point is that to give such a map \( \phi \) is equivalent to producing a collection of continuous, \( \Lambda \)-balanced, \( G \)-equivariant, left \( \Omega \)-linear, and right \( \Sigma \)-linear pairings

\[
\langle -, - \rangle_{mn}: M^m \times N^n \to L^{m+n}
\]

that are compatible with coboundaries in the sense of [Lim2, Section 3.3]. The cup product then arises from the induced cup products

\[
\cup_{ij}^{ij}: C^i(G,M^m) \otimes_{\Lambda} C^j(G,N^n) \to C^{i+j}(G,L^{m+n})
\]

by combining them to a map of the total complexes with appropriate signs as in [Nek, (3.4.5.2)]; \( \cup = (-1)^{imjn} \cup_{ij}^{ij} \).

Let us fix some notation. Let \( \iota: \Lambda \to \Lambda \) denote the unique continuous \( R \)-algebra homomorphism that satisfies \( \iota(\gamma) = \gamma^{-1} \) for all \( \gamma \in \Gamma \). If \( A \) is a \( \Lambda[G] \)-module, we let \( A^\iota \) denote the \( \Lambda^\iota[G] \)-module that is \( A \) as an \( R[G] \)-module but on which any \( \lambda \in \Lambda \) now acts by left multiplication by \( \iota(\lambda) \). We extend this to complexes in the obvious fashion. Of particular interest to us is \( \mathcal{F}_T(T)^\iota \) for a complex \( T \) of objects in \( C_{R,G} \). Since \( \iota \) induces a continuous isomorphism of \( \Lambda^\iota[G] \)-modules \( \chi^\iota \Lambda \xrightarrow{\sim} \Lambda^\iota \), we can and will make the identification

\[
\mathcal{F}_T(T)^\iota = \chi^\iota \Lambda \otimes_R T,
\]

of complexes in \( C_{\Lambda,\iota,G} \) from this point forward.

**Lemma 3.3.2.** For objects \( M \) and \( N \) in \( C_{R,G} \), the pairing

\[
\langle -, - \rangle: \mathcal{F}_T(M) \times \mathcal{F}_T(N)^\iota \to \Lambda \otimes_R (M \otimes_R N)
\]

\[
(\lambda \otimes m, \mu \otimes n) \mapsto \lambda \mu \otimes m \otimes n
\]

is continuous, \( R \)-balanced, \( G \)-equivariant with respect to the trivial action on \( \Lambda \) in the target, and is left and right \( \Lambda \)-linear. Moreover, if \( M \) and \( N \) are bounded above complexes of objects in \( C_{R,G} \), we have a morphism

\[
\phi: \mathcal{F}_T(M) \otimes_R \mathcal{F}_T(N)^\iota \to \Lambda \otimes_R (M \otimes_R N)
\]

of complexes in \( C_{\Lambda-\Lambda,G} \).
Proof. We begin by remarking that by Lemma 2.4.1 it makes sense to define
the pairing on pairs of tensors, identified with their images in the completed
tensor product. The pairing is clearly \( R \)-balanced, and the reader can check the
statements on the \( \Lambda, \Lambda^\circ \), and \( G \)-actions. The continuity of the pairing reduces
immediately to the continuity of multiplication on \( \Lambda \). The second statement
follows easily.

We next apply these pairings to study a duality between induced modules. As
described in the introduction, the dualizing complex for \( R \) is an object
\( \omega_R \) of \( D_{R-\text{ft}}(\text{Mod}_R) \) with the property that for every object \( M \) of \( D(\text{Mod}_{R-\text{ft}}) \), the object \( R\text{Hom}_R(M,\omega_R) \) lies in \( D_{R-\text{ft}}(\text{Mod}_R) \) and the canonical morphism
\[
M \longrightarrow R\text{Hom}_R(R\text{Hom}_R(M,\omega_R),\omega_R)
\]
is an isomorphism in \( D(\text{Mod}_R) \). A dualizing complex exists for \( R \) and is unique
up to translation and isomorphism in \( D_{R-\text{ft}}(\text{Mod}_R) \) (see [Har1, Ch. V]). We
choose a bounded complex \( J_R \) of injective \( R \)-modules which represents a choice
of the dualizing complex in \( D(\text{Mod}_R) \). (See [Nek, Section 0.4], or Section 4.3
below, for the construction of such a complex.)

Let \( T \) be a bounded complex of objects in \( C_{R,G} \). Then \( \text{Hom}_R(T,J_R) \) is a
bounded complex of “admissible” \( R[G] \)-modules with cohomology groups that
are finitely generated over \( R \) (see [Nek, (4.3.2)]). By [Nek, Proposition 3.3.9],
there is a subcomplex \( T^* \) of \( \text{Hom}_R(T,J_R) \) which is a complex of objects in \( C_{R,G}^{R-\text{ft}} \) (giving said objects the \( \mathfrak{m} \)-adic topology) and is quasi-isomorphic to
\( \text{Hom}_R(T,J_R) \) via the inclusion map.

We have a composite morphism
\[
\pi: T \otimes_R T^* \longrightarrow T \otimes_R \text{Hom}_R(T,J_R) \longrightarrow J_R
\]
of complexes of \( R[G] \)-modules, the first morphism being induced by the inclusion
and the second being the usual evaluation map. By Lemma 3.3.2, this in turn induces a composite map
\[
\varpi: \mathcal{F}_T(T) \otimes_R \mathcal{F}_T(T^*)^i \overset{\delta}{\longrightarrow} \Lambda \otimes_R T \otimes_R T^* \otimes_R J_R
\]
of complexes of \( (\Lambda \otimes_R \Lambda^\circ)[G] \)-modules. Twisting \( \varpi \) by the identity map on
an auxiliary bounded above complex of objects of \( \mathcal{C}_{\Omega-\Lambda}^{\Lambda^\circ-\text{ft}} \), Lemma 3.3.1 now
provides the cup product morphisms of the following lemma, which will be
used in the next section. For a continuous character \( \kappa: G \rightarrow R^\times \) and an \( R[G] \)-
module \( M \), we let \( M(\kappa) \) denote the \( R \)-module \( M \) with the new commuting
\( G \)-action given by the twist of the original by \( \kappa \). (In addition, \( M(\kappa) \) will be
taken to maintain any topology and actions of profinite \( R \)-algebras with which
\( M \) may be endowed.)

Lemma 3.3.3. Let \( A \) be a bounded above complex of objects in \( \mathcal{C}_{\Omega-\Lambda}^{\Lambda^\circ-\text{ft}} \), and let
\( \kappa: G \rightarrow R^\times \) be a continuous homomorphism. The map \( \varpi_A \) defined above induces a map
\[
\varpi_A: A \otimes_R \mathcal{F}_T(T) \otimes_R \mathcal{F}_T(T^*)^i \longrightarrow A \otimes_R J_R
\]
and, in turn, a well-defined cup product
\[ C(G, A \otimes \mathcal{F}_T(T)) \otimes_R C(G, \mathcal{F}_T(T^*)^i(\kappa)) \to C(G, A \otimes_R J_R(\kappa)), \]

which is a map of complexes of \( \Omega \otimes_R \Lambda^\circ\)-modules.

Note that, by definition, the adjoint maps (defined as in [Lim2, Lemma 2.2])
\[ \text{adj}(\pi): T^* \to \text{Hom}_R(T, J_R) \quad \text{and} \quad \text{adj}'(\pi): T \to \text{Hom}_R(T^*, J_R) \]
of \( \pi \) are quasi-isomorphisms, the first being simply the inclusion of complexes already defined. Since the terms of \( J_R \) are injective \( R \)-modules, it follows from Proposition 2.2.1 that the derived adjoint maps
\[ \text{adj}(\pi): T^* \to \text{RHom}_R(T, \omega_R) \quad \text{and} \quad \text{adj}'(\pi): T \to \text{RHom}_R(T^*, \omega_R) \]
are isomorphisms in \( D^b_{R,G}(\text{Mod}_R[G]) \), hence in \( D^b_{R,G}(\mathcal{F}_R,G) \), considering the restriction of \( \text{RHom}_R(-, -) \) to a bifunctor
\[ D^+(C^{R-R}_{R,G}) \times D^+(\text{Mod}_R) \to D^+(\mathcal{F}_R,G). \]

(see [Nek (3.5.9)]).

Though we shall not use it later, we feel it important to note that the derived adjoint maps of \( \pi \) are also isomorphisms. While this is easy enough to prove in the bounded below derived category \( D^+(\text{Mod}_{\Lambda^*}[G]) \) of abstract modules (in the case of \( \text{adj}(\pi) \)), we are interested in \( G \)-cohomology groups, so we want such an isomorphism in \( D^+(\mathcal{A}^*, G) \).

**Theorem 3.3.4.** For any bounded complexes \( T \) and \( T^* \) of objects in \( C_{R,G} \) such that \( T^* \) sits quasi-isomorphically as a subcomplex of the dual \( \text{Hom}_R(T, J_R) \), the derived adjoint maps
\[ \text{adj}(\pi): \mathcal{F}_T(T^*)^i \to \text{RHom}_{\Lambda,\text{cts}}(\mathcal{F}_T(T), \Lambda \otimes_R \omega_R) \]
\[ (\text{resp., } \text{adj}'(\pi): \mathcal{F}_T(T) \to \text{RHom}_{\Lambda,\text{cts}}(\mathcal{F}_T(T^*)^i, \Lambda \otimes_R \omega_R)) \]
are isomorphisms in \( D^+(\mathcal{A}^*, G) \) (resp., \( D^+(\mathcal{A}, G) \)). Moreover, the derived object \( \text{RHom}_{\Lambda,\text{cts}}(\mathcal{F}_T(T), \Lambda \otimes_R \omega_R) \) (resp., \( \text{RHom}_{\Lambda,\text{cts}}(\mathcal{F}_T(T^*)^i, \Lambda \otimes_R \omega_R) \)) can be represented by \( \text{Hom}_{\Lambda}(\mathcal{T}_T(T), \Lambda \otimes_R J_R) \) (resp., \( \text{Hom}_{\Lambda}(\mathcal{T}_T(T^*)^i, \Lambda \otimes_R J_R) \)).

**Proof.** Let \( X = \text{RHom}_{\Lambda,\text{cts}}(\mathcal{F}_T(T), \Lambda \otimes_R \omega_R) \). Since \( J_R \) is a complex of \( R \)-injectives and \( \chi \Lambda \) is flat over \( R \), Lemma 2.5.9 implies that the functor \( F \) given by
\[ \text{Hom}_{\Lambda,\text{cts}}(\Lambda \otimes_R T, \Lambda \otimes_R J_R): \text{Ch}^-(C_{R,G})^\circ \to \text{Ch}^+(\mathcal{A}^*, G) \]
is exact. In particular, if \( P \to T \) is a quasi-isomorphism with \( P \) a bounded above complex of projective objects in \( C_{R,G} \), then \( F(T) \to F(P) \) is a quasi-isomorphism. As \( \Lambda \otimes_R P \) is a complex of objects in \( C_{A,G} \) that are projective in \( C_{R,G} \), we then have that \( F(T) \) represents \( X \) by Proposition 2.5.8.
By Lemmas 2.4.2(b) and 2.5.7, the functor $F$ equals $\text{Hom}_\Lambda (\Lambda \chi \otimes_R -, \Lambda \otimes_R J_R)$ on $\text{Ch}^-(\mathcal{R}_R)^{\otimes}$. Therefore, $X$ is represented by $\text{Hom}_\Lambda (\mathcal{T}_R(T), \Lambda \otimes_R J_R)$ in $\mathcal{D}^+(\mathcal{I}_\Lambda, G)$. Finally, since the map $T^* \to \text{Hom}_R(T, J_R)$ is a quasi-isomorphism, the adjoint map of complexes $\chi \Lambda \otimes_R T^* \to \text{Hom}_\Lambda (\Lambda \chi \otimes_R T, \Lambda \otimes_R J_R)$ is a quasi-isomorphism of complexes of $\Lambda \circ\llbracket G\rrbracket$-modules by Lemma 2.2.3. Therefore, we have the result for $\text{adj}(\pi)$, and the proof for $\text{adj}'(\pi)$ is analogous.

4 Duality over $p$-adic Lie extensions

We now turn to arithmetic. Here, we fix the notation that we shall use throughout this section. To start, let $p$ be a prime. We let $F$ be a global field of characteristic not equal to $p$. Let $S$ be a finite set of primes of $F$ that, in the case that $F$ is a number field, contains all primes above $p$ and all real places. Let $S_f$ (resp., $S_R$) denote the set of finite places (resp., real places) in $S$. Let $G_{F,S}$ denote the Galois group of the maximal unramified outside $S$ extension of $F$. For a place $v$ of $F$, let $F_v$ denote the completion of $F$ at $v$, and let $G_v$ denote a fixed decomposition group for $v$ in the absolute Galois group of $F$.

We fix a $p$-adic Lie extension $F_{\infty}$ of $F$ that is unramified outside $S$, and we let $\Gamma$ denote its Galois group. We let $R$ denote a complete commutative local Noetherian ring with finite residue field of characteristic $p$, and we set $\Lambda = R[\Gamma]$. We take the homomorphism $\chi: G_{F,S} \to \Gamma$ of Section 3.1 to be restriction.

4.1 Iwasawa cohomology

We recall the following facts, all of which can all be found in [NSW, Chapters VII-VIII]. The $G_{F,S}$-cohomology groups of a finite $G_{F,S}$-module of $p$-power order are all finite. If $p$ is odd or $F$ has no real places, then $G_{F,S}$ has $p$-cohomological dimension at most 2 and so is $p$-cohomologically finite in the sense of Section 3.2. If $v$ is a nonarchimedean place of $F$, then $G_v$ has $p$-cohomological dimension equal to 2, and the $G_v$-cohomology groups of a finite $G_v$-module of $p$-power order are finite as well. Of course, the $G_v$-cohomology groups of a finite module are also finite for archimedean $v$, since $G_v$ is of order dividing 2 for such places.

Recall (e.g., from [Nek, (5.7.2)]) that, for any profinite ring $\Omega$, the (Tate) compactly supported $G_{F,S}$-cochain complex of a complex of objects $M$ in $\mathcal{C}_\Omega$ is defined as

$$C_{(\partial)}(G_{F,S}, M) = \text{Cone} \left( C(G_{F,S}, M) \to \bigoplus_{v \in S_f} C(G_v, M) \oplus \bigoplus_{v \in S_R} \widetilde{C}(G_v, M) \right)[-1],$$

where $\widetilde{C}(G_v, M)$ is the Tate cohomology complex of $G_v$-modules.
where $\hat{C}(G_v, M)$ is defined as in \cite[Section 3.4]{Lim2} and denotes the standard complete complex of Tate $G_v$-cochains for $M$.

We remark that, in Nekovář’s notation, $C_{(c)}(G_{F,S}, M)$ is denoted $\hat{C}_{(c)}(G_{F,S}, M)$. We use the notation of \cite{FK1}, where parentheses are used to distinguish the latter group from the compactly supported cochains $C_c(G_{F,S}, M)$, for which one uses the usual cohomology groups at real places. For archimedean $v$, we will abuse notation and use $RΓ(G_v, M)$ to denote the derived object of the Tate cochains $\hat{C}(G_v, M)$. These cochains may as well be taken to be zero for complex places or for real places if $p$ is odd, $RΓ(G_v, M)$ being a zero object in the derived category.

We denote the derived object corresponding to $C_{(c)}(G_{F,S}, M)$ by $RΓ_{(c)}(G_{F,S}, M)$ and its $i$th cohomology group by $H^i_{(c)}(G_{F,S}, M)$. By definition, we have an exact triangle

$$RΓ_{(c)}(G_{F,S}, M) \rightarrow RΓ(G_{F,S}, M) \rightarrow \bigoplus_{v \in S} RΓ(G_v, M)$$

in $\mathcal{D}(\text{Mod}_\Lambda)$. The results of and methods used in Section 3.1 allow us to prove the following descent result for such an exact triangle with induced coefficients.

**Proposition 4.1.1.** Let $Γ' = \text{Gal}(F'_\infty/F)$ be a quotient of $Γ$ by a closed normal subgroup, and set $Λ' = R[Γ/Γ'][Γ]$. Let $T$ be a bounded above complex of objects in $C_{R,G_{F,S}}$. Suppose that at least one of the following holds: (i) $p$ is odd, (ii) $F'_\infty$ has no real places, or (iii) $T$ is bounded and $F'_\infty$ has no real places that become complex in $F_\infty$. Then we have an isomorphism of exact triangles

\[ \Lambda' \otimes^L RΓ_{(c)}(G_{F,S}, \mathcal{F}_T'(T)) \cong RΓ_{(c)}(G_{F,S}, \mathcal{F}_{Γ'}(T)) \]

\[ \Lambda' \otimes^L RΓ(G_{F,S}, \mathcal{F}_T'(T)) \cong RΓ(G_{F,S}, \mathcal{F}_{Γ'}(T)) \]

\[ \Lambda' \otimes^L \bigoplus_{v \in S} RΓ(G_v, \mathcal{F}_T'(T)) \cong \bigoplus_{v \in S} RΓ(G_v, \mathcal{F}_{Γ'}(T)) \]

in $\mathcal{D}(\text{Mod}_Λ)$.

**Proof.** We need only show that the lower and upper horizontal morphisms are isomorphisms. The upper morphism exists and is an isomorphism for $T$ bounded above without additional assumption. That is, recall that the compactly supported cohomology groups of a module vanish in dimension greater than $3$ \cite[Lemma 5.7.3]{Nek}. The hypercohomology spectral sequence for compactly supported cohomology therefore converges for bounded above complexes, yielding the existence of a functor

$$RΓ_{(c)}(G_{F,S}, -): \mathcal{D}^-(C_{Λ,G_{F,S}}) \rightarrow \mathcal{D}^-(\text{Mod}_Λ).$$

\[ \text{We write an exact triangle } A \rightarrow B \rightarrow C \rightarrow A[1] \text{ more compactly as } A \rightarrow B \rightarrow C \text{ throughout.} \]
The analogous result to Proposition \[4.1.3\] then holds by the original argument (in the case that (1) holds and \(Y\) is bounded above). That the upper morphism exists and is an isomorphism follows immediately as in Theorem \[3.1.8\]. The next proposition shows in particular that if \(\hat{\Gamma}(\mathbb{L}), S_f\) then holds with “bounded above” removed if we suppose that the kernel of \(\Gamma \rightarrow \Gamma'\) has no elements of order \(p\). This follows quickly if \(p\) is odd or \(F\) has no real places, but it requires some work in the remaining case that \(F_{\infty}^p\) has no real places but \(D\) does. We omit this for purposes of brevity.

The analogous result to Proposition \[4.1.3\] then holds by the original argument (in the case that (1) holds and \(Y\) is bounded above). That the upper morphism exists and is an isomorphism follows immediately as in Theorem \[3.1.8\]. The analogous result to Proposition \[3.1.3\] then holds by the original argument (in the case that (1) holds and \(Y\) is bounded above). That the upper morphism exists and is an isomorphism follows immediately as in Theorem \[3.1.8\]. For a real place \(v\) of \(F\), the functor \(\Gamma(\mathbb{L}, -)\) is well-defined only on the bounded derived category. However, if \(v\) becomes complex in \(F_{\infty}^p\), then the relevant hypercohomology spectral sequence implies that the composite functor \(\Pi(\mathbb{L}, \mathbb{F})\) is both well-defined and zero on \(D(\mathbb{L})\). So, the \(v\)-summand of the lower horizontal morphism is trivially an isomorphism. If \(v\) splits completely in \(F_{\infty}^p\), then the modules \(\Lambda\) and \(\chi\Lambda\) have trivial \(G_v\)-actions, so we have isomorphisms \(\mathbb{F}(T) \cong \Lambda \otimes_R T\) and \(\mathbb{F}(T)^* \cong \Lambda \otimes_R T^*\) of complexes in \(C_{\Lambda, G_v}\) and \(C_{\Lambda', G_v}\), respectively. Since \(G_v\) is finite, the terms of \(\mathbb{C}(\mathbb{L}, \mathbb{M})\) for a \(G_v\)-module \(M\) are each naturally isomorphic to a finite direct sum of copies of \(M\). With these identifications, the canonical map

\[
\Lambda \otimes_R \mathbb{C}(\mathbb{L}) \rightarrow \mathbb{C}(\mathbb{L}, \Lambda \otimes_R T)
\]

agrees with the identity map, so is an isomorphism. Of course, we have the corresponding result for \(\Lambda\), and the isomorphism

\[
\Lambda' \otimes_R (\Lambda \otimes_R \mathbb{C}(\mathbb{L})) \rightarrow \Lambda' \otimes_R \mathbb{C}(\mathbb{L})
\]

in \(D(\mathbb{M})\) provides the \(v\)-summand of the lower isomorphism. □

**Remark 4.1.2.** Proposition \[4.1.1\] holds with “bounded above” removed if we suppose that the kernel of \(\Gamma \rightarrow \Gamma'\) has no elements of order \(p\). This follows quickly if \(p\) is odd or \(F\) has no real places, but it requires some work in the remaining case that \(F_{\infty}^p\) has no real places but \(D\) does. We omit this for purposes of brevity.

The next proposition shows in particular that if \(T\) is a bounded complex of objects of \(C_{R, G_{\infty}}\-\mathbb{R}\), then the global, local, and compactly supported cohomology groups of \(\mathbb{F}(T)\) are finitely generated \(\Lambda\)-modules. (Here, we implicitly identify cochain complexes with their quasi-isomorphic truncations in the derived category.)

**Proposition 4.1.3.** Let \(T\) be a complex of objects of \(C_{R, G_{\infty}}\-\mathbb{R}\).

(a) The complex \(\mathbb{R}\Gamma(G, \mathbb{F}(\mathbb{S}))\) lies in \(D_{\mathbb{L}}(\mathbb{M})\) if \(F_{\infty}^p\) has no real places or \(p\) is odd and in \(D_{\mathbb{L}}(\mathbb{M})\) if \(T\) is bounded above.

(b) The complex \(\mathbb{R}\Gamma(G, \mathbb{F}(\mathbb{S}))\) for \(v \in S_f\) lies in \(D_{\mathbb{L}}(\mathbb{M})\) if \(F_{\infty}^p\) has no real places or \(p\) is odd and in \(D_{\mathbb{L}}(\mathbb{M})\) if \(T\) is bounded below.

(c) If \(T\) is bounded and \(\mathbb{R}\Gamma(G, \mathbb{F}(\mathbb{S}))\) lie in \(D_{\mathbb{L}}(\mathbb{M})\) if \(T\) is bounded above.

\(\mathbb{R}\Gamma(G, \mathbb{F}(\mathbb{S}))\) and \(\mathbb{R}\Gamma(G, \mathbb{F}(\mathbb{S}))\) lie in \(D_{\mathbb{L}}(\mathbb{M})\).
(d) For $v \in S_f$, the complex $R\Gamma(G_v, \mathcal{F}_T(T))$ lies in $D_{\Lambda,-}(\text{Mod}_\Lambda)$ and in $D^b_{\Lambda,-}(\text{Mod}_\Lambda)$ if $T$ is bounded.

(e) The complex $R\Gamma(G_v, \mathcal{F}_T(T))$ for $v \in S_R$ lies in $D_{\Lambda,-}(\text{Mod}_\Lambda)$ if $T$ is bounded.

Proof. For real $v$, any $i \in \mathbb{Z}$ and $T$ concentrated in degree 0, the Tate cohomology group $\check{H}^i(G_v, \mathcal{F}_T(T))$ is a 2-torsion subquotient of $\mathcal{F}_T(T)$, as $G_v$ has order 2. Hence, it is finitely generated over $\Lambda$ as $\mathcal{F}_T(T)$ is. The hypercohomology spectral sequence allows one to pass to the case of a bounded complex $T$, and therefore $R\Gamma(G_v, \mathcal{F}_T(T))$ lies in $D_{\Lambda,-}(\text{Mod}_\Lambda)$, hence (e). Recall also that $R\Gamma(G_v, \mathcal{F}_T(T))$ is a zero object for all complexes $T$ if $v$ extends to a complex place of $F_\infty$.

If $v \in S$ is nonarchimedean, then $G_v$ is $p$-cohomologically finite by our above remarks, and hence Theorem 3.2.4 implies that $R\Gamma(G_v, \mathcal{F}_T(T))$ lies in $D_{\Lambda,-}(\text{Mod}_\Lambda)$. In the case that $p$ is odd or $F_\infty$ has no real places, Theorem 3.2.4 again implies that $R\Gamma(G_v, \mathcal{F}_T(T))$ lies in $D_{\Lambda,-}(\text{Mod}_\Lambda)$. (For this, note that if $p = 2$ and $F$ has a real place that becomes complex in $F_\infty$, it will already have become complex in any extension with Galois group $\Gamma/\Gamma_0$, where $\Gamma_0$ is an open uniform pro-$p$ subgroup of $\Gamma$.) In this case, it follows from the above-mentioned exact triangle that $R\Gamma(G_v, \mathcal{F}_T(T))$ is in $D_{\Lambda,-}(\text{Mod}_\Lambda)$.

Finally, if $T$ is bounded, then clearly all of the above complexes will additionally lie in the bounded derived category. In particular, we have (c) and (d) and part of each of (a) and (b).

The analogue of Theorem 3.2.4 for compactly supported cohomology holds for bounded above $T$ since $H^i(G_{F,S}, \mathcal{F}_T(T))$ vanishes for sufficiently large $i$. The proof is essentially identical but uses the semilocal version of Shapiro's lemma [Nek 8.5.3.2] in the first step and the relevant hypercohomology spectral sequence in the last. Thus we have (a).

Finally, if $T$ is concentrated in degree 0, the exact triangle tells us that $H^i(G_{F,S}, \mathcal{F}_T(T))$ is finitely generated over $\mathcal{F}_T(T)$ for sufficiently large $i$. Thus, we are still able to perform the inductive step in the proof of Theorem 3.2.4 to obtain that $R\Gamma(G_{F,S}, \mathcal{F}_T(T))$ sits in $D^+_{\Lambda,-}(\text{Mod}_\Lambda)$ for such $T$, and then for bounded below $T$ via the hypercohomology spectral sequence. Thus, we have (b).

Remark 4.1.4. Part (d) of Proposition 4.1.3 makes sense and holds more generally for any complex $T$ in $C^{R,-1}_{\Lambda,\text{Mod}_\Lambda}$, as $G_v$-action through the composite with the canonical map $G_v \to G_{F,S}$.

Remark 4.1.5. We may also consider the setting in which we take $F$ itself to be a nonarchimedean local field of characteristic not equal to $p$. We then let $G_F$ be its absolute Galois group, let $\Gamma$ denote the Galois group of a $p$-adic Lie extension of $F$, and as before, set $\Lambda = \mathbb{F}_p[\Gamma]$. In this case, we obtain immediately from Theorem 3.2.4 that the cohomology groups $H^j(G_F, \mathcal{F}_T(T))$ are finitely generated $\Lambda$-modules for all $j$ and any complex of objects of $T$ of

Documenta Mathematica 18 (2013) 621–678
for any complex $T$ of objects in $\mathcal{C}_{R,G_F}^{R\text{-ft}}$.

4.2 Duality over local fields

In this subsection, we will state a version of Tate duality for $\Gamma$-induced modules with $\Lambda \otimes R \omega_R$ for a dualizing complex $\omega_R$ replacing $\mathbb{Q}_p/\mathbb{Z}_p$. Fix a nonarchimedean prime $v \in S$. We continue to suppose that $\Gamma$ is a compact $p$-adic Lie group that is a quotient of $G_{F,S}$ defining an extension $F_\infty$, though as we remark later, we could just as well assume it to be a quotient of $G_v$. Let $T$ be a bounded complex of objects in $\mathcal{C}_{R,G_F}^{R\text{-ft}}$. As before, we let $J_R$ denote a complex of injective $R$-modules that represents $\omega_R$, and we let $T^*$ be a quasi-isomorphic subcomplex of $\text{Hom}_R(T,J_R)$ consisting of objects of $\mathcal{C}_{R,G_F}^{R\text{-ft}}$.

The cup product of Lemma 4.3.9 yields a map

$$C(G_v, \mathcal{F}_T(T)) \otimes_R C(G_v, \mathcal{F}_T(T^*)^i(1)) \longrightarrow \tau_{\geq 2}^{I_I} C(G_v, \Lambda \otimes_R J_R(1))$$

of $\Lambda \otimes_R \Lambda^\circ$-modules. Taking adjoints, we have the following maps of complexes of $\Lambda$-modules and $\Lambda^\circ$-modules, respectively:

$$C(G_v, \mathcal{F}_T(T)) \longrightarrow \text{Hom}_\Lambda \left( C(G_v, \mathcal{F}_T(T^*)^i(1)), \tau_{\geq 2}^{I_I} C(G_v, \Lambda \otimes_R J_R(1)) \right),$$

$$C(G_v, \mathcal{F}_T(T^*)^i(1)) \longrightarrow \text{Hom}_\Lambda \left( C(G_v, \mathcal{F}_T(T)), \tau_{\geq 2}^{I_I} C(G_v, \Lambda \otimes_R J_R(1)) \right).$$

From now on, we use $M^\circ$ to denote the Pontryagin dual of a locally compact abelian group $M$. We have the following analogue of [Nek] (5.2.1).

**Lemma 4.2.1.** Let $A$ be a bounded complex in $\mathcal{C}_{\Lambda,\Lambda}^{\Lambda^\circ -$ft with terms that are flat as $R$-modules. Then there exists a quasi-isomorphism

$$q_A: (A \otimes_R J_R)[-2] \longrightarrow \tau_{\geq 2}^{I_I} C(G_v, A \otimes_R J_R(1)),$$

natural in $A$, of chain complexes of $\Lambda \otimes_R \Lambda^\circ$-modules.

**Proof.** We first note that if $M$ is any finitely generated $R$-module with a trivial action of $G_v$, then Tate local duality as in [Lim2] Theorem 4.1.2 yields isomorphisms

$$H^2(G_v, A \otimes_R M(1)) \cong H^0(G_v, (A \otimes_R M)^\circ)^\circ = A \otimes_R M,$$

of $\Lambda \otimes_R \Lambda^\circ$-modules. Now suppose that $M$ is any (ind-admissible) $R[G_v]$-module with trivial $G_v$-action, and write $M = \bigcup M_i$, where the $M_i$ are finitely generated $R$-submodules of $M$. Since the terms of $A$ are $R$-flat, $A^i \otimes_R M$ is the union of the $A^i \otimes_R M_i$ for any $i \in \mathbb{Z}$, so

$$H^2(G_v, A \otimes_R M(1)) \cong \lim_{\longrightarrow} H^2(G_v, A \otimes_R M_i(1)) \cong \lim_{\longrightarrow} A \otimes_R M_i \cong A \otimes_R M.$$
As $G_v$ is of $p$-cohomological dimension 2, we therefore have a quasi-isomorphism

$$A \otimes_R M[-2] \to \tau_{\geq 2} \mathcal{C}(G_v, A \otimes_R M(1)),$$

of complexes of $\Lambda \otimes_R \Lambda^\circ$-modules. The case of a bounded complex $\mathcal{M}$ of ind-admissible $R$-modules, e.g. $M = J_R$, then follows easily. Naturality is immediate from the construction.

Combining Lemma 4.2.1 for $A = \Lambda$ with the morphisms constructed above and passing to the derived category, we obtain morphisms as in the following theorem.

**Theorem 4.2.2.** Let $T$ be a bounded complex in $\mathcal{C}_{R,G_v}$. Then the morphisms

$$\mathcal{R}\Gamma(G_v, \mathcal{F}_T(T)) \to \mathcal{R}\text{Hom}_\Lambda\left(\mathcal{R}\Gamma(G_v, \mathcal{F}_T(T^\ast)(1)), \Lambda \otimes^I_R \omega_R\right)[-2]
\mathcal{R}\Gamma(G_v, \mathcal{F}_T(T^\ast)(1)) \to \mathcal{R}\text{Hom}_\Lambda\left(\mathcal{R}\Gamma(G_v, \mathcal{F}_T(T)), \Lambda \otimes^I_R \omega_R\right)[-2]$$

in $\mathcal{D}_{\Lambda-R}(\text{Mod}_\Lambda)$ and $\mathcal{D}_{\Lambda-R}(\text{Mod}^\ast_{\Lambda})$, respectively, are isomorphisms.

Over the course of the next two subsections, we will show that the first of the morphisms of Theorem 4.2.2 is an isomorphism. The proof that the second is an isomorphism is completely analogous. By an argument similar to that in [Nek, Lemma 5.2.5], it suffices to prove the theorem for a particular translate of the dualizing complex. In view of this, we will assume that our choice of the dualizing complex satisfies [Nek, (2.5)(i)] throughout the next two subsections.

**Remark 4.2.3.** In the setting that $F$ is a nonarchimedean local field of characteristic not equal to $p$, that $F_\infty$ is any $p$-adic Lie extension of $F$, and that $T$ is a bounded complex of objects in $\mathcal{C}_{R,G_F}$, the argument we are about to describe also carries over as in Remark 4.1.5 to prove the direct analogue of Theorem 4.2.2 with $G_v$ replaced by $G_F$.

### 4.3 Change of rings

Recall that $R$ is a complete commutative local Noetherian ring with maximal ideal $m$ and finite residue field $k$. In this subsection, we shall show that it suffices to prove Theorem 4.2.2 for any complete commutative local Noetherian ring $S$ that has $R$ as a quotient. (In this subsection only, $S$ denotes such a ring, rather than the set of primes chosen above, which is used only implicitly.)

We suppose given a surjection $\psi: S \to R$, which necessarily induces an isomorphism on residue fields. We use $m_S$ to denote the maximal ideal of $S$, and we set

$$d = \dim_k m_S/m_S^2 - \dim_k m/m^2.$$

Note that if $A$ is any complex of $S$-modules, then $\text{Hom}_S(R, A)$ is isomorphic to the subcomplex of $A$ with terms the torsion submodules under $\ker\psi$. In
particular, if $B$ is a complex of $R$-modules, then the resulting inclusion map induces an isomorphism

$$\text{Hom}_R(B, \text{Hom}_S(R, A)) \xrightarrow{\sim} \text{Hom}_S(B, A),$$

as is easy enough to see by the usual adjointness principle. In particular, we have

$$\text{Hom}_R(B, R^\vee) \xrightarrow{\sim} \text{Hom}_S(B, S^\vee).$$

With this comparison of Matlis duals with respect to $R$ and $S$ in hand, we now compare Grothendieck duals with respect to these rings (see also [Har2]). For this, we choose particular dualizing complexes $J_R$ and $J_S$ of injectives as follows. Let $\bar{x}_1, \ldots, \bar{x}_r \in \mathfrak{m}$ with images forming a $k$-basis of $\mathfrak{m}/\mathfrak{m}^2$. Recall (cf. [Nek, Section 0.4]) that we may take $J_R$ to be the Matlis dual of $C_R[r]$, where $C_R$ is defined to be the complex

$$[R \to \bigoplus_i R_{\bar{x}_i} \to \bigoplus_{i<j} R_{\bar{x}_i \bar{x}_j} \to \cdots \to R_{\bar{x}_1 \cdots \bar{x}_r}]$$

in degrees $[0, r]$, with the usual Čech differentials, which is to say that

$$J_R = \text{Hom}_R(C_R[r], R^\vee).$$

Note that $J_R$ is a complex of $R$-injectives by the $R$-flatness of the terms of $C_R$ and the $R$-injectivity of $R^\vee$. We lift $\bar{x}_1, \ldots, \bar{x}_r$ to elements $x_1, \ldots, x_r$ of $\mathfrak{m}_S$ and extend to a sequence $x_1, \ldots, x_s$ such that the images of $x_1, \ldots, x_s$ form a basis of $\mathfrak{m}_S/\mathfrak{m}_S^2$ and the elements $x_{r+1}, \ldots, x_s$ map trivially to $\mathfrak{m}/\mathfrak{m}^2$. Modifying each $x_i$ with $i > r$ by an element of $\mathfrak{m}_S^2$ if necessary, we may choose these elements so that they in fact map trivially to $\mathfrak{m}$. We use this sequence to define $C_S$, and we set $J_S = \text{Hom}_S(C_S[s], S^\vee)$.

**Lemma 4.3.1.** Let $T$ be a bounded complex in $C_{R,G}^{0}$, which we view also as a complex in $C_{S,G}^{-r}$ via $\psi$. Then $\psi$ induces a natural map $J_R \to J_S[d]$ and in turn an isomorphism

$$\text{Hom}_R(T, J_R) \xrightarrow{\sim} \text{Hom}_S(T, J_S)[d]$$

of complexes in $\mathcal{A}_{S,G}$. 

**Proof.** The natural map $C_S \to C_R$ factors through an isomorphism $R \otimes_S C_S \xrightarrow{\sim} C_R$. This induces an isomorphism

$$J_R \xrightarrow{\sim} \text{Hom}_S(R, J_S)[d],$$

which is given by the following composition of isomorphisms

$$\text{Hom}_R(C_R[r], R^\vee) \xrightarrow{\sim} \text{Hom}_S(C_R[r], S^\vee) \xrightarrow{\sim} \text{Hom}_S(R \otimes_S C_S[r], S^\vee)$$

$$\xrightarrow{\sim} \text{Hom}_S(R, \text{Hom}_S(C_S[r], S^\vee)) \xrightarrow{\sim} \text{Hom}_S(R, \text{Hom}_S(C_S[s], S^\vee))[d],$$

\[\text{Documenta Mathematica} 18 (2013) 621–678\]
where the latter natural map is defined as in [Nek (1.2.15)] (noting [Nek
(1.2.5)]). The result then follows from the isomorphisms

\[ \text{Hom}_R(T, J_R) \sim \text{Hom}_R(T, \text{Hom}_S(R, J_S))[d] \sim \text{Hom}_S(T, J_S)[d]. \]

We now proceed to the reductive step.

**Proposition 4.3.2.** Let \( T \) be a bounded complex in \( \mathcal{C}^{R-\text{ft}}_{R, G_v} \). Setting \( \Omega = S[\Gamma] \), there is a commutative diagram of natural morphisms

\[
\begin{array}{ccc}
\mathbf{R}\Gamma(G_v, \mathcal{F}(T)) & \longrightarrow & \mathbf{R}\text{Hom}_{\Lambda^c}(\mathbf{R}\Gamma(G_v, \mathcal{F}(T)^{\ast}(1)), \Lambda \otimes_R \omega_R)[-2] \\
\downarrow \iota & & \downarrow \iota \\
\mathbf{R}\Gamma(G_v, \Omega \otimes_S T) & \longrightarrow & \mathbf{R}\text{Hom}_{\Omega^c}(\mathbf{R}\Gamma(G_v, \Omega \otimes_S T)^{\ast}(1), \Omega \otimes_S \omega_S)[d] [-2]
\end{array}
\]

in which the horizontal morphisms are as in Theorem 4.2.2 and the vertical arrows are isomorphisms in \( \mathbf{D}(\text{Mod}_\Omega) \). In particular, Theorem 4.2.2 holds for \( R \) if it holds for \( S \).

**Proof.** Note that we have

\( \Omega \otimes_S T \cong \lambda \otimes_R T \) (resp., \( \Omega \otimes_S T^\ast \cong \lambda \otimes_R T^\ast \))

so \( \mathbf{R}\Gamma(G_v, \mathcal{F}(T)) \) (resp., \( \mathbf{R}\Gamma(G_v, \mathcal{F}(T)^{\ast}(1)) \)) is the same complex for \( R \) and for \( S \). By Lemma 4.3.1 and employing the \( S \)-flatness of \( \Omega \) in applying Lemma 2.1.4, we have canonical isomorphisms

\[ \lambda \otimes_R J_R \sim \Omega \otimes_S J_R \sim \Omega \otimes_S \text{Hom}_S(R, J_S)[d] \sim \text{Hom}_S(R, \Omega \otimes_S J_S)[d], \]

of complexes of \( \Omega^c \otimes_R R \cong \mathcal{A}^c \)-modules, and the composition of the composite isomorphism with the natural inclusion

\[ \text{Hom}_S(R, \Omega \otimes_S J_S)[d] \longrightarrow \Omega \otimes_S J_S[d] \]

induces the right-hand vertical map in the proposition. That the diagram commutes then follows directly from the definition of the cup product in Lemma 3.3.3.

Let us fix a quasi-isomorphism \( \iota: \Omega \otimes_S J_S \rightarrow K \), where \( K \) is a bounded below complex of injective \( \Omega^c \)-modules. We check that the map

\[ \iota_R: \text{Hom}_S(R, \Omega \otimes_S J_S) \longrightarrow \text{Hom}_S(R, K) \]

induced by \( \iota \) is a quasi-isomorphism, which implies that \( \text{Hom}_S(R, \Omega \otimes_S J_S) \) represents \( \mathbf{R}\text{Hom}_S(R, \Omega \otimes_S \omega_S) \) in \( \mathbf{D}(\text{Mod}_{\Lambda^c}) \). Let \( \varepsilon: P \rightarrow R \) be a resolution of
R by a complex $P$ of finitely generated free $S$-modules. We have a commutative diagram

$$
\begin{array}{cccccc}
\Omega \otimes_S \text{Hom}_S(R, J) & \overset{\phi_R}{\longrightarrow} & \text{Hom}_S(R, \Omega \otimes_S J) & \overset{\psi_R}{\longrightarrow} & \text{Hom}_\Omega^\circ(\Omega \otimes_S R, K) \\
\downarrow \epsilon \otimes_S \iota & & \downarrow \iota & & \downarrow (\iota \otimes_S) K \\
\Omega \otimes_S \text{Hom}_S(P, J) & \overset{\phi_P}{\longrightarrow} & \text{Hom}_S(P, \Omega \otimes_S J) & \overset{\psi_P}{\longrightarrow} & \text{Hom}_\Omega^\circ(\Omega \otimes_S P, K)
\end{array}
$$

in which the vertical maps are all induced by $\epsilon$. Moreover, note that the upper horizontal morphisms are of complexes of $\Lambda^\circ$-modules and the others are of complexes of $\Omega^\circ$-modules. The maps $\phi_P$ and $\phi_R$ are isomorphisms, and the map $\iota$ is a quasi-isomorphism by the $\Lambda^\circ$-injectivity of $J$ and the $\Lambda^\circ$-flatness of $\Omega$, so $\epsilon_{\Omega \otimes_S, P}$ is also a quasi-isomorphism. The maps $\psi_R$ and $\psi_P$ are isomorphisms by the usual adjointness of $\text{Hom}$ and the tensor product, and the map $(\iota \otimes_S) K$ is a quasi-isomorphism by the $\Omega^\circ$-injectivity of $K$, so $\epsilon_K$ is a quasi-isomorphism as well. Finally, the map $\iota$ is a quasi-isomorphism by the $\Lambda^\circ$-projectivity of $P$, and it follows that $\iota_R$ is a quasi-isomorphism.

To finish the proof, we need only show that the morphism

$$\text{RHom}_{\Lambda^\circ}(X, \text{RHom}_S(R, \Omega \otimes_S \omega)) \longrightarrow \text{RHom}_{\Omega^\circ}(X, \Omega \otimes_S \omega_S)$$

is an isomorphism for $X = C(G_v, \mathcal{F}_T(T^*)^!(1))$. It is easy to see that $\text{Hom}_S(R, K)$ is a complex of injective $\Lambda^\circ$-modules. Moreover, every $\Omega^\circ$-homomorphism from the complex $X$ of $\Lambda^\circ$-modules to $K$ must factor through $\text{Hom}_\Omega^\circ(\Lambda, K) \cong \text{Hom}_\Omega^\circ(\Omega \otimes_S R, K) \cong \text{Hom}_S(R, K)$ of $K$, so the map

$$\text{Hom}_{\Lambda^\circ}(X, \text{Hom}_S(R, K)) \longrightarrow \text{Hom}_{\Omega^\circ}(X, K)$$

induced by inclusion is an isomorphism, as desired.

4.4 Duality over flat $\mathbb{Z}_p$-algebras

In this subsection, we will prove Theorem 4.2.2. As remarked in the proof of Proposition 4.3.1, every complete commutative Noetherian local ring with finite residue field of characteristic $p$ is a quotient of a power series ring in finitely many variables over an unramified extension of $\mathbb{Z}_p$. Proposition 4.3.2 then allows us to reduce Theorem 4.2.2 to the case of such power series rings. We therefore can and do assume that $R$ is $\mathbb{Z}_p$-flat throughout this subsection.

We will require a bit more general of a derived adjoint map than the one we wish to prove is an isomorphism. The construction is found in the following easy lemma.

**Lemma 4.4.1.** Let $A$ be a bounded complex of $R$-flat objects in $\mathcal{C}^{\Lambda^\circ\text{-}\otimes_R}_{\Lambda\text{-}}$. Then we have a morphism

$$\text{RHom}(G_v, A \otimes_\Lambda \mathcal{F}_T(T)) \longrightarrow \text{RHom}_{\Lambda^\circ}(\text{RHom}(G_v, \mathcal{F}_T(T^*)^!(1)), A \otimes_R \omega_R)[{-2}]$$
in $\mathbf{D}(\text{Mod}_\Lambda)$ that is natural in the complex $A$.

Proof. This is simply the derived adjoint of the composition of the cup product in Lemma 3.3.3 with the map to the truncation $\tau_{\geq 2}^1 C(G_v, A \otimes_R J_R(1))$, which is quasi-isomorphic to $(A \otimes_R J_R)[-2]$ by Lemma 4.2.1. Naturality is immediate. □

We will mostly be interested in complexes of length 2, so we introduce the following notation. Suppose that $B$ is an object in $C_{\Lambda^0, \Lambda}$ and that $A$ is a subobject of $B$, which is to say that it is a closed $\Lambda \hat{\otimes}_R \Lambda^0$-submodule with the subspace topology. Then for any bounded complex $T$ of objects in $C_{R, G_v}$, we define a complex $F_{B/A}(T) = [A \rightarrow B] \otimes_{\Lambda} F_T(T)$, of objects in $C_{\Lambda, G_v}$, where $A$ and $B$ are in degree -1 and 0 respectively.

Let $I = I(\Gamma)$ denote the augmentation ideal of $\Lambda$. We then have the following lemmas. Recall for the first that we assume that $R$ is $\mathbb{Z}_p$-flat in this subsection.

Lemma 4.4.2. The following statements hold for any $n \geq 0$.

(a) The ideal $I^n$ is a flat $R$-module.

(b) The module $I^n/I^{n+1}$ is finitely generated and of finite projective dimension over $R$.

Proof. (a) Suppose first that $\Gamma$ is finite, and let $I$ denote the augmentation ideal in $\mathbb{Z}_p[\Gamma]$. As $\mathbb{Z}_p[\Gamma]$ is finitely generated and free over $\mathbb{Z}_p$, so is $I^n$. Since $R$ is $\mathbb{Z}_p$-flat, the natural surjection $R \otimes_{\mathbb{Z}_p} I^n \rightarrow I^n$ is an isomorphism, and therefore $I^n$ is free over $R$.

In the general case, $\Gamma \cong \lim_{\leftarrow} \Gamma'$, where $\Gamma'$ runs over the Galois groups of the finite Galois extensions of $F$ in $F_\infty$, and $I^n$ is the inverse limit of the $n$th powers of the augmentation ideals $I(\Gamma')$ of the $R[\Gamma']$. As $I(\Gamma')^n \cong I^n \otimes_R R[\Gamma']$, an application of Lemma 2.4.2 yields an isomorphism

$$I^n \otimes_R B \xrightarrow{\sim} \lim_{\Gamma'} (I(\Gamma')^n \otimes_R B)$$

for any ideal $B$ of $R$. The left exactness of the inverse limit and the $R$-flatness of $I(\Gamma')^n$ then imply that the canonical map $I^n \otimes_R B \rightarrow I^n$ is an injection, proving the flatness of $I^n$.

(b) Let $I$ denote the augmentation ideal in $\mathbb{Z}_p[\Gamma]$. For each $n$, the composition

$$R \otimes_{\mathbb{Z}_p} I^n \rightarrow R \otimes_{\mathbb{Z}_p} \mathbb{Z}_p[\Gamma] \rightarrow R[\Gamma]$$

induces a surjection $R \otimes_{\mathbb{Z}_p} I^n \rightarrow I^n$ that fits into the following commutative diagram with exact rows:

$$\begin{array}{ccc}
R \otimes_{\mathbb{Z}_p} I^{n+1} & \rightarrow & R \otimes_{\mathbb{Z}_p} I^n \rightarrow R \otimes_{\mathbb{Z}_p} (I^n/I^{n+1}) \rightarrow 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \rightarrow & I^{n+1} & \rightarrow & I^n & \rightarrow & I^n/I^{n+1} & \rightarrow & 0.
\end{array}$$

Documenta Mathematica 18 (2013) 621–678
Since the two vertical maps on the left are surjections, so is the one on the right. Note that $I^n/I^{n+1}$ is a quotient of $\mathbb{Z}_p$-module. Therefore, the existence of the above surjection implies that $I^n/I^{n+1}$ is finitely generated over $R$. To see that the last assertion holds, we note that it follows from (a) that $I^n/I^{n+1}$ has finite flat dimension. Since flat dimension coincides with projective dimension for every finitely generated module over a Noetherian ring (see [Wei, Proposition 4.1.5]), we have our assertion.

**Lemma 4.4.3.** Let $T$ be a bounded complex of objects in $\mathcal{C}_{R,G_\Lambda}^{\mathbb{N}-f}$. Then we have a quasi-isomorphism

$$\text{Cone}\left(\mathcal{F}_{I^n/I^{n+1}}(T) \longrightarrow \mathcal{F}_{\Lambda/I^{n+1}}(T)\right) \sim \mathcal{F}_{\Lambda/I^n}(T)$$

of complexes in $\mathcal{C}_{\Lambda,G_\Lambda}^{\mathbb{N}-f}$. Moreover, we have an exact triangle

$$I^n/I^{n+1} \otimes_R \omega_R \longrightarrow \Lambda/I^n \otimes_R \omega_R \longrightarrow \Lambda/I^n \otimes \Lambda$$

in $D^b(\text{Mod}_{\Lambda \otimes_R \Lambda})$.

**Proof.** Since the powers of $I$ are $R$-flat by Lemma 4.4.2(a), it suffices in both cases to show that there is a quasi-isomorphism

$$\text{Cone}\left([I^{n+1} \rightarrow I^n] \longrightarrow [I^{n+1} \rightarrow \Lambda]\right) \sim [I^n \rightarrow \Lambda].$$

Note that the latter cone is precisely the complex

$$I^{n+1} \xrightarrow{f} I^{n+1} \oplus I^n \xrightarrow{g} \Lambda,$$

where $f(x) = (x, -x)$ and $g(x, y) = x + y$ for $x \in I^{n+1}$ and $y \in I^n$. One can now easily check that the diagram

$$\begin{array}{c}
I^{n+1} \xrightarrow{f} I^{n+1} \oplus I^n \xrightarrow{g} \Lambda \\
\downarrow \quad \downarrow \nu \quad \downarrow \nu \\
0 \xrightarrow{} I^n \xrightarrow{} \Lambda
\end{array}$$

commutes, where $\nu$ is given by $\nu(x, y) = x + y$ for $x \in I^{n+1}$ and $y \in I^n$, and the vertical maps induce isomorphisms on cohomology. We are now able to prove the following proposition, which is an important ingredient in the proof of Theorem 4.2.2.
Lemma 4.4.4. Let $T$ be a bounded complex of objects in $\mathcal{C}_{\Lambda,G}$. Then we have the following isomorphisms

\[
\begin{array}{c}
\text{Proof.}\quad \text{By Lemma 4.4.2(a), we see that} \quad I^n/I^{n+1} \otimes_R \omega_R \text{ and } \Lambda/I^n \otimes_R \omega_R \text{ are represented by } [I^{n+1} \to I^n] \otimes_R J_R \text{ and } [I^n \to \Lambda] \otimes_R J_R \text{ respectively. Therefore, the commutativity of the diagram in the proposition follows from the naturality in Lemma 4.4.1. By Lemma 4.4.3 both columns are exact triangles.}
\end{array}
\]

We now describe the idea of the proof of Theorem 4.2.2. We shall first prove that the morphism

\[
\text{Proposition 4.4.4 is a quasi-isomorphism, so is the third one. Therefore, by an inductive argument, we are reduced to showing that the morphism}
\]

\[
\begin{array}{c}
\text{is an isomorphism for all } n. \text{ Then, Theorem 4.2.2 will follow from this by a limit argument. To show that above morphism is an isomorphism, we will utilize Proposition 4.4.4. Note that if any two of the horizontal morphisms in Proposition 4.4.4 is a quasi-isomorphism, so is the third one. Therefore, by an inductive argument, we are reduced to showing that the morphism}
\end{array}
\]

\[
\begin{array}{c}
\text{is an isomorphism for all } n \geq 0. \text{ Note that } \Gamma \text{ acts trivially on } I^n/I^{n+1}. \text{ Therefore, one may view } I^n/I^{n+1} \text{ as a } \Lambda^0\text{-module via the augmentation map } \Lambda \to R. \text{ We now have the following lemma.}
\end{array}
\]

Lemma 4.4.5. Let $T$ be a bounded complex of objects in $\mathcal{C}_{\Lambda,G}$. Then we have the following isomorphisms

\[
\begin{array}{c}
\text{in } \mathbf{D}(\mathcal{C}_{\Lambda,G}). \text{ Therefore, we have an isomorphism}
\end{array}
\]

\[
\text{in } \mathbf{D}^b(\mathcal{C}_{\Lambda,G}).
\]
Proof. Let $P$ be a resolution of $I^n/I^{n+1}$ consisting of projective objects in $\mathcal{C}_{\Lambda^{-\Lambda}}$. Then there is a quasi-isomorphism $P \to [I^{n+1} \to I^n]$ of complexes in $\mathcal{C}_{\Lambda^{-\Lambda}}$ which lifts the identity map on $I^n/I^{n+1}$. By Lemmas 2.4.3 and 4.4.2(a) this induces a quasi-isomorphism

$$P \otimes \Lambda T \to [I^{n+1} \to I^n] \otimes \Lambda T$$

of bounded above complexes in $\mathcal{C}_{\Lambda,G_v}$, proving the first isomorphism. The second isomorphism is a special case of Lemma 3.1.6 with $N = I^n/I^{n+1}$ and $\Gamma'$ trivial. Note that both quasi-isomorphisms of complexes factor through a quasi-isomorphic truncation of $P \otimes \Lambda T$, so they may be seen in the bounded derived category.

**Lemma 4.4.6.** For each $n$, there is a commutative diagram

$$\begin{array}{cccc}
I^n/I^{n+1} \otimes_R \mathbf{R}\Gamma(G_v, T) & \longrightarrow & I^n/I^{n+1} \otimes_R \mathbf{R}\mathrm{Hom}_R\left(\mathbf{R}\Gamma(G_v, T^*(1)), \omega_R\right)[-2] \\
\downarrow & & \downarrow \\
\mathbf{R}\Gamma(G_v, I^n/I^{n+1} \otimes_R T) & \longrightarrow & \mathbf{R}\mathrm{Hom}_R\left(\mathbf{R}\Gamma(G_v, T^*(1)), I^n/I^{n+1} \otimes_R \omega_R\right)[-2] \\
\downarrow & & \downarrow \\
\mathbf{R}\mathrm{Hom}_R\left(\mathbf{R}\Gamma(G_v, \mathcal{F}_T(T^*(1)) \otimes_R J_R(1)), I^n/I^{n+1} \otimes_R \omega_R\right)[-2] & & \mathbf{R}\mathrm{Hom}_{\Lambda} \left(\mathbf{R}\Gamma(G_v, \mathcal{F}_T(T^*(1))), I^n/I^{n+1} \otimes_R \omega_R\right)[-2],
\end{array}$$

where the vertical morphisms are isomorphisms in $\mathbf{D}(\text{Mod}_R)$.

**Proof.** By Lemma 4.4.2(b), we may choose a bounded resolution $Q$ of $I^n/I^{n+1}$ by finitely generated projective $R$-modules. By [Nek] Proposition 3.4.4], we have an isomorphism of complexes

$$\alpha : Q \otimes R C(G_v, T) \simto C(G_v, Q \otimes R T)$$

that fits into the commutative diagram

$$\begin{array}{cccc}
Q \otimes R C(G_v, T) \otimes R C(G_v, T^*(1)) & \longrightarrow & Q \otimes R C(G_v, J_R(1)) \\
\alpha \otimes \text{id} & & \alpha' \\
C(G_v, Q \otimes R T) \otimes R C(G_v, T^*(1)) & \longrightarrow & C(G_v, Q \otimes R J_R(1)),
\end{array}$$

where $\alpha'$ is defined analogously to $\alpha$ and is also an isomorphism. Since $J_R$ is a bounded complex of $R$-injectives, we may find homotopy inverses to the quasi-
isomorphisms $q_R$ and $q_Q$ of Lemma 4.2.1 that fit in a commutative diagram

\[
\begin{array}{ccc}
Q \otimes_R C(G_v, J_R(1)) & \xrightarrow{\alpha'} & Q \otimes_R \tau^{II}_{\geq 2}C(G_v, J_R(1)) \\
\downarrow & & \downarrow \\
C(G_v, Q \otimes_R J_R(1)) & \xrightarrow{\tau^{II}_{\geq 2}} & Q \otimes_R J_R[2].
\end{array}
\]

We obtain the top commutative square in the lemma and the fact that the vertical morphisms therein are isomorphisms, the one on the right by Lemma 2.1.8 and [Nek, Proposition 4.2.3].

Let $P$ be a resolution of $I_n/I_{n+1}$ consisting of finitely generated projective $\Lambda$-modules. We may view $Q$ as a resolution of $\Lambda^\circ$-modules via the augmentation map $\Lambda \rightarrow R$. Then, by Proposition 2.4.6, the map

\[
P \otimes \Lambda F(T) = P \otimes \Lambda^\circ T \rightarrow Q \otimes \Lambda^\circ T \cong Q \otimes_R T
\]

is a quasi-isomorphism of complexes of objects in $C_{R, G}$, and we let $f$ denote its induced map on cochains. Let $L$ be a resolution of $R$ consisting of finitely generated projective $\Lambda$-modules. Then, by an opposite version of Theorem 3.1.8, we have a quasi-isomorphism

\[
g: C(G_v, F_T(T^\ast)(1)) \otimes_A L \rightarrow C(G_v, T^\ast(1))
\]

of complexes of $R$-modules. Noting Lemma 5.3.3 we have a commutative diagram

\[
\begin{array}{ccc}
C(G_v, P \otimes_A F_T(T)) \otimes_R C(G_v, F_T(T^\ast)(1)) \otimes_A L & \xrightarrow{f \otimes \varepsilon} & C(G_v, P \otimes_R J_R(1)) \otimes_A L \\
\downarrow & & \downarrow \\
C(G_v, Q \otimes_R T) \otimes_R C(G_v, T^\ast(1)) & \xrightarrow{\varepsilon} & C(G_v, Q \otimes_R J_R(1)),
\end{array}
\]

where $\varepsilon$ is induced by the augmentation $L \rightarrow R$ and the map $P \rightarrow Q$. Taking adjoints and applying the homotopy inverse to $q_Q$ as above, we obtain the commutative diagram

\[
\begin{array}{ccc}
C(G_v, Q \otimes_R T) & \xrightarrow{\text{Hom}_R} & \text{Hom}_R\left(C(G_v, T^\ast(1)), Q \otimes_R J_R\right)[-2] \\
\downarrow & & \downarrow \\
C(G_v, P \otimes_A F_T(T)) & \xrightarrow{\text{Hom}_R} & \text{Hom}_R\left(C(G_v, F_T(T^\ast)(1)) \otimes_A L, Q \otimes_R J_R\right)[-2] \\
\downarrow & & \downarrow \\
C(G_v, P \otimes_A F_T(T)) & \xrightarrow{\text{Hom}_A} & \text{Hom}_R\left(C(G_v, F_T(T^\ast)(1)), \text{Hom}_R(L, Q \otimes_R J_R)\right)[-2],
\end{array}
\]

which yields upon passage to the derived category the lower part of the diagram in the statement of the lemma, the vertical morphisms therein being isomorphisms by Theorem 3.1.8, Lemma 4.4.5 and Lemma 2.1.6.
**Lemma 4.4.7.** The morphisms

\[ R\Gamma(G_v, \mathcal{F}_{I^n/I^{n+1}}(T)) \to R\text{Hom}_{\Lambda^c}\left(R\Gamma(G_v, \mathcal{F}_T(T^*)^!(1)), I^n/I^{n+1} \otimes_R \omega_R\right)[-2] \]

in \( D(\text{Mod}_\Lambda) \) are isomorphisms for every \( n \geq 0 \).

**Proof.** By Corollary 2.1.3 it suffices to show that the above morphism is an isomorphism in \( D(\text{Mod}_R) \). The morphism

\[ R\Gamma(G_v, T) \to R\text{Hom}_R\left(R\Gamma(G_v, T^*(1)), \omega_R\right)[-2] \]

in \( D(\text{Mod}_R) \) is an isomorphism by [Nek, Proposition 5.2.4(ii)], and so the top morphism of the diagram in Lemma 4.4.6 is an isomorphism. Since all the vertical morphisms in the diagram are isomorphisms, it follows that the bottom morphism is also an isomorphism, as required.

**Proposition 4.4.8.** The morphisms

\[ R\Gamma(G_v, \mathcal{F}_{A/I^n}(T)) \to R\text{Hom}_{\Lambda^c}\left(R\Gamma(G_v, \mathcal{F}_T(T^*)^!(1)), \Lambda/I^n \otimes_R \omega_R\right)[-2] \]

in \( D(\text{Mod}_\Lambda) \) are isomorphisms for every \( n \geq 1 \).

**Proof.** As seen in the above discussion, the preceding lemma allows us to perform an inductive argument using the morphism of exact triangles in Proposition 4.4.4 to obtain the required conclusion.

We finish the proof of Theorem 4.2.2 by passing to the inverse limit.

**Proof of Theorem 4.2.2.** By Remark 4.1.4 there exists a quasi-isomorphism

\[ W \to C(G_v, \mathcal{F}_T(T^*)^!(1)) \]

with \( W \) a bounded above complex of finitely generated projective \( \Lambda^c \)-modules. Since \( J_R \) has cohomology groups which are finitely generated over \( R \), Lemma 2.3.1 implies the existence of a subcomplex \( C \) of \( J_R \) such that \( C \) is a complex of finitely generated \( R \)-modules and the inclusion \( i: C \hookrightarrow J_R \) is a quasi-isomorphism. We fix such a \( C \) and write \( X_n \) for \([I^n \to \Lambda]\). The complex \( \text{Hom}_{\Lambda^c}(W, X_n \otimes_R C) \) represents

\[ R\text{Hom}_{\Lambda^c}\left(R\Gamma(G_v, \mathcal{F}_T(T^*)^!(1)), \Lambda/I^n \otimes_R \omega_R\right) \]

and \( \text{Hom}_{\Lambda^c}(W, \Lambda \otimes_R C) \) represents

\[ R\text{Hom}_{\Lambda^c}\left(R\Gamma(G_v, \mathcal{F}_T(T^*)^!(1)), \Lambda \otimes_R \omega_R\right) \].
since $X_n$ is a complex of flat $R$-modules by Lemma 4.4.2(a). Now, for each $n$, we have a commutative diagram

$$
\begin{array}{cccc}
\text{R}^\Gamma(G_v, \mathcal{F}_T(T)) & \longrightarrow & \text{RHom}_{\Lambda^e} \left( \text{R}^\Gamma(G_v, \mathcal{F}_T(T^*)^i(1)), \Lambda \otimes_R \omega_R \right)[-2] \\
\downarrow & & \downarrow \\
\text{R}^\Gamma(G_v, \mathcal{F}_{\Lambda/I^n}(T)) & \longrightarrow & \text{RHom}_{\Lambda^e} \left( \text{R}^\Gamma(G_v, \mathcal{F}_T(T^*)^i(1)), \Lambda/I^n \otimes_R \omega_R \right)[-2]
\end{array}
$$

which induces a commutative diagram

$$
\begin{array}{cc}
H^i(G_v, \mathcal{F}_T(T)) & \longrightarrow H^{2-i}(\text{Hom}_{\Lambda^e}(W, \Lambda \otimes R C)) \\
\downarrow & \downarrow \\
H^i(G_v, \mathcal{F}_{\Lambda/I^n}(T)) & \longrightarrow H^{2-i}(\text{Hom}_{\Lambda^e}(W, X_n \otimes R C))
\end{array}
$$

of cohomology groups. Since the maps in this diagram are compatible as we vary $n$, we obtain the commutative diagram

$$
\begin{array}{ccc}
\lim H^i(G_v, \mathcal{F}_T(T)) & \longrightarrow & H^{2-i}(\text{Hom}_{\Lambda^e}(W, \Lambda \otimes R C)) \\
\downarrow & & \downarrow \\
\lim H^i(G_v, \mathcal{F}_{\Lambda/I^n}(T)) & \longrightarrow & \lim H^{2-i}(\text{Hom}_{\Lambda^e}(W, X_n \otimes R C)).
\end{array}
$$

It remains to show that the upper horizontal map in the latter diagram is an isomorphism. By Proposition 4.4.3, we know that the lower horizontal map is one. Noting that the maps $X_{n+1} \otimes R C \rightarrow X_n \otimes R C$ are injections of complexes with intersection $\Lambda \otimes R C$ over all $n$, we have an isomorphism

$$
\lim H^i(G_v, \mathcal{F}_{\Lambda/I^n}(T)) \cong H^i(G_v, \mathcal{F}_T(T))
$$

of complexes of finitely generated $\Lambda$-modules. Since inverse limits are exact for finitely generated $\Lambda$-modules (being that they are compact), after taking cohomology groups, we have that the vertical map on the right is an isomorphism. On the other hand, we also have an isomorphism

$$
\lim (X_n \otimes_\Lambda \mathcal{F}_T(T)) \cong \mathcal{F}_T(T)
$$

of complexes of objects in $\mathcal{C}_{\Lambda^e}^{\Lambda^e}$ and hence an isomorphism

$$
\lim C(G_v, X_n \otimes_\Lambda \mathcal{F}_T(T)) \cong C(G_v, \mathcal{F}_T(T)).
$$
where \( \bar{\xi} \) have an isomorphism \( S \) in which are given by the formulas

\[
\otimes_{\Lambda} (\text{ complexes of } \Lambda)
\]

map is an isomorphism, as required. Therefore, the vertical map on the left is also an isomorphism. Hence, the top map is an isomorphism, as required. 

\[\square\]

4.5 Duality over global fields

We end this paper by describing the global analog of Theorem 4.2.2 that is our main result. Let \( T \) be a bounded complex of objects in \( \mathcal{C}^{R,G}_{R,G,F,S} \) and we choose \( T^* \) in \( \mathcal{C}^{R-G}_{R,G,F,S} \) as in Section 3.3. As in [Nek, (5.3.3)], we define two morphisms of complexes of \( \Lambda \otimes_R \Lambda^\omega \)-modules

\[
\cup: C_c(G,F,S, \mathcal{F}_T(T)) \otimes_R C(G,F,S, \mathcal{F}_T(T^*)^!(1)) \to C_c(G,F,S, \Lambda \otimes_R J_R(1))
\]

\[
\cup_c: C(G,F,S, \mathcal{F}_T(T)) \otimes_R C_c(G,F,S, \mathcal{F}_T(T^*)^!(1)) \to C_c(G,F,S, \Lambda \otimes_R J_R(1))
\]

which are given by the formulas

\[
(a, a_S) \cup b = (a \cup b, a_S \cup_S \res_S(b))
\]

\[
a \cup_c (b, b_S) = (a \cup b, (-1)^a \res_S(a) \cup_S b_S),
\]

where \( \bar{\alpha} \) denotes the degree of \( a \), the direct sum of the restrictions to the primes in \( S \) is denoted \( \res_S \), the symbol \( \cup \) is the total cup product

\[
C(G,F,S, \mathcal{F}_T(T)) \otimes_R C(G,F,S, \mathcal{F}_T(T^*)^!(1)) \to C(G,F,S, \Lambda \otimes_R J_R(1))
\]

of Lemma 3.3.3 and \( \cup_S \) is the direct sum of the corresponding local cup products. Much as in Lemma 4.2.1 (but now analogously to [Nek, Lemma 5.7.3]), we have a quasi-isomorphism

\[
\Lambda \otimes_R J_R \to \tau_{R,F}^{-1} C_{(\mathcal{F}_T(T^*)^!(1), \Lambda \otimes_R \omega_R)}[1]
\]

of complexes of \( \Lambda \otimes_R \Lambda^\omega \)-modules, which allows us to use adjoints to define the morphisms in the following theorem.

**Theorem 4.5.1.** Let \( T \) be a bounded complex of objects in \( \mathcal{C}^{R-G}_{R,G,F,S} \). Then we have an isomorphism

\[
\otimes_{\Lambda} (\text{ complexes of } \Lambda)
\]

\[
\begin{array}{cccc}
\mathbf{R} \Gamma_{(\mathcal{F}_T(T^*)^!(1), \Lambda \otimes_R \omega_R)} & \longrightarrow & \mathbf{R} \Hom_{\Lambda^*} \left( \mathbf{R} \Gamma_{(\mathcal{F}_T(T^*)^!(1), \Lambda \otimes_R \omega_R)}[-3] \right) \\
\downarrow & & \downarrow \\
\mathbf{R} \Gamma_{(\mathcal{F}_T(T^*)^!(1), \Lambda \otimes_R \omega_R)} & \longrightarrow & \mathbf{R} \Hom_{\Lambda^*} \left( \mathbf{R} \Gamma_{(\mathcal{F}_T(T^*)^!(1), \Lambda \otimes_R \omega_R)}[-3] \right) \\
\downarrow & & \downarrow \\
\bigoplus_{v \in S} \mathbf{R} \Gamma_{(\mathcal{F}_T(T^*)^!(1), \Lambda \otimes_R \omega_R)}[-2] & \longrightarrow & \bigoplus_{v \in S} \mathbf{R} \Hom_{\Lambda^*} \left( \mathbf{R} \Gamma_{(\mathcal{F}_T(T^*)^!(1), \Lambda \otimes_R \omega_R)}[-3] \right)
\end{array}
\]

of exact triangles in \( \mathbf{D}_{\Lambda - R}(\text{Mod}_\Lambda) \).
Proof. The morphism of exact triangles can be constructed as in \cite[Theorem 4.2.6]{Lim2}. We remark that it suffices to prove that any two of the three horizontal morphisms is an isomorphism, and we focus on the second and third. By an analogous result to Proposition 4.3.2, we may assume that $R$ is regular and $\mathbb{Z}_p$-flat. That the second of the morphisms is an isomorphism follows as in proof of Theorem 4.2.2. In fact, aside from the obvious changes of notation for cochains and other objects, there are no significant changes to the proof.

As for the third, note that for a nonarchimedean prime $v$ in $\mathcal{S}$, the map\[
\tau_{\geq 2}^{I_1} C(G_v, \Lambda \otimes_R J_R(1)) \longrightarrow \tau_{\geq 2}^{I_1} \left(C(\mathcal{O}(G_{F,S}), \Lambda \otimes_R J_R(1))[1]\right)
\]
is a quasi-isomorphism that induces the identity maps on the cohomology groups of each complex. Under this identification, the map on the $v$-summand in the lower horizontal morphism in the statement is exactly the first isomorphism of Theorem 4.2.2.

It remains to prove that, for a real place $v$, the morphisms\[
R \Gamma(G_v, \mathcal{F}_T(T)) \longrightarrow R \text{Hom}_{\Lambda^\circ} \left(R \Gamma(G_v, \mathcal{F}_T(T^*)^\vee(1)), \Lambda \otimes_R \omega_R\right)[-2]
\]
are isomorphisms in $D_{\Lambda^\circ-\text{R-Mod}}$. There are two cases: either $v$ extends to a complex place or splits completely in $F_\infty$. However, if $v$ becomes complex, then both $\mathcal{C}(G_v, \mathcal{F}_T(T))$ and $\mathcal{C}(G_v, \mathcal{F}_T(T^*)^\vee(1))$ are acyclic, and there is nothing to prove.

If $v$ splits completely in $F_\infty$, then the adjoint map of interest is identified (as in the proof of Proposition 4.1.1) with\[
\Lambda \otimes_R \mathcal{C}(G_v, T) \longrightarrow \text{Hom}_{\Lambda^\circ} \left(\Lambda \otimes_R \mathcal{C}(G_v, T^*(1)), \Lambda \otimes_R J_R\right)[-2],
\]
where\[
\cup : \mathcal{C}(G_v, T) \otimes_R \mathcal{C}(G_v, T^*(1)) \longrightarrow J_R[-2]
\]
is given by cup product followed by the maps\[
\mathcal{C}(G_v, J_R(1)) \longrightarrow \tau_{\geq 2}^{I_1} \mathcal{C}(G_v, J_R(1)) \longrightarrow \tau_{\geq 2}^{I_1} \left(C(\mathcal{O}(G_{F,S}), J_R(1))[1]\right) \longrightarrow J_R[-2],
\]
the latter being a homotopy inverse to the inclusion of complexes.

Suppose now that we have a quasi-isomorphism $D \rightarrow \mathcal{C}(G_v, T^*(1))$ with $D$ a $g$-projective complex of finitely generated $R$-modules. In the commutative
of canonical maps, the lower right vertical map is an isomorphism by Lemma 2.1.7 (the terms of $D$ being finitely generated over $R$), and the lower horizontal morphism is a quasi-isomorphism by the $R$-injectivity of the terms of $J_R$ and the $R$-flatness of $\Lambda$. It follows that the derived adjoint morphism is represented by a morphism

\[ \Lambda \otimes_R \tilde{C}(G_v, T^*(1)) \longrightarrow \Lambda \otimes_R \text{Hom}_R(\tilde{C}(G_v, T^*(1)), J_R)[{-2}] \]

that by construction takes $\lambda \otimes f$ to $\lambda \otimes (g \mapsto f \cup g)$, which is the tensor product of the identity map with the derived adjoint that is already known to be a quasi-isomorphism by [Nek, (5.7.5)].

We are left to construct $D$, the existence of which is not so obvious as $\tilde{C}(G_v, T^*(1))$ need not have bounded cohomology. However, as is seen in [Spa, Lemma 3.3] (see also [Kel, Appendix]), such a complex may be constructed as a direct limit $\lim \rightarrow \lim D_n$ using maps $\alpha_n : D_n \to D_{n+1}$ that are split injective in each degree, and where

\[ D_0 \longrightarrow \tau_{\leq 0} \tilde{C}(G_v, T^*(1)) \]

is a quasi-isomorphism from a bounded above complex of finitely generated $R$-projectives and $E_n = D_n/\alpha_{n-1}(D_{n-1})$ for $n \geq 1$ is chosen to be quasi-isomorphic to

\[ \text{Cone} \left( \tau_{\leq n-1} \tilde{C}(G_v, T^*(1)) \longrightarrow \tau_{\leq n} \tilde{C}(G_v, T^*(1)) \right), \]

which has bounded, $R$-finitely generated cohomology. Since $R$ has finite global dimension, we may choose the $E_n$ as bounded complexes of finitely generated projective $R$-modules such that, for each $k$, only finitely many $E^n_k$ are nonzero, which is what was needed.

We end with a straightforward remark.

**Remark 4.5.2.** There is also an analogous diagram and isomorphisms for the other adjoint, with morphisms as in the second map in Theorem 4.2.2, and this follows, for instance, by a nearly identical argument.
References


ElliPtic CuRvE SOn Some Homogeneous SPlaces

B. Pasquier and N. Perrin

RecEived: July 18, 2011
Revised: November 19, 2012

Communicated by Wolfgang Soergel

Abstract. Let $X$ be a minuscule homogeneous space, an odd-dimensional quadric, or an adjoint homogeneous space of type different from $A$ and $G_2$. Let $C$ be an elliptic curve. In this paper, we prove that for $d$ large enough, the scheme of degree $d$ morphisms from $C$ to $X$ is irreducible, giving an explicit lower bound for $d$ which is optimal in many cases.

2010 Mathematics Subject Classification: 14M15, 14N35

Keywords and Phrases: Elliptic curves, minuscule homogeneous spaces, adjoint homogeneous spaces.

Introduction

In this paper we study the scheme $\text{Hom}_d(C, X)$ of degree $d$ morphisms from an elliptic curve $C$ to a rational homogeneous space $X$. Specifically we will assume that $X$ is either a minuscule homogeneous space or an adjoint homogeneous space. Minuscule homogeneous spaces\(^1\) are a natural generalisation of Grassmann varieties (see Table 1). Adjoint homogeneous spaces are also called quasi-minuscule and are obtained as the unique closed orbits in $\mathbb{P}g$ under the adjoint action. These varieties are also called minimal nilpotent orbits (see Table 2).

We prove that the scheme $\text{Hom}_d(C, X)$ is irreducible as soon as $d$ is large enough and we give an explicit bound $d(X)$ for $d$ (see Tables 1 and 2).

---

\(^1\)A homogeneous space $X$ is minuscule if the minimal representation $V$ defining an embedding $X \subset \mathbb{P}(V)$, also called the Plücker embedding, is a minuscule representation i.e. a representation of highest weight $\varpi$ with $|\langle \alpha^\vee, \varpi \rangle| \leq 1$ for all roots $\alpha$. The space $X$ is cominuscule if the above highest weight $\varpi$ is fundamental and its associated fundamental coweight is minuscule.
**Theorem 0.1** Let $C$ be a smooth elliptic curve and let $X$ be a minuscule or an adjoint homogeneous space. If $X$ is of adjoint type, assume furthermore that the group $G$ is not of type $A$ or $G_2$. For $d \geq d(X)$, the scheme of morphisms $\text{Hom}_d(C,X)$ is irreducible of dimension $c_1(X)d$.

In many cases, we also prove that the bound $d(X)$ is optimal in the following sense: for $d < d(X)$, the scheme $\text{Hom}_d(C,X)$ has dimension strictly bigger than the expected dimension $c_1(X)d$. Note that it may happen that for $d < d(X)$ the scheme $\text{Hom}_d(C,X)$ is irreducible. This is the case for all $d$ when $X$ is isomorphic to the maximal isotropic Grassmann variety $G_Q(n,2n)$ (see Table 1 and Corollary 3.9).

**Remark 0.2** (ı) Note that B. Kim and R. Pandharipande proved the connectedness of the moduli space of stable maps $\overline{M}_{d,g}(X)$ to any homogeneous space and for any genus. However, this space is almost never irreducible since there will be irreducible components of unexpected dimension (see Proposition 4.2).

(ii) Note also that for $d = 2$, the scheme of morphisms $\text{Hom}_d(C,X)$ to any homogeneous space with Picard rank 1 is irreducible of dimension $c_1(X) + \dim X + 1$. Indeed, any degree 2 morphism factors through a line.

Minuscule and adjoint homogeneous spaces are of the form $G/P$ for a semisimple algebraic group $G$. Except for the adjoint homogeneous space of type $A$, the parabolic subgroup $P$ is a maximal parabolic subgroup. This is the reason why we will assume that if $X$ is of adjoint type the group $G$ is not of type $A$. Recall that the Dynkin diagram of the group $G$ has vertices indexed by simple roots. In the following tables, we give the list of all minuscule homogeneous spaces. Note that we also include odd dimensional quadrics which are not minuscule but cominuscule. The above statement is still true for odd dimensional quadrics.

In Table 1, we give the list of all adjoint homogeneous spaces for a group of type different from $A$. In this table we also include the adjoint variety of type $G_2$ even though we have no result for this variety.

In these tables, we followed the notation of [Bou54, Tables] and we depicted the set $\Delta(P)$ of simple roots not in $P$ with plain vertices. The minuscule and adjoint varieties $X = G/P$ are described in the second column. By convention we denote by $G(k,n)$ (resp. $G_Q(k,n)$) the Grassmann variety of $k$-dimensional subspaces in $\mathbb{C}^n$ (resp. isotropic $k$-dimensional subspaces in $\mathbb{C}^n$ for a non-degenerate quadratic form $Q$ in $\mathbb{C}^n$). For $G_Q(n,2n)$ we only consider one of the two connected components of the above Grassmann variety. We denoted by $Q_m$ any smooth $m$-dimensional quadric. The varieties $\mathbb{O}P^2 = E_6/P_1$ and $E_7/P_7$ are the Cayley plane and the Freudenthal variety. Recall that the index of $X = G/P$ is the integer $c_1(X)$ such that the anticanonical divisor $-K_X$ equals $c_1(X)H$ where $H$ is an ample generator of the Picard group. The last column gives the bound $d(X)$ of our main theorem.

**Remark 0.3** Note that the above statement was already known for Grassmann varieties by results of A. Bruguieres [Bru87] and for quadrics by results of E.
### Table 1: Minuscule homogeneous spaces and odd quadrics

<table>
<thead>
<tr>
<th>Type</th>
<th>Variety</th>
<th>Diagram</th>
<th>Dim</th>
<th>Index</th>
<th>$d(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{n-1}$</td>
<td>$G(k, n)$</td>
<td>$\circ \cdots \circ$</td>
<td>$k(n-k)$</td>
<td>$n$</td>
<td>$n$</td>
</tr>
<tr>
<td>$B_n$</td>
<td>$\mathbb{Q}_{2n-1}$</td>
<td>$\bullet \cdots \bullet$</td>
<td>$2n - 1$</td>
<td>$2n - 1$</td>
<td>$3$</td>
</tr>
<tr>
<td>$D_n$</td>
<td>$\mathbb{Q}_{2n-2}$</td>
<td>$\circ \cdots \circ$</td>
<td>$2n - 2$</td>
<td>$2n - 2$</td>
<td>$3$</td>
</tr>
<tr>
<td>$D_n$</td>
<td>$G_Q(n, 2n)$</td>
<td>$\circ \cdots \circ$</td>
<td>$\frac{n(n-1)}{4}$</td>
<td>$2n - 2$</td>
<td>$n - 1$</td>
</tr>
<tr>
<td>$E_6$</td>
<td>$\mathbb{O}\mathbb{P}^2$</td>
<td>$\circ \cdots \circ$</td>
<td>$16$</td>
<td>$12$</td>
<td>$3$</td>
</tr>
<tr>
<td>$E_7$</td>
<td>$E_7/P_7$</td>
<td>$\circ \cdots \circ$</td>
<td>$27$</td>
<td>$18$</td>
<td>$8$</td>
</tr>
</tbody>
</table>

### Table 2: Adjoint homogeneous spaces of Picard number one

<table>
<thead>
<tr>
<th>Type</th>
<th>Variety</th>
<th>Diagram</th>
<th>Dim</th>
<th>Index</th>
<th>$d(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_n$</td>
<td>$G_Q(2, 2n + 1)$</td>
<td>$\circ \cdots \circ$</td>
<td>$4n - 5$</td>
<td>$2n - 2$</td>
<td>$2n$</td>
</tr>
<tr>
<td>$C_n$</td>
<td>$\mathbb{P}^{2n-1}$</td>
<td>$\bullet \cdots \bullet$</td>
<td>$2n - 1$</td>
<td>$2n$</td>
<td>$2$</td>
</tr>
<tr>
<td>$D_n$</td>
<td>$G_Q(2, 2n)$</td>
<td>$\circ \cdots \circ$</td>
<td>$4n - 7$</td>
<td>$2n - 3$</td>
<td>$2n - 1$</td>
</tr>
<tr>
<td>$E_6$</td>
<td>$E_6/P_2$</td>
<td>$\circ \cdots \circ$</td>
<td>$21$</td>
<td>$11$</td>
<td>$9$</td>
</tr>
<tr>
<td>$E_7$</td>
<td>$E_7/P_7$</td>
<td>$\circ \cdots \circ$</td>
<td>$33$</td>
<td>$17$</td>
<td>$11$</td>
</tr>
<tr>
<td>$E_8$</td>
<td>$E_8/P_8$</td>
<td>$\circ \cdots \circ$</td>
<td>$57$</td>
<td>$29$</td>
<td>$15$</td>
</tr>
<tr>
<td>$F_4$</td>
<td>$F_4/P_3$</td>
<td>$\bullet \cdots \bullet$</td>
<td>$15$</td>
<td>$8$</td>
<td>$8$</td>
</tr>
<tr>
<td>$G_2$</td>
<td>$G_2/P_1$</td>
<td>$\bullet \circ$</td>
<td>$5$</td>
<td>$3$</td>
<td></td>
</tr>
</tbody>
</table>
Ballico [Bal89]. Note also that for orthogonal Grassmann varieties, the above result was obtained by another technique by the second author in the paper [Per12]. This case was the very first motivation for our study. It answered a question of D. Markushevitch and has been used in [IM07].

Remark 0.4 Note that the choice of minuscule and adjoint varieties for studying elliptic curves is not a random one. Indeed, already in the study of rational curves and their associated enumerative properties encoded in quantum cohomology, these variety have very special properties, see for example [Buc03, BKT03, BM11, BCMP10, CMP08, CMP07, CMP10] for minuscule homogeneous spaces and [CP11b] for adjoint varieties.

Let us briefly describe the content of the paper. The strategy of the proof relies on a description given in [Per02] of a big open cell $U$ in any homogeneous space $X$ as a tower of affine bundles $\phi : U \to Y$ (see Definition 1.2) where $Y$ is again homogeneous under a smaller group. In section 1 we explain how to restrict the study of the irreducibility of $\text{Hom}_d(C, X)$ to the one of $\text{Hom}_d(C, U)$ and we explain, under some conditions on the fibration $\phi : U \to Y$, how to deduce the irreducibility of $\text{Hom}_d(C, U)$ from the irreducibility of $\text{Hom}_d(C, Y)$. In section 1.3, we recall the construction of $U$ and of the the fibration $\phi : U \to Y$ and describe the tower of affine fibrations. In section 2, we prove some results on the cohomology of the restriction to elliptic curves of equivariant locally free sheaves on projective spaces quadrics. In section 3, we gather all these results to prove Theorem 0.1. In section 4, we include some remarks on the moduli space of stable maps to minuscule or adjoint homogeneous spaces with source an elliptic curve. Finally, in an appendix we deal with the quadric of dimension 3, which we cannot approach by the same method as all other homogeneous spaces.

Acknowledgments. The second author would like to thank Christian Peskine for enlightening discussions on the paper [GLP83] and Piotr Achinger for useful comments on spinor bundles. Part of this work was done while both authors were supported by Hausdorff Center for Mathematics in Bonn. We thank the HCM for perfect working conditions.

Contents

1 Reduction of the theorem 683

1.1 Restriction to an open subset 683

1.2 Passing through towers of affine bundles 684

1.3 Decomposition of $X$ 685

2 Restriction of some homogeneous vector bundles to curves 687

2.1 Projective spaces 688

2.2 Spinor bundles on quadrics 690
3 IRREDUCIBILITY 692

3.1 Projective space .................................. 692
3.2 Grassmann varieties ................................ 692
3.3 Quadrics ........................................... 693
3.4 Maximal orthogonal Grassmann varieties ....... 694
3.5 Adjoint varieties of type B and D ............... 695
3.6 Type $E_6$ ........................................... 696
  3.6.1 Cayley plane ................................... 696
  3.6.2 Adjoint variety ................................ 697
3.7 Type $E_7$ ........................................... 697
  3.7.1 Freudenthal variety ............................ 697
  3.7.2 Adjoint variety ................................ 697
3.8 Adjoint variety of type $E_8$ ....................... 698
3.9 Adjoint variety of type $F_4$ ....................... 698

4 STABLE MAPS 699

A THE QUADRIC OF DIMENSION 3 700

1 REDUCTION OF THE THEOREM 700

1.1 Restriction to an open subset

Let $C$ be a smooth curve, let $X$ be a homogeneous space with Picard number 1 and let $\alpha \in A_1(X)$. The strategy of the proof will be similar to the one in [Per02]. We prove that there exists an open subset $U$ of $X$ whose complement is of codimension at least 2 and such that $U$ can be realised as a tower of affine bundles $\phi : U \to Y$ (see Definition 1.2 below) over a homogeneous space $Y$ of smaller dimension. The next result from [Per02, Proposition 2] proves that the irreducibility of $\text{Hom}_i(\alpha, U)$ is equivalent to the irreducibility of $\text{Hom}_i(\alpha, C)$ (where $i : U \to X$ is the inclusion morphism). This result is an easy application of Kleiman-Bertini Theorem [Kle74].

PROPOSITION 1.1 Let $C$ be a smooth curve and let $X$ be a homogeneous space under a group $G$. Assume that $U$ be an open subset of $X$ such that $\text{Codim}_X(X \setminus U) \geq 2$ and let $i : U \to X$ be the inclusion morphism. For $\alpha \in A_1(X)$, if $\text{Hom}_{i\ast \alpha}(C, U)$ is irreducible, then so is $\text{Hom}_{i\ast \alpha}(C, X)$.

We then need to study the open subset $U$. It will be the open orbit of the action of some parabolic subgroup of $G$ on $X$. The second author proved in [Per02] that $U$ can be realised as a tower $\phi : U \to Y$ of affine bundles (see Definition 1.2) over a homogeneous space $Y$ of smaller dimension. However, if this was enough to study rational curves, we need to be more precise if we want to study higher genus curves. If $\phi : U \to Y$ is an affine bundle associated to a vector bundle $E$, we explain in the next subsection how a good understanding of the restriction of $E$ to higher genus curves in $Y$ enables us to proceed by induction on the dimension.
1.2 Passing through towers of affine bundles

Let us fix some notation. In this section \( X \) is a homogeneous space of Picard number 1, we denote by \( U \) an open subset whose complement \( Z \) is of codimension 2 and we assume that \( U \) can be realised as a tower of affine bundles \( \phi : U \to Y \) over a homogeneous space \( Y \) of smaller dimension in the following sense.

**Definition 1.2** Let \( n \) be a non negative integer and for all \( i \in [1, n-1] \), let \( \phi_i : X_i \to X_{i+1} \) be an affine bundle. Let \( U = X_1 \) and \( Y = X_n \). The composition morphism \( \phi : U \to Y \) is called a tower of affine bundles over \( Y \).

Note that because of the inequality \( \text{Codim}_X Z \geq 2 \), we have \( \text{Pic}(U) = \text{Pic}(X) \). Furthermore since \( \phi : U \to Y \) is a tower of affine bundles, then we also have \( \text{Pic}(U) = \text{Pic}(Y) \). For \( f : C \to U \) a morphism, we may define a linear form \([f]\) on \( \text{Pic}(X) \) by \( \mathcal{L} \mapsto \deg(f^* \mathcal{L}) \). We shall say that for \( \alpha \in \text{Pic}(X)^\vee \) a curve \( f : C \to U \) is of class \( \alpha \) if \([f] = \alpha \). We shall write \( \langle \alpha, \mathcal{L} \rangle \) for the evaluation of \( \alpha \) at \( \mathcal{L} \in \text{Pic}(X) \). For such a class \( \alpha \) in \( \text{Pic}(X)^\vee \), we denote by \( \text{Hom}_\alpha(C, U) \) resp. \( \text{Hom}_\alpha(C, Y) \) the scheme of morphisms from \( f : C \to X \) resp. \( f : C \to U, f : C \to Y \) such that \([f] = \alpha \). Note that for \( X \) (or for \( Y \)) we have identifications \( A_1(X) \simeq \text{Pic}(X)^\vee \) and \( A_1(Y) \simeq \text{Pic}(Y)^\vee \) and that these schemes of morphisms are the classical schemes of morphisms as defined in [Gro61].

We now prove a general result on the scheme of morphisms to an affine bundle. For this we introduce the following notation. Let \( \phi : U \to Y \) be an affine bundle with direction vector bundle \( E \) on \( Y \) and let \( \alpha \in \text{Pic}(U)^\vee \). We denote by \( H_b \) the locally closed subset of \( \text{Hom}_{\phi, \alpha}(C, Y) \) defined by

\[
H_b = \{ f \in \text{Hom}_{\phi, \alpha}(C, Y) \mid \dim H^1(C, f^* E) = b \}.
\]

**Proposition 1.3** Let \( \phi : U \to Y \) be an affine bundle over a variety \( Y \) with direction vector bundle \( E \). Let \( \alpha \in \text{Pic}(U)^\vee \) and let \( C \) be a smooth curve such that the scheme \( \text{Hom}_{\phi, \alpha}(C, Y) \) is irreducible of dimension \( \langle \phi, \alpha, c_1(T_Y) \rangle \).

(i) Assume that there exists \( f \in \text{Hom}_{\phi, \alpha}(C, Y) \) satisfying the vanishing \( H^1(C, f^* E) = 0 \). If for all integer \( b > 0 \) we have the inequality \( \text{Codim} H_b > b \), then \( \text{Hom}_{\alpha}(C, U) \) is irreducible of dimension \( \langle \alpha, c_1(T_U) \rangle \).

(ii) Assume that \( \phi \) is a vector bundle and let \( a \) be the smallest integer such that there exists an element \( f \in \text{Hom}_{\phi, \alpha}(C, Y) \) with \( \dim H^1(C, f^* E) = a \). If there exists an integer \( b > a \) satisfying the inequality \( \text{Codim} H_b \leq b - a \), then the scheme \( \text{Hom}_{\alpha}(C, U) \) is reducible.

**Proof.** (i) Let us consider the natural morphism \( \text{Hom}_{\alpha}(C, U) \to \text{Hom}_{\phi, \alpha}(C, Y) \). As explained in [Per02] for rational curves, the fiber of this map over an element \( f \in \text{Hom}_{\phi, \alpha}(C, Y) \) is given by the sections \( s \in H^0(C, f^* E) \). In particular, such a section exists if and only if the affine bundle \( C \times Y U \to C \) obtained by pull-back to \( C \) is a vector bundle. For a
general element \( f \in \text{Hom}_{\Phi,\alpha}(C,Y) \) this is the case thanks to the vanishing \( H^1(C,f^*E) = 0 \).

For \( b > 0 \) the fibers over the locally closed subset \( H_b \) are either empty (if the pull-back is not a vector bundle) or of dimension \( \dim H^b(C,f^*E) = \chi(C,f^*E) + \dim H^1(C,f^*E) + \langle \phi_\alpha, c_1(E) \rangle + b \). The dimension of the inverse image of \( H_b \) in \( \text{Hom}_{\alpha}(C,U) \) is therefore strictly less than \( \langle \phi_\alpha, c_1(T_Y) + c_1(E) \rangle = \langle \alpha, c_1(T_Y) \rangle \), which is the expected dimension of \( \text{Hom}_{\alpha}(C,U) \). As any irreducible component is at least of this expected dimension (cf. [Mor79]), these locally closed subsets do not form irreducible components. Therefore the inverse image of the open subset \( H_0 \) is dense in \( \text{Hom}_{\alpha}(C,U) \). As explained in [Per02] for rational curves, this inverse image is an affine bundle over the base and is thus irreducible of dimension \( \langle \phi_\alpha, c_1(T_Y) + c_1(E) \rangle = \langle \alpha, c_1(T_Y) \rangle \).

(1) In this case, the inverse image of the locally closed subset \( H_b \) is of dimension \( \langle \alpha, c_1(T_Y) \rangle + a \) while the inverse image of the locally closed subset \( H_b \) is of dimension at least \( \langle \alpha, c_1(T_Y) \rangle + a \). This second locally closed subset is therefore not contained in the closure of the first one, thus they have to be contained in two different irreducible components.

\[ \square \]

1.3 Decomposition of \( X \)

The aim of this subsection is to describe a big open cell \( U \) in the homogeneous space \( X \) and to realise it as a fibration \( \phi : U \to Y \) over a space \( Y \) homogeneous under a group of smaller rank. This fibration can be decomposed in general as a tower of affine bundles (see [Per02, Proposition 5]). Let us recall this result. Let \( T \) be a maximal torus in \( G \) and let \( B \) be a Borel subgroup of \( G \) containing \( T \). Denote by \( w_0 \) the longest element of the Weyl group \( W \) of \((G,T)\). For any subgroup \( H \) of \( G \), we set \( H^{w_0} := w_0 H w_0^{-1} \) and for any parabolic subgroup \( P \) of \( G \) containing \( B \), we denote by \( L_P \) the Levi subgroup of \( P \) containing \( T \) and by \( U_P \) the unipotent radical of \( P \). We also denote by \( \{1\} = U_P^0 \subset U_P^1 \subset \cdots \subset U_P^r \) the upper central series of the group \( U_P \). For any parabolic subgroup \( P \) containing \( B \), we denote by \( \Sigma(P) \) the set of simple roots which are not roots of \( P \). Finally, for a simple root \( \alpha \), we denote by \( \iota(\alpha) \) its image via the Weyl involution \( i.e. \, \iota(\alpha) = -w_0(\alpha) \).

**Lemma 1.4 (Proposition 5 and 6, [Per02])** Let \( P \) and \( Q \) be two parabolic subgroups of \( G \) containing \( B \) and assume that \( \iota(\Sigma(P)) \cap \Sigma(Q) = \emptyset \). Set \( U := Q^{w_0} P / P \). Then \( U \) is an open subset of \( G / P \) with complement in codimension at least 2 and admits an \( L_{Q^{w_0}} \) equivariant fibration \( \phi : U \to Y \) over the flag variety \( Y = L_{Q^{w_0}} / (L_{Q^{w_0}} \cap P) \). Moreover this fibration decomposes in a tower of affine bundles

\[
U = X_1 \xrightarrow{\phi_1} X_2 \xrightarrow{\phi_2} \cdots \xrightarrow{\phi_{r-2}} X_{r-1} \xrightarrow{\phi_{r-1}} X_r = Y
\]

\( ^2\)Recall that the upper central series \( (Z^k(G))_{k \geq 0} \) of a group \( G \) is defined by induction by \( Z^0(G) = \{1\}, \, Z^1(G) = Z(G) \) the center of \( G \) and \( Z^{k+1}(G) \) is the (normal) subgroup of \( G \) containing \( Z^k(G) \) such that \( Z^{k+1}(G) / Z^k(G) = Z(G/Z^k(G)) \).
such that the vector bundle associated to the affine bundle \( \phi_i : X_i \rightarrow X_{i+1} \) is obtained by pull-back from \( Y \) of the \( L_{Q^0} \)-equivariant vector bundles associated to the \( L_{Q^0} \cap P \) representations \( U_{Q^0} / U_{Q^0} \cdot (U_{Q^0} \cap P) \).

**Remark 1.5** Note that it is easy to determine the weights of \( T \) on the representation \( U_{Q^0} / U_{Q^0} \cdot (U_{Q^0} \cap P) \) defining the vector bundle. For example for maximal parabolic subgroups \( P \) and \( Q \) with \( \Sigma(P) = \{ \alpha_P \} \) and \( \Sigma(Q) = \{ \alpha_Q \} \), the weights of the above representation are the negative roots \( \alpha \) such that the coefficient of \( \alpha \) on \( \alpha_P \) is negative and the coefficient of \( \alpha \) on \( \alpha_Q \) is exactly \(-i\). Keeping the previous notation we deduce the existence of the following fibrations. In the next result we will use the description of simple roots given in [Bou54, Tables]

**Corollary 1.6** We have the following fibrations.

1. If \( G \) is of type \( A_n \), if \( \Sigma(P) = \{ \alpha_i \} \) with \( 2 \leq i \leq n \) and if \( \Sigma(Q) = \{ \iota(\alpha_{i-1}) \} \), then \( Y \) is isomorphic to \( \mathbb{P}^{n-i+1} \) and \( \phi : U \rightarrow Y \) is the vector \((T_{P_{n-i}}(-1))(\mathbb{P}^{n-i+1}) \oplus (\Omega_{P_{n-i}}^1(n-i+1))^{\oplus i-1} \).

2. If \( G \) is of type \( B_n \), if \( \Sigma(P) = \{ \alpha_1 \} \) and if \( \Sigma(Q) = \{ \alpha_n \} \), then \( Y \) is isomorphic to \( \mathbb{P}^{n-1} \) and \( \phi : U \rightarrow Y \) is a tower of affine fibrations whose direction vector bundles are pull-backs of \( \mathcal{Q}_{\mathbb{P}^{n-1}}(1) \) and \( \Omega_{\mathbb{P}^{n-1}}^1(2) \).

3. If \( G \) is of type \( D_n \), if \( \Sigma(P) = \{ \alpha_2 \} \) and if \( \Sigma(Q) = \{ \alpha_1 \} \), then \( Y \) is isomorphic to \( \mathbb{P}^{2n-3} \) and \( \phi : U \rightarrow Y \) is the vector bundle \((\mathcal{Q}_{\mathbb{P}^{2n-3}}(1))^\perp \) i.e. the restriction of the tautological quotient bundle on \( \mathbb{P}^{2n-3} \) to \( \mathbb{Q}_{2n-4} \).

4. If \( G \) is of type \( D_n \), if \( \Sigma(P) = \{ \alpha_1 \} \) and if \( \Sigma(Q) = \{ \alpha_n \} \), then \( Y \) is isomorphic to \( \mathbb{P}^{n-2} \) and \( \phi : U \rightarrow Y \) is the vector bundle \( \Omega_{\mathbb{P}^{n-2}}^1(2) \).

5. If \( G \) is of type \( D_n \), if \( \Sigma(P) = \{ \alpha_2 \} \) and if \( \Sigma(Q) = \{ \alpha_1 \} \), then \( Y \) is isomorphic to \( \mathbb{P}^{2n-4} \) and \( \phi : U \rightarrow Y \) is the vector bundle \((\mathcal{Q}_{\mathbb{P}^{2n-4}}(1))^\perp \) i.e. the restriction of the tautological quotient bundle on \( \mathbb{P}^{2n-3} \) to \( \mathbb{Q}_{2n-4} \).

6. If \( G \) is of type \( D_n \), if \( \Sigma(P) = \{ \alpha_n \} \) and if \( \Sigma(Q) = \{ \iota(\alpha_{n-1}) \} \), then \( Y \) is isomorphic to \( \mathbb{P}^{n-1} \) and \( \phi : U \rightarrow Y \) is the vector bundle \( \Omega_{\mathbb{P}^{n-2}}^1(n-2) \).

7. If \( G \) is of type \( E_6 \), if \( \Sigma(P) = \{ \alpha_1 \} \) and if \( \Sigma(Q) = \{ \iota(\alpha_6) \} = \{ \alpha_1 \} \), then \( Y \) is isomorphic to \( \mathbb{Q}_8 \) and \( \phi : U \rightarrow Y \) is one of the two spinor bundles on \( \mathbb{Q}_8 \).

8. If \( G \) is of type \( E_6 \), if \( \Sigma(P) = \{ \alpha_2 \} \) and if \( \Sigma(Q) = \{ \alpha_1, \alpha_6 \} \), then \( Y \) is isomorphic to \( \mathbb{Q}_6 \) and \( \phi : U \rightarrow Y \) is a tower of affine bundles whose direction vector bundles are pull-backs of \( E = (\mathcal{Q}_{\mathbb{Q}_6}(1))^\perp \) and \( E' = S \oplus S' \) where \( E \) is the restriction of the tautological quotient bundle of \( \mathbb{P}^7 \) to \( \mathbb{Q}_6 \) and \( S \) and \( S' \) are the two spinor bundles on \( \mathbb{Q}_6 \).
Elliptic Curves on Some Homogeneous Spaces

9. If $G$ is of type $E_7$, if $\Sigma(P) = \{\alpha_7\}$ and if $\Sigma(Q) = \{\alpha_2\}$, then $Y$ is isomorphic to $\mathbb{P}^6$ and $\phi : U \to Y$ is a tower of affine bundles whose direction vector bundles are pull-backs of $E = \Omega^2_{\mathbb{P}^6}(6)$ and $E' = \Omega^2_{\mathbb{P}^6}(3)$.

10. If $G$ is of type $E_7$, if $\Sigma(P) = \{\alpha_2\}$ and if $\Sigma(Q) = \{\alpha_6\}$, then $Y$ is isomorphic to $\mathbb{P}^6$ and $\phi : U \to Y$ is a tower of affine bundles whose direction vector bundles are pull-backs of $E = (\mathcal{O}_{\mathbb{P}^6}(1))^\perp$ and $E' = S^2$ where $E$ is the restriction of the tautological quotient bundle of $\mathbb{P}^6$ to $\mathbb{P}^6$ and $S$ is one of the two spinor bundles on $\mathbb{P}^6$.

11. If $G$ is of type $E_8$, if $\Sigma(P) = \{\alpha_8\}$ and if $\Sigma(Q) = \{\alpha_1\}$, then $Y$ is isomorphic to $\mathbb{P}^{12}$ and $\phi : U \to Y$ is a tower of affine bundles whose direction vector bundles are pull-backs of $E = (\mathcal{O}_{\mathbb{P}^{12}}(1))^\perp$ and $E' = S$ where $E$ is the restriction of the tautological quotient bundle of $\mathbb{P}^{13}$ to $\mathbb{P}^{12}$ and $S$ is one of the two spinor bundles on $\mathbb{P}^{12}$.

12. If $G$ is of type $F_4$, if $\Sigma(P) = \{\alpha_1\}$ and if $\Sigma(Q) = \{\alpha_4\}$, then $Y$ is isomorphic to $\mathbb{P}^5$ and $\phi : U \to Y$ is a tower of affine bundles whose direction vector bundles are pull-backs of $E = (\mathcal{O}_{\mathbb{P}^5}(1))^\perp$ and $E' = S$ where $E$ is the restriction of the tautological quotient bundle of $\mathbb{P}^{10}$ to $\mathbb{P}^5$ and $S$ is the spinor bundle on $\mathbb{P}^5$.

Proof. According to the previous Lemma, we only need to identify in each case the direction vector bundles of the affine bundles and to notice (see [Per02] proof of Proposition 5) that the first affine bundle over $Y$ is always a vector bundle.

To identify the vector bundles, we only have to identify the corresponding representations. Indeed, recall from [Per02, Proposition 5] that the direction vector bundles are equivariant and correspond to the $L_{\mathbb{P}^n} \cap P$-representations $U_{\mathbb{P}^n}^i \cap U_{\mathbb{P}^n}^{i-1}$, $(U_{\mathbb{P}^n}^i \cap P)$. We therefore only have to compute their highest weight which is an easy check using [Bou54, Tables]. Note that the representation corresponding to the sheaf $(\mathcal{O}_{\mathbb{P}^5}(1))^\perp$ on an $n$-dimensional quadric is not irreducible as a representation of the semisimple part of $L_{\mathbb{P}^n} \cap P$ but is an irreducible non simple representation of $L_{\mathbb{P}^n} \cap P$. Note also that the above description of vector bundles for the adjoint varieties were given in [CP11]. $\Box$

2 Restriction of some homogeneous vector bundles to curves

In this section we gather some known results and some new ones on the geometry of curves in the projective space and in smooth quadrics. As explained in the previous sections, all the affine bundles $\phi : U \to Y$ we have to consider are homogeneous vector bundles over these varieties. Therefore the main aim of the section is to control the cohomology of the pull-back to curves of some equivariant vector bundles. Even if we only need these results for elliptic curves, we state them for any curve of genus $g$.
2.1 Projective spaces

In this subsection we recall some results on secants to a curve. The following result can be found in [Har77, Proposition IV.3.8 and Theorem IV.3.9].

**Proposition 2.1** Let $C$ be a reduced and irreducible curve in $\mathbb{P}^r$ such that any secant line to $C$ is a multisecant, then $C$ is a line.

We will use this statement via the following corollary.

**Corollary 2.2** Let $C$ be a reduced and irreducible curve in $\mathbb{P}^r$ which is not a line, then there exists a point $x$ in $C$ such that the projection from $x$ is birational from $C$ onto its image.

Let $C$ be a smooth curve of genus $g$ and let $f : C \to \mathbb{P}^r$ be a morphism such that $f^*\mathcal{O}_{\mathbb{P}^r}(1)$ is of degree $d$. We want to prove that the pull back of the homogeneous bundles $\Omega^k_{\mathbb{P}^r}(k)$ through $f$ admits nice filtrations. The following lemma can be found in [GLP83, Remark (2) page 498].

**Lemma 2.3** Assume that $f$ is locally injective and that $f(C)$ is non degenerate i.e. not contained in an hyperplane. Then, for almost all choices of $r-1$ points $(x_i)_{i \in [1,r-1]}$ in $C$, there exists a filtration

$$0 = F^{r+1} \subset F^r \subset \cdots \subset F^1 = f^*\Omega^1_{\mathbb{P}^r}(1)$$

such that $F^i/F^{i+1} = \mathcal{O}_C(-x_i)$ for $i \in [1,r-1]$ and $F^r/F^{r+1} = f^*\mathcal{O}_{\mathbb{P}^r}(-1) \otimes \mathcal{O}_C(\sum_{i} x_i)$.

**Proof.** Let us set $M = f^*\Omega^1_{\mathbb{P}^r}(1)$ and $L = f^*\mathcal{O}_{\mathbb{P}^r}(1)$. Let us also denote by $V^\vee$ the vector space $H^0(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(1))$. Because $f(C)$ is non degenerate, we may consider $V^\vee$ as a subspace of $H^0(C,L)$. We have an exact sequence $0 \to M \to V^\vee \otimes \mathcal{O}_C \to L \to 0$.

We proceed by induction on $r$. Let us choose a point $x_1$ on $C$. Let us denote by $W^\vee$ the intersection $V^\vee \cap H^0(C, L(-x_1))$ in the vector space $H^0(C,L)$. If $f(x_1)$ is smooth, the map $W^\vee \otimes \mathcal{O}_C \to L(-x_1)$ is surjective and we have the commutative diagram:

$$
\begin{array}{cccccc}
W^\vee \otimes \mathcal{O}_C & \to & L(-x_1) \\
0 & \to & M & \to & V^\vee \otimes \mathcal{O}_C & \to & L & \to & 0 \\
0 & \to & \mathcal{O}_C(-x_1) & \to & \mathcal{O}_C & \to & \mathcal{O}_{x_1} & \to & 0.
\end{array}
$$

The snake lemma implies that the left most vertical map is surjective and we get the term $F^2$ of the filtration by taking the kernel of this map which is also
the kernel of the map $W^\vee \otimes \mathcal{O}_C \to L(-x_1)$. Replacing $V$ by $W$ and $L$ by $L(-x_1)$, we get a morphism $f_2$ from $C$ to $\mathbb{P}^{r-1}$ and $F^2 = f_2^* \mathcal{O}_{\mathbb{P}^{r-1}}$. The morphism $f_2$ is the composition of $f$ with the projection with center $x_1$ and because, in characteristic zero, general secants are not multisecant (see Corollary 2.2), $f_2$ is also generically injective for $x_1$ general. We may apply then the induction hypothesis. For $r = 1$, the result is true. □

**Remark 2.4** Note that in the above filtration, all the vector bundles $E^i$ are constructed as tautological subbundles, therefore their dual are globally generated.

**Corollary 2.5** Assume that $f(C)$ is non degenerate, then for almost all choices of $r - 1$ points $(x_i)_{i \in [1, r-1]}$ in $C$, there exists a filtration

$$0 = G^{s+1} \subset G^s \subset \cdots \subset G^1 = f^* \Omega_{\mathbb{P}^r}^1(k)$$

such that $G^i/G^{i+1} = \mathcal{O}_C(-(x_{i_1} + \cdots + x_{i_k}))$ with $\{i_1, \cdots, i_k\} \subset [1, r-1]$ or $G^i/G^{i+1} = f^* \mathcal{O}_{\mathbb{P}^r}(-1) \otimes \mathcal{O}_C(x_{i_1} + \cdots + x_{i_{r-k}})$ with $\{i_1, \cdots, i_{r-k}\} \subset [1, r-1]$.

**Proof.** We simply take the $k$-th exterior product of the previous filtration. □

**Corollary 2.6** Assume that $f(C)$ is non degenerate. Let $g$ be the genus of $C$ and $d$ the degree of $f^* \mathcal{O}_{\mathbb{P}^r}(-1)$. If $g \leq r - k$, then we have the vanishing $H^1(C, f^* \Omega_{\mathbb{P}^r}^1(k+1)) = 0$.

**Proof.** Indeed, for $g \leq r - k$ and for general points $(x_i)_{i \in [1, r-1]}$ on $C$, we have the vanishing $H^1(C, \mathcal{O}_C(x_{i_1} + \cdots + x_{i_{r-k}})) = 0$. Note that because $f(C)$ is non degenerate, we must have the inequality $g \leq d - r \leq d - k$ and this condition implies the following vanishing $H^1(C, f^* \mathcal{O}_{\mathbb{P}^r}(-1) \otimes \mathcal{O}_C(-(x_{i_1} + \cdots + x_{i_k}))) = 0$. □

Let us assume that $C$ is elliptic. Then even if $f(C)$ is degenerate we may obtain results. Indeed, if $f(C)$ is degenerate, there is a minimal linear space $H$ in $\mathbb{P}^r$ of codimension say $a$ containing $f(C)$ and such that $f(C)$ is non degenerate in $H$. Furthermore we have the exact sequence $0 \to \mathcal{O}_C \to f^* \Omega_{\mathbb{P}^r}^1 \to f^* \Omega_{\mathbb{P}^r}^1(k+1) \to 0$ from which we deduce a filtration of $f^* \Omega_{\mathbb{P}^r}^1(1)$. In particular we get the following result.

**Proposition 2.7** Assume that $C$ is elliptic. If $f(C)$ is non degenerate in a linear subspace of codimension $a$, then we have the equality $\dim H^1(C, f^* \Omega_{\mathbb{P}^r}^1(k+1)) = \binom{a}{r-a}$. (with $\binom{a}{r} = 0$ if $a < r - k$).

In particular, if we have the inequality $\dim H^1(C, f^* \Omega_{\mathbb{P}^r}^1(k+1)) > 0$, then there exists an integer $a \geq r - k$ such that $\dim H^1(C, f^* \Omega_{\mathbb{P}^r}^1(k+1)) = \binom{a}{r-a}$ and $f(C)$ in non degenerate in a linear subspace of codimension $a$. 

Documenta Mathematica 18 (2013) 679–706
**Proof.** Taking the k-th exterior power of the exact sequence 0 → \( \mathcal{O}_C^n \) → \( f^*\Omega_{P^r}^n(1) \) → 0 and tensoring it by \( f^*\mathcal{O}_P(1) \), we get a filtration of \( f^*\Omega_{P^r}^n(k + 1) \) by the vector bundles \( \Lambda^u(\mathcal{O}_C^n) \otimes f^*\Omega_H^v(1) \) for \( u + v = k \) and \( u, v \) non negative. By Corollary 2.6, the first cohomology groups of these bundles vanish except maybe for \( v = 0 \) or \( v = r - a \). For \( v = 0 \) we have \( \Lambda^a(\mathcal{O}_C^n) \otimes f^*\Omega_H^v(1) = \Lambda^k(\mathcal{O}_C^n) \otimes f^*\mathcal{O}_P(1) \) and for \( v = r - a \) we have \( \Lambda^a(\mathcal{O}_C^n) \otimes f^*\Omega_H^v(1) = \Lambda^{k+a-r}(\mathcal{O}_C^n) \). The first cohomology group in the first case vanishes while the first cohomology group in the second case has dimension \( \binom{n-a}{r-k} \). We thus have an inequality \( \dim H^1(C, f^*\Omega_{P^r}^n(k + 1)) \leq \binom{n-a}{r-k} \).

For \( a < r - k \), we are done. If \( a \geq r - k \), let us consider the identification \( \Omega_{P^r}^n(k + 1) = (\Omega_{P^r}^{r-k}(r - k))^\vee \). By the above exact sequence, we have a trivial subbundle of rank \( a \) of \( f^*\Omega_{P^r}^{r-k}(1) \) therefore we have a trivial subbundle of rank \( \binom{n-a}{r-k} \) in \( f^*\Omega_{P^r}^{r-k}(r - k) \) (take the \( r - k \) exterior power). By duality, there is a trivial quotient bundle of rank \( \binom{n-a}{r-k} \) of \( f^*\Omega_{P^r}^n(k + 1) \).

**Lemma 2.8** Let \( E \) be a globally generated vector bundle on \( C \) such that \( E \) has a trivial quotient of rank \( n \), then \( \dim H^1(C, E) \geq n \).

**Proof.** By Serre duality we have an isomorphism \( H^1(C, E) \cong H^0(C, E^\vee)^\vee \). We therefore need to prove that \( \dim H^0(C, E^\vee) \geq n \). But we have a surjection \( E \to \mathcal{O}_C^n \), giving rise to an injection \( \mathcal{O}_C^n \to E^\vee \) and the result follows. \( \square \)

We conclude using the previous lemma and because \( \Omega_{P^r}^n(k + 1) \) is globally generated. \( \square \)

### 2.2 Spinor bundles on quadrics

In this section we consider the spinor bundles on quadrics and their restriction to elliptic curves.

**Proposition 2.9** Let \( Q_n \) be a smooth \( n \)-dimensional quadric and let \( E \) be a spinor bundle on \( Q_n \). Let \( f : C \to Q_n \) be a morphism from a smooth irreducible curve to \( Q_n \). Then the pull-back \( f^*E \) is isomorphic to a direct sum \( \mathcal{O}_C^{2^h} \oplus F \) where \( F \) has no trivial factor if and only if \( f \) factors through an isotropic subspace of (linear) dimension \( [n/2] + 1 - \alpha \) with

\[
\alpha = \begin{cases} 
\beta & \text{for } n \text{ odd}, \\
\beta + 1 & \text{for } n \text{ even and } \beta > 0, \\
0 & \text{for } n \text{ even and } \beta = 0
\end{cases}
\]

and in the last case, the maximal isotropic subspace has to be an element of a fixed component of the two connected components of maximal isotropic subspaces.

**Proof.** We shall here use the results of G. Ottaviani [Ott88] on the spinor bundles. Let us write \( h = [n/2] - 1 \) and \( p = [n/2] + 1 \). There is an embedding \( \varphi : Q_n \to G(2^h, 2^{h+1}) \) such that \( E = \varphi^*K \) with \( K \) the tautological subbundle in the Grassmann variety \( G(2^h, 2^{h+1}) \) of \( 2^h \) dimensional subspaces in \( \mathbb{C}^{2^{h+1}} \).
Lemma 2.10 Let $x$ and $y$ be points on $Q_n$, then we have the equality $C^{2h+1} = K_{\varphi(x)} \oplus K_{\varphi(y)}$ if and only if $x$ and $y$ are not orthogonal.

Proof. We shall consider the dual assertion: the map $C^{2h+1} \to K^\vee_{\varphi(x)} \oplus K^\vee_{\varphi(y)}$ is an isomorphism if and only if $x$ and $y$ are not orthogonal. To test this assertion, we can restrict the bundle $K$ to some subvariety containing both $x$ and $y$. If $x$ and $y$ are not orthogonal, then we may choose a smooth conic $c : \mathbb{P}^1 \to Q_n \to G(2^h, 2h+1)$ passing through $x$ and $y$. By [Ott88, Theorem 1.4], the pull back $c^*K^\vee$ is isomorphic to $O_{\mathbb{P}^1}(1)$. The surjectivity condition holds.

Conversely, if $x$ and $y$ are orthogonal, then there exists a maximal isotropic subspace $V_p$ (if $n$ is even we can choose $V_p$ to be in any of the two connected components of maximal isotropic subspaces) such that $x$ and $y$ are contained in $P(V_p)$. By [Ott88, Theorems 2.5 and 2.6] the restriction of $K$ to $P(V_p)$ contains a trivial factor (if $n$ is even this is true only for $V_p$ in one of the two connected components). Therefore the surjectivity condition does not hold. □

Let us return to the proof of the proposition. Remark that $f^*E$ has a trivial factor if and only if for all $x \in C$ the subspaces $K_{\varphi(x)}$ have a common intersection in $C^{2h+1}$. This occurs only if for all $x$ and $y$ in $C$, the points $f(x)$ and $f(y)$ are orthogonal. In particular the linear subspace spanned by $f(C)$ has to be isotropic. Therefore if the curve $f(C)$ is not contained in an isotropic subspaces then $f^*E$ has no trivial factor.

Let us assume that $f^*E$ has a trivial factor. Then $f$ factors through an isotropic subspace $V_p$. Furthermore by [Ott88, Theorems 2.5 and 2.6], we know that the restriction $E_{\mathbb{P}(V_p)}$ is isomorphic to one of the following direct sums (the first case occurs for $n$ odd while the other two cases occur for $n$ even):

$$\bigoplus_{k=0}^p \Omega^k_{\mathbb{P}(V_p)}(k), \bigoplus_{k=0, k \text{ odd}}^p \Omega^k_{\mathbb{P}(V_p)}(k) \text{ or } \bigoplus_{k=0, k \text{ even}}^p \Omega^k_{\mathbb{P}(V_p)}(k).$$

By the previous results on the projective space, we conclude that if $f(C)$ factors through a linear subspace of linear dimension $p - \alpha$ and if there is a unique such factor then $V_p$ is in the connected component of $G_{Q}(p, 2p)$ corresponding to the last decomposition above. The number of trivial factors in $f^*E$ is then $2^\beta$ with $\beta$ and $\alpha$ as above. □

For elliptic curves, applying the same techniques as in the case of the projective space, we obtain the following result.

Corollary 2.11 Let $Q_n$ be a smooth $n$-dimensional quadric and let $E$ be a spinor bundle on $Q_n$. Let $f : C \to Q_n$ be a morphism from a smooth irreducible elliptic curve to $Q_n$. Then we have the equality $\dim H^1(C, f^*E) = 2^\beta$ if and only if $f$ factors through an isotropic subspace of (linear) dimension $[n/2] + 1 - \alpha$. 

Documenta Mathematica 18 (2013) 679–706
with
\[
\alpha = \begin{cases} 
\beta & \text{for } n \text{ odd,} \\
\beta + 1 & \text{for } n \text{ even and } \beta > 0, \\
0 & \text{for } n \text{ even and } \beta = 0
\end{cases}
\]
and in the last case, the maximal isotropic subspace has to be an element of a fixed component of the two connected components of maximal isotropic subspaces.

3 Irreducibility

We are now in position to prove our main result. Let \( C \) be a smooth elliptic curve. We proceed via a case-by-case analysis to finish the proof of Theorem 0.1.

3.1 Projective space

Our proof is based on the fact that the irreducibility of the scheme of morphisms to a projective space is well known. Let us first recall this fact. Let \( X = \mathbb{P}(V) \) with \( V \) of dimension \( n \).

**Fact 3.1** Let \( d \geq 2 \) be an integer, the scheme \( \text{Hom}_d(C, X) \) is irreducible of dimension \( dn \).

**Proof.** The scheme of morphisms is described by simple Brill-Noether data: a line bundle \( \mathcal{L} \) on \( C \) of degree \( d \) and a linear map from \( V^\vee \) to \( H^0(C, \mathcal{L}) \). But for \( d \) positive, we have \( H^1(C, \mathcal{L}) = 0 \) therefore \( H^0(C, \mathcal{L}) \) is of constant dimension \( d \) for all \( \mathcal{L} \) and the result follows from the fact that \( \text{Jac}^d(C) \), the Jacobian of degree \( d \) line bundles on \( C \), is irreducible: the variety \( \text{Hom}_d(C, \mathbb{P}(V)) \) is a locally trivial bundle over \( \text{Jac}^d(C) \) with fiber \( \mathbb{P}(V \otimes H^0(C, \mathcal{L})) \) over \( \mathcal{L} \). \( \square \)

3.2 Grassmann varieties

We now deal with Grassmann varieties. Note that in this case the result was proved by A. Bruguères [Bru87]. We include here a proof that will serve as a model for the other cases. Let \( X \) be the Grassmann variety \( G(p, n) \) of \( p \)-dimensional vector spaces in a vector space of dimension \( n \).

**Proposition 3.2** The scheme \( \text{Hom}_d(C, X) \) is irreducible of dimension \( nd \) for \( d \geq n \).

**Proof.** Consider the open subset \( U \) given by Theorem 1.6. We have a morphism \( \phi : U \to Y \) with \( Y \simeq \mathbb{P}^{n-p} \) which is the vector bundle
\[
E = T_{\mathbb{P}^{n-p}}(-1) \otimes \mathcal{O}_{\mathbb{P}^{n-p}}^{p-1} = (\Omega_{\mathbb{P}^{n-p}}^{n-p-1}(n-p))^{p-1}.
\]
By Proposition 1.1, we only need to prove the irreducibility of the scheme \( \text{Hom}_d(C, U) \). For this we study the map \( \phi \). Let us denote by \( H_a \) the locally
closed subset of $\text{Hom}_d(C,Y)$ of maps whose image is contained in a linear subspace of codimension $a$ and not in a smaller linear subspace. The dimension of this locally closed subset is $a(n - p + 1 - a) + d(n - p + 1 - a)$. Indeed, there is a morphism $H_a \to G(n - p + 1 - a, n - p + 1)$ sending to a morphism the unique linear subspace of codimension $a$ containing its image. The fibers are isomorphic to $\text{Hom}_d(C, \mathbb{P}^{n-p-a})$.

Now the inverse image under the map $\Phi : \text{Hom}_d(C,U) \to \text{Hom}_d(C,Y)$ induced by $\phi$ of $H_a$ is irreducible of dimension $\dim H_a + \dim H^0(C, f^*E)$ for $f \in H_a$. By Proposition 2.7 we have $\dim H^0(C, f^*E) = (d \deg(E) + a)(p-1) = d(p-1) + a(p-1)$. In particular for $d \geq n$, the dimension of $\Phi^{-1}(H_a)$ is $(d + a)(n - a) = dn - a(d - n) - a^2 < dn$ for $a > 0$. The locally closed subsets $\Phi^{-1}(H_a)$ for $a > 0$ are therefore of dimension smaller than the dimensions (at least $dn$) of irreducible components of $\text{Hom}_d(C, U)$ thus the irreducible components of $\text{Hom}_d(\mathbb{P}^1, U)$ are those of $\Phi^{-1}(H_0)$ which is irreducible of dimension $dn$.

**Remark 3.3** For smaller degrees, the scheme $\text{Hom}_d(C,X)$ is not irreducible (except in degree 2). Its irreducible components were described in [Bru87].

### 3.3 Quadrics

We shall now deal with smooth quadrics. Note that in this case the irreducibility of the scheme of morphisms was proved by E. Ballico [Bal89] for quadrics of dimension $n \geq 6$ for curves of genus $g$ with $g \leq \lfloor \frac{n-1}{2} \rfloor$ and for the degree $d \geq 2g - 1$. Let $X$ be a smooth quadric of dimension $n$.

**Proposition 3.4** Assume that $\dim X \neq 3$, then the scheme $\text{Hom}_d(C,X)$ is irreducible of dimension $nd$ for $d \geq 3$.

**Proof.** We proceed as for the Grassmann variety and consider the open subset $U$ given by Theorem 1.6. We have a tower of affine bundles $\phi : U \to Y$ where $Y \simeq \mathbb{P}^r$ with $r = \lfloor n/2 \rfloor$ and the direction vector bundles are the pullbacks of $E = \Omega_{\mathbb{P}^r}^1(2)$ for $n$ even (here $\phi$ is a vector bundle) and to $E = \Omega_{\mathbb{P}^r}^1(2)$ and $E' = \mathcal{O}_{\mathbb{P}^r}(1)$ for $n$ odd.

By Proposition 2.7 the vector bundle $E$ satisfies $H^1(C, f^*E) = 0$ for any map $f : C \to Y$ of degree $d \geq 3$ as soon as $f$ does not factor through a line. In this latter case we have $\dim H^1(C, f^*E) = 1$ and we conclude by Proposition 1.3.

**Remark 3.5** (i) Note that if $\dim X \leq 2$, then $X$ is a product of projective lines and the result follows from that case. However, if $\dim X = 3$, then in the above proof we get $r = 1$ and $\Omega_{\mathbb{P}^2}^1(2) = \mathcal{O}_{\mathbb{P}^2}$, and we do not get the vanishing condition.

(ii) Note that for $d = 2$ the scheme $\text{Hom}_d(C,X)$ is also irreducible but not of the expected dimension since any degree 2 morphism factors through a line. The dimension is $2n + 1$. 

*Documenta Mathematica* 18 (2013) 679–706
Even if we cannot prove, with our method, Proposition 3.4 when $X$ is 3-dimensional, the result is still true.

**Proposition 3.6** Let $X = \mathbb{Q}^3$ be a smooth quadric of dimension 3, then the scheme $\text{Hom}_d(C, X)$ is irreducible of dimension $3d$ for $d \geq 3$.

We give a proof of this result for $d \geq 4$ in the appendix, using a different method. The case where $d = 3$ is given by the following remark.

**Remark 3.7** Any morphism of degree 2 or 3 from an elliptic curve to $\mathbb{Q}^3$ has to factor through a line. This is clear for degree 2, and for degree 3 the morphism factors through a plane. But as $\mathbb{Q}^3$ contains no plane, the image is contained in the intersection of $\mathbb{Q}^3$ with a plane. It has to be a line for degree reasons. In particular the scheme of morphisms $\text{Hom}_d(C, X)$ is irreducible of dimension $3 + 2d$ for $d = 2$ or $d = 3$.

### 3.4 Maximal orthogonal Grassmann varieties

Let $X$ be a maximal orthogonal Grassmann variety $G_{\mathbb{Q}}(n, 2n)$. The following result already appeared in the work of the second author [Per12] where it was proved by a different method.

**Proposition 3.8** The scheme $\text{Hom}_d(C, X)$ is irreducible of dimension $2(n - 1)d$ for $d \geq n - 1$.

**Proof.** The map $\phi : U \to Y$ is the vector bundle $E = \Lambda^2(T_{\mathbb{P}^{n-1}}(-1)) = \mathcal{O}_{\mathbb{P}^{n-1}}^{n-3}(n - 2)$ over $Y \simeq \mathbb{P}^{n-1}$. By Proposition 2.7, the fibers of $\Phi : \text{Hom}_d(C, U) \to \text{Hom}_d(C, Y)$ over the locally closed subset $H_a$ of morphisms whose image is non-degenerate in a linear subspace of codimension $a$ are of dimension $d(n - 2) + \binom{n}{2}$. Therefore we have $\dim p^{-1}(H_a) = 2d(n - 1) + a(n - a) + \binom{n}{2} - ad$. This dimension is strictly smaller than the expected dimension $2d(n - 1)$ for $a > 0$ and $d \geq n - 1$. The result follows. □

**Corollary 3.9** For $d \in [2, n]$, the scheme $\text{Hom}_d(C, X)$ is irreducible of dimension

$$2(n - 1)d + \frac{1}{2}(n - d)(n - d - 1).$$

**Proof.** Let $f : C \to X$ be a morphism of degree $d$, then the kernel $\ker(f)$ of $f$ i.e. the intersection of all subspaces $f(x)$ for $x \in C$ is of dimension at least $n - d$ (see [Buc03] for other applications of this definition). Indeed, consider the tautological exact sequence

$$0 \to K \to \mathcal{O}_X^{2n} \to Q \to 0$$

on $X$ where $K$ and $Q$ are the tautological subbundle and quotient bundle. On $X$ we have the identification $Q \simeq K^\vee$. Pulling back this sequence to $C$ via $f$ we get an exact sequence on the curve. Taking cohomology we obtain:

$$0 \to H^0(C, K) \to \mathcal{C}^{2n} \to H^1(C, K) \to H^1(C, K) \to \mathcal{C}^{2n} \to H^0(C, K) \to 0.$$
Let us denote by $h^0$ and $h^1$ the dimensions of $H^0(C, K)$ and $H^1(C, K)$. Since the map $f$ is of degree $d$ in $X$, the degree of $K$ is $-2d$ and we have the equality $h^0 = h^1 - 2d$. We deduce the inequality $h^0 \geq n - d$. This implies that $\ker(f)$ is of dimension at least $n - d$. Note also that for a general curve the kernel is of dimension exactly $n - d$.

Let us consider the following incidence variety $I = \{(f, V_{n-d}) \in \text{Hom}_d(C, X) \times \mathcal{G}_Q(n - d, 2n) / \text{ for all } x \in C \text{ we have } V_{n-d} \subset f(x) \subset V_{n-d}^\perp\}$. Here $\mathcal{G}_Q(n - d, 2n)$ is the Grassmann variety of isotropic subspaces of $\mathbb{C}^{2n}$ of dimension $n - d$. The above condition on $f$ simply translates in $V_{n-d} \subset \ker(f)$. Note that the second inclusion is implied by the first one since we have $f(x) = f(x)^\perp$. The projection $p : I \to \text{Hom}_d(C, X)$ is surjective and birational (because for a general curve the kernel is of dimension $n - d$) while the map $q : I \to \mathcal{G}_Q(n - d, 2n)$ is an equivariant locally trivial fibration with fibers $\text{Hom}_d(C, Y)$ with $Y \simeq \mathcal{G}_Q(d, 2d)$. By the previous proposition these fibers are irreducible of dimension $2d(d - 1)$. The result follows. □

3.5 Adjoint varieties of type B and D

Let us consider the variety $X = \mathcal{G}_Q(2, n)$ of isotropic planes in a vector space of dimension $n$ endowed with a non-degenerate bilinear form.

**Proposition 3.10** The scheme $\text{Hom}_d(C, X)$ is irreducible of dimension $(n - 3)d$ for $d \geq n - 1$.

**Proof.** The map $\phi : U \to Y$ is the vector bundle $E = (\mathcal{O}_{\mathbb{P}^{n-4}}(1))^\perp$ over $Y \simeq \mathbb{Q}_{n-4}$, where $\mathcal{O}_{\mathbb{P}^{n-4}}(-1)$ is the tautological subbundle on $\mathbb{Q}_{n-4}$.

Here we have to compute the dimension of the scheme of morphisms from an elliptic curve to a cone (the intersection of $\mathbb{Q}_{n-4}$ with a linear subspace of $\mathbb{P}^{n-3}$ in our case). Let $\mathfrak{C}$ be a cone in $\mathbb{P}^{k+r-1}$ associated to a quadratic form of rank $r$ (and therefore with kernel $K$ of dimension $k$).

**Lemma 3.11** Consider the open subset in the scheme of degree $d$ morphisms from an elliptic curve $C$ to $\mathfrak{C}$ consisting of morphisms whose image is non-degenerate. Then this open subset is of dimension at most $d(k + r - 2)$.

Remark that, if $k = 0$, the cone $\mathfrak{C}$ is a smooth quadric of dimension $(k + r - 2)$ and we already know the result with equality.

**Proof.** Consider the incidence variety $I = \{(V_1, V_{k+1}) \in \mathfrak{C} \times \mathcal{G}_Q(k + 1, k + r) / V_1 \subset V_{k+1} \supset K\}$. There is a diagram

$$
\begin{array}{ccc}
I & \xrightarrow{p} & \mathcal{G}_{r-2} \\
\downarrow q & & \\
\mathfrak{C} & \end{array}
$$

**Documenta Mathematica 18 (2013) 679–706**
The map $q$ is simply the blow-up of the vertex of the cone. If $f : C \to \mathcal{C}$ has a non degenerate image then its image is not contained in the vertex $\mathbb{P}(K)$ of the cone and we may consider the composition of $f$ with the projection from $K$. This defines a morphism $f' : C \to \mathbb{Q}_{r-2}$ which can be obtained by lifting $f$ to $f_I : C \to I$ and composing with $p$. The degree of this morphism is $d - x$ where $x$ is the multiplicity of the intersection of $f(C)$ with the vertex of the cone.

To compute the dimension of our open set, we consider the map defined by $f \mapsto f'$. Its image is contained in $\text{Hom}_{d-a}(\mathcal{C}, \mathbb{Q}_{r-2})$ which is of dimension $(d - x)(r - 2)$. The fiber is given by the morphisms $\mathcal{L} \to f''V_{k+1}$ where $\mathcal{L}$ is of degree $d$ and $V_{k+1}$ is the tautological subbundle in $\mathbb{Q}_{r-2} \simeq \{V_{k+1} \in \mathcal{G}_Q(k+1,k+r) / K \subset V_{k+1}\}$. Because the projection by $K$ gives a curve of degree $d - x > 0$ we have an extension $0 \to K \otimes \mathcal{O}_C \to f''V_{k+1} \to \mathcal{L}' \to 0$ with $\mathcal{L}'$ of degree $d - x$. This extension has to split, thus the fibers of the map are given by $\text{Pic}(\mathcal{L}, K \otimes \mathcal{O}_C \oplus \mathcal{L}')$ which is of dimension $kd + x$. The dimension is therefore at most $d(k + r - 2) - x(r - 3)$ proving the result for $r \geq 3$. But if $r \leq 2$, then the cone is a union of hyperplanes, thus the curve is degenerate.

By Proposition 2.7, since $E$ is the restriction to $\mathbb{Q}_{n-4}$ of $\Omega_{\mathbb{P}^{n-3}}^n(n-3)$, the fibers of $\Phi : \text{Hom}_d(\mathcal{C}, U) \to \text{Hom}_d(\mathcal{C}, Y)$ over the locally closed subset $H_a$ of morphisms whose image is non-degenerate in a linear subspace of codimension $a$ are of dimension $d + a$. Therefore the dimension of $\Phi^{-1}(H_a)$ is at most the sum of $d + a$, the dimension of the Grassmann variety $\mathcal{G}(n - 2 - a, n - 2)$ and the dimension of the scheme of degree $d$ morphisms from an elliptic curve $C$ to a cone $\mathcal{C}$ in $\mathbb{P}^{n-3-a}$, so that $\dim \Phi^{-1}(H_a) \leq d(n - a - 3) + a(n - 2 - a) + a$. This dimension is strictly smaller than the expected dimension $(n - 4)d$ for $a > 0$ and $d \geq n - 1$. The result follows.

Remark 3.12 The above argument shows that for $d = n - 2$, the scheme of morphisms has 2 irreducible components of the expected dimension.

3.6 Type $E_6$

3.6.1 Cayley plane

Let $X$ be isomorphic to $E_6/P_1$.

Proposition 3.13 The scheme $\text{Hom}_d(\mathcal{C}, X)$ is irreducible of dimension $12d$ for $d \geq 3$.

Proof. The map $\phi : U \to Y$ is one of the two spinor bundles $S$ over $Y \simeq \mathbb{Q}_8$. By Corollary 2.11, the fibers of $p : \text{Hom}_d(\mathcal{C}, U) \to \text{Hom}_d(\mathcal{C}, Y)$ over the locally closed subset $H_a$ of morphisms whose image is non degenerate in an isotropic linear subspace of codimension $a$ are of dimension $4d + 2^{\text{max}(0,a-6)}$ with $a \geq 5$. Therefore we have the equality $\dim \Phi^{-1}(H_a) = 12d + a(10 - a) - \frac{(10-a)(11-a)}{2} + 2^{\text{max}(0,a-6)} - ad$. This dimension is strictly smaller than the expected dimension $12d$ for $a > 0$ and $d \geq 3$. The result follows.
3.6.2 Adjoint variety

Let $X$ be isomorphic to $E_8/P_2$.

**Proposition 3.14** The scheme $\text{Hom}_d(C, X)$ is irreducible of dimension $11d$ for $d \geq 9$.

**Proof.** The map $\phi : U \to Y$ is a tower of affine bundles over $Y \simeq \mathbb{Q}_6$ with direction vector bundles $E = (\mathcal{O}_{\mathbb{Q}_6}(1))^\perp$ (where $\mathcal{O}_{\mathbb{Q}_6}(-1)$ is the tautological subbundle on $\mathbb{Q}_6$) and $E' = S \oplus S'$ where $S$ and $S'$ are the two spinor bundles on $\mathbb{Q}_6$. Note that the bundle $E$ is the restriction of the tautological quotient bundle of $P_7$ to $\mathbb{Q}_6$. By Proposition 2.7 and Corollary 2.11, the fibers of $\Phi : \text{Hom}_d(C, U) \to \text{Hom}_d(C, Y)$ over the locally closed subset $H_a$ of morphisms whose image is non degenerate in a linear subspace of codimension $a$ are of dimension at most $5d + a$ for $a \leq 3$ or $a \geq 4$ and the subspace containing the curve is non isotropic and $5d + a + 2^{\max(0, a-4)}$ for $a \geq 4$ and the subspace containing the curve is isotropic. Therefore we have the equality $\dim \Phi^{-1}(H_a) = d(11 - a) + a(8 - a) + a$ for $a \leq 3$ or $a \geq 4$ and the subspace containing the curve is non isotropic and $d(13 - a) + a(8 - a) - \frac{(5-a)(a-6)}{2} + a + 2^{\max(0, a-4)}$ for $a \geq 4$ and the subspace containing the curve is isotropic. This dimension is strictly smaller than the expected dimension $11d$ for $a > 0$ and $d \geq 9$. The result follows. □

3.7 Type $E_7$

3.7.1 Freudenthal variety

Let $X$ be isomorphic to $E_7/P_7$.

**Proposition 3.15** The scheme $\text{Hom}_d(C, X)$ is irreducible of dimension $18d$ for $d \geq 8$.

**Proof.** The map $\phi : U \to Y$ is a tower of affine bundles with direction vector bundles $E = \Omega^6_{\mathbb{P}_6}(6)$ and $E' = \Omega^3_{\mathbb{P}_6}(3)$. By Proposition 2.7, the fibers of $\Phi : \text{Hom}_d(C, U) \to \text{Hom}_d(C, Y)$ over the locally closed subset $H_a$ of morphisms whose image is non degenerate in a linear subspace of codimension $a$ are of dimension at most $11d + a + \left(\begin{smallmatrix}a \\ 2\end{smallmatrix}\right)$. Therefore we have $\dim \Phi^{-1}(H_a) = 18d + a(7 - a) + a + \left(\begin{smallmatrix}a \\ 3\end{smallmatrix}\right) - ad$. This dimension is strictly smaller than the expected dimension $18d$ for $a > 0$ and $d \geq 8$. The result follows. □

3.7.2 Adjoint variety

Let $X$ be isomorphic to $E_7/P_1$.

**Proposition 3.16** The scheme $\text{Hom}_d(C, X)$ is irreducible of dimension $17d$ for $d \geq 11$. 

*Documenta Mathematica 18 (2013) 679–706*
Proposition 3.17 The scheme $\text{Hom}_d(C, U) \rightarrow \text{Hom}_d(C, Y)$ over the locally closed subset $H_d$ of morphisms whose image is non degenerate in a linear subspace of codimension $a$ are of dimension at most $9d + a$ for $a \leq 4$ or $a \geq 5$ and the subspace containing the curve is non isotropic and $9d + a + 2^{\max(0, a - 6)}$ for $a \geq 5$ and the subspace containing the curve is isotropic. Therefore we have the equality $\dim \Phi(H_d) = d(17 - a) + a(10 - a) + a$ for $a \leq 4$ or $a \geq 5$ and the subspace containing the curve is non isotropic and $d(19 - a) + a(10 - a) - \frac{(10 - a)(11 - a)}{2} + a + 2^{\max(0, a - 6)}$ for $a \geq 5$ and the subspace containing the curve is isotropic. This dimension is strictly smaller than the expected dimension $17d$ for $a > 0$ and $d \geq 11$. The result follows.

3.8 Adjoint variety of type $E_8$

Let $X$ be isomorphic to $E_8/P_8$.

**Proposition 3.17** The scheme $\text{Hom}_d(C, X)$ is irreducible of dimension $29d$ for $d \geq 15$.

Proof. The map $\phi : U \rightarrow Y$ is a tower of affine bundles over $Y \simeq \mathbb{Q}_8$ with direction vector bundles $E = (\mathcal{O}_{\mathbb{Q}_8}(1))^\perp$ where $\mathcal{O}_{\mathbb{Q}_8}(-1)$ is the tautological subbundle on $\mathbb{Q}_8$ and $E' = S^2$ where $S$ is one of the two spinor bundles on $\mathbb{Q}_8$. By Proposition 2.7 and Corollary 2.11, the fibers of $\Phi : \text{Hom}_d(C, U) \rightarrow \text{Hom}_d(C, Y)$ over the locally closed subset $H_d$ of morphisms whose image is irreducible of dimension $29d$ for $d \geq 15$.

3.9 Adjoint variety of type $F_4$

Let $X$ be isomorphic to $F_4/P_1$.

**Proposition 3.18** The scheme $\text{Hom}_d(C, X)$ is irreducible of dimension $11d$ for $d \geq 8$.

Proof. The map $\phi : U \rightarrow Y$ is a tower of affine bundles over $Y \simeq \mathbb{Q}_8$ with direction vector bundles $E = (\mathcal{O}_{\mathbb{Q}_8}(1))^\perp$ where $\mathcal{O}_{\mathbb{Q}_8}(-1)$ is the tautological subbundle on $\mathbb{Q}_8$ and $E' = S$ where $S$ is the spinor bundle on $\mathbb{Q}_8$. By Proposition 2.7 and Corollary 2.11, the fibers of $\Phi : \text{Hom}_d(C, U) \rightarrow \text{Hom}_d(C, Y)$
over the locally closed subset $H_a$ of morphisms whose image is non degenerate in a linear subspace of codimension $a$ are of dimension at most $6d + a$ for $a \leq 3$ or $a \geq 4$ and the subspace containing the curve is non isotropic and $6d + a + 2\max(0, a - 5)$ for $a \geq 4$ and the subspace containing the curve is isotropic. Therefore we have the equality $\dim \Phi^{-1}(H_a) = d(11 - a) + a(7 - a) + a$ for $a \leq 3$ or $a \geq 4$ and the subspace containing the curve is non isotropic and $d(13 - a) + a(7 - a) - \frac{(7 - a)(8 - a)}{2} + a + 2\max(0, a - 5)$ for $a \geq 4$ and the subspace containing the curve is isotropic. This dimension is strictly smaller than the expected dimension $11d$ for $a > 0$ and $d \geq 8$. The result follows.

**Remark 3.19** Note that, in Propositions 3.14, 3.15, 3.16, 3.17 and 3.18, it is possible that one can improve the bound on the dimension of $\Phi^{-1}(H_a)$. Indeed if the affine bundle restricted to a curve in $H_a$ is not a vector bundle then this bound will be strictly smaller. We therefore expect a better bound $d(X)$ in these situations.

4 Stable maps

Let us consider the moduli space $\overline{M}_{1,d}(X)$ of stable maps with source an elliptic curve. Recall the following general result obtained in [KP01].

**Theorem 4.1** ([KP01]) The moduli space $\overline{M}_{1,d}(X)$ is connected for any rational homogeneous space $X$ and any degree $d$.

Note, in contrast, that for any $X$ in the tables of the introduction, we have the proposition.

**Proposition 4.2** For $d \geq d(X)$, the moduli space $\overline{M}_{1,d}(X)$ is never irreducible.

**Proof.** Indeed, consider the locally closed subset of $\overline{M}_{1,d}(X)$ where the degree on the unique elliptic irreducible component of the source of the map is 2. The dimension of this locally closed subset is $dc_1(X) + \dim X$ which is always strictly bigger than the expected dimension $dc_1(X)$ which is the dimension of the component containing irreducible curves for $d \geq d(X)$. □

However, there is a natural decomposition of this space into locally closed subsets $M^\tau_{1,d_\tau}(X)$ parametrised by their combinatorial graph $\tau$ and combinatorial degree $d_\tau$ (see [KM94] or also [KP01]). Note that for stable maps of genus one, at most one irreducible component of the source map is an elliptic curve. We call this component (if it exists) the elliptic component of $\tau$. As a consequence of our results we obtain the following proposition.

**Proposition 4.3** If the degree $d_\tau$ is bigger that $d(X)$ on the elliptic component of $\tau$, then $M^\tau_{1,d_\tau}(X)$ is irreducible.
Proof. If there is no elliptic component, then the result follows from the fact (see [BCMP10]) that the moduli space of stable maps of genus 0 passing through at most 2 fixed points is irreducible for any homogeneous space $X$.

If there is an elliptic component, then we claim that the natural forgetful map to the moduli space of elliptic curves is flat with irreducible fibers. Indeed, the fiber over $C$ is isomorphic to the product of $\text{Hom}_d(C, X)$ with some moduli spaces of rational stable maps passing through at most 2 fixed points. These last moduli spaces are irreducible of the expected dimension by [BCMP10, Corollary 3.3]. Now by dimension count, any irreducible component of $M^\tau_d(X)$ has to dominate the moduli space of elliptic curves (the fibers have constant dimension 1 less than the expected dimension) thus the map is flat. The result follows by [Gro65, Corollaire 2.3.5.(ıı)].

□

A The quadric of dimension 3

In this section, we prove Proposition 3.6 for $d \geq 4$ using Bott-Samelson resolutions (the reader could see [Dem74] for more details on Bott-Samelson resolutions). This method was already used by the second author in [Per12] to study elliptic curves in spinorial varieties.

First we introduce some notation. Let $V$ be a 5-dimensional vector space endowed with a non degenerate quadratic form $Q$. Then $X$ is the subvariety of $P(V)$ consisting of isotropic lines in $V$.

For any isotropic flag $W_\bullet = (W_1, W_2)$ of $V$, we define a Bott-Samelson resolution $\pi : \tilde{X} \rightarrow X$ by

$$\tilde{X} = \tilde{X}_{W_\bullet} = \{(V'_1, V_2, V_1) \in \mathbb{Q}_3 \times \mathbb{G}_Q(2, 5) \times \mathbb{Q}_3 \mid V'_1 \subset W_2, V'_1 \subset V_2, V_1 \subset V_2\},$$

and $\pi(V'_1, V_2, V_1) = V_1$. The map $\pi$ is birational, in particular it is an isomorphism over the open $B$-orbit of $X$, where $B$ is the Borel subgroup of $\text{SO}(5)$ associated to $W_\bullet$. Moreover if we define $X_0 = \{\text{pt}\}$, $X_1 = \{V'_1 \in \mathbb{Q}_3 \mid V'_1 \subset W_2\}$ $X_2 = \{(V'_1, V_2) \in \mathbb{Q}_3 \times \mathbb{G}_Q(2, 5) \mid V_2 \supset V'_1 \subset W_2\}$ and $X_3 = \tilde{X}$, there is sequence of $\mathbb{P}^1$-bundles

$$\tilde{X} = X_3 \rightarrow X_2 \rightarrow X_1 \rightarrow X_0 = \{\text{pt}\}.$$ 

The proof of Proposition 3.6 uses the 3 following facts.

Fact 1 Under some conditions $(*)_d$ on the class $\tilde{\alpha}$ of 1-cycles in $\tilde{X}$ (see Definition A.4), the scheme $\text{Hom}_d(C, \tilde{X})$ of morphisms from $C$ to $\tilde{X}$ of class $\tilde{\alpha}$ is irreducible of dimension at most $3d - 1$ with equality for a unique class $\tilde{\alpha}$.

Fact 2 For all $f \in \text{Hom}_d(C, X)$ such that $f(C)$ is not contained in an isotropic projective line of $P(V)$, we can choose a flag $W_\bullet$ such that $f$ lifts into a unique $\tilde{f} : C \rightarrow \tilde{X}_{W_\bullet}$, such that $\tilde{\alpha} := [\tilde{f}_*(C)]$ satisfies the conditions $(*)_d$. And, for each such $f$, the set of flags that we can choose is a subvariety (not closed) of $\text{SO}(5)/B$ of dimension 3.
Fact 3 For $d \geq 4$ we prove that the scheme of morphisms from $C$ to $X$ of degree $d$ such that $f(C)$ is contained in an isotropic projective line of $\mathbb{P}(V)$ cannot be an irreducible component of $\text{Hom}_d(C, X)$. Assuming these facts for the moment, let us prove Proposition 3.6. Let $I$ be the set of couples $(W_\bullet, f) \in \text{SO}(5)/B \times \text{Hom}_n(C, \tilde{X}_{W_\bullet})$ where $\text{Hom}_n(C, \tilde{X}_{W_\bullet})$ is the scheme of morphisms $\tilde{f}$ from $C$ to $\tilde{X}_{W_\bullet}$ such that $[\tilde{f}_*(C)]$ satisfies $(*)_{d}$. Call $p$ and $q$ the natural projections from $I$ to $\text{Hom}_d(C, X)$ and to $\text{SO}(5)/B$ respectively. Then, by the second and third facts, $p$ is dominant as soon as $d \geq 4$ and general fibers of $p$ have dimension 3. Now, for any flag $W_\bullet$, the fiber $q^{-1}(W_\bullet)$ is the disjoint union, over the $\tilde{\alpha}$ satisfying conditions $(*)_d$, of the schemes $\text{Hom}_d(C, \tilde{X}_{W_\bullet})$ of morphisms $\tilde{f}$ from $C$ to $\tilde{X}_{W_\bullet}$ such that $[\tilde{f}_*(C)] = \tilde{\alpha}$. By the first fact, exactly one of these latter schemes has maximal dimension $3d - 1$ and moreover this scheme is irreducible, so that the fiber $q^{-1}(W_\bullet)$ is irreducible. Thus $I$ is a union of irreducible connected components of dimension at most $3d - 1 + 3d + 3$ with equality for a unique component. The images in $\text{Hom}_d(C, X)$ of these components are irreducible of dimension at most $3d + 3 - 3 = 3d$ which is the expected dimension. Therefore all these images are contained in the closure of the image of the maximal one proving the irreducibility. Moreover the dimension of $\text{Hom}_d(C, X)$ is therefore $4 + (3d - 1) - 3 = 3d$.

Let us prove the three facts one by one. We begin with a general lemma.

**Lemma A.1** Let $\phi : X \to Y$ be a $\mathbb{P}^1$-fibration with a section $\sigma$. Let $\tilde{\alpha}$ be a class of 1-cycles in $X$. Denote by $T_\phi$ the relative tangent bundle and by $\xi$ the divisor $\sigma(Y)$. Suppose that $\tilde{\alpha} \cdot \xi \geq 0$ and that $\tilde{\alpha} \cdot (T_\phi - \xi) > 0$. Then, if $\text{Hom}_{\tilde{\alpha}, \phi}(C, Y)$ is irreducible, $\text{Hom}_{\tilde{\alpha}}(C, X)$ is also irreducible, and

$$\dim(\text{Hom}_{\tilde{\alpha}}(C, X)) = \dim(\text{Hom}_{\tilde{\alpha}, \phi}(C, Y)) + \tilde{\alpha} \cdot T_\phi.$$ 

**Proof.** Let $E$ be a rank two vector bundle on $Y$ such that $X = \mathbb{P}_Y(E)$. The section $\sigma$ is given by a surjection $E \to L$ where $L$ is an invertible sheaf on $Y$. Let $N$ be the kernel of this map. We first study the fiber of the map $\text{Hom}_{\tilde{\alpha}}(C, X) \to \text{Hom}_{\tilde{\alpha}, \phi}(C, Y)$ given by composition of morphisms over a morphism $f : C \to Y$. An element in this fiber is given by a lift of $f$ i.e. a surjective map $f^* E \to M$ of vector bundles on $C$ where $M$ is an invertible sheaf with $2 \deg(M) - \deg(f^* E) = \tilde{\alpha} \cdot T_\phi$. An element in the fiber is therefore given by an invertible sheaf $M$ on $C$ with degree $d = \frac{1}{2}(\deg(f^* E) + \tilde{\alpha} \cdot T_\phi)$ and a surjective map in $\mathbb{P}(\text{Hom}(f^* E, M))$. Note that we have the equalities $\tilde{\alpha} \cdot \xi = \deg(M) - \deg(f^* N)$ and $\tilde{\alpha} \cdot (T_\phi - \xi) = \deg(M) - \deg(f^* L)$. We discuss two cases.

If $\tilde{\alpha} \cdot \xi > 0$ then $\text{Hom}(f^* E, M)$ is isomorphic to $\text{Hom}(f^* N, M) \oplus \text{Hom}(f^* L, M)$ and its dimension $\tilde{\alpha} \cdot T_\phi$ does not depend on $f$. For any invertible sheaf $M$, there exist surjections $f^* E \to M$ and they form an open subset of $\text{PHom}(f^* E, M) \times \text{Pic}_d(C)$. Doing this construction in family as in [Per02, Proposition 4] we get...
a smooth fibration or relative dimension $\alpha \cdot T_0$ over $\text{Hom}_{E,M}(C,Y)$ and the result follows.

If $\alpha \cdot \xi = 0$, then if $M \not= f^*N$ we have $\text{Hom}(f^*N, M) = 0$ and any map $f^*E \to M$ factors through $f^*L$ and is never surjective because $\text{deg}(M) - \text{deg}(f^*L) = \alpha \cdot (T_0 - \xi) > 0$. Therefore any pair $(M, p : f^*E \to M)$ of the fiber satisfies $M \simeq f^*N$. In that case, because of the equality $\text{deg}(M) - \text{deg}(f^*L) = \alpha \cdot (T_0 - \xi) > 0$, the sheaf $f^*E$ is isomorphic to $M \oplus f^*L$ and we have an isomorphism $\text{Hom}(f^*E, M) \simeq \text{Hom}(f^*N, M) \oplus \text{Hom}(f^*L, M)$. The dimension of $\text{Hom}(f^*E, M)$ is $\alpha \cdot T_0 + 1$ and does not depend on $f$. The fiber is therefore a non empty open subset of $\text{FHom}(f^*E, M)$. We therefore again get a smooth fibration of relative dimension $\alpha \cdot T_0$ over $\text{Hom}_{E,M}(C,Y)$. 

Applying this lemma to the $\mathbb{P}^1$-fibrations $\phi_i : X_i \to X_{i-1}$ for $i \in \{1, 2, 3\}$ given above by a Bott-Samelson resolution $\tilde{X}$ of $\mathcal{Q}_3$, we obtain the following corollary. We denote by $T_i$ the relative tangent space of $\phi_i$, we define three divisors $\xi_1$, $\xi_2$, $\xi_3$ of $\tilde{X}$, which are given by natural sections of the $\phi_i$, by

$$\xi_1 = \{(V'_1, V_2, V_1) \in \tilde{X} \mid V'_1 = W_1\}$$
$$\xi_2 = \{(V'_1, V_2, V_1) \in \tilde{X} \mid V_2 = W_2\}$$
$$\xi_3 = \{(V'_1, V_2, V_1) \in \tilde{X} \mid V_1 = V'_1\}.$$

We will again denote by $T_i$ the pull-back of $T_i$ under the morphisms $\phi_j$ for $j > i$.

**Corollary A.2** Let $\alpha$ be a class of 1-cycles in $\tilde{X}$ satisfying for all $i \in \{1, 2, 3\}$, $\alpha \cdot \xi_i \geq 0$ and $\alpha \cdot (T_i - \xi_i) > 0$.

Then $\text{Hom}_{\tilde{X}}(C, \tilde{X})$ is irreducible of dimension $\alpha \cdot (T_1 + T_2 + T_3)$.

To obtain the first fact, we use the following result that can be deduced from [Per05, Corollary 3.8 and Fact 3.7] by a short computation.

**Proposition A.3** (i) The relative tangent bundles $T_i$ are expressed in terms of the $\xi_i$ as follows:

$$T_1 = 2\xi_1, \quad T_2 = 2\xi_2 + \xi_1 \quad \text{and} \quad T_3 = 2\xi_3 + 2\xi_2.$$

In particular, we have $\alpha \cdot \xi_i \geq 0$ and $\alpha \cdot (T_i - \xi_i) > 0$ for all $i \in \{1, 2, 3\}$ as soon as $\alpha \cdot \xi_1 > 0$, $\alpha \cdot \xi_2 > 0$ and $\alpha \cdot \xi_3 \geq 0$, or $\alpha \cdot \xi_1 > 0$, $\alpha \cdot \xi_2 \geq 0$ and $\alpha \cdot \xi_3 > 0$.

(ii) The pull-back $\pi^*\mathcal{Q}_3(1)$ of the ample generator of the Picard group of $\mathcal{Q}_3$ equals $\xi_1 + 2\xi_2 + \xi_3$. In particular, if $\alpha = \pi^*(\alpha)$ where $\alpha$ is the class of 1-cycles of degree $d$ in $\mathcal{Q}_3$, we have $d = \alpha \cdot (\xi_1 + 2\xi_2 + \xi_3)$.

**Definition A.4** We say that $\alpha$ satisfies conditions $(\ast)_d$ if

- $\alpha \cdot \xi_1 > 0$, $\alpha \cdot \xi_2 > 0$ and $\alpha \cdot \xi_3 \geq 0$, or $\alpha \cdot \xi_1 > 0$, $\alpha \cdot \xi_2 \geq 0$ and $\alpha \cdot \xi_3 > 0$;
- $d = \alpha \cdot (\xi_1 + 2\xi_2 + \xi_3)$. 

Documenta Mathematica 18 (2013) 679–706
Remark that, under conditions (\(\ast\))\(_d\), we have \(\hat{\alpha} \cdot (T_1 + T_2 + T_3) = 3d - \hat{\alpha} \cdot (2\xi_2 + \xi_3)\) and then the dimension of the scheme \(\text{Hom}_d(C, \bar{X})\) is at most \(3d - 1\) with equality if an only if \(\hat{\alpha} \cdot \xi_1 = d - 1\), \(\hat{\alpha} \cdot \xi_2 = 0\) and \(\hat{\alpha} \cdot \xi_3 = 1\). We deduce the first fact from above results.

To prove the second fact, fix a curve \(f \in \text{Hom}_d(C, \mathbb{Q}_3)\) such that \(f(C)\) is not contained in an isotropic projective line of \(\mathbb{P}(V)\). We begin by proving the following lemma. If \(E\) is a vector subspace of \(V\), we denote by \(E^\perp\) the orthogonal subspace to \(E\) in \(V\) (with respect to \(Q\)). For any point \(x\) in \(\mathbb{P}(V)\), we denote by \(\pi\) the corresponding line in \(V\).

**Lemma A.5** There exist 3 points \(x_0, x_1, x_2\) of \((f(C)\) and an isotropic flag \(W_* = (W_1, W_2)\) such that \(\pi_0\) is not contained in \(W_1^\perp\), \(\pi_1\) is contained in \(W_1^\perp\) but not in \(W_2\), and \(\pi_2\) is contained in \(W_2\).

**Proof.** First remark that, because \(f(C) \subset \mathbb{Q}_3\), the condition saying that \(f(C)\) is not contained in an isotropic projective line of \(\mathbb{P}(V)\) means that \(f(C)\) is not contained in a projective line of \(\mathbb{P}(V)\).

Then there exist 3 non-collinear points \(x_0, x_1, x_2\) in \(f(C)\). Let \(E := \pi_1^\perp \oplus \pi_2^\perp\) and \(F := \pi_1^\perp \oplus \pi_2\). We can assume that \(Q|_F\) is non-degenerate, in particular \(V = F \oplus F^\perp\) and \(Q\) is non-degenerate on \(F^\perp\). Then there exists an isotropic line \(W_1\) in \(F^\perp\setminus E^\perp\). And we can define \(W_2 = W_1 \oplus \pi_2\), because \(\pi_2 \subset F^\perp\).

It is obvious that \(\pi_2\) is contained in \(W_2\). And we can also easily check that \(\pi_0\) is not contained in \(W_1^\perp\) and that \(\pi_1\) is contained in \(W_1^\perp\) but not in \(W_2\). \(\square\)

**Corollary A.6** There exists a flag \(W_*\) such that \(f\) lifts into a unique \(\tilde{f} : C \to \tilde{X}_{W_*}\), and \([\tilde{f}_*(C)]\) satisfies the conditions (\(\ast\))\(_d\).

**Proof.** Let \(W_*\) be an isotropic flag and \((x_0, x_1, x_2) \in f(C)^3\) as in Lemma A.5. Denote by \(B\) the Borel subgroup of \(\text{SO}(5)\) corresponding to \(W_*\). Then the open \(B\)-orbit \(\Omega\) in \(\mathbb{Q}_3\) is the set of points \(x\) such that \(\pi\) is not contained in \(W_1^\perp\).

Recall that, over \(\Omega\), the morphism \(\pi : \tilde{X}_{W_*} \to \mathbb{Q}_3\) is an isomorphism. Thus, since \(C\) is smooth and \(f(C)\) intersects \(\Omega\) (at least in \(x_0\)), \(f\) lifts into a unique \(\tilde{f} : C \to \tilde{X}_{W_*}\).

Moreover, we can compute that \(\pi(\xi_1) = \{ x \in \mathbb{Q}_3 \mid \pi \subset W_1^\perp \}\) and \(\pi(\xi_2) = \pi(\xi_3) = \{ x \in \mathbb{Q}_3 \mid \pi \subset W_2 \}\). So, for any \(i \in \{1, 2, 3\}\), \(\pi^{-1}(\Omega)\) does not intersect \(\xi_i\) so that \([\tilde{f}_*(C)] \cdot \xi_i \geq 0\). Also, the existence of \(x_1\) implies that \([\tilde{f}_*(C)] \cdot \xi_1 > 0\) and the existence of \(x_2\) implies that \([\tilde{f}_*(C)] \cdot \xi_2 > 0\) or \([\tilde{f}_*(C)] \cdot \xi_3 > 0\). \(\square\)

Let us now explain why the set of flags that we can choose is of dimension 3. In the proof of Lemma A.5, we can note that, when \(x_0, x_1, x_2\) are fixed in \(f(C)^3\), we choose \(W_1\) in a quadric of dimension 1 and \(W_2\) is uniquely determined. Moreover, if \(W_*\) is fixed, then we have finitely many choices for \(x_1\) and \(x_2\) whereas the set of possible \(x_0\) is one-dimensional. At the end, the dimension we are looking for is \(3 + 1 - 1 = 3\).
Finally, let us prove the third fact. Let $J$ be the scheme of morphisms from $C$ to $X$ of degree $d$ such that $f(C)$ is contained in an isotropic projective line of $\mathbb{P}(V)$. We have a natural projection from $J$ to the variety $G_Q(2,5)$ of isotropic projective lines in $\mathbb{P}(V)$, which is of dimension 3. The fibers are isomorphic to the scheme $\text{Hom}_d(C, \mathbb{P}^1)$ of morphisms of degree $d$ from $C$ to $\mathbb{P}^1$ and are of dimension $2d$. The dimension of $J$ is therefore $2d + 3$. If $d \geq 4$, we have $2d + 3 < 3d$ and then $J$ cannot be an irreducible component of $\text{Hom}_d(C,X)$.

**References**


Quillen Homology for Operads via Gröbner Bases

Vladimir Dotsenko and Anton Khoroshkin

Received: March 23, 2012
Revised: December 23, 2012
Communicated by Stefan Schwede

Abstract. The main goal of this paper is to present a way to compute Quillen homology of operads. The key idea is to use the notion of a shuffle operad we introduced earlier; this allows to compute, for a symmetric operad, the homology classes and the shape of the differential in its minimal model, although does not give an insight on the symmetric groups action on the homology. Our approach goes in several steps. First, we regard our symmetric operad as a shuffle operad, which allows to compute its Gröbner basis. Next, we define a combinatorial resolution for the “monomial replacement” of each shuffle operad (provided by the Gröbner bases theory). Finally, we explain how to “deform” the differential to handle every operad with a Gröbner basis, and find explicit representatives of Quillen homology classes for a large class of operads. We also present various applications, including a new proof of Hoffbeck’s PBW criterion, a proof of Koszulness for a class of operads coming from commutative algebras, and a homology computation for the operads of Batalin–Vilkovisky algebras and of Rota–Baxter algebras.

2010 Mathematics Subject Classification: 18G10 (Primary); 13P10, 16E05, 18D50, 18G55 (Secondary)

1The first author’s research was supported by Grant GeoAlgPhys 2011-2013 awarded by the University of Luxembourg and by an IRCSET research fellowship. The second author’s research was supported by grants NSh3349.2012.2, RFBR-10-01-00836, RFBR-CNRS-10-01-93111, RFBR-CNRS-10-01-93113, and a grant Ministry of Education and Science of the Russian Federation under the contract 14.74011.081.
INTRODUCTION

CONTEXT AND THE MAIN GOAL OF THE PAPER

Quillen’s philosophy of homotopical algebra [53, 54] suggests to study invariants of associative algebras (and their variants in various monoidal categories) within the homotopical category obtained from the usual category of (differential graded) algebras via the localisation inverting all quasi-isomorphisms. One of the key invariants of that sort is what is now called Quillen homology, the left derived functor of the functor of indecomposables $A \mapsto A^{ab} := A/A^2$ (this functor is also somewhat informally called “abelianisation”). Since we work in the homotopical category, studying an algebra is the same as studying its cofibrant replacement, which in many known model categories (associative algebras, operads etc.) is given by a (quasi-)free resolution. One such resolution, given by the cobar-bar construction, is readily available, however, sometimes it is preferable (and possible) to have a smaller resolution. The so called minimal resolution (if it exists) has Quillen homology of the algebra as its space of generators. In general, given an algebra $A$ and a free resolution $(F_\bullet, d)$ of $A$, Quillen homology $H^Q(A)$ is isomorphic to the homology of the differential induced on the space of indecomposable elements of $F_\bullet$, i.e., on the space of generators of our resolution. As a vector space, it can also be identified with the homology of the differential induced on the space of generators of a resolution of the trivial $A$-module by free right $A$-modules, that is the appropriate Tor groups, also called syzygies of the given algebra. One general way of computing the space of syzygies is a step-by-step procedure which is usually referred to as the Koszul–Tate method [36, 60]. However, in some cases it is possible to visualise the whole space of syzygies “in one go”, like the Koszul duality theory [27, 51] suggests. A question raised by Jean-Louis Loday in [42, Question 7] is to compare the computations via the Koszul duality theory (when available) with those by the Koszul–Tate approach. In this paper, we give a method that brings those two approaches together, applying the machinery of Gröbner bases and thus understanding the intrinsic structure of relations between relations in the spirit of Koszul duality.

One of most important practical results provided (in many different frameworks) by Gröbner bases is that when dealing with various linear algebra information (bases, dimensions etc.) one can replace an algebra with complicated relations by an algebra with monomial relations without losing any information of that sort. When it comes to questions of homological algebra, things become more subtle, since homology may “jump up” for a monomial replacement of an algebra. However, the idea of applying Gröbner bases to problems of homological algebra is far from hopeless. It turns out that for monoids with monomial relations it is often possible to construct very neat resolutions that can be used for various computations; furthermore, the data computed by these resolutions can be used to obtain results in the general...
Proposed methods

Operads that usually arise naturally in various topics are symmetric operads; they encode intrinsic properties of operations with several arguments acting on certain algebraic or geometric objects, and as such are equipped with the action of symmetric groups. The presence of such an action instantly implies that there is no meaningful notion of a monomial operad: any notion of that sort is not rich enough for every operad to have a monomial replacement. A way to deal with this problem proposed in [13] is to include symmetric operads in a larger universe of shuffle operads, where every object has a monomial replacement. On shuffle operads, no symmetries are allowed to act directly: the only way symmetries enter the game is through the formulae for compositions of operations. In that category, it is possible to make the Gr"obner bases machinery work, and hence there is hope that it can be applied to questions of homological algebra. It is indeed possible, along the following lines.

We begin with a resolution which generally not minimal even in the monomial case, but has the advantage of being purely combinatorial and not using much information about the underlying monoidal category. It is based on the inclusion–exclusion principle, and is in a sense a version of the cluster method of enumerative combinatorics due to Goulden and Jackson [26]. The resolution obtained is not always minimal, and we also discuss how to use it to compute the Quillen homology as a vector space, using algebraic Morse theory. This is followed, by an explanation of how to “deform” the differential of our resolution to incorporate lower terms of relations and handle arbitrary algebras with known Gr"obner bases. Note that the Quillen homology of a symmetric operad is, as a homotopy shuffle co-operad, isomorphic to the Quillen homology of that operad considered in the shuffle category. Since our results allow to compute the homotopy co-operad structure on the Quillen homology via the homotopy transfer theorem for homotopy co-operads [17], in principle we recover most of the information on Quillen homology in the symmetric category (and of course the shape the differential in the minimal model). The reader will see that in some of the examples we discuss.

As we mentioned earlier, our approach generalises the Koszul duality theory for defining relations of arbitrary degrees. In the case of a Koszul operad, one is able to write down a formula for its minimal resolution right away: such a resolution has the Koszul dual co-operad as its space of generators. Most known examples of Koszul algebras and operads actually satisfy the PBW condition [31, 51], or equivalently have a quadratic Gr"obner basis [13, 50]. In that case, the minimal resolution provided by the Koszul duality theory coincides with that obtained by our methods. Generally, Gr"obner bases allow to choose a “good” system of relations that captures the structure of relations between relations (higher syzygies).
Related results

In the case of usual associative algebras and right modules, the approach we discuss has been known since the celebrated paper of Anick [1] where for a monomial algebra a minimal right module resolution of the trivial module was computed, and an explicit way to deform the differential was presented to handle the general case. Later, Anick’s resolution was generalised to the case of categories by Malbos [44] who also asked whether this work could be extended to the case of operads. Results of this paper give such an extension, and suggest a way to handle associative algebras presented via generators and relations in many different monoidal categories (e.g. commutative associative algebras, associative dialgebras, (shuffle) coloured operads, dioperads, 4PROP’s) in a uniform way. If, in addition, we assume that our algebras are linear spans of algebras in sets, our constructions are closely related to those of free polygraphic resolutions for (∞, 1)-categories obtained by methods of rewriting theory [49]; this relationship is currently investigated by the first author in a joint work with Yves Guiraud and Philippe Malbos, and will be discussed elsewhere.

Overview of applications

There are various applications of our approach; some of them are presented in this paper. Two interesting theoretical applications are a new short proof of Hoffbeck’s PBW criterion for operads [31], and an upper bound on the homology for operads obtained from commutative algebras; in particular, we prove that an operad obtained from a Koszul commutative algebra is Koszul. Some interesting concrete examples where all steps of our construction can be completed are the case of the operad $RB$ of Rota–Baxter algebras, its noncommutative analogue $ncRB$, and, the last but not the least, the operad $BV$ of Batalin–Vilkovisky algebras. Using our methods, we were able to compute Quillen homology of the operad $BV$ and relate it to the gravity operad of Getzler [25]. After completing the first version of this paper in 2009 [12], we learned that these (and other) results concerning the operad $BV$ were also obtained independently by Drummond-Cole and Vallette [17]. Our methods appear to be completely different; we also believe that our approach to the operad $BV$ is of independent interest as an illustration of a rather general method to compute Quillen homology. It also shows how to use information coming “from the symmetric world” to partly understand the shape of trees that appear in the formula for the differential of the minimal model. This idea (to move to the universe of shuffle operads, compute the vector space structure there, and then to use known information on our operad to obtain results about the symmetries of homology) also belongs to the core of the shuffle operad approach.
Plan of the paper

This paper is organised as follows. 
In Section 1, we recall necessary background information on shuffle operads, and provide references for definitions and results that are relevant for the paper but are not discussed in detail. 
In Section 2, we present an “inclusion-exclusion” resolution for an arbitrary shuffle operad with monomial relations, which we then use in the subsequent sections. 
In Section 3, we use algebraic Morse theory to construct representatives for Quillen homology classes under a minor assumption on the combinatorics of defining relations. This section includes brief recollections of algebraic Morse theory, as well as full proofs of existence of Morse matchings; whereas results of that section are important for some of our examples, a reader primarily interested in applications may skip the proofs without any disadvantages for understanding the rest of the paper. 
In Section 4, we use a version of homological perturbation to obtain a resolution for a general shuffle operad with a Gröbner basis. 
In Section 5, we exhibit applications of our results outlined above. Those are a new proof of the PBW criterion, homology estimates for operads coming from commutative algebras, and a computation of Quillen homology for the operads $RB$, $ncRB$, and $BV$. 

Acknowledgments

Work on this paper ended up being spread over several years, and we benefited a lot from discussing preliminary versions of this paper with our colleagues. We thank Frédéric Chapoton, Yves Guiraud, Eric Hoffbeck, Muriel Livernet, Jean–Louis Loday, Philippe Malbos and Dmitri Piontkovski for useful discussions, remarks and references to literature. Special thanks are due to Li Guo for drawing our attention to Rota–Baxter algebras as an example to which our methods could be applied, and especially to Bruno Vallette for many stimulating discussions and in particular for explaining the approach to the operad $BV$ used in his joint work with Gabriel Drummond-Cole. The first author gratefully acknowledges the hospitality of the University Lyon 1 during the final stage of working on this paper. The second author wishes to thank Dublin Institute for Advanced Studies and the University of Luxembourg, where parts of this paper were completed, for invitations to visit and excellent working conditions. 

1 Recollections

All vector spaces and (co)chain complexes throughout this work are defined over an arbitrary field $k$. To handle suspensions, we introduce a formal symbol $s$ of degree 1, and define, for a graded vector space $V$, its suspension $sV$. 
Vladimir Dotsenko and Anton Khoroshkin

as \( k \Sigma \otimes V \). All algebras and operads are assumed non-unital; to adapt our results for unital algebras, one has to restrict the setup to augmented algebras, and consider the abelianisation \( A^\phi = A_+/A_+^2 \), where \( A_+ \) is the augmentation ideal.

The main thing about shuffle operads that is crucial for our constructions is the relevant combinatorics of trees. Hence, in this section we pay most attention in explaining that combinatorics in detail. We also discuss the precise relationship between homological/homotopical results obtained in the symmetric and in the shuffle category, since most of the applications we have in mind concern symmetric operads. For information on operads in general, we refer the reader to the book [40], for information on shuffle operads and Gröbner bases for operads in not necessarily quadratic case — to our paper [13]. Throughout this paper by an operad, unless otherwise specified, we mean a shuffle operad: there is no machinery of Gröbner bases available in the symmetric case, so we have to sacrifice the symmetric groups action. Of course, in some cases we deal with non-symmetric operads, and in that case there is nothing to sacrifice, and in fact the story is somewhat richer since one can include constants (operations of arity 0) in the picture and avail of Gröbner bases at no additional cost. For details on that, see [16]. For the sake of brevity, in this paper we shall discuss shuffle operads in detail, while non-symmetric operads will only appear in some examples.

1.1 Shuffle operads, bar construction and homology

Let us denote by \( \text{Ord} \) the category whose objects are non-empty finite ordered sets (with order-preserving bijections as morphisms). Also, we denote by \( \text{Vect} \) the category of vector spaces (with linear operators as morphisms). It is usually enough to assume vector spaces to be finite-dimensional, though sometimes more generality is needed, and one assumes, for instance, that they are graded with finite-dimensional homogeneous components.

**Definition 1.1.** A (non-symmetric) collection is a contravariant functor from the category \( \text{Ord} \) to the category \( \text{Vect} \). Because of functoriality, a nonsymmetric collection \( \mathcal{P} \) is completely determined by its components \( \mathcal{P}(n) := \mathcal{P}([1, \ldots, n]), \ n \geq 1 \).

For two nonsymmetric collections \( \mathcal{P} \) and \( \mathcal{D} \), the shuffle composition product of \( \mathcal{P} \) and \( \mathcal{D} \) is the non-symmetric collection \( \mathcal{P} \circ_{sh} \mathcal{D} \) defined by the formula

\[
(\mathcal{P} \circ_{sh} \mathcal{D})(l) := \bigoplus_{k \geq 1} \mathcal{P}(k) \otimes \left( \bigoplus_{\phi: [1] \rightarrow [k]} \mathcal{D}(\phi^{-1}(1)) \otimes \cdots \otimes \mathcal{D}(\phi^{-1}(k)) \right),
\]

where the sum is taken over all shuffling surjections \( f \), that is surjections for which \( \min \phi^{-1}(i) < \min \phi^{-1}(j) \) whenever \( i < j \).

**Proposition 1.2** ([13]). The shuffle composition product equips the category of non-symmetric collections with a structure of a monoidal category.
Definition 1.3. A **shuffle operad** is a monoid in the category of non-symmetric collections equipped with the shuffle composition product.

For the monoidal category of shuffle operads, it is possible to define the bar complex of an operad $O$. The bar complex $B^\ast(O)$ is a dg co-operad freely generated by the suspension $sO$; the differential comes from operadic compositions in $O$. Similarly, for a co-operad $Q$, it is possible to define the cobar complex $\Omega^\ast(Q)$, which is a dg operad freely generated by $s^{-1}Q$, with the appropriate differential. The bar-cobar construction $\Omega^\ast(B^\ast(O))$ gives a free resolution of $O$. This can be proved in a rather standard way, similarly to known proofs in the case of operads, properads etc. \[21, 27, 62\]. The general homotopical algebra philosophy mentioned in the introduction is applicable in the case of operads as well; various checks and justifications needed to ensure that are quite standard and similar to the ones available in the literature; we refer the reader to \[6, 22, 30, 47, 55, 59\] where symmetric operads are handled. Thus, the Quillen homology of an operad can be computed as homology of its bar complex (since the abelianisation of the bar-cobar construction is the bar complex), though sometimes this complex is too big to handle, so it is important to seek more economic free resolutions. Our approach allows to build free resolutions for shuffle operads with known Gröbner bases, thus giving an alternative way to compute Quillen homology.

1.2 Symmetric vs shuffle

Let us explain precisely what information on Quillen homology for symmetric operads “survives” in the shuffle world, and what is lost. Of course, the information on the symmetric group actions does get lost. However, we argue that all other relevant structures on the homology do survive. Recall that the forgetful functor $\mathcal{P} \rightarrow \mathcal{P}'$ from the category of symmetric collections to the category of nonsymmetric collections (with the shuffle product) is monoidal \[13, Prop. 3\]. This easily implies the following

**Proposition 1.4.** For a symmetric operad $\mathcal{P}$, we have

$$B^\ast(\mathcal{P}') \cong B^\ast(\mathcal{P}),$$

that is the (symmetric) bar complex of $\mathcal{P}$ is isomorphic, as a shuffle dg co-operad, to the (shuffle) bar complex of $\mathcal{P}'$.

Appropriate homotopy transfer for homotopy co-operads \[17\] (together with the observation that homotopy co-operad maps on the Quillen homology are up to suspension equal to components of the differential in the minimal model) implies

**Corollary 1.5.** For a symmetric operad $\mathcal{P}$, we have

$$H^Q(\mathcal{P}') \cong H^Q(\mathcal{P}),$$
that is the (symmetric) Quillen homology $P$ is isomorphic, as a shuffle homotopy co-operad, to the (shuffle) Quillen homology of $P^f$. Also, if $\mathcal{R}_P$ denotes the minimal model of an operad $\mathcal{O}$ in the appropriate category (symmetric or shuffle), we have

$$(\mathcal{R}_P)^f \cong \mathcal{R}_{P^f}$$

as shuffle dg operads.

In particular, this means that on the minimal model in the shuffle category it is in principle possible to introduce a symmetric groups action compatible with the differential so that it becomes precisely the minimal model in the symmetric category.

1.3 Tree monomials

Let us recall tree combinatorics used to describe monomials in shuffle operads. See [13] for more details.

Basis elements of the free operad are represented by (decorated) trees. A (rooted) tree is a non-empty connected directed graph $T$ of genus 0 for which each vertex has at least one incoming edge and exactly one outgoing edge. Some edges of a tree might be bounded by a vertex at one end only. Such edges are called external. Each tree should have exactly one outgoing external edge, its output. The endpoint of this edge which is a vertex of our tree is called the root of the tree. The endpoints of incoming external edges which are not vertices of our tree are called leaves.

Each tree with $n$ leaves should be (bijectively) labelled by the standard $n$-element set $[n] = \{1, 2, \ldots, n\}$. For each vertex $v$ of a tree, the edges going in and out of $v$ will be referred to as inputs and outputs at $v$. A tree with a single vertex is called a corolla. There is also a tree with a single input and no vertices called the degenerate tree. Trees are originally considered as abstract graphs but to work with them we would need some particular representatives. For a tree with labelled leaves, its canonical planar representative is defined as follows. In general, an embedding of a (rooted) tree in the plane is determined by an ordering of inputs for each vertex. To compare two inputs of a vertex $v$, we find the minimal leaves that one can reach from $v$ via the corresponding input. The input for which the minimal leaf is smaller is considered to be less than the other one.

Let us introduce an explicit realisation of the free operad generated by a collection $\mathcal{M}$. The basis of this operad will be indexed by canonical planar representatives of trees with decorations of all vertices. First of all, the simplest possible tree is the degenerate tree; it corresponds to the unit of our operad. The second simplest type of trees is given by corollas. We shall fix a basis of $\mathcal{M}$ and decorate the vertex of each corolla with a basis element; for a corolla with $n$ inputs, the corresponding element should belong to the basis of $\mathcal{Y}(n)$. The basis for whole free operad consists of all canonical planar representatives of trees built from these corollas (explicitly, one starts with this collection of
corollas, defines compositions of trees in terms of grafting, and then considers all trees obtained from corollas by iterated shuffle compositions). We shall refer to elements of this basis as tree monomials. Vice versa, if we forget the labels of vertices and leaves of a tree monomial $\alpha \in \mathcal{F}_M$, we obtain a planar tree. We shall refer to this planar tree as the underlying tree of $\alpha$.

For example, if $O = \mathcal{F}_M$ is the free operad for which the component $\mathcal{M}(n)$ is only non-zero for $n = 2$, and $\mathcal{M}(2) = k\langle \circ \rangle$, the basis of $\mathcal{F}_M(3)$ is given by the tree monomials

\[
\begin{array}{c}
\circ \\
\circ \\
\circ
\end{array}
\quad,
\begin{array}{c}
\circ \\
\circ \\
\circ
\end{array}
\quad,
\begin{array}{c}
\circ \\
\circ \\
\circ
\end{array}
\quad,
\begin{array}{c}
\circ \\
\circ \\
\circ
\end{array}
\quad.
\]

There are two standard ways to think of elements of an operad defined by generators and relations: using either tree monomials or operations. For example, the above tree monomials correspond to operations

\[
(a_1 \circ a_2) \circ a_3, \quad (a_1 \circ a_3) \circ a_2, \quad \text{and} \quad a_1 \circ (a_2 \circ a_3).
\]

Our approach is somewhere in the middle between the two viewpoints: we strongly encourage the reader to think of tree monomials, but to write down the formulas required for definitions and proofs we prefer the language of operations since it makes things more compact.

Example 1.6. The following is a tree monomial in the free operad $\mathcal{F}_M$ generated by some collection $\mathcal{M}$ with $\circ, \bullet \in \mathcal{M}(2)$:

\[
\begin{array}{c}
1 \\
3 \\
2 \\
7 \\
4 \\
6
\end{array}
\quad.
\]

In the language of operations, it corresponds to the operation

\[
((a_1 \circ a_3) \bullet a_5) \circ ((a_2 \bullet a_7) \circ (a_4 \circ a_6)).
\]

Divisors of a tree monomial $\alpha$ in the free operad correspond to a special kind of subgraphs of its underlying tree. Allowed subgraphs contain, together with each vertex, all its incoming and outgoing edges (but not necessarily other endpoints of these edges). Throughout this paper we consider only this kind of subgraphs, and we refer to them as subtrees hoping that it does not lead to any confusion. Clearly, a subtree $T'$ of every tree $T$ is a tree itself. Let us define the tree monomial $\alpha'$ corresponding to $T'$. To label vertices of $T'$, we recall the labels of its vertices in $\alpha$. We immediately observe that these labels match the restriction labels of a tree monomial should have: each vertex has the same number of inputs as it had in the original tree, so for a vertex with $n$ inputs its label does belong to the basis of $\mathcal{M}(n)$. To label leaves of $T'$, note that each such leaf is either a leaf of $T$, or is an output of some vertex of $T$. This
allows us to assign to each leaf \( l' \) of \( T' \) a leaf \( l \) of \( T \), which we call the smallest descendant of \( l' \): if \( l' \) is a leaf of \( T \), put \( l = l' \), otherwise let \( l \) be the smallest leaf of \( T \) that can be reached through \( l' \). We then number the leaves according to their smallest descendants: the leaf with the smallest possible descendant gets the label 1, the second smallest — the label 2 etc.

**Example 1.7.** Let us consider the following two choices of subtrees of the tree from Example 1.6:

\[
\begin{array}{c}
1 & 3 & 2 & 7 & 4 & 6 \\
\text{and} \\
1 & 5 & 2 & 4
\end{array}
\]

In the first case, the subtree marked by bold lines yields the tree monomial \( \begin{array}{c}
1 & 4 & 2 & 3 \\
\end{array} \), and the ”standardisation” re-labelling, as above, gives the tree monomial \( \begin{array}{c}
2 & 7 & 4 & 6 \\
\end{array} \). In the second case, the subtree marked by bold lines yields the tree monomial \( \begin{array}{c}
1 & 4 & 2 & 3 \\
\end{array} \), and the ”standardisation” re-labelling gives the same tree monomial \( \begin{array}{c}
2 & 7 & 4 & 6 \\
\end{array} \). In the language of operations,

\[
(a_1 \bullet a_3) \circ (a_2 \circ a_4) \approx (a_2 \bullet a_7) \circ (a_4 \circ a_6) \approx (a_1 \bullet a_4) \circ (a_2 \circ a_3).
\]

Thus, \( \begin{array}{c}
1 & 4 & 2 & 3 \\
\end{array} \) occurs as a divisor of the original tree monomial at two different places.

For two tree monomials \( \alpha, \beta \) in the free operad \( \mathcal{T}_{\mathcal{M}} \), we say that \( \alpha \) is divisible by \( \beta \), if there exists a subtree of the underlying tree of \( \alpha \) for which the corresponding tree monomial \( \alpha' \) is equal to \( \beta \).

There exist several ways to introduce a total ordering of tree monomials in such a way that the operadic compositions are compatible with that total ordering. A Gröbner basis of an ideal \( I \) of the free operad is a system \( S \) of generators of \( I \) for which the leading monomial of every element of the ideal is divisible by one of the leading terms of elements of \( S \). Such a system of generators allows to perform “long division” modulo \( I \), computing for every element
its canonical representative. There exists an algorithmic way to compute a Grobner basis starting from any given system of generators (“Buchberger’s algorithm for shuffle operads”). For our purposes, it is important to note that if the tree monomials of our operad have additional internal grading, and the relations are homogeneous with respect to that grading, then the corresponding reduced Grobner basis is also homogeneous, as well as all our homological constructions.

1.4 Operads in the Differential Graded Setting

The above description of the free shuffle operad works almost literally when we work with operads whose components are chain complexes (as opposed to vector spaces), and the symmetric monoidal structure on the corresponding category involves signs. The only difference is that every tree monomial should carry an ordering of its internal vertices, so that two different orderings contribute appropriate signs. In this section, we give an example of a shuffle dg operad that should help a reader to understand the graded case better; this operad was introduced and explored in [48].

Definition 1.8. The odd \((2k+1)\)-associative operad is a non-symmetric operad with one generator \(μ\) of arity \(2k+1\) and odd homological degree, and relations

\[ μ \circ p \, μ = μ \circ_{2k+1} μ \quad \text{for all } p \leq 2k. \]

Let us show that the Buchberger algorithm for operads from [13] discovers a cubic relation in the Grobner basis for this operad, thus showing that this operad fails to be PBW in the sense of Hoffbeck [31] (for this particular ordering). We use the path-lexicographic ordering of monomials.

From the common multiple \((μ \circ_1 μ) \circ_1 μ\) of the leading term \(μ \circ_1 μ\) with itself, we compute the S-polynomial

\[ (μ \circ_{2k+1} μ) \circ_1 μ - μ \circ_1 (μ \circ_{2k+1} μ). \]

We can perform the following chain of reductions (with leading monomials underlined):

\[
\begin{align*}
& (μ \circ_{2k+1} μ) \circ_1 μ - μ \circ_1 (μ \circ_{2k+1} μ) = (μ \circ_{2k+1} μ) \circ_1 μ - (μ \circ_1 μ) \circ_{2k+1} μ \mapsto
& \mapsto (μ \circ_{2k+1} μ) \circ_1 μ - (μ \circ_{2k+1} μ) \circ_{2k+1} μ = -(μ \circ_1 μ) \circ_{4k+1} μ - (μ \circ_{2k+1} μ) \circ_{2k+1} μ \mapsto
& \mapsto -(μ \circ_{2k+1} μ) \circ_{4k+1} μ - (μ \circ_{2k+1} μ) \circ_{2k+1} μ = -(μ \circ_{2k+1} μ) \circ_{4k+1} μ - μ \circ_{2k+1} (μ \circ_1 μ) \mapsto
& \mapsto -(μ \circ_{2k+1} μ) \circ_{4k+1} μ - μ \circ_{2k+1} (μ \circ_{2k+1} μ) = -2(μ \circ_{2k+1} μ) \circ_{4k+1} μ.
\end{align*}
\]

Note that we used the formula \((μ \circ_{2k+1} μ) \circ_1 μ = -(μ \circ_1 μ) \circ_{4k+1} μ\) which reflects the fact that the operation \(μ\) is of odd homological degree.

The monomial \((μ \circ_{2k+1} μ) \circ_{4k+1} μ\) cannot be reduced further, and we recover the relation \((μ \circ_{2k+1} μ) \circ_{4k+1} μ = 0\) discovered in [48]. Furthermore, we arrive at the following proposition (note the similarity with the computation of the Grobner basis for the operad AntiCom in [13]).
Proposition 1.9. Elements \(\mu \circ_p \mu - \mu \circ_{2k+1} \mu\) with \(1 \leq p \leq 2k\) and \((\mu \circ_{2k+1} \mu) \circ_{4k+1} \mu\) form a Gröbner basis for the operad of odd \((2k+1)\)-associative algebras.

2 Resolution for monomial relations

Assume that the operad \(O = \mathcal{F}_\mathcal{M}/(\mathcal{G})\) is generated by a collection of finite sets \(\mathcal{M} = \{\mathcal{M}(n)\}\), and that \(\mathcal{G}\) consists of tree monomials, so we are dealing with an operad that only has monomial relations. We shall explain how to construct a free resolution of \(O\).

Our first step is to construct a free shuffle dg operad \(A\) which does not take the account relations of \(O\); it is a somewhat universal object for operads generated by \(\mathcal{M}\), various suboperads of \(A\) will be used as resolutions for various choices of \(\mathcal{G}\).

2.1 The inclusion–exclusion operad

Let \(T\) be a tree monomial, and let the symbols \(S_1, \ldots, S_q\) be in one-to-one correspondence with all the divisors of \(T\). We denote by \(\mathcal{A}(T)\) the vector space \(\mathbb{k}T \otimes \Lambda(S_1, \ldots, S_q)\). We shall say that underlying tree monomial for elements of this vector space is \(T\). The degree \(-1\) derivations \(\partial_i\) on the exterior algebra defined by the rule \(\partial_i(S_j) = \delta_{ij}\) anticommute, and the differential \(d = \sum_{i=1}^q \partial_i\) makes \(\mathcal{A}(T)\) into a chain complex isomorphic to the augmented chain complex of a \((q-1)\)-dimensional simplex \(\Delta^{q-1}\).

By definition, the chain complex \(\mathcal{A}(n)\) is the direct sum of complexes \(\mathcal{A}(T)\) over all tree monomials \(T\) with \(n\) leaves. There is a natural operad structure on the collection \(\mathcal{A} = \{\mathcal{A}(n)\}\); the operadic composition composes the trees, and computes the wedge product of symbols labelling their divisors. Overall, we defined a shuffle dg operad, which we shall call the inclusion–exclusion operad.

Let us emphasize that the symbols \(S_i\) correspond to divisors, i.e. mark occurrences of tree monomials in \(T\) rather than monomials themselves, so in particular the Koszul sign rule does not imply that a composition of an element of our operad with itself is equal to zero. Basically, when computing products, the \(S\)-symbols “remember” which divisors of factors they come from. Graphically, it is convenient to think of basis elements of our chain complex as tree monomials with some of the occurrences of relations additionally marked, as in Example 1.7.

The following example should make our construction more clear.

Example 2.1. Assume that the operad \(O\) has two binary generators \(\circ\) and \(\cdot\).
Then the corresponding operad \( \mathcal{A} \) contains, among others, two elements
\[
\begin{array}{c}
\begin{array}{ccccc}
2 & 5 & 3 & 4 & \\
1 & & & & \\
\end{array}
\end{array}
\quad\text{and}\quad
\begin{array}{c}
\begin{array}{ccccc}
1 & 3 & 2 & \\
& & & & \\
\end{array}
\end{array}
\]

An appropriate shuffle composition of these two produces the element
\[
\begin{array}{c}
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 6 & \\
& 5 & & & & \\
\end{array}
\end{array}
\]

where two different divisors are marked. Incidentally, the underlying tree monomial for each of them is
\[
\begin{array}{c}
\begin{array}{ccccc}
1 & 3 & 2 & \\
& & & & \\
\end{array}
\end{array}
\].

Let us call the first tree \( T \), the second tree \( T' \), the first divisor \( S \), and the second divisor \( S' \). On the level of formulas, we have
\[
(T \otimes S) \circ \phi (T' \otimes S') = (T \circ \phi T') \otimes S_1 \wedge S_2,
\]
where \( \phi: [1,2,3,4,5,6,7] \to [1,2,3,4,5] \) is the shuffle surjection with \( \phi(1) = \phi(3) = \phi(5) = 1, \phi(2) = 2, \phi(4) = 3, \phi(6) = 4, \phi(7) = 5 \), and \( S_1 \) and \( S_2 \) indicate the two different divisors of \( T \circ \phi T' \) equal to
\[
\begin{array}{c}
\begin{array}{ccccc}
1 & 3 & 2 & \\
& & & & \\
\end{array}
\end{array}
\].

Note that even though the underlying tree monomials of the two divisors coincide, the wedge product is not equal to zero, since the letters \( S \) correspond to distinct divisors, that is occurrences of tree monomials, not tree monomials themselves.

In fact, the underlying operad of the dg operad \( \mathcal{A} \) is free. Indeed, let us call a \( \text{“monomial”} \) \( T \otimes S_{i_1} \wedge \cdots \wedge S_{i_q}, q \geq 0, \) indecomposable, if it is not a composition in the operad \( \mathcal{A}/q \) of two monomials of the same type. (This means that each edge between the two internal vertices of \( T \) is an edge between two internal vertices of at least one of the divisors \( S_{i_1}, \ldots, S_{i_q} \); note that some of the internal vertices of \( T \) are leaves of its divisors, and hence are not considered internal vertices of the respective divisors.) It is easy to see that \( \mathcal{A} \) is freely generated by indecomposable elements; those are elements \( m \otimes 1 \) with \( m \in \mathcal{M} \) being a generator of \( \mathcal{O} \), and indecomposable monomials \( T \otimes S_{i_1} \wedge \cdots \wedge S_{i_q}, q \geq 1 \).

For each monomial \( T \otimes S_{i_1} \wedge \cdots \wedge S_{i_q} \) different from the generators described above, and each internal edge of \( T \) that is not an internal edge of either of \( S_{i_1}, \ldots, S_{i_q} \), the endpoint of that edge which is further from the root of \( T \) is a "grafting point": the subtree growing from this vertex (and its divisors among \( S_{i_1} \)) factors out in our operad. This factorisation procedure gives a unique way to factorise elements as compositions of generators.
So far we did not use the relations of our operad. Let us incorporate relations in the picture.

### 2.2 Suboperads of the inclusion–exclusion operad

Let $\mathcal{G}$ be the set of relations of our operad $\mathcal{O}$. The dg operad $(\mathcal{A}_G, d)$ is defined similarly to $\mathcal{A}$, but with the additional restriction that every symbol $S_k$ corresponds to a divisor of $T$ for which the underlying tree monomial is a relation. The differential $d$ is the restriction of the differential defined above. Informally, an element of the operad $\mathcal{A}_G$ is a tree with some distinguished divisors that are relations from the given set.

**Theorem 2.2.** The dg operad $(\mathcal{A}_G, d)$ is a free resolution (as a shuffle operad) of the corresponding operad with monomial relations $\mathcal{O} = \mathcal{F}_G/(\mathcal{G})$.

**Proof.** Similarly to the case of the operad $\mathcal{A}$, the operad $\mathcal{A}_G$ is freely generated by its elements $m \otimes 1$ with $m \in \mathcal{M}$ and all indecomposable monomials $T \otimes S_1 \wedge \cdots \wedge S_q, q \geq 1$, where each of the divisors $S_i$ is a relation of $\mathcal{O}$.

Let us prove that $\mathcal{A}_G$ provides a resolution for $\mathcal{O}$. Since the differential $d$ only omits wedge factors but does not change the tree monomial, the chain complex $\mathcal{A}_G$ is isomorphic to the direct sum of chain complexes $\mathcal{A}_T$ spanned by the elements for which the first tensor factor is the given tree monomial $T$. If $T$ is not divisible by any relation, the complex $\mathcal{A}_T$ is concentrated in degree 0 and is spanned by $T \otimes 1$. Thus, to prove the theorem, we should show that $\mathcal{A}_T$ is acyclic whenever $T \otimes 1$ is divisible by some relation $g_i$.

Assume that there are exactly $k$ divisors of $T$ which are relations of $\mathcal{O}$. We immediately see that the complex $\mathcal{A}_T$ is isomorphic to the chain complex of a simplex $\Delta_{k-1}$ which is acyclic whenever $k > 0$. □

**Remark 2.3.** Using the machinery of twisting cochains [7], one can obtain from the free dg operad resolution a resolution of the trivial module by free right modules whose spaces of generators of various homological degrees are the same as the spaces of generators of the original operad resolution. More precisely, the differential of every generator in our free operad resolution is a sum of compositions of generators; this provides the space of generators with a structure of a homotopy co-operad, and the twisting cochain method applies. See [39, 52] for details in the (simpler) case of associative algebras, and [17] for details in the operad case.

### 3 Homology classes for monomial operads

In general, the fact that “trees grow in several different directions”, means that it is more difficult to describe representatives of homology classes combinatorially in the same way as it can be done for the case of associative algebras [1]. However, in some cases it is possible to come up with a reasonable description.
In this section, we shall describe homology classes under a minor restriction on the combinatorics of defining relations.

3.1 Algebraic Morse theory: recollections

To obtain our description, we use the algebraic Morse theory developed independently in [32, 37, 58]. We refer the reader to those references for details; for our purposes, the algebraic Morse theory is a way to describe a smaller subcomplex of a chain complex having the same homology. It is done as follows. Suppose that our chain complex \((C_\bullet, d)\) has a basis \(X = \bigsqcup_{i \geq 1} X_i\), where \(X_i\) is the basis of the space \(C_i\) of our complex; this basis should be finite or satisfy some local finiteness condition (i.e., internal grading). We consider a directed graph \(\Gamma\) whose vertex set \(V\) coincides with \(X\), and the edge set \(E\) reflects the combinatorics of the differential: there is an edge from \(x \in X_i\) to \(y \in X_{i-1}\) if \(y\) appears in \(d(x)\) with a non-zero coefficient. A set of edges \(M \subset E\) is called a Morse matching if two conditions are satisfied:

1. every vertex of \(\Gamma\) belongs to at most one edge from \(M\);
2. the graph \(\Gamma'\) on the vertex set \(V\) whose edge set is the union of \(E \setminus M\) with the set of all edges of \(M\) reversed has no directed cycles.

The vertices that do not belong to any edge of \(M\) are called critical. The subset of critical vertices of \(X_i\) is denoted by \(X_i^M\). The key result of algebraic discrete Morse theory states that there is a way to define a new differential \(d^M\) on the linear span \(C^M_\bullet\) of critical vertices so that the chain complex \((C^M_\bullet, d^M)\) is quasi-isomorphic to \((C_\bullet, d)\). The only property of this differential that we shall really need is that its “structure constants”, that is the coefficients \([v : v']\) in the formula

\[
d^M(v) = \sum_{v'} [v : v']v'
\]

are defined as sums over paths from \(v\) to \(v'\) in the graph \(\Gamma'\).

3.2 Homology classes via algebraic Morse theory

Let \(O = \mathcal{F}_M / (\mathcal{G})\) be an operad generated by the collection \(M\) with monomial relations. Quillen homology \(H^Q(\mathcal{G})\) is isomorphic to the homology of the differential induced on the space of indecomposable elements \((\mathcal{A}_d)^{\mathit{adm}}\) of the operad \(\mathcal{A}_d\). That space of generators, as a chain complex, can be decomposed into a direct sum of chain complexes \((\mathcal{A}_d)^{\mathit{adm}}\) spanned by the elements for which the underlying tree monomial is the given tree monomial \(T\), and therefore the Quillen homology acquires a direct sum decomposition

\[
H^Q(\mathcal{G}) = \bigoplus_{T \text{ a tree monomial}} H^Q_\mathcal{T}(\mathcal{G}).
\]
We shall define Morse matchings of chain complexes \((A_G^ab)_{T^b}\) under some technical conditions which we believe to be not very restrictive; at least in all naturally arising examples that we discuss throughout the paper these conditions are fulfilled.

**Definition 3.1.** For a tree monomial \(T\), we call a numbering of the set of all divisors of \(T\) that are relations of \(O\) an Anick numbering if whenever \(i < j < k\) and \(S_i \cap S_j \neq \emptyset\) we have \(S_i \cap S_k \subset S_j \cap S_k\). (Here and below by intersection we mean the most naive combinatorial intersection of divisors inside \(T\).)

Let us give two examples of Anick numberings. The first one, which we shall use as a toy model in this section, explains the term we chose: we shall see that in the case of associative algebras it corresponds to the numbering of divisors used by Anick [1].

**Example 3.2.** Suppose that our operad \(O\) with monomial relations is generated by unary operations. In this case, it is nothing but an associative algebra with monomial relations. If we number subwords of the given word which are relations according to the position of the first letter, the corresponding ordering is manifestly an Anick ordering.

The second example is new, and has a genuine operadic meaning to it; we shall discuss an application of this result in Section 5.2.

We assume that our operad is generated by elements of arity 2. Let us use the usual terms “left combs” and “right combs” for tree monomials corresponding to the operations of the form

\[
a_1(a_2(\ldots (a_k(1, i_2), i_3), \ldots, i_k), i_{k+1}) \quad \text{and} \quad a_1(1, a_2(2, \ldots a_k(1, k, k + 1) \ldots))
\]

respectively: left combs are obtained from the generators by iterated compositions in the first slot, and right combs are obtained from the generators by iterated compositions in the last slot.

**Proposition 3.3.** Suppose that \(O = \mathcal{F}_M/\langle \mathcal{G} \rangle\) is an operad with binary generators and monomial relations, \(\mathcal{G} = \mathcal{G}_l \cup \mathcal{G}_r\), where \(\mathcal{G}_l\) consists of left combs and \(\mathcal{G}_r\) consists of right combs, and at least one of the sets \(\mathcal{G}_l, \mathcal{G}_r\) is contained in \(\mathcal{F}_M(3)\). Then for each tree monomial \(T\), the set of all divisors of \(T\) that are relations admits an Anick numbering.

**Proof.** Without loss of generality, \(\mathcal{G}_l \subset \mathcal{F}_M(3)\). Let us define a partial ordering of the set of all divisors of \(T\) that are relations of \(O\) as follows: \(S < S'\) if the root of \(S'\) is on the path from the root of \(T\) to the root of \(S\), or if \(S = S'\) share the same root, \(S\) is a left comb and \(S'\) is a right comb. Let us prove that if we extend this partial ordering to a total ordering in any way, and consider the numbering of the divisors according to that total ordering in the increasing order, the numbering thus obtained is an Anick numbering. Indeed, suppose that \(i < j < k\) and \(S_i \cap S_j \neq \emptyset\). There are three different situations when this can happen:
- $S_i$ and $S_j$ are left combs, and the root of $S_j$ is on the path from the root of $T$ to the root of $S_i$.

- $S_i$ and $S_j$ are right combs, and the root of $S_j$ is on the path from the root of $T$ to the root of $S_i$.

- $S_i$ is a left comb, $S_j$ is a right comb, and they share a vertex which is a root vertex of $S_i$.

The only situation when the Anick numbering condition $S_i \cap S_k \subset S_j \cap S_k$ may fail is when $S_i \cap S_k \neq \emptyset$.

In the first case, if $S_k$ is a left comb as well, the condition is manifestly fulfilled. Otherwise, if $S_k$ is a right comb, its intersection with $S_i$ consists of exactly one vertex. If that vertex is the root of $S_j$, then it is contained in $S_j$, and the condition is fulfilled. If it is not the root of $S_j$, it has to be the root of $S_k$, so the root of $S_j$ is manifestly on the path from the root of $T$ to the root of $S_k$, so $S_k \prec S_j$, a contradiction.

In the second case, if $S_k$ is a right comb as well, the condition is manifestly fulfilled. Otherwise, if $S_k$ is a left comb, its intersection with $S_i$ consists of exactly one vertex. If that vertex is the root of $S_j$, then it is contained in $S_j$, and the condition is fulfilled. If it is not the root of $S_j$, it has to be the root of $S_k$, so $S_k \prec S_j$, a contradiction.

In the third case, if $S_k$ is a right comb, then since $S_i \cap S_k \neq \emptyset$, we instantly conclude that $S_k \prec S_j$, a contradiction. If $S_k$ is a left comb, then its intersection with $S_i$ may only consist of one vertex, which is precisely the root vertex of $S_j$, that is the intersection of $S_i$ and $S_j$. $\square$

Throughout this section, we always assume that we are dealing with monomial relations for which for each $T$ the set of its divisors that are relations admits an Anick numbering. Under this assumption, we shall prove the following result.

**Theorem 3.4.** A basis of $H^Q_T(O)$ is in one-to-one correspondence with basis elements $v \in (\mathcal{A}_Q)^{ab}_T$ for which the following two properties hold:

1. For each $S_j$ present in $v$, we have $\partial_j(v) = 0$ in $(\mathcal{A}_Q)^{ab}_T$; in other words, after removing $S_j$, $v$ becomes decomposable,

2. For each $S_j$ not present in $v$, there exists $i < j$ for which $\partial_i(v \wedge S_j) \neq 0$ in $(\mathcal{A}_Q)^{ab}_T$; in other words, after marking $S_j$ in $v$ it is possible to remove the mark from the divisor $S_i$ for some $i < j$ so that the result is indecomposable.

**Proof.** Let us denote by $X$ the natural basis of the chain complex $(\mathcal{A}_Q)^{ab}_T$. It gives rise to a graph $\Gamma$ reflecting the combinatorics of the differential. We shall now describe inductively a matching of the vertices of $\Gamma$, and demonstrate that under our assumptions it is a Morse matching. Let us put $M_1$ to be the set of
edges \( v \rightarrow w \), where \( v, w \in X \), \( w = \pm \partial_j(v) \). For \( k > 1 \) we denote by \( X^{(k-1)} \) the set of critical vertices with respect to the matching \( M_1 \cup \ldots \cup M_{k-1} \), and let

\[
M_k = \{ v \rightarrow w : v, w \in X^{(k-1)}, w = \pm \partial_k(v) \}.
\]

The following proposition is a “bounded version” of Theorem 3.4.

**Proposition 3.5.** The set \( X^{(k)} \) consists of the basis elements \( v \in (\mathcal{A}_q)^{ab} \) for which the following two properties hold:

1. \( (I_k) \) for each \( j \leq k \) such that \( S_j \) is present in \( v \), the monomial \( \partial_j(v) \) is decomposable,
2. \( (II_k) \) for each \( j \leq k \) such that \( S_j \) is not present in \( v \), there exists \( i < j \) for which \( \partial_i(v \wedge S_j) \neq 0 \) in \( (\mathcal{A}_q)^{ab} \).

**Proof.** We prove this statement by induction on \( k \). For \( k = 1 \), it is obvious. Let us explain the step of induction. Let \( v \in X^{(k)} \). By induction, Conditions \( (I_{k-1}) \) and \( (II_{k-1}) \) hold for \( v \).

Let us examine Condition \( (I_k) \) for a basis element \( v \), and for \( j = k \). Assume that it does not hold, so that \( \partial_k(v) \neq 0 \) in \( (\mathcal{A}_q)^{ab} \). We shall now prove that in this case the elements \( v \) and \( \pm \partial_k v \) will have been matched when forming the matching \( M_k \). Basically, it follows from

**Lemma 3.6.** Suppose that \( v \in X^{(k-1)} \), and that \( \partial_k(v) \) is indecomposable. Then \( \pm \partial_k v \) belongs to the set of critical vertices \( X^{(k-1)} \) as well.

**Proof.** First, Condition \( (I_{k-1}) \) for \( v \) implies the same condition for \( \partial_k(v) \), since removing a factor \( S_k \) does not ruin decomposability. Let us prove that Condition \( (II_{k-1}) \) holds for \( \partial_k(v) \). We should check that for each \( j < k \) such that \( S_j \) is not present in \( \partial_k(v) \), there exists \( i < j \) for which \( \partial_i(\partial_k(v) \wedge S_j) \) is indecomposable. Since Condition \( (II_{k-1}) \) holds for \( v \), for each \( j < k \) not present in \( v \) we can find \( i < j \) such that \( \partial_i(v \wedge S_j) \) is indecomposable. By \( (I_{k-1}) \), \( \partial_i(v) \) is decomposable. Therefore, \( S_i \cap S_j \neq \emptyset \), and since we work with an Anick numbering we have \( S_i \cap S_k \subset S_j \cap S_k \). This means that \( \partial_i(\partial_k(v) \wedge S_j) \) is indecomposable (the only reason for \( \partial_i(v \wedge S_j) \) to become decomposable after removing \( S_k \) would be that something covered by both \( S_i \) and \( S_k \) wasn’t covered anymore, but the intersection of these divisors is covered by \( S_j \)).

Let us examine Condition \( (II_k) \) for a basis element \( v \), and for \( j = k \). Assume that it does not hold, so that \( S_k \) does not occur in \( v \), and for all \( i < k \) such that \( S_i \) present in \( v \wedge S_k \) the element \( \partial_i(v \wedge S_i) \) is decomposable. We shall now show that \( \pm v \wedge S_k \) and \( v \) will have been matched when forming the matching \( M_k \). Indeed, by our assumption Condition \( (I_{k-1}) \) holds for the monomial \( \pm v \wedge S_k \). Condition \( (II_{k-1}) \) holds for this monomial trivially, since it holds for \( v \), and indecomposability is preserved by the operators of wedge multiplication by \( S_p \). Thus, the monomial \( \pm v \wedge S_k \) belongs to \( X^{(k-1)} \) by induction, and we found an edge of the matching \( M_k \).
Vice versa, let us assume that Conditions (I\(_k\)) and (II\(_k\)) hold. Then Conditions (I\(_{k-1}\)) and (II\(_{k-1}\)) also hold, and so \(v \in X^{(k-1)}\) by induction. If \(v \notin X^{(0)}\), \(v\) is used in one of the edges of the matching \(M_k\). Condition (I\(_k\)) guarantees that no edge \(v \rightarrow w\) can appear on that step, so the only option is an edge \(\pm v \wedge S_k \rightarrow v\).

But by condition (II\(_k\)), there exists \(i < k\) for which \(\partial_i(v \wedge S_k)\) is indecomposable, which shows that \(v \wedge S_k\) would have been used at an earlier stage.

\[\square\]

**Proposition 3.7.** The matching \(M = \bigcup M_i\) is a Morse matching.

**Proof.** The only condition we need to check is acyclicity, since every vertex is involved in at most one edge by the construction. Suppose that there is a directed cycle in the graph \(\Gamma'\). Because it is a cycle, it has the same number of “increasing edges”, that is reversed \(M\)-edges, and decreasing edges, that is edges from \(E \setminus M\). For the rest of the proof, we choose an edge \(e = (v \rightarrow \pm v \wedge S_k)\) of our cycle with the largest possible \(k\).

Suppose that the edge following \(e\) is a decreasing one, that is we have a fragment \(v \rightarrow \pm v \wedge S_k \rightarrow \pm \partial(v \wedge S_k)\) in our cycle. Clearly, \(l \neq k\), since otherwise we would have the same edge belonging both to \(M\) and the reversion of \(M\), a contradiction. Also, it cannot be \(l > k\), since otherwise we would have found an edge \(u \rightarrow \pm u \wedge S_l\), which would contradict the definition of \(k\). Therefore, \(l < k\). But in this case, applying Condition (I\(_{k-1}\)) with \(j = l\) to the monomial \(v \wedge S_k\) we obtain a decomposable element, which is a contradiction.

Now suppose that the edge following \(e\) is an increasing one, that is we have a fragment \(v \rightarrow \pm v \wedge S_k \rightarrow \pm v \wedge S_k \wedge S_l\) in our cycle. Clearly, \(l \neq k\), since otherwise we have \(v \wedge S_k \wedge S_l = 0\). Then, according to the definition of \(k\), we have \(l < k\). Let us look at the element \(v \wedge S_k\). Since \(S_l\) is not present in it, we apply Condition (II\(_{k-1}\)) with \(j = l\) to this element, concluding that for some \(r < l\) the monomial \(\partial_r(v \wedge S_k \wedge S_l)\) is indecomposable. If we choose the smallest \(r\) for which \(\partial_r(v \wedge S_k \wedge S_l)\) is indecomposable, we observe that the monomials \(v \wedge S_k \wedge S_l\) and \(\pm \partial_r(v \wedge S_k \wedge S_l)\) were matched on the step \(r\). This contradicts the fact that \(v \wedge S_k \wedge S_l\) is matched with \(\pm S_k\).

\[\square\]

To complete the proof of Theorem 3.4, it is enough to show that the Morse differential \(d^M\) on the critical vertices is identically zero. This would mean that the critical vertices are precisely the homology classes of the chain complex \((\mathcal{A}_G)^{[0]}\).

The former statement can be proved as follows. Every path between two vertices in the graph \(\Gamma'\) starts either with an edge from \(E \setminus M\) or with a reversed edge from \(M\). No edge from \(E\), in particular an edge from \(E \setminus M\) can start with a critical vertex \(v\), since Conditions (I\(_k\)) altogether mean that for every \(k\) the monomial \(\partial_k v\) is decomposable, and consequently \(d(v) = 0\) in \((\mathcal{A}_G)^{[0]}\). No reversed edge from \(M\) can contain a critical vertex either, for tautological reasons.

\[\square\]

We proceed with our examples of Anick numberings. In the case discussed in Example 3.2, we shall, as we already mentioned, obtain Anick chains for...
monomial algebras [1, 61]. Let us recall their definition. Every chain is a monomial of the free algebra $k(x_1, \ldots, x_n)$. For $q \geq 0$, $q$-chains and their tails are defined inductively as follows:

- each generator $x_i$ is a 0-chain; it coincides with its tail;
- each $q$-chain is a monomial $m$ equal to a product $nst$ where $t$ is the tail of $m$, and $ns$ is a $(q-1)$-chain whose tail is $s$;
- in the above decomposition, the product $st$ has exactly one divisor which is a relation of $R$; this divisor is a right divisor of $st$.

In other words, a $q$-chain is a monomial formed by linking one after another $q$ relations so that only neighbouring relations are linked, the first $(q-1)$ of them form a $(q-1)$-chain, and no proper left divisor is a $q$-chain. In our notation above, such a monomial $m$ corresponds to the generator $m \otimes S_1 \wedge \cdots \wedge S_q$ where $S_1, \ldots, S_q$ are the relations we linked.

**Proposition 3.8.** For the Anick numbering of divisors from Example 3.2, the representatives for homology classes suggested by Theorem 3.4 are precisely Anick chains.

**Proof.** Indeed, condition (I) means that only neighbours are linked, and condition (II) means that no proper beginning of a $q$-chain forms a $q$-chain. □

**Remark 3.9.** If we consider the numbering of subwords according to the position of their last letters, we obtain another Anick numbering. The fact that both of the numberings are Anick numberings can be used to obtain a conceptual proof of a result of Bardzell [2, 3] who observed that “Anick left chains” and “Anick right chains” have the same set of underlying monomials, and used it to obtain a resolution of $A$ as an $A - A$-bimodule for an algebra $A$ with monomial relations.

In the setup of our second example, we shall in fact obtain a combinatorial picture modelled on Anick chains as well. Recall that we are dealing with an operad $\mathcal{O} = \mathcal{F}_\mathcal{A}/(\mathcal{G})$ is an operad with binary generators and monomial relations all of which are left and right combs, and assume that we fix an Anick numbering of the kind described in Proposition 3.3.

**Definition 3.10.** To a tree monomial $T$ made up of generators of $\mathcal{O}$, we associate a set of maximal combs. A maximal left comb of $T$ is a sequence of internal vertices $a_1, \ldots, a_q$ of $T$ for which the left child of $a_q$ is a leaf, the left child of $a_l$ is $a_{l+1}$ for $1 \leq l \leq q - 1$, and the parent of $a_l$ (if any) has $a_l$ as its right vertex. Maximal right combs are defined similarly.

Clearly, for every indecomposable monomial of the operad $\mathcal{A}_\mathcal{G}$ each maximal left comb of the underlying tree monomial must be covered by left combs from $\mathcal{G}$, and each maximal right comb of the underlying tree monomial must be covered by right combs from $\mathcal{G}$.
**Definition 3.11.** A monomial in $\mathcal{A}_G$ is said to be an *Anick chain* for $\mathcal{O}$ if for each of its maximal combs its covering by combs from $\mathcal{G}$ obeys the pattern governing Anick chains for associative algebras.

The definitions are given in such a way that following result is proved completely analogously to Proposition 3.8.

**Proposition 3.12.** Suppose that $\mathcal{O} = \mathcal{F}_M / (\mathcal{G})$ is an operad with binary generators and monomial relations, $\mathcal{G} = \mathcal{G}_l \cup \mathcal{G}_r$ where $\mathcal{G}_l$ consists of left combs and $\mathcal{G}_r$ consists right combs, and at least one of the sets $\mathcal{G}_l$, $\mathcal{G}_r$ is contained in $\mathcal{F}_M(3)$. The representatives for homology classes of $\mathcal{O}$ suggested by Theorem 3.4 are precisely Anick chains.

### 4 Resolution for general relations

Let $\widetilde{\mathcal{O}} = \mathcal{F}_\mathcal{M} / (\mathcal{G})$ be an operad, and let $\mathcal{O} = \mathcal{F}_\mathcal{M} / (\mathcal{G})$ be its monomial replacement, that is, $\mathcal{F}$ is a Gröbner basis of relations, and $\mathcal{G}$ consists of all leading monomials of $\mathcal{F}$. In Section 2, we defined a free resolution $(\mathcal{A}_G, d)$ for $\mathcal{O}$, so that $H(A_G, d) \cong \mathcal{O}$.

Let $\phi$ be the canonical homomorphism from $\mathcal{A}_G$ to its homology $\mathcal{O}$ (it kills all generators of positive homological degree, and on elements of homological degree 0 is the canonical projection from $\mathcal{F}_M$ to its quotient). Tree monomials that are not divisible by any of the monomial relations $\mathcal{G}$ form a basis of $\mathcal{O}$, and we define a map $\pi$ as the composition of $\phi$ with the corresponding section; it sends elements of homological degree zero to their residues modulo $\mathcal{G}$, represented as linear combinations of tree monomials not divisible by $\mathcal{G}$ in our resolution. Since $(\mathcal{A}_G, d)$ is a resolution of $\mathcal{O}$, there exists a contracting homotopy $h$ for this resolution, so that $(dh)|_{\ker d} = \text{id} - \pi$ (in fact, below we shall specify a particular choice for such homotopy). Our goal is to “deform” this statement in the following sense. Let $\widetilde{\phi}$ be the homomorphism from $\mathcal{A}_G$ to $\widetilde{\mathcal{O}}$ that kills all generators of positive homological degree, and on elements of homological degree 0 is the canonical projection from $\mathcal{F}_\mathcal{M}$ to its quotient $\widetilde{\mathcal{O}} = \mathcal{F}_\mathcal{M} / (\mathcal{G})$. By general results on Gröbner bases, tree monomials that are not divisible by any of the leading terms $\mathcal{G}$ of relations $\mathcal{F}$ form a basis of $\mathcal{O}$ (each element $f$ of the free operad $\mathcal{F}_\mathcal{M}$ is represented as its residue $\overline{f}$ modulo the Gröbner basis $\overline{\mathcal{G}}$), and we define a map $\overline{\pi}$ as the composition of $\overline{\phi}$ with the corresponding section; it sends elements of homological degree zero to their residues modulo $\mathcal{G}$, represented as linear combinations of tree monomials not divisible by $\mathcal{G}$ in our resolution.

We shall prove the following result, which is essentially nothing but homological perturbation in the same way as it is used in the case of free resolutions of trivial modules over augmented associative algebras in [1, 35, 38].
**Theorem 4.1.** There exists a “deformed” differential $D$ on $\mathcal{A}_g$ and a homotopy 

$$H : \ker D \to \mathcal{A}_g$$

such that 

$$H(\mathcal{A}_g, D) \simeq \overline{\partial} \quad \text{and} \quad (DH)_{\ker D} = \text{id} - \overline{\eta}.$$ 

**Proof.** We shall construct $D$ and $H$ simultaneously by induction. Let us introduce a partial ordering of basis elements in $\mathcal{A}_g$ which just compares the underlying tree monomials. This partial ordering suggests the following definition: for an element $u \in \mathcal{A}_g$, its leading term $\hat{u}$ is the part of the expansion of $u$ as a combination of basis elements where we keep only basis elements $T \otimes S_1 \wedge \cdots \wedge S_q$ with maximal possible $T$.

If $L$ is a homogeneous linear operator on $\mathcal{A}_g$ of some fixed (homological) degree of homogeneity (like $D, H, d, h$), we denote by $L_k$ the operator $L$ acting on elements of homological degree $k$. We shall define the operators $D$ and $H$ by induction: we define the pair $(D_{k+1}, H_k)$ assuming that all previous pairs are defined. At each step, we shall also be proving that

$$D(x) = d(x) + \text{lower terms}, \quad H(x) = h(x) + \text{lower terms},$$

where the words “lower terms” refers to the partial order we defined above, meaning a linear combination of basis elements whose underlying tree monomial is smaller than the underlying tree monomial of $\hat{x}$.

**Basis of induction:** $k = 0$, so we have to define $D_1$ and $H_0$ (note that $D_0 = 0$ because there are no elements of negative homological degrees). In general, to define $D_k$, we should only consider the case where our element is a generator of $\mathcal{A}_g$, since in a dg operad the differential is defined by images of generators. For $l = 1$, this means that we should consider the case where our generator corresponds to a leading monomial $T = \text{lt}(g)$ of some relation $g$, and is of the form $T \otimes S$ where $S$ corresponds to the only divisor of $T$ which is a leading term, that is $T$ itself. Letting $D_1(T \otimes S) = \frac{1}{c_g} g$, where $c_g$ is the leading coefficient of $g$, we see that $D_1(T \otimes S) = T + \text{lower terms}$, as required. To define $H_0$, we use a yet another inductive argument, decreasing the monomials on which we want to define $H_0$. First of all, if a tree monomial $T$ is not divisible by any of the leading terms of relations, we put $H_0(T) = 0$. Assume that $T$ is divisible by some leading terms of relations, and $S_1, \ldots, S_p$ are the corresponding divisors. Then on $\mathcal{A}_g$ we can use $S_1 \wedge \cdots$ as a homotopy, so $h_0(T) = T \otimes S_1$. We put

$$H_0(T) = h_0(T) + H_0(T - D_1 h_0(T)).$$

Here the leading term of $T - D_1 h_0(T)$ is smaller than $T$ (since we already know that the leading term of $D_1 h_0(T)$ is $d_1 h_0(T) = T$), so induction on the leading term applies. Note that by induction the leading term of $H_0(T)$ is $h_0(T)$. 

Suppose that $k > 0$, that we know the pairs $(D_{l+1}, H_l)$ for all $l < k$, and that in these degrees

$$D(x) = d(\hat{x}) + \text{lower terms}, \quad H(x) = h(\hat{x}) + \text{lower terms}.$$
To define $D_{k+1}$, we should, as above, only consider the case of generators. In this case, we put

$$D_{k+1}(x) = d_{k+1}(x) - H_{k-1}D_k d_{k+1}(x).$$

The property $D_{k+1}(x) = d_{k+1}(x) + \text{lower terms}$ now easily follows by induction.

To define $H_k$, we proceed in a way very similar to what we did for the induction basis. Assume that $u \in \ker D_k$, and that we know $H_k$ on all elements of $\ker D_k$ whose leading term is less than $\hat{u}$. Since $D_k(u) = d_k(\hat{u}) + \text{lower terms}$, we see that $u \in \ker D_k$ implies $\hat{u} \in \ker d_k$. Then $h_k(\hat{u})$ is defined, and we put

$$H_k(u) = h_k(\hat{u}) + H_k(u - D_{k+1}h_k(\hat{u})).$$

Here $u - D_{k+1}h_k(\hat{u}) \in \ker D_k$ and its leading term is smaller than $\hat{u}$, so induction on the leading term applies (and it is easy to check that by induction $H_{k+1}(x) = h_{k+1}(x) + \text{lower terms}$).

Let us check that the mappings $D$ and $H$ defined by these formulas satisfy, for each $k > 0$, $D_k D_{k+1} = 0$ and $(D_{k+1}H_k)_{\ker D_k} = \text{id} - \bar{\pi}$. A computation checking that is somewhat similar to the way $D$ and $H$ were constructed. Let us prove both statements simultaneously by induction. If $k = 0$, the first statement is obvious. Let us prove the second one and establish that $D_1 H_0(T) = (\text{id} - \bar{\pi})(T)$ for each tree monomial $T$. Slightly rephrasing that, we shall prove that for each tree monomial $T$ we have $D_1 H_0(T) = T - \bar{T}$ where $\bar{T}$ is the residue of $T$ modulo $\mathcal{G}$ [13]. We shall prove this statement by induction on $T$. If $T$ is not divisible by any leading terms of relations, we have $H_0(T) = 0 = T - \bar{T}$. Let $T$ have divisors $S_1, \ldots, S_p$. We have $H_0(T) = h_0(T) + H_0(T - D_1 h_0(T))$, so

$$D_1 H_0(T) = D_1 h_0(T) + D_1 H_0(T - D_1 h_0(T)).$$

By induction, we may assume that

$$D_1 H_0(T - D_1 h_0(T)) = T - D_1 h_0(T) - \bar{(T - D_1 h_0(T))}.$$

Also,

$$D_1 h_0(T) = D_1 (T \otimes S_1) = \frac{1}{c_g} m_{T, S_1}(g) = T - r_g(T).$$

Here we use the usual notation for Gröbner bases computations [13]: $r_g(T)$ is the result of reduction of $T$ modulo $g$, and $m_{T, S_1}(g)$ denotes the result of the substitution of $g$ into $T$ at that place (we have $D_1 (T \otimes S_1) = \frac{1}{c_g} m_{T, S_1}(g)$ since it is true when $T$ is a relation, and the differential agrees with operadic compositions).

Combining the three previous equations, we obtain,

$$D_1 H_0(T) = T - r_g(T) + \bar{(T - D_1 h_0(T))} - \bar{(T - D_1 h_0(T))} = T - r_g(T) + (r_g(T) - r_g(T)) = T - r_g(T) = T - \bar{T},$$

$$\text{(id} - \bar{\pi})(T).$$
since for a Gröbner basis the residue does not depend on a choice of reductions. Assume that \( k > 0 \), and that our statement is true for all \( l < k \). We have

\[ D_k D_{k+1}(x) = 0 \]

since

\[
D_k D_{k+1}(x) = D_k(d_{k+1}(x) - H_{k-1}D_k d_{k+1}(x)) = \\
= D_k d_{k+1}(x) - D_k D_k d_{k+1}(x) = D_k d_{k+1}(x) - D_k d_{k+1}(x) = 0,
\]

because \( D_k d_{k+1} k \in \ker D_{k-1} \), and so \( D_k H_{k-1}(D_k(y)) = D_k(y) \) by induction. Also, for \( u \in \ker D_k \) we have

\[ D_{k+1} H_k(u) = D_{k+1} h_k(\tilde{u}) + D_{k+1} H_k(u - D_{k+1} h_k(\tilde{u})), \]

and by the induction on \( \tilde{u} \) we may assume that

\[ D_{k+1} H_k(u - D_{k+1} h_k(\tilde{u})) = u - D_{k+1} h_k(\tilde{u}) \]

(on elements of positive homological degree, \( \pi = 0 \)), so

\[ D_{k+1} H_k(u) = D_{k+1} h_k(\tilde{u}) + u - D_{k+1} h_k(\tilde{u}) = u, \]

which is exactly what we need. \( \square \)

5 Applications

5.1 Another proof of the PBW criterion for Koszulness

The goal of this section is to give a new proof of the Gröbner bases formulation [13] version of the PBW criterion of Hoffbeck [31] (generalising the PBW criterion of Priddy [51] for associative algebras).

**Theorem 5.1.** An operad with a quadratic Gröbner basis is Koszul.

**Proof.** First of all, it is enough to prove it in the monomial case, since it gives an upper bound on the homology: for the deformed differential, the cohomology may only decrease. In the monomial quadratic case, every divisor of a tree monomial covers one internal edge, and every internal edge is covered by precisely one divisor, so all the generators of our free resolution are of homological degree one less than the number of corollas used in them, hence the homology of the bar complex is concentrated on the diagonal, and our operad is Koszul. \( \square \)
5.2 Operads and commutative algebras

Recall a construction of an operad from a graded commutative algebra described in [33].
Let $A$ be a connected graded associative commutative algebra. Define an operad $\mathcal{O}_A$ as follows. We put
\[ \mathcal{O}_A(I) := A|_{|I| - 1}, \]
and define, for each shuffle surjection $\phi: I \rightarrow \{1, \ldots, k\}$, the composition map
\[ o_\phi: \mathcal{P}(k) \otimes \mathcal{P}(\phi^{-1}(1)) \otimes \cdots \otimes \mathcal{P}(\phi^{-1}(k)) \rightarrow \mathcal{P}(I) \]
to be the product in $A$:
\[ a \circ_\phi (b_1, \ldots, b_k) = ab_1 \cdots b_k. \]
The arities of the elements match: since $I = \bigsqcup_{i=1}^{k} \phi^{-1}(i)$, we have $|I| - 1 = k - 1 + (|\phi^{-1}(1)| - 1) + \cdots + (|\phi^{-1}(k)| - 1)$. (In the symmetric case, we have also to define the actions of symmetric groups; by definition, all the components of $\mathcal{O}_A$ are trivial representations of the respective symmetric groups.)
As we remarked in [13], a basis of the algebra $A$ leads to a basis of the operad $\mathcal{O}_A$: product of generators of the polynomial algebra is replaced by the iterated composition of the corresponding generators of the free operad where each composition is substitution into the last slot of an operation. Assume that we know a Gröbner basis for the algebra $A$ (as an associative algebra). It leads to a Gröbner basis for the operad $\mathcal{O}_A$ as follows: we first impose the quadratic relations defining the operad $\mathcal{O}_A/CZ_{\{x_1, \ldots, x_n\}}$ coming from the polynomial algebra (stating that the result of a composition depends only on the operations composed, not on the order in which we compose operations), and then use the identification of relations in the polynomial algebra with elements of the corresponding operad, as above. Our next goal is to explain how to use the Anick resolution of the trivial module for $A$ to construct a small resolution of the trivial module for $\mathcal{O}_A$.
Let us define a collection $\mathcal{R}$ which will then use to construct a resolution. We take the free operad generated by the collection $\mathcal{H}$ with $s^{-1} \mathcal{H}(k) = H^Q_{k-2}(A)$. We take $\mathcal{R}$ to be the quotient of that operad by the relations $c_1 \circ_2 c_2 = 0$, where $c_1 \in \mathcal{H}(k)$. In other words, in $\mathcal{R}$ all compositions are allowed except for those using the last slot of an operation. We shall show that $\mathcal{R}$ gives an upper bound on the Quillen homology of $\mathcal{O}_A$ by proving the following theorem.

**Theorem 5.2.** There exists a free resolution $(\mathcal{F}_{\mathcal{R}}, d) \rightarrow \mathcal{O}_A$.

**Proof.** This statement is almost immediate from our previous results. Indeed, we know how to obtain a Gröbner basis for $\mathcal{O}_A$ from a Gröbner basis of $A$. The leading terms of that Gröbner basis are all left combs
\[ a(\beta(1, 2), 3) \quad \text{and} \quad a(\beta(1, 3), 2) \]
with three leaves and some right combs

\[ \alpha_1(1, \alpha_2(2, \ldots \alpha_k(k, k+1)), \ldots) \].

By Proposition 3.12, the corresponding homology classes of the associated monomial operad can be described by elements of the same shape as defined above, but we should start with the operad generated by Anick chains [1], not by the homology. To understand what happens in the transition from the monomial replacement to \( \mathcal{O}_A \), let us look carefully into the general reconstruction scheme from the previous section. It recovers lower terms of differentials and homotopies by recalling lower terms of elements of the Gröbner basis. Let us do the reconstruction in two steps. At first, we shall recall all lower terms of relations except for those starting with \( \alpha(\beta(-, -), -) \); the latter are still assumed to vanish. On the next step we shall recall all lower terms of those quadratic relations. Note that after the first step we model many copies of the associative algebra resolution and the differential there; so we can compute the homology explicitly. At the next step, a differential will be induced on this homology we computed, and we end up with a resolution of the required type.

In some cases the existence of such a resolution is enough to compute Quillen homology of \( \mathcal{O}_A \); for example, it is so when the algebra \( A \) is Koszul, as we shall see now. In general, the differential of this resolution incorporates lots of information, including the higher operations (Massey (co)products) on the homology of \( A \).

Recall that if the algebra \( A \) is quadratic, then the operad \( \mathcal{O}_A \) is quadratic as well. In [13], we proved that if the algebra \( A \) is PBW, then the operad \( \mathcal{O}_A \) is PBW as well, and hence is Koszul. Now we shall prove the following substantial generalisation of this statement (substantially simplifying the proof of this statement given in [33]).

**Theorem 5.3.** If the algebra \( A \) is Koszul, then the operad \( \mathcal{O}_A \) is Koszul as well.

**Proof.** Koszulness of our algebra implies that the homology of the bar resolution is concentrated on the diagonal. Consequently, the operad \( \mathcal{R} \) constructed above is automatically concentrated on the diagonal, and so is its homology, which completes the proof.

5.3 **The operads of Rota–Baxter algebras**

The main goal of this section is to compute Quillen homology for the operad of Rota–Baxter algebras, and the operad of noncommutative Rota–Baxter algebras. Those are among the simplest examples of operads which are not covered by the Koszul duality theory, being operads with nonhomogeneous relations. Note that it is even not clear that these operads have minimal models: being operads with nontrivial unary operations, they are not covered by results of [47], and indeed some operads with nontrivial unary operations do not admit minimal models.
5.3.1 The operads \( \text{RB} \) and \( \text{ncRB} \) and their Gröbner bases

**Definition 5.4.** A commutative Rota–Baxter algebra of weight \( \lambda \) is a vector space with an associative commutative product \( a, b \mapsto a \cdot b \) and a unary operator \( P \) which satisfy the following identity:

\[
P(a) \cdot P(b) = P(P(a) \cdot b + a \cdot P(b) + \lambda a \cdot b).
\]

We denote by \( \text{RB} \) the operad of Rota–Baxter algebras. We view it as a shuffle operad with one binary and one unary generator.

Commutative Rota–Baxter algebras were defined in [57] with a motivation coming from probability theory [4]. Various constructions of free commutative Rota–Baxter algebras appear in [57, 9, 28]. The latter paper also contains extensive bibliography and information on various applications of those algebras.

**Definition 5.5.** A noncommutative Rota–Baxter algebra of weight \( \lambda \) is a vector space with an associative product \( a, b \mapsto a \cdot b \) and a unary operator \( P \) which satisfy the same identity as above:

\[
P(a) \cdot P(b) = P(P(a) \cdot b + a \cdot P(b) + \lambda a \cdot b).
\]

We denote by \( \text{ncRB} \) the operad of noncommutative Rota–Baxter algebras. Somehow, it is a bit simpler than the operad in the commutative case, because it can be viewed as a non-symmetric operad with one binary and one unary generator.

Noncommutative Rota–Baxter algebras has been extensively studied in the past years. We refer the reader to the paper of Ebrahimi–Fard and Guo [18] for an extensive discussion of applications and occurrences of those algebras in various areas of mathematics, and a combinatorial construction of the corresponding free algebras.

Let us consider the path-lexicographic ordering of the free operad; we assume that \( P > \cdot \).

**Proposition 5.6.** The defining relations for operads \( \text{RB} \) and \( \text{ncRB} \) form a Gröbner basis.

**Proof.** Here we present a proof for the case of \( \text{ncRB} \), the proof for \( \text{RB} \) is essentially the same, with the only exception that there are two S-polynomials to be reduced, as opposed to one S-polynomial in the case of \( \text{ncRB} \) (which, as we pointed above, is easier because we are dealing with a non-symmetric operad). For the associative suboperad of \( \text{ncRB} \), the defining relations form a Gröbner basis, so the S-polynomials coming from the small common multiples the leading term of the associativity relation has with itself clearly can be reduced to zero. The leading term of the Rota–Baxter relation is \( P(P(a_1)a_2) \). This term only has a nontrivial overlap with itself, not with the leading term of the
associativity relation, and that overlap is $P(P(a_1 \cdot a_2) \cdot a_3)$. From this overlap, we compute the S-polynomial

$$-P(P(a_1 \cdot P(a_2)) \cdot a_3) - \lambda P(P(a_1 \cdot a_2) \cdot a_3) + P((P(a_1) \cdot P(a_2)) \cdot a_3) + P((P(a_1) \cdot a_2) \cdot P(a_3)) + \lambda P((P(a_1) \cdot a_2) \cdot a_3) - P(P(a_1) \cdot a_2) \cdot P(a_3),$$

and it can be reduced to zero by a lengthy sequence of reductions which we omit here (but which in fact can be read from the formula for $d_{\nu_3}$ in Proposition 5.11 below). By Diamond Lemma [13], our relations form a Gröbner basis. □

**Remark 5.7.** In the case of the operad $n\text{cRB}$, our computation immediately provides bases for free noncommutative Rota–Baxter algebras. Indeed, since our operad is non-symmetric, the degree $n$ part of the free noncommutative Rota–Baxter algebra generated by the set $B$ is nothing but $n\text{cRB}(n) \otimes V^\otimes n$, where $V = \text{span}(B)$, so we can use the above Gröbner basis to describe that part. More precisely, we first define the set of admissible expressions on a set $B$ recursively as follows:

- elements of $B$ are admissible expressions;
- if $b$ is an admissible expression, then $P(b)$ is an admissible expression;
- if $b_1, \ldots, b_k$ are admissible expressions, and for each $i$ either $b_i$ is an element of $B$ or $b_i = P(b'_i)$ with $b'_i$ an admissible expression, then their associative product $b_1 \cdot b_2 \cdots b_k$ is an admissible expression.

Based on this definition, we shall call some of admissible expressions the Rota–Baxter monomials, tracing the construction of an admissible expression and putting some restrictions. Namely,

- elements of $B$ are Rota–Baxter monomials;
- if $b$ is a Rota–Baxter monomial, which, as an admissible expression, is either $b = P(b')$ or $b = b_1 \cdot b_2 \cdots b_k$ with $b_1 \in B$, then $P(b)$ is a Rota–Baxter monomial;
- if $b_1, \ldots, b_k$ are Rota–Baxter monomials, and for each $i$ either $b_i \in B$ or $b_i = P(b'_i)$ for some $b'_i$, then their associative product $b_1 \cdot b_2 \cdots b_k$ is a Rota–Baxter monomial.

Our previous discussion means that the set of all Rota–Baxter monomials forms a basis in the free noncommutative Rota–Baxter algebra generated by the set $B$. It would be interesting to compare this basis with the basis from [19].
5.3.2 Quillen homology of the operads $RB$ and $ncRB$

**Proposition 5.8.** For each of the operads $RB$ and $ncRB$, the resolution for its monomial version from Section 2 is minimal, that is the differential induced on the space of generators is zero.

*Proof.* In the case of the operad $RB$, the overlaps obtained from the leading monomials $(a_1 \cdot a_2) \cdot a_3$, $(a_1 \cdot a_3) \cdot a_2$, and $P(a_1) \cdot a_2$ are, in arity $n$,

$$(\ldots((a_i \cdot a_k) \cdot a_l)\ldots) \cdot a_m$$

and

$$P(P(\ldots P(P(a_i) \cdot a_k) \cdot a_l)\ldots) \cdot a_m),$$

for all permutations $i_2, i_3, \ldots, i_n$ of integers $2, 3, \ldots, n$. It is easy to see that for each of them there exists only one indecomposable covering by relations, so the differential maps such a generator to the space of decomposable elements, and the statement follows.

Similarly, in the case of the operad $ncRB$, the only overlaps obtained from the leading monomials $(a_1 \cdot a_2) \cdot a_3$ and $P(a_1) \cdot a_2$ are, in arity $n$,

$$(\ldots((a_1 \cdot a_2) \cdot a_3)\ldots) \cdot a_{n-1} \cdot a_n$$

and

$$P(P(\ldots P(P(a_1) \cdot a_2) \cdot a_3)\ldots) \cdot a_{n-1} \cdot a_n).$$

It is easy to see that for each of them there exists only one indecomposable covering by relations, so the differential maps such a generator to the space of decomposable elements, and the statement follows. 

**Theorem 5.9.** We have

$$\dim H^Q_l(RB)(k) = \begin{cases} (k - 1)!, & l = k \geq 1, \\ (k - 1)! , & l = k + 1 \geq 2. \end{cases}$$

$$\dim H^Q_l(ncRB)(k) = \begin{cases} 1, & l = k \geq 1, \\ 1, & l = k + 1 \geq 2. \end{cases}$$

*Proof.* In both cases, the subspace of generators of the free resolution splits into two parts: the part obtained as overlaps of the leading terms of the associativity relations, and the part obtained as overlaps of the leading term of the Rota–Baxter relation with itself. In arity $k$, the former are all of homological degree $k - 1$, while the latter — of homological degree $k$. This means that when we compute the homology of the differential of our resolution restricted to the space of generators, the only cancellations can happen if some of the elements resolving the associativity relation appear as differentials of some elements resolving the Rota–Baxter relation. However, since all monomials in the Rota–Baxter relation are of degree at least 1 in $P$, the way the deformed differential
is constructed in the proof of Theorem 4.1 shows that all the terms appearing in the formulas for the respective differentials are also of degree at least 1 in $P$, so no cancellations are possible.

In addition to Quillen homology computation, one can ask for explicit formulas for differentials in the free resolutions. It is not difficult to write down formulas for small arities (see the example below), but in general compact formulas are yet to be found. We expect that they incorporate the Spitzer’s identity and its noncommutative analogue [20]. However, the following statement is immediate.

**Corollary 5.10.**

- The minimal model $RB_{\infty}$ for the operad $RB$ is a quasi-free operad whose space of generators has a $(k-1)!$-dimensional space of generators of homological degree $(k-2)$ in each arity $k \geq 2$, and a $(k-1)!$-dimensional space of generators of homological degree $k-1$ in each arity $k \geq 1$.

- The minimal model $ncRB_{\infty}$ for the operad $ncRB$ is a quasi-free operad generated by operations $\mu_k$, $k \geq 2$ of arity $k$ and homological degree $k-2$, and $\nu_l$, $l \geq 1$ of arity $l$ and homological degree $l-1$.

Let us conclude this section with formulas for low arities differentials in $ncRB_{\infty}$, to give the reader a flavour of what sort of formulas to expect.

**Example 5.11.** We have $d\nu_1 = d\mu_2 = 0$, and

$$
d\nu_2 = P(\mu_2(P(\cdot), \cdot)) + P(\mu_2(\cdot, P(\cdot))) - \mu_2(P(\cdot), P(\cdot)) + \lambda P(\mu_2(\cdot, \cdot)),
\quad d\mu_3 = \mu_2(\mu_2(\cdot, \cdot), \cdot) - \mu_2(\cdot, \mu_2(\cdot, \cdot)),
$$

$$
d\nu_3 = \mu_3(P(\cdot), P(\cdot), P(\cdot)) - P(\mu_2(\nu_2(\cdot, \cdot), \cdot) - \mu_2(\cdot, \nu_2(\cdot, \cdot))) -
\quad - P(\mu_3(P(\cdot), P(\cdot), \cdot)) + \mu_3(\cdot, P(\cdot), P(\cdot)) + \mu_3(\cdot, P(\cdot), P(\cdot)) +
\quad + \nu_2(\mu_2(P(\cdot), \cdot) - \nu_2(\cdot, \mu_2(\cdot, \cdot)) + \nu_2(\cdot, P(\cdot), \cdot) + \nu_2(\cdot, \cdot, P(\cdot))) +
\quad + \mu_2(\nu_2(\cdot, \cdot), P(\cdot)) - \mu_2(P(\cdot), \nu_2(\cdot, \cdot)) + \lambda [\nu_2(\mu_2(\cdot, \cdot), \cdot) - \nu_2(\cdot, \mu_2(\cdot, \cdot)) -
\quad - P(\mu_3(P(\cdot), \cdot, \cdot) + \mu_3(\cdot, P(\cdot), \cdot) + \mu_3(\cdot, \cdot, P(\cdot))) - \lambda^2 P(\mu_3(\cdot, \cdot, \cdot)).
$$

### 5.4 The operad $BV$ and hypercommutative algebras

The main goal of this section is to explain how our results can be used to study the operad $BV$ of Batalin–Vilkovisky algebras. The key result below (Theorem 5.18) is also proved in [17]; our proofs are based on entirely different methods.

#### 5.4.1 The operad $BV$ and its Gröbner basis.

Batalin–Vilkovisky algebras show up in various questions of mathematical physics. In [23], a cofibrant resolution for the corresponding operad was presented. However, that resolution is a little bit more that minimal. In this
section, we present a minimal resolution for this operad in the shuffle category. The operad $BV$, as defined in most sources, is an operad with quadratic–linear relations: the odd Lie bracket can be expressed in terms of the product and the unary operator. However, alternatively one can say that a $BV$-algebra is a dg commutative algebra with a unary degree 1 operator $\Delta$ with $\Delta^2 = 0$ which is a differential operator of order at most 2. This definition of a $BV$-algebra is certainly not new, see, e. g., [24]. With this presentation, the corresponding operad becomes an operad with homogeneous relations (of degrees 2 and 3). Our choice of degrees and signs is taken from [23] where it is explained how to translate between this convention and other popular definitions of $BV$-algebras.

We write identities for operations evaluated on elements of degree 0, assuming the usual Koszul sign rule for evaluating operations on elements of arbitrary degrees. We want to emphasize that when computing Gröbner bases, we are dealing with operations only, and all signs arise from evaluating operations on elements. The representatives of the identities are chosen in such a way that they can be viewed as elements of the free shuffle operad; we use the language of operations, as opposed the language of tree monomials: for each $i$, the argument $a_i$ of an operation corresponds to the leaf $i$ of the corresponding tree monomial.

**Definition 5.12 (BV-algebras with homogeneous relations).** A Batalin-Vilkovisky algebra, or $BV$-algebra for short, is a differential graded vector space $(A, d_A)$ endowed with

- a symmetric binary product $\bullet$ of degree 0,
- a unary operator $\Delta$ of degree +1,

such that $(A, d_A, \Delta)$ is a mixed complex, $d_A$ is a derivation with respect to the product, and such that

- the product $\bullet$ is associative, $(a_1 \bullet a_2) \bullet a_3 = a_1 \bullet (a_2 \bullet a_3)$ and $(a_1 \bullet a_3) \bullet a_2 = a_1 \bullet (a_2 \bullet a_3)$
- the operator $\Delta$ satisfies $\Delta^2(a_1) = 0$,
- the operations satisfy the cubic identity

$$\Delta((a_1 \bullet a_2) \bullet a_3) = \Delta(a_1 \bullet a_2) \bullet a_3 + \Delta(a_1 \bullet a_3) \bullet a_2 + a_1 \bullet \Delta(a_2 \bullet a_3) - (\Delta(a_1) \bullet a_2) \bullet a_3 - (a_1 \bullet \Delta(a_2)) \bullet a_3 - (a_1 \bullet \Delta(a_3)) \bullet a_2,$$

Let us consider the ordering of the free operad where we first compare lexicographically the operations on the paths from the root to leaves, and then the planar permutations of leaves; we assume that $\Delta > \bullet$. 
Proposition 5.13. The above relations together with the degree 4 relation
\[
\Delta(a_1 \bullet \Delta(a_2 \bullet a_3)) + \Delta(\Delta(a_1 \bullet a_2) \bullet a_3) + \Delta(\Delta(a_1 \bullet a_3) \bullet a_2) - \\
\Delta((\Delta(a_1) \bullet a_2) \bullet a_3) - \Delta((a_1 \bullet \Delta(a_2)) \bullet a_3) - \Delta((a_1 \bullet \Delta(a_3)) \bullet a_2) = 0 \tag{1}
\]
form a Gröbner basis of relations for the operad of BV-algebras.

Proof. With respect to our ordering, the leading monomials of our original relations are 
\((a_1 \bullet a_2) \bullet a_3\), \((a_1 \bullet a_3) \bullet a_2\), \(\Delta^2(a_1)\), and 
\(\Delta(a_1 \bullet (a_2 \bullet a_3))\). The only small common multiple of \(\Delta^2(a_1)\) and 
\(\Delta(a_1 \bullet (a_2 \bullet a_3))\) gives a nontrivial S-polynomial which, is precisely the relation (1). The leading term of that relation is 
\(\Delta(\Delta(a_1 \bullet a_2) \bullet a_3)\).

It is well known that \(\dim BV(n) = 2^n n!\) [24], so to verify that our relations form a Gröbner basis, it is sufficient to show that the restrictions imposed by these leading monomials are enough. In other words, we may check that the number of arity \(n\) tree monomials that are not divisible by any of these is equal to \(2^n n!\). Moreover it is sufficient to check that for \(n \leq 4\), since all S-polynomials of our relations will be elements of arity at most 4. This can be easily checked by hand, or by a computer program [15]. □

5.4.2 Quillen homology of the operad BV

Let us denote by \(\mathcal{G}\) the Gröbner basis from the previous section.

Proposition 5.14. For the monomial replacement of BV, the resolution \(\mathcal{A}/\mathcal{G}\) from Section 2 is minimal, that is the differential induced on the space of generators is zero.

Proof. Let us describe explicitly the space of generators, that is possible indecomposable coverings of monomials by leading terms of relations (all monomials below are chosen from the basis of the free shuffle operad, so the correct ordering of subtrees is assumed). These are

- all monomials \(\Delta^k(a_1)\), \(k \geq 2\) (covered by several copies of \(\Delta^2(a_1)\)),
- all “left combs”
  \[
  \lambda = \left( (a_1 \bullet a_2) \bullet a_3 \right) \bullet \cdots \bullet a_n \tag{2}
  \]
  where \((k_2, \ldots, k_n)\) is a permutation of numbers 2, \ldots, \(n\), \(n \geq 3\) (only the leading terms \((a_1 \bullet a_2) \bullet a_3\) and \((a_1 \bullet a_3) \bullet a_2\) are used in the covering),
- all the monomials
  \[
  \Delta^2(\Delta(\lambda_1 \bullet (\lambda_2 \bullet a_i)))
  \]
  where \(k \geq 1\), each \(\lambda_i\) is a left comb as described above (so that several copies of \(\Delta^2\), the leading term of degree 3, and several leading terms of the associativity relations are used in the covering),
Quillen Homology for Operads via Gröbner Bases

- all monomials

\[ \Delta^k(\Delta(\ldots \Delta(\Delta(\lambda_1 \bullet \lambda_2) \bullet \lambda_3) \bullet \ldots) \bullet \lambda_n) \] (3)

where \( k \geq 0, \ n \geq 3, \) and \( \lambda_i \) are left combs (several copies of all leading terms are used, including at least one copy of the degree 4 leading term).

This is a complete list of tree monomials \( T \) for which \((\mathcal{A}_g)^P\) is nonzero in positive homological degrees. It is easy to see that for each of them there exists only one indecomposable covering by relations, that is only one generator of \( \mathcal{A}_g \) of shape \( T \). Consequently, the differential maps such a generator to a combination of decomposable elements, so the differential induced on generators is identically zero.

The resolution of the operad \( BV \) which one can derive by our methods from this one is quite small (in particular, smaller than the one of [23]) but still not minimal. However, we now have enough information to compute Quillen homology of the operad \( BV \).

**Theorem 5.15.** The basis of \( H^0(BV) \) is formed by monomials

\[ \Delta^k(a_1), \quad k \geq 1, \]

and all monomials of the form

\[ \Delta(\ldots \Delta(\Delta(\lambda_1 \bullet \lambda_2) \bullet \ldots) \bullet (\lambda_n \bullet a_j)), \quad n \geq 1 \]

from the resolution of the monomial replacement of \( BV \) discussed above. Here all \( \lambda_i \) are left combs.

**Proof.** First of all, let us notice that since \( \Omega(B(BV)) \), a free operad generated by \( B(BV)[-1] \), provides a resolution for \( BV \), the space \( H^0(BV)[-1] \) is the space of generators of the minimal free resolution, and we shall study the resolution provided by our methods.

It is easy to check that the element \( \Delta(\Delta(a_1 \bullet a_2) \bullet a_3) \) that corresponds to the leading term of the only contributing S-polynomial will be killed by the differential of the element \( \Delta^2(a_1 \bullet (a_2 \bullet a_3)) \) (covered by corresponding overlapping leading terms \( \Delta^2(a_1) \) and \( \Delta(a_1 \bullet (a_2 \bullet a_3)) \)) in the deformed resolution. This observation goes much further, namely we have for \( k \geq 1 \)

\[
D(\Delta^k(\Delta(\ldots \Delta(\Delta(\lambda_1 \bullet \lambda_2) \bullet \lambda_3) \bullet \ldots) \bullet (\lambda_n \bullet a_j)) = \\
= \Delta^{k-1}(\Delta(\ldots \Delta(\Delta(\lambda_1 \bullet \lambda_2) \bullet \lambda_3) \bullet \ldots) \bullet \lambda_n) \bullet a_j) + \text{lower terms}
\]

in the sense of the partial ordering we discussed earlier. So, if we retain only leading terms of the differential, the resulting homology classes are represented by all the monomials of arity \( m \)

\[ \Delta(\ldots \Delta(\Delta(\lambda_1 \bullet \lambda_2) \bullet \ldots) \bullet \lambda_n) \]
with \( \lambda_n \) having at least two leaves. They all have the same homological degree \( m - 2 \) in the resolution (that is, formed by overlapping \( m - 2 \) leading terms), and so there are no further cancellations.

□

So far we have not been able to describe a minimal resolution of the operad \( BV \) by relatively compact closed formulas, even though in principle our proof, once processed by a version of Brown’s machinery \([8, 10]\), would clearly yield such a resolution (in the shuffle category).

5.4.3 The gravity operad and the Quillen homology of \( BV \)

The gravity operad \( \text{Grav} \) and its Koszul dual \( \text{Hycom} \) were originally defined in terms of moduli spaces of curves of genus 0 with marked points \( \mathcal{M}_{0,n+1} \) \([25, 27]\). However, we are interested in the algebraic aspects of the story, and we use the following descriptions of the gravity operad as a quadratic algebraic operad \([25]\).

An algebra over the operad \( \text{Grav} \) is a chain complex with graded antisymmetric products

\[
[x_1, \ldots, x_n] : A^\otimes n \to A
\]

of degree \( 2 - n \), which satisfy the relations:

\[
\sum_{1 \leq i < j < k} \pm[[a_i, a_j], a_1, \ldots, a_i, \ldots, a_j, \ldots, a_k, b_1, \ldots, b_l] = \\
\begin{cases} 
[[a_1, \ldots, a_k], b_1, \ldots, b_l], & l > 0, \\
0, & l = 0, 
\end{cases}
\]

for all \( k > 2, l \geq 0, \) and \( a_1, \ldots, a_k, b_1, \ldots, b_l \in A \). For example, setting \( k = 3 \) and \( l = 0 \), we obtain the Jacobi relation for \([a, b]\).

Let us define an admissible ordering of the free operad whose quotient is \( \text{Grav} \) as follows. We introduce an additional weight grading, putting the weight of the corolla corresponding to the binary bracket equal to 0, all other weights of corollas equal to 1, and extending it to compositions by additivity of weight. To compare two monomials, we first compare their weights, then the root corollas, and then path sequences \([13]\) according to the reverse path-lexicographic order. For both of the latter steps, we need an ordering of corollas; we assume that corollas of larger arity are smaller. Then for the relation \((k, l)\) in (4) (written in the shuffle notation with variables in the proper order), its leading monomial is equal to the monomial in the right hand side for \( l > 0 \), and to the monomial \([a_1, \ldots, a_{n-2}, [a_{n-1}, a_n]]\) for \( l = 0 \).

The following theorem, together with the PBW criterion, implies that the operads \( \text{Grav} \) and \( \text{Hycom} \) are Koszul, the fact first proved by Getzler \([24]\).

**Theorem 5.16.** For our ordering, the relations of \( \text{Grav} \) form a Gröbner basis of relations.
Proof. The tree monomials that are not divisible by leading terms of relations are precisely

\[ [\lambda_1, \lambda_2, \ldots, \lambda_{n-1}, a_j], \]

where all \( \lambda_i, 1 \leq i \leq (n - 1) \) are left combs as in (2) (but made from brackets, not products).

Lemma 5.17. The graded character of the space of such elements of arity \( n \) is

\[ (2 + t^{-1})(3 + t^{-1}) \cdots (n - 1 + t^{-1}). \]

Proof. To compute the number of basis elements where the top degree corolla is of arity \( k + 1 \) (or, equivalently, degree \( 1 - k \)), \( k \geq 1 \), let us notice that this number is equal to the number of basis elements \( [\lambda_1, \lambda_2, \ldots, \lambda_k] \) where the arity of \( \lambda_k \) is at least 2 (a simple bijection: join \( \lambda_{n-1} \) and \( a_j \) into \( [\lambda_{n-1}, a_j] \)). The latter number is equal to

\[
\sum_{m_1 + \cdots + m_k = n, \quad m_i \geq 1, m_k \geq 2} \frac{(m_1 - 1)! \cdots (m_k - 1)! m_1 m_2 \cdots m_k}{(m_1 + m_2 + \cdots + m_k)(m_2 + \cdots + m_k) \cdots (m_1, m_2, \ldots, m_k)}
\]

where each factor \( (m_i - 1)! \) counts the number of left combs of arity \( m_i \), and the remaining factor is known [15] to be equal to the number of shuffle permutations of the type \( (m_1, \ldots, m_k) \). This can be rewritten in the form

\[
\sum_{m_1 + \cdots + m_k = n, m_i \geq 1, m_k \geq 2} \frac{(m_1 + \cdots + m_k - 1)!}{(m_2 + \cdots + m_k)(m_3 + \cdots + m_k) \cdots m_k}
\]

and if we introduce new variables \( p_i = m_i + \cdots + m_k \) it takes the form

\[
\sum_{2 \leq p_1 < \cdots < p_k \leq n-1} \frac{(n - 1)!}{p_2 \cdots p_k}
\]

which clearly is the coefficient of \( t^{1-k} \) in the product

\[
(n - 1)! \left( 1 + \frac{1}{2t} \right) \left( 1 + \frac{1}{3t} \right) \cdots \left( 1 + \frac{1}{(n-1)t} \right) = \]

\[
= (2 + t^{-1})(3 + t^{-1}) \cdots (n - 1 + t^{-1}).
\]

Since the graded character of Grav is given by the same formula [25], we indeed see that the leading terms of defining relations give an upper bound on dimensions of homogeneous components of Grav that coincides with the actual dimensions, so there is no room for further Gröbner basis elements. \( \square \)
Using the basis of Grav we just constructed, we are able to prove the following result (which gets a conceptual explanation in the next section):

**Theorem 5.18.** On the level of collections of graded vector spaces, we have

\[ sH^Q(BV) \cong \text{Grav}^* \otimes \text{End}_{s \mathbb{C}^{s-1}} \oplus \delta[s][\delta], \]  

(5)

where Grav* is the co-operad dual to Grav, End_{s \mathbb{C}^{s-1}} is the endomorphism operad of the graded vector space \( s \mathbb{C}^{s-1} \), and \( \delta[s][\delta] \) is a cofree coalgebra generated by an element \( \delta \) of degree 2.

**Proof.** As above, instead of looking at the bar complex, we shall study the basis of the space of generators of the minimal resolution obtained in Theorem 5.15. In arity 1, the element \( \delta^k \) (of degree 2k) corresponds to \( s \Delta^k(a_1) \) (of degree \( k + (k-1) + 1 = 2k \)), the first summand coming from the fact that \( \Delta \) is of degree 1, the second from the fact that \( \Delta^k \) is an overlap of \( k-1 \) relations, and the last one is the degree shift given by \( s \). The case of elements of internal degree 0 (which in both cases are left combs) is also obvious; a left comb of arity \( n \) in the space of generators of the free resolution is of homological degree \( n - 2 + 1 = n - 1 \), the second summand coming from the degree shift given by \( s \), and this matches the degree shift given by \( \text{End}_{s \mathbb{C}^{s-1}}(n) \). For elements of internal degree \( k - 1 \), let us extract from a typical monomial

\[ T = \Delta(\ldots \Delta(\Delta(\lambda_1 \bullet \lambda_2) \bullet \cdots) \bullet (\lambda_k \bullet a_j)), \]

of this degree and of arity \( n \) the left combs \( \lambda_1, \lambda_2, \ldots, \lambda_{k-1}, \lambda_k, a_j \), and assign to \( T \) the element of \( \text{Grav}^* \otimes \text{End}_{s \mathbb{C}^{s-1}} \) corresponding, via the degree shift, to the element dual to the monomial \( [\lambda_1, \lambda_2, \ldots, \lambda_{k-1}, \lambda_k, a_j] \in \text{Grav} \). This establishes a degree-preserving bijection, because if arities of \( \lambda_1, \ldots, \lambda_k \) are \( n_1, \ldots, n_k \), the total (internal plus homological) degree of the former element is

\[ (k - 1) + (k - 2 + 1 + (n_1 - 1) + \cdots + (n_k - 1)) + 1 = n + k - 2, \]

(where we add up the \( \Delta \) degree, the overlap degree, and the degree shift), and the total degree of the latter one is \( (k - 1) + (n - 1) = n + k - 2 \). \( \square \)

5.4.4 Relationship to Frobenius manifold construction of Baranikov and Kontsevich

We conclude with a brief discussion on how our results match those of Baranikov and Kontsevich ([5], see also [43, 45]) who proved in a rather indirect way that for a dg BV-algebra that satisfies the "\( \partial - \overline{\partial} \)-lemma", there exists a Hycom-algebra structure on its cohomology. Their result hints that our isomorphism (5) exists not just on the level of graded vector spaces, but rather has some deep operadic structure behind it. For precise statements and more details we refer the reader to [17, 34]; the point we are trying to make here is that
From Theorem 5.16, it follows that the operads Grav and Hycom are Koszul, so the cobar construction $\Omega(\text{Grav}^* \otimes \text{End}_{d[1]})$ is a minimal model for Hycom. We shall now show that the differential of $BV^\infty$ on generators coming from Grav deforms the differential of Hycom in a certain sense. Let $D$ and $d$ denote the differentials of $BV^\infty$ and Hycom respectively. We can decompose $D = D_2 + D_3 + \cdots$ according to the homotopy co-operad structure it provides on the space of generators (note that $d = d_2$ since the operad Hycom is Koszul). Also, let $m^*$ denote the obvious coalgebra structure on $\delta / \text{CZ}[\delta]$. We shall call a tree monomial in $BV^\infty$ mixed, if it contains both corollas from Grav $^* \otimes \text{End}_{d[1]}$ and from $(\delta / \text{CZ}[\delta])$. Then $D_2 = d_2 + m^*$, while for $k \geq 3$ the co-operation $D_k$ is zero on the generators $\delta / \text{CZ}[\delta]$, and maps generators from Grav into linear combinations of mixed tree monomials. Indeed, Hycom-algebras are closely related to formal Frobenius manifolds, and the result of Barannikov and Kontsevich [5] essentially implies that there exists a mapping from Hycom to the homotopy quotient $BV / \Delta$. In fact, it is an isomorphism, which can be proved in several different ways, both using Gröbner bases and geometrically; see [46] for a short geometric argument proving that. This means that the following maps exist (the vertical arrows are quasiisomorphisms between the operads and their minimal models):

$$
\begin{array}{ccc}
BV^\infty & \xrightarrow{\pi} & \text{Hycom}^\infty \\
\downarrow & & \downarrow \\
BV & \xrightarrow{\pi} & \text{Hycom} \\
\text{BV} / \Delta & \xleftarrow{\sim} & \text{Hycom}
\end{array}
$$

Lifting $\pi$: $BV^\infty \to BV / \Delta \cong \text{Hycom}$ to the minimal model $\text{Hycom}^\infty$ of Hycom, we obtain the commutative diagram

$$
\begin{array}{ccc}
BV^\infty & \xrightarrow{\psi} & \text{Hycom}^\infty \\
\downarrow & & \downarrow \\
BV & \xrightarrow{\pi} & \text{Hycom} \\
\text{BV} / \Delta & \xleftarrow{\sim} & \text{Hycom}
\end{array}
$$

so there exists a map of dg operads (and not just graded vector spaces, as it follows from our previous computations) between $BV^\infty$ and $\text{Hycom}^\infty$. Commutativity of our diagram together with simple degree considerations yields what we need.
References


[33] Anton Khoroshkin, Koszul operads and distributive lattices, Preprint ITEP-TH-24/06.


[46] Nikita Markarian, Hycom = BV/Δ, see http://nikitamarkarian.wordpress.com


AN EQUIVARIANT MAIN CONJECTURE IN IWASAWA THEORY
AND THE COATES-SINNOTT CONJECTURE

REZA TALEB

Received: December 12, 2012
Revised: April 4, 2013
Communicated by Otmar Venjakob

Abstract. We formulate and prove an Equivariant Main Conjecture (EMC) for all prime numbers $p$ under the assumptions $\mu = 0$ and the validity of the 2-adic Main Conjecture in Iwasawa theory [47]. This equivariant version coincides with the version, which Ritter and Weiss formulated and proved for odd $p$ under the assumption $\mu = 0$ in [35]. Our proof combines the approach of Ritter and Weiss with ideas and techniques used by Greither and Popescu in [15] in a recent proof of an equivalent formulation of the above EMC under the same assumptions ($p$ odd and $\mu = 0$) as in [35]. As an application of the EMC we prove the Coates-Sinnott Conjecture, again assuming $\mu = 0$ and the 2-adic Main Conjecture.

2010 Mathematics Subject Classification: 11R23, 11R42, 14F42, 11R70, 11R33, 11R34
Keywords and Phrases: Iwasawa theory, global and $p$-adic $L$-functions, motivic cohomology, algebraic $K$-theory, Fitting ideals

Contents

1 Introduction 750
2 Algebraic Construction 756
3 Analytic Construction 761
1 Introduction

One of the most fascinating discoveries in Arithmetic Algebraic Geometry is the still mysterious relationship between certain algebraic and analytic data attached to a given arithmetic object. Classical examples include the Conjecture of Birch and Swinnerton-Dyer, which conjecturally relates the order of vanishing of the $L$-function attached to an elliptic curve at 1 to the rank of the algebraically defined Mordell-Weil group of the curve, and Dirichlets Analytic Class Number Formula, which gives a precise algebraic interpretation of the residue of the zeta-function of a number field at 1. This last connection has been generalised to yield interpretations of special values of zeta-functions at arbitrary negative integers in terms of algebraic $K$-theory and motivic cohomology. One of the main tool to understand the deep relations between algebraic and analytic objects is Iwasawa Theory. In this theory a precise formulation of such a relationship is called the Main Conjecture. We first recall the formulation of the Main Conjecture in Iwasawa theory in the classical form:

Let $p$ be a prime number, let $F$ be a totally real number field, and let $\psi$ be a 1-dimensional $p$-adic Artin character for $F$ with $F_\psi$ totally real, where $F_\psi$ denotes the fixed field of the kernel of $\psi$. Let $F_\infty$ denote the cyclotomic $\mathbb{Z}_p$-extension of $F$. We recall Greenberg’s terminology about the different types of the characters $\psi$: $\psi$ is of type S, if $F_\psi \cap F_\infty = F$, and $\psi$ is of type W, if $F_\psi \subseteq F_\infty$. Let $O_\psi$ denote the ring obtained by adjoining all $\psi$-values to the ring $\mathbb{Z}_p$. Let $F_\psi, \infty$ be the cyclotomic $\mathbb{Z}_p$-extension of $F_\psi$ with Galois group $\Gamma$ over $F_\psi$. Throughout we fix a topological generator $\gamma$ of $\Gamma$. We denote by $S$ a finite set of primes of $F$ containing the set $S_p$ of primes above $p$ and the set of the infinite primes, and by $S_f$ the set of finite primes in $S$. Deligne and Ribet [8] and independently Cassou-Nogués [4] showed the existence of a $p$-adic $L$-function for the character $\psi$, which is continuous for $s \in \mathbb{Z}_p \setminus \{1\}$, and even at $s = 1$, if $\psi$ is not trivial. This satisfies the following interpolation property for any integer $n \geq 1$:

$$L_p(1 - n, \psi) = L(1 - n, \psi \omega^{-n}) \prod_{p \in S_p} (1 - \psi \omega^{-n}(p)Nm(p)^{1-n}).$$

Here $L(1 - n, \psi \omega^{-n})$ is the usual Artin $L$-function with respect to the character $\psi \omega^{-n}$, where $\omega : F(\mu_{2p}) \to \mathbb{Z}_p^\times$ is the Teichmüller character. Let $H_\psi \in O_\psi[T]$ be defined as $\psi(\gamma)(T + 1) - 1$ if $\psi$ is of type W, and 1 otherwise. Deligne and Ribet showed that there exists a power series $G_{\psi, S}(T) \in O_\psi[[T]]$ so that

$$L_p^S(1 - s, \psi) = \frac{G_{\psi, S}(\kappa(\gamma)^s - 1)}{H_\psi(\kappa(\gamma)^s - 1)},$$

(1.1)

**Documenta Mathematica** 18 (2013) 749–783
where $L^S_p(1 - s, \psi)$ denotes the $p$-adic $L$-function with Euler factors removed at the primes in $S$, and $\kappa$ is the restriction of the cyclotomic character to $\Gamma$. By the Weierstrass Preparation Theorem (cf. §7.1 in [16]) we have the following decomposition:

$$G_{\psi,S}(T) = \pi^u(G_{\psi,S}) g^*_{\psi,S}(T) u_{\psi,S}(T),$$

where $g^*_{\psi,S}(T)$ is a distinguished polynomial in $O_\psi[T]$, and $u_{\psi,S}(T)$ is a unit power series in $O_\psi[[T]]$. This power series represents the analytic object in the Main Conjecture.

For the cyclotomic $\mathbb{Z}_p$-extension $F_{\psi,\infty}$ of $F_\psi$ let $M^S_{\psi,\infty}$ be the maximal abelian pro-$p$-extension of $F_{\psi,\infty}$, which is unramified outside the primes in $S$, with Galois group $\mathcal{X}^S_{\infty} := Gal(M^S_{\psi,\infty}/F_{\psi,\infty})$. The pro-$p$-group $\mathcal{X}^S_{\infty}$ is equipped with a (torsion) $O_\psi[[T]]$-module structure, as well as a $Gal(F_{\psi}/F)$-action given by inner automorphisms. Serre showed that the completed group ring $O_\psi[[T]]$ can be identified with the one variable power series $O_\psi[[T]]$, by mapping $\gamma - 1$ to $T$. By the Structure Theorem of Iwasawa theory (cf. §13.2 in [46]) for $O_\psi[[T]]$-modules, the $\psi$-eigenspace

$$\mathcal{X}^S_{\infty} := \{ x \in \mathcal{X}^S_{\infty} \otimes_{\mathbb{Z}_p} O_\psi | \sigma(x) = \psi(\sigma)x \text{ for all } \sigma \in Gal(F_{\psi}/F) \}$$

of $\mathcal{X}^S_{\infty}$ is pseudo-isomorphic, as an $O_\psi[[T]]$-module, to a unique $O_\psi[[T]]$-module of the form

$$\bigoplus_{i=1}^m O_\psi[[T]]/p_1^{n_1}$$

for $m \geq 1$ and $n_i \geq 1$. Here $p_i$ is the ideal generated by either a fixed uniformizer $\pi \in O_\psi$ or a monic irreducible polynomial in $O_\psi[T]$. We call the ideal $\prod_{i=1}^m p_i^{n_i}$ the characteristic ideal. By the Weierstrass Preparation Theorem (cf. §7.1 in [16]) we can choose a unique generator for the characteristic ideal of the following form

$$F_{\psi,S}(T) = \pi^u(F_{\psi,S}) f^*_{\psi,S}(T),$$

where $f^*_{\psi,S}(T)$ is a distinguished polynomial in $O_\psi[T]$. The polynomial $F_{\psi,S}(T)$ is called the characteristic polynomial of $\mathcal{X}^S_{\infty,\psi}$. The classical Main Conjecture in Iwasawa theory is formulated as follows: If the character $\psi$ is of type $S$, then

$$f^*_{\psi,S}(T) = g^*_{\psi,S}(T).$$

This was proved by Wiles in [47] for any totally real number field $F$ and an odd prime $p$. He also proved the conjecture for the prime 2 and the character $\psi$ provided $F_\psi$ is an abelian extension of $\mathbb{Q}$. He also showed the equality of the $\mu$-invariants $\mu(F_{\psi,S}) = \mu(G_{\psi,S})$ for odd primes $p$ and characters $\psi$. For odd primes $p$, both invariants $\mu(F_{\psi,S})$ and $\mu(G_{\psi,S})$ are known to be zero in the case $F_{\psi}/\mathbb{Q}$ is abelian (cf. [10]), and are conjectured to be zero in general.

The $O_\psi[[T]]$-torsion module $\mathcal{X}^S_{\infty,\psi}$ is of projective dimension at most one and has a principal Fitting ideal generated by the characteristic polynomial.
If we denote by \( G \) the set of primes of \( F \), and let \( \mathbb{Z} \) be the filed of totally real number fields, the Main Conjecture gives an equality of ideals over the power series ring \( \mathbb{O}_E[[T]] \). As we see, for any character \( \psi \) of \( G \), the Galois group of the abelian \( p \)-extension of \( \mathbb{Z}_p \), unramified outside the primes in \( S_f \), over \( \mathbb{F}_\psi,\infty \). For an odd prime \( p \) the two \( \mathbb{O}_E[[T]] \)-modules are the same, since \( \psi \)_primes are unramified in \( p \)-extensions for \( p \) odd. However, for \( p = 2 \) the two modules are related by Lemma 5.9, which shows that they may differ in their \( \mu \)-invariants. If we assume that \( X_{\infty}^f \) has trivial \( \mu \)-invariant, then the analogous formulation of the Main Conjecture in terms of ideals for the prime \( 2 \) reads as follows:

\[
Fitt_{\mathbb{O}_E[[T]]}(X_{\infty}^{S_f}) = (G_{\psi,S}(T)).
\]

To obtain a similar formulation of the Main Conjecture in terms of ideals for the prime \( 2 \), we replace \( X_{\infty}^f \) by \( X_{\infty}^f \), where \( X_{\infty}^f := X_{\infty}^{S_f} \) is the Galois group of the maximal abelian \( p \)-extension of \( \mathbb{F}_\psi,\infty \). In this article the assumption \( \mu = 0 \) refers to the following assumption:

\[
\mu = 0 : \quad \text{The } \mu \text{-invariant of } X_{\infty}^f := X_{\infty}^{S_f} \text{ is zero, i.e. the } \mathbb{Z}_p \text{-module } X_{\infty}^f \text{ is finitely generated.}
\]

We note that \( X_{\infty}^f \) maps by Galois restriction to the Galois group \( \text{Gal}(M_{\psi,\infty}^{S_f}/\mathbb{F}_\psi,\infty) \), where \( M_{\psi,\infty}^{S_f} \) is the maximal abelian \( p \)-extension of \( \mathbb{F}_\psi,\infty \) unramified outside the primes in \( S_f \), with a finite cokernel for any character \( \psi \) of \( G \). Here the cokernel is \( \text{Gal}(\mathbb{F}_\infty \cap M_{\psi,\infty}^{S_f}/\mathbb{F}_\psi) \), which is a
quotient of a subgroup of $H$, and whence finite. As a result the assumption 
$\mu = 0$ implies that the $\mu$-invariant of $Gal(M^{S_f}_{\psi, \infty}/F_{\psi, \infty})$ is zero, and therefore the $\mu$-invariant of $X^{\psi}_{\infty}$ vanishes for any (even) character $\psi$ of $G$.

The pro-$p$ group $X_{\infty}$ has a (torsion) $A$-module structure, whose projective dimension is not necessarily at most one. However to formulate an Equivariant Main Conjecture, similar to the classical Main Conjecture, one needs a finitely generated $A$-torsion module of projective dimension at most one. Let $d_{\infty}$ be a non-zero divisor of the augmentation ideal $\Delta G_{\infty}$ of $A$, let $c_{\infty}$ be an invertible element of the total ring of fraction of $A$, so that $d_{\infty} = c_{\infty}((\gamma - 1)e + (1 - e))$, where $e$ is the idempotent attached to the trivial character of $H$. We denote by $L$ the fixed field of $E/F$ under the action of the $p$-Sylow subgroup of $G$, by $G$ the Galois group of the maximal algebraic extension $\Omega^S_L$ of $L$ unramified outside the primes in $S$, over $F$, and by $H$ the Galois group of $\Omega^S_L/E_{\infty}$. There is a commutative diagram of $A$-modules

\[
\begin{array}{cccccc}
0 & 0 & 0 & 0 & 0 \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
X_{\infty} & Y_{\infty} & \Delta G_{\infty} & 0 & 0 \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
Z_{\infty} & z_{\infty} & 0 & 0 & 0 \\
\end{array}
\]

where $\psi$ maps 1 to $d_{\infty}$, $\Psi$ maps 1 to a pre-image $y_{\infty}$ of $d_{\infty}$, and $Y_{\infty} = H_{\infty}(H, \Delta G)$. Here $\Delta G$ denotes the augmentation ideal of $\mathbb{Z}_p[[G]]$. We will see that this definition of $Y_{\infty}$ is the same as the definition of Ritter-Weiss in [35]. The $A$-torsion module $Z_{\infty}$ in the diagram above, whose projective dimension is at most one, shows up as the algebraic object in the Equivariant Main Conjecture of Ritter-Weiss. We note that the construction of $Z_{\infty}$ depends on the choice of $d_{\infty}$. Before stating the algebraic object we remark that there exists a subgroup $\Gamma \leq G_{\infty}$, topologically generated by $\gamma$, so that $G_{\infty} = H \times \Gamma$ for the abelian group $G_{\infty}$. The analytic object is defined as follows:

$$G_S := \sum_{\psi \in H} G_{\psi, \psi}(\gamma - 1) \cdot e_\psi \in \frac{1}{|H|} O[H][[\Gamma]],$$

where $e_\psi$ is the idempotent attached to the character $\psi$ of $H$, i.e.

$$e_\psi := \frac{1}{|H|} \sum_{\sigma \in G} \psi(\sigma)\sigma^{-1}.$$

One version of the Equivariant Main Conjecture of Ritter-Weiss [35] for odd primes is as follows (cf. [28], §2, (CPE2)):

$$Fitt_A(Z_{\infty}) = (c_{\infty}G_S),$$

Documenta Mathematica 18 (2013) 749–783
which was verified under the assumption of the vanishing of the $\mu$-invariant of $X_\infty$, i.e. assuming that $X_\infty$ is a finitely generated $\mathbb{Z}_p$-module. It is worth mentioning that they have generalized and proved their Equivariant Main Conjecture in the non-commutative case, still assuming the vanishing of the $\mu$-invariant of a certain Iwasawa module (see [36]). In [15], Greither and Popescu have recently formulated and proved an Equivariant Main Conjecture in Iwasawa theory in the abelian case in terms of the Tate module of a certain Iwasawa-theoretic abstract 1-motive again under the assumptions $\mu = 0$ and $p$ odd. More recently, Nickel [29] showed that this formulation is equivalent to the formulation of Ritter-Weiss.

We now describe our Equivariant Main Conjecture for an arbitrary prime $p$. For an abelian extension $E/F$, by applying the algebraic construction of the Equivariant Main Conjecture of Ritter-Weiss to the set $S_f$ of finite primes in $S$, we construct the $A$-torsion module $Z_{f,\infty}$, which is of projective dimension at most one. We show that it satisfies the following exact sequence:

$$0 \to \mathbb{Z}_p \to (A/d_\infty A)^\# \to \alpha(Z_{f,\infty}) \to \alpha(X_{f,\infty}) \to 0,$$

in which $pd_A((A/d_\infty A)^\#) \leq 1$ and $pd_A(\alpha(Z_{f,\infty})) \leq 1$.

Here $\alpha$ is the adjoint functor in Iwasawa theory with the contravariant action and $M^\#$, for any $A$-module $M$, denotes the same underlying module but with $g$ acting as $g^{-1}$ for any $g \in G_\infty$. The Equivariant Main Conjecture is then formulated as follows (cf. Conjecture 4.1):

$$Fitt_A(Z_{f,\infty}) = (c_\infty G_S^*),$$

where

$$G_S^* := \sum_{\psi \in H} G_{\psi,S}^*(\gamma - 1) \cdot e_\psi \in \frac{1}{|H|} \mathcal{O}[H][[\Gamma]].$$

Here we recall that $G_{\psi,S}^*(T) = g_{\psi,S}(T)u_{\psi,S}(T)$. In Section 4, we prove that this conjecture follows from the classical Main Conjecture under the assumption $\mu = 0$ by taking advantage of the idea of determinantal ideals used by Greither and Popescu [15].

In the last section we show that the Coates-Sinnott Conjecture follows from the Equivariant Main Conjecture assuming $\mu = 0$ (cf. Theorem 5.10). After some fundamental work of Coates-Sinnott in [6] and more recent results by Ritter-Weiss, Nguyen Quang Do, Burns-Greither, Greither-Popescu et al. the Coates-Sinnott Conjecture is completely known up to powers of 2, assuming $\mu = 0$. However, the 2-primary information was neglected more or less completely due to various technical problems. For example, there was no formulation of an Equivariant Main Conjecture in Iwasawa theory for the prime 2 at the time.
An Equivariant Main Conjecture

The Coates-Sinnott Conjecture is a generalization of the classical Stickelberger Theorem, which provides elements annihilating the class group of a cyclotomic field, using special values of analytic functions. To make it more precise, let $E/F$ be an abelian extension with Galois group $G$, and let $S$ be a finite set of primes in $F$ containing the primes ramified in $E$ and the infinite primes. Let

$$\Theta_{E/F}^S(s) := \sum_{\chi \in \hat{G}} L_{E/F}^S(s, \chi^{-1}) \cdot e_\chi \in \mathbb{C}[G]$$

be the $S$-incomplete equivariant $L$-function, where $e_\chi$ is the idempotent attached to any character $\chi$ of $G$. Deligne and Ribet [8] and independently Cassou-Noguès [4] proved that

$$\text{Ann}_{\mathbb{Z}[G]}(H^0(E, \mathbb{Q}/\mathbb{Z}(n))) \cdot \Theta_{E/F}^S(1 - n) \subseteq \mathbb{Z}[G]$$

for any integer $n \geq 1$. Stickelberger’s Theorem shows that the following analytic object is in the annihilator ideal of the class group $\text{Cl}(\mathcal{O}_E)$ of the field $E$ in the case $F = \mathbb{Q}$:

$$\text{Ann}_{\mathbb{Z}[G]}(H^0(E, \mathbb{Q}/\mathbb{Z}(1))) \cdot \Theta_{E/F}^S(0) \subseteq \text{Ann}_{\mathbb{Z}[G]}(\text{Cl}(\mathcal{O}_E)).$$

This setup has been generalized in two directions: First of all one looks at an arbitrary relative abelian extension $E/F$ of number fields. Here the analogue of Stickelberger’s theorem (Brumer’s Conjecture) is still not completely known. In a different direction one replaces the class group by algebraic $K$-groups or motivic cohomology groups and studies annihilators of these groups as Galois modules for relative abelian extensions. In [6], Coates and Sinnott formulated the relevant conjecture in terms of higher Quillen $K$-groups as

$$\text{Ann}_{\mathbb{Z}[G]}(H^0(E, \mathbb{Q}/\mathbb{Z}(n))) \cdot \Theta_{E/F}^S(1 - n) \subseteq \text{Ann}_{\mathbb{Z}[G]}(K_{2n - 2}(\mathcal{O}_E))$$

for any integer $n \geq 2$. As a result of the recent work of Voevodsky in [45] the relation between algebraic $K$-theory, étale cohomology for all prime numbers and motivic cohomology is known. This yields the motivic formulation of the Coates-Sinnott Conjecture, which implies the $K$-theoretic version. Moreover, it enables us to study each $p$-primary part of the conjecture separately for any prime number $p$ as follows:

$$\text{Ann}_{\mathbb{Z}_p[G]}(H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n))) \cdot \Theta_{E/F}^S(1 - n) \subseteq \text{Ann}_{\mathbb{Z}_p[G]}(H_{\text{ét}}^2(\mathcal{O}_E', \mathbb{Z}_p(n)))$$

for any integer $n \geq 2$, where $H_{\text{ét}}^2(\mathcal{O}_E', \mathbb{Z}_p(n)) = \lim_{\rightarrow m} H_{\text{ét}}^2(\mathcal{O}_E', \mu_p^m)$ and $H_{\text{ét}}^2(\mathcal{O}_E', \mu_p^m)$ refers to the étale cohomology of the scheme $\text{Spec}(\mathcal{O}_E[1/p])$ with values in the étale sheaf $\mu_p^m$. In the last section we complete the proof of the Coates-Sinnott Conjecture under the assumption $\mu = 0$ by proving it for the prime 2.
Acknowledgement. The material in this article forms a part of the author’s Ph.D. thesis [43] at McMaster university. The author would like to express his gratitude to his Ph.D. supervisor Manfred Kolster for introducing these subjects, for his encouragement and helpful discussions and suggestions. The author would also like to thank the referee for his detailed comments and suggestions.

2 Algebraic Construction

Let $E/F$ be a finite abelian extension of totally real number fields with Galois group $G$, and let $p$ be an arbitrary prime. Let $E_\infty$ (resp. $F_\infty$) be the cyclotomic $\mathbb{Z}_p$-extension of $E$ (resp. $F$). We denote the multiplicative group $\text{Gal}(E_\infty/E)$ (resp. $\text{Gal}(F_\infty/F)$) by $\Gamma_E$ (resp. $\Gamma_F$). Let $H$ denote the Galois group of the finite abelian extension $E_\infty/F_\infty$. We denote by $G_\infty := \text{Gal}(E_\infty/F)$ the Galois group of the abelian extension $E_\infty/F$. We let $S$ denote a finite set of primes of $F$ which ramify in $E_\infty$, and the infinite primes. In particular $S$ contains the set $S_p$ of the primes above $p$. The set of finite primes in $S$ is also denoted by $S_f$. We use the same notations for the set of primes above the primes in $S$ and $S_f$, respectively, in any intermediate field of $E_\infty/F$. Since $\Gamma_F$ is topologically generated by one element, the exact sequence

$$0 \rightarrow H \rightarrow G_\infty \xrightarrow{\imath} \Gamma_F \rightarrow 0$$

(2.1)

splits. We denote by $\Gamma \leq G_\infty$ the image of $\Gamma_F$, so that $G_\infty \simeq H \times \Gamma$, and by $\Lambda$ the completed group ring $\mathbb{Z}_p[[\Gamma]]$. Let $E'$ be the fixed field of $E_\infty$ under the action of the closed subgroup $\Gamma$. Then $E' \cap F_\infty = F$, $E_\infty = E' \cdot F_\infty$, $\text{Gal}(E'/F) \simeq H$ and $E_\infty/E'$ is also a cyclotomic $\mathbb{Z}_p$-extension.

Let $M_\infty^S$ and $M_\infty^{S_f}$ be the maximal abelian pro-$p$-extensions of $E_\infty$ unramified outside the primes in $S$ and $S_f$, respectively. We recall that $\mathcal{X}_\infty = \mathcal{X}_\infty^S$ and $\mathcal{X}_\infty^{S_f} := \mathcal{X}_\infty^S/\mathcal{X}_\infty^{S_f}$ denote the Galois group of the extensions $M_\infty^S/E_\infty$ and $M_\infty^{S_f}/E_\infty$, respectively. Since $E$ is totally real, the $\Lambda$-module $\mathcal{X}_\infty$ is a torsion module with no non-trivial finite submodule by Propositions 10.3.22 and 10.2.25 in [31]. The $\Lambda$-module $\mathcal{X}_\infty^{S_f}$, which is a quotient of $\mathcal{X}_\infty$, is also torsion and has no non-trivial finite submodule (cf. [38], §6.4). Finally, we set $\mathcal{A} := \mathbb{Z}_p[[G_\infty]]$ and we freely use the identification

$$\mathcal{A} \simeq \mathbb{Z}_p[H][[T]],$$

(2.2)

which is given by mapping the topological generator $\gamma$ of $\Gamma$ to $1 + T$.

Remark 2.1. For an odd prime $p$, infinite primes of $F$ are unramified in a pro-$p$-extension. Hence $M_\infty^S$ and $M_\infty^{S_f}$ coincide and therefore, $\mathcal{X}_\infty = \mathcal{X}_\infty^{S_f}$ for odd primes.
The following diagram illustrates the situation:

We recall that the Main Conjecture in Iwasawa theory for a character \( \psi \) of \( G \) can be written in the form (1.2), assuming \( \mu = 0 \). Here the Fitting ideal of the (finitely generated) \( \mathcal{O}_\psi[[T]] \)-torsion module \( \mathcal{X}^I,\psi \) is principal, because it has projective dimension at most one, and is generated by the \( p \)-adic \( L \)-function associated to \( \psi \). Hence to formulate an Equivariant Main Conjecture we construct an appropriate (finitely generated) \( \mathcal{H} \)-torsion module of projective dimension at most one. The resulting Fitting ideal is then principal, and conjecturally generated by an equivariant \( p \)-adic \( L \)-function. This was done by Ritter-Weiss for odd primes in [35]. The strategy of this part is as follows: Since the \( \mathcal{H} \)-torsion module \( \mathcal{X}_f \) is not necessarily of projective dimension at most one, we first construct an \( \mathcal{H} \)-module \( \mathcal{Y}_f \) of projective dimension at most one. Since this module is not necessarily \( \mathcal{H} \)-torsion, we pass to a quotient \( \mathcal{Z}_f \), which is then shown to be a (finitely generated) \( \mathcal{H} \)-torsion module of projective dimension at most one, whose principal Fitting ideal is conjecturally generated by an equivariant \( L \)-function.

Let \( P \) be the \( p \)-Sylow subgroup of \( G \) and let \( L \) be the fixed field of \( E \) under the action of \( P \) with Galois group \( Q \) over \( F \). Let \( \Omega^{S_f}_L \) be the maximal algebraic pro-\( p \)-extension of \( L \), which is unramified outside the primes in \( S_f \). We denote by \( \mathcal{H} \) the Galois group of \( \Omega^{S_f}_L \) over \( E_\infty \), and by \( \mathcal{G} \) the Galois group of \( \Omega^{S_f}_L \) over \( F \). The finitely generated group \( \mathcal{G} \) has a presentation of the form \( \mathcal{G} \simeq \mathcal{F}/\mathcal{W} \), where \( \mathcal{F} \) is an appropriate free profinite group of rank \( d \) and \( \mathcal{W} \) is a relation subgroup of \( \mathcal{F} \) of rank \( r \). For a certain relation subgroup \( \mathcal{R} \) of \( \mathcal{F} \) we then have an isomorphism \( \mathcal{G}_\infty \simeq \mathcal{F}/\mathcal{R} \). The following diagram illustrates the situation:
We apply Proposition 5.6.7 in [31] (see also Lemma 4.3 in [17]) to the profinite groups in the commutative diagram

\[
\begin{array}{ccc}
1 & 1 \\
\uparrow & \uparrow \\
W & = & W \\
\downarrow & \downarrow \\
1 & \to & R \\
\downarrow & \downarrow \\
1 & \to & H \\
\downarrow & \downarrow \\
1 & \to & G & \to & G_{\infty} & \to & 1 \\
\end{array}
\]

and obtain a commutative diagram:

\[
\begin{array}{cccccc}
0 & 0 \\
\downarrow & \downarrow \\
0 & \to & H_2(H, \mathbb{Z}_p) & \to & H_0(H, \mathcal{W}_{ab}) & \to & \mathcal{R}_{ab}(p) & \to & X_{\infty}^f & \to & 0 \\
\| & \| & \downarrow & \downarrow \\
0 & \to & H_2(H, \mathbb{Z}_p) & \to & H_0(H, \mathcal{W}_{ab}) & \to & \mathfrak{A}_d & \to & Y_{\infty}^f & \to & 0 \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\Delta G_{\infty} & = & \Delta G_{\infty} \\
0 & 0,
\end{array}
\]

(2.3)

where $\Delta G_{\infty}$ denotes the augmentation ideal of $A$. Here $Y_{\infty}^f := H_0(H, \Delta \mathcal{G})$ and $\Delta \mathcal{G}$ and $\Delta G_{\infty}$ denote the augmentation ideals of $\mathcal{G}$ and $G_{\infty}$, respectively.

**Remark 2.2.** The same construction leads to a similar diagram for an arbitrary intermediate field of $\Omega_S^f/L$.

Since the cyclotomic $\mathbb{Z}_p$-extension $E_{\infty}/E$ satisfies the weak Leopoldt Conjecture by Proposition 10.3.25 in [31], the group $H_2(H, \mathbb{Z}_p)$ in diagram (2.3) vanishes (cf. Proposition 10.3.22 in [31]). Moreover we have the following proposition:

**Proposition 2.3.** The $\mathfrak{A}$-module $H_0(H, \mathcal{W}_{ab})$ is projective.

**Proof.** Since $|Q|$ is prime to $p$, we have the equality of cohomological dimensions

$$cd_p(\mathcal{G}) = cd_p(Gal(\Omega^{S_f}_L/L)).$$

The $p$-cohomological dimension of the pro-$p$ group $Gal(\Omega^{S_f}_L/L)$ is at most 2 by Proposition 8.3.17 in [31] for odd primes $p$ (note that infinite primes are unramified in any $p$-extension for $p$ odd), and by Theorem 1 in [38] for $p = 2$, i.e.

$$cd_p(\mathcal{G}) \leq 2$$

for any prime $p$. Now Proposition 5.6.7 in [31] completes the proof. \qed

**Remark 2.2.** The same construction leads to a similar diagram for an arbitrary intermediate field of $\Omega_S^f/L$. 

Since the cyclotomic $\mathbb{Z}_p$-extension $E_{\infty}/E$ satisfies the weak Leopoldt Conjecture by Proposition 10.3.25 in [31], the group $H_2(H, \mathbb{Z}_p)$ in diagram (2.3) vanishes. (cf. Proposition 10.3.22 in [31]). Moreover we have the following proposition:

**Proposition 2.3.** The $\mathfrak{A}$-module $H_0(H, \mathcal{W}_{ab})$ is projective.

**Proof.** Since $|Q|$ is prime to $p$, we have the equality of cohomological dimensions

$$cd_p(\mathcal{G}) = cd_p(Gal(\Omega^{S_f}_L/L)).$$

The $p$-cohomological dimension of the pro-$p$ group $Gal(\Omega^{S_f}_L/L)$ is at most 2 by Proposition 8.3.17 in [31] for odd primes $p$ (note that infinite primes are unramified in any $p$-extension for $p$ odd), and by Theorem 1 in [38] for $p = 2$, i.e.

$$cd_p(\mathcal{G}) \leq 2$$

for any prime $p$. Now Proposition 5.6.7 in [31] completes the proof. \qed
Let $\chi$ be a $C_p$-valued character of the group $Q$, let

$$e_\chi := \frac{1}{|Q|} \sum_{\sigma \in Q} \chi(\sigma)\sigma^{-1}$$

be the idempotent of $Q$ attached to the character $\chi$, and let $A_\chi := O_\chi[[G_\infty(p)]]$, where $O_\chi$ is the ring obtained by adjoining all character values of $\chi$ to $\mathbb{Z}_p$. Since $G_\infty(p) \simeq P \times \Gamma$ is a pro-$p$ group, $A_\chi$ is a local ring and therefore, $e_\chi H_0(H, W^ab)$ is a free $A_\chi$-module of rank $r_\chi$:

$$e_\chi H_0(H, W^ab) \simeq A_\chi^{r_\chi}. \tag{2.4}$$

Now by applying $e_\chi$ to the exact sequence in the second row of diagram (2.3) we obtain:

$$0 \to A_\chi^{r_\chi} \to A_\chi^d \to e_\chi Y_f^\infty \to 0.$$

From the last column of diagram (2.3) we have

$$0 \to e_\chi X_f^\infty \to e_\chi Y_f^\infty \to A_\chi \to e_\chi \mathbb{Z}_p \to 0.$$

This implies that $e_\chi Y_f^\infty$ has rank one and as a result $r_\chi = d - 1$ for any character $\chi$. Now by taking the direct sum over all characters of $Q$ in equality (2.4) we obtain:

$$H_0(H, W^ab) \simeq A^r$$

for $r = d - 1$. Therefore, diagram (2.3) can be rewritten as

\[
\begin{array}{cccccc}
0 & \to & A^r & \xrightarrow{f} & \mathcal{R}^{ab}(p) & \to \mathcal{X}_\infty^f & \to 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \to & \mathcal{A}^{r+1} & \xrightarrow{g} & \mathcal{Y}_\infty^f & \to & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
\Delta G_\infty & = & \Delta G_\infty & \to & 0 & .
\end{array}
\tag{2.5}
\]

So far we have constructed the module $Y_\infty^f$, which fits into the exact sequence

$$0 \to \mathcal{X}_\infty^f \to \mathcal{Y}_\infty^f \to \Delta G_\infty \to 0. \tag{2.6}$$

The second row in diagram (2.5) implies that the $A$-module $Y_\infty^f$ is of projective dimension at most one. To replace it by a torsion $A$-module we now take a quotient of $Y_\infty^f$ by a certain submodule as follows:

Let $d_\infty \in \Delta G_\infty$ be a non-zero divisor in the augmentation ideal of $A$ and let $c_\infty$ be an invertible element in $Q(A)$ such that

$$d_\infty = c_\infty((\gamma - 1)e + (1 - e)), \tag{2.7}$$
where \( \gamma \) is the fixed (topological) generator of \( \Gamma \leq G_\infty \) and \( e = \frac{1}{|H|} \sum_{h \in H} h \) is the idempotent of \( \mathbb{Q}_p[H] \) attached to the trivial character of \( H \). Here we note that \( \gamma - 1 \) and \( 1 - e \) generate \( \Delta G_\infty \otimes \mathbb{Q}_p \), and that \( \gamma - 1 \) and \( 1 - e \) can be written in the form (2.7) as follows:

\[
\begin{align*}
\gamma - 1 &= (e + (\gamma - 1)(1 - e))((\gamma - 1)e + (1 - e)), \\
1 - e &= (1 - e)((\gamma - 1)e + (1 - e)).
\end{align*}
\]

Let \( y_\infty \) be a pre-image of \( d_\infty \) in \( Y_\infty \) in diagram (2.5). We have the following diagram:

\[
\begin{array}{ccccccccc}
0 & 0 & 0 \\
\downarrow & \downarrow & \downarrow \\
A & = & A \\
\downarrow \Phi & \downarrow \phi \\
0 \to X_\infty^f & \to Y_\infty^f & \to \Delta G_\infty & \to 0 \\
\downarrow & \downarrow & \downarrow & \downarrow \\
0 \to X_\infty^f & \to Z_\infty^f & \to z_\infty^f & \to 0
\end{array}
\]

(2.8)

where \( \Phi \) and \( \phi \) are defined by mapping \( 1 \in A \) to \( y_\infty \) and to \( d_\infty \), respectively, and \( Z_\infty^f \) and \( z_\infty^f \) are the quotients of \( Y_\infty^f \) and \( \Delta G_\infty \) by the images of \( \Phi \) and \( \phi \), respectively. We note that the vertical maps are injective since \( d_\infty \in A \) is a non-zero divisor. By a diagram chase in the diagram

\[
\begin{array}{ccccccccc}
A & = & A \\
\downarrow \phi & \downarrow \phi \\
0 \to \Delta G_\infty & \to A & \to Z_p & \to 0,
\end{array}
\]

we obtain:

**Lemma 2.4.** The sequence

\[
0 \to z_\infty^f \to A/d_\infty A \to Z_p \to 0
\]

(2.9)

is exact, where the middle term is of projective dimension one and

\[
\text{Fitt}_A(A/d_\infty A) = (d_\infty).
\]

By using the middle column of diagram (2.8) and the first row of diagram (2.5) we also obtain the commutative diagram

\[
\begin{array}{ccccccccc}
0 & 0 & 0 \\
\downarrow & \downarrow & \downarrow \\
\mathcal{A} & = & \mathcal{A} \\
\downarrow \Psi & \downarrow \phi \\
0 \to \mathcal{A}_r & \to \mathcal{A}_r+1 & \to \mathcal{A} & \to 0 \\
\downarrow & \downarrow & \downarrow & \downarrow \\
0 \to \mathcal{R}^{ab}(p) & \to \mathcal{A}_r+1 & \to \Delta G_\infty & \to 0 \\
\downarrow & \downarrow & \downarrow & \downarrow \\
0 \to X_\infty^f & \to Z_\infty^f & \to z_\infty^f & \to 0
\end{array}
\]

(2.10)
which implies the following proposition:

**Proposition 2.5.** $Z_f^\infty$ is a finitely generated $A$-torsion module of projective dimension at most one:

$$\text{pd}_A(Z_f^\infty) \leq 1.$$  

3 Analytic Construction

Again let $E/F$ be an abelian extension of totally real fields, let $S$ be a finite set of primes in $F$ containing the primes above $p$, the primes ramified in $E$ and the infinite primes, and let $S_f$ denote the set of finite primes in $S$. As before, we denote by $E_\infty$ and $F_\infty$ the cyclotomic $\mathbb{Z}_p$-extensions of $E$ and $F$, respectively, by $H$ the Galois group of $E_\infty/F_\infty$, and by $G_\infty \cong H \times \Gamma$ the Galois group of $E_\infty/F$. We define equivariant versions of $G_{\psi,S}$ and $H_\psi$ as follows (cf. [34], Proposition 5.4): For a character $\psi$ of $G_{\infty}$, let $G_{\psi,S}^*(T), H_\psi(T) \in \mathcal{O}_{\psi}[\![T]\!]$ be the power series defined in (1.1). Let

$$G_S := \sum_{\psi \in \hat{H}} G_{\psi,S}(\gamma - 1) \cdot e_\psi \in \left[\frac{1}{|H|}\mathcal{O}[H][[\Gamma]]\right]$$

$$H_S := \sum_{\psi \in \hat{H}} H_\psi(\gamma - 1) \cdot e_\psi \in \left[\frac{1}{|H|}\mathcal{O}[H][[\Gamma]]\right]$$

be the equivariant versions of $G_{\psi,S}$ and $H_\psi$. For any character $\chi$ of $G_{\infty}$, they satisfy the following:

$$\chi(G_S) = G_{\chi,S}(0) \quad , \quad \chi(H_S) = H_\chi(0).$$

We recall that for any character $\psi$ of $G$ one has

$$G_{\psi,S}(T) = \pi^{\mu(G_{\psi,S})} \cdot g_{\psi,S}^*(T) \cdot u_{\psi,S}(T)$$

by the Weierstrass Preparation Theorem, where $\pi$ is a fixed uniformizer in $\mathcal{O}_\psi$, $g_{\psi,S}^*(T) \in \mathcal{O}_{\psi}[\![T]\!]$ is a distinguished polynomial, and $u_{\psi,S}(T) \in \mathcal{O}_\psi[\![T]\!]$ is a unit. The modified equivariant $L$-function $G_S^*$ is now defined as follows:

$$G_S^* := \sum_{\psi \in \hat{H}} G_{\psi,S}^*(\gamma - 1) \cdot e_\psi \in \left[\frac{1}{|H|}\mathcal{O}[H][[\Gamma]]\right],$$

where $G_{\psi,S}^*(T) = g_{\psi,S}^*(T)u_{\psi,S}(T)$. The following lemma relates $G_S$ and $G_S^*$, assuming $\mu = 0$ (cf. (1.3)):

**Lemma 3.1.** Under the assumption $\mu = 0$ we have the following equalities:


Documenta Mathematica 18 (2013) 749–783
2. \( G_S = 2^{\nu_1(F)}G^*_S \) for \( p = 2 \).

Proof. Part 1 follows from the result of Wiles [47] that \( \mu(G_{\psi, S}) \) is the same as the Iwasawa \( \mu \)-invariant of \( X_{\infty}^L \psi \) for all odd primes \( p \). Part 2 follows from the fact that \( \pi \mu(G_{\psi, S}) = 2^{\nu_1(F)} \) under the assumption \( \mu = 0 \), where \( \pi \in O_\psi \) is a uniformizer, for any character \( \psi \) of \( G \) (see [12], pages 82 and 87).

We note that Lemma 3.1 holds unconditionally for abelian extensions \( E \) of \( \mathbb{Q} \), since in this case \( \mu = 0 \) (cf. [10]).

To prove the next lemma we briefly review the definition of a \( p \)-adic pseudo-measure on a certain Galois group and its relation to the \( p \)-adic \( L \)-function. For more properties one can consult [39]. For a commutative profinite group \( \mathcal{G} \), \( \lambda_S \in Q(\mathbb{Z}_p[[\mathcal{G}]]) \) is called a pseudo-measure on \( \mathcal{G} \) if \((g-1)\lambda_S \) is a measure, i.e. \((g-1)\lambda_S \in \mathbb{Z}_p[[\mathcal{G}]]\), for any \( g \in \mathcal{G} \), where \( Q(\mathbb{Z}_p[[\mathcal{G}]]) \) denotes the quotient ring of \( \mathbb{Z}_p[[\mathcal{G}]] \). Let \( \mathcal{X} \) denote the Galois group of the maximal abelian extension of \( F \) unramified outside the primes in \( S_F \) over \( F \). By a theorem of Deligne and Ribet there is a unique pseudo-measure on \( \mathcal{X} \) denoted by \( \lambda_S \in Q(\mathbb{Z}_p[[\mathcal{X}]]) \), which satisfies the following relation for any finite order character \( \chi \) of \( \mathcal{X} \):

\[
L_{p, S}(1 - s; \chi) = \langle \chi h^e, \lambda_S \rangle.
\]

For the definition of this inner product see [39]. Equivalently, if we let \( \varepsilon : \mathcal{X} \to \mathbb{Z}_p \) be the locally constant function defined by \( \varepsilon(g) = 1 \) if \( g \) has image 1 in \( H = \text{Gal}(\mathbb{E}_\infty / F_\infty) \), and zero otherwise, then

\[
\zeta_p^S(\varepsilon_h, 1 - n) = \langle \varepsilon_h h^n, \lambda_S \rangle.
\]

Here \( \rho \) is the cyclotomic character, \( \varepsilon_h \) is the locally constant function satisfying \( \varepsilon_h(x) = \varepsilon(hx) \) and \( \zeta_p^S(\varepsilon_h, s) \) is the \( S \)-incomplete \( p \)-adic partial zeta function associated to \( \varepsilon_h \). The image of \( \lambda_S \) under the natural surjection \( \pi : \mathcal{X} \to G_\infty \) is a \( p \)-adic pseudo-measure on \( G_\infty \) which is denoted by \( \theta_S \in Q(\mathbb{A}) \). So if \( \hat{\gamma} \in \mathbb{X} \) denotes a pre-image of \( \gamma \in G_\infty \) under the surjection above, then

\[
Z_S := \pi((\hat{\gamma} - 1)\lambda_S) \in \mathbb{A}.
\]

In fact \( \theta_S = (\gamma - 1)^{-1}Z_S \in Q(\mathbb{A}) \). With notations as above we have the following lemma:

**Lemma 3.2.** Let \( d_\infty \) be a non-zero divisor in the augmentation ideal \( \Delta G_\infty \), and let \( c_\infty \) be an invertible element in \( Q(\mathbb{A}) \) so that \( d_\infty = c_\infty((\gamma - 1)e + (1 - e)) \in \Delta G_\infty \), where \( e \) is the idempotent associated to the trivial character of \( H \). Then

\[
c_\infty G_S = d_\infty \theta_S \in \mathbb{A}.
\]

Proof. A calculation in Proposition 12 in [35], which works for any character \( \chi \) of \( G_\infty \) satisfying \( \chi(\gamma) = 1 \) and for any prime \( p \), provides the following equality:

\[
\frac{G_{\chi, S}(T)}{T^{\chi(1)}} = \sum_{h \in H} \chi(h) \frac{Z_S(h, T)}{T}
\]
Here $Z_S(h, T)$ is given by the relation $Z_S = \sum_{h \in H} Z_S(h, \gamma - 1) h \in A$. As a result,
\[
G_S / H_S = \sum_{\chi \in H} G_{\chi, S}(T) \mathbb{F}_{\chi, T} e_{\chi} = \theta_S,
\]
for any prime $p$. Since $H_S = (\gamma - 1)e + (1 - e)$, we obtain $c_{\infty} G_S = d_{\infty} \theta_S$. Hence, for the $p$-adic pseudo-measure $\theta_S$ on $G_{\infty}$, we have $c_{\infty} G_S = d_{\infty} \theta_S \in \mathbb{A}$. 

Let $\bar{E} = E(\zeta_{2p})$ be the field obtained by adjoining a primitive $2p$-th root of unity $\zeta_{2p}$ to $E$, and let $\bar{E}_{\infty} := E_{\infty}(\zeta_{2p}) = E(\mu_{p^{\infty}})$ be the $\mathbb{Z}_p$-cyclotomic extension of $\bar{E}$, where $\mu_{p^{\infty}}$ is the group of all $p$-power roots of unity. We denote by $\tilde{G}_{\infty}$ the Galois group of $\bar{E}_{\infty}/F$. Since $\bar{E}_{\infty}$ contains all $p$-power roots of unity, we have the cyclotomic character
\[
\rho : \tilde{G}_{\infty} \to \mathbb{Z}_p^* = \text{Aut}(\mu_{p^{\infty}})
\]
of $\tilde{G}_{\infty}$. We extend the definitions of a Tate twisted module and an inverse module to the following:

- Let $t_n$ be the unique continuous isomorphism of $\mathcal{O}$-algebras
\[
t_n : \mathcal{O}[[\tilde{G}_{\infty}]] \to \mathcal{O}[[\tilde{G}_{\infty}]],
\]
which satisfies $t_n(g) = \rho(g)^n \cdot g$ for all $g \in \tilde{G}_{\infty}$ and $n \in \mathbb{Z}$. For a $\mathcal{O}[[\tilde{G}_{\infty}]]$-module $M$ let the Tate twisted module $M(n)$ be the same underlying group $M$ with a new $\mathcal{O}[[\tilde{G}_{\infty}]]$-action given by $\sigma * m := t_n(\sigma)m$ for $\sigma \in \mathcal{O}[[\tilde{G}_{\infty}]]$ and $m \in M$. For even $n$, we note that it is enough to replace $\bar{E}_{\infty}$ by its maximal real subfield $\tilde{E}_{\infty}^+$ to have the isomorphisms $t_n$.

- Let $\iota$ be the unique continuous isomorphism of $\mathcal{O}$-algebras
\[
\iota : \mathcal{O}[[G_{\infty}]] \to \mathcal{O}[[G_{\infty}]],
\]
which satisfies $\iota(g) = g^{-1}$ for all $g \in G_{\infty}$. For a $\mathcal{O}[[G_{\infty}]]$-module $M$, let the inverse module $M^\#$ be the same underlying group $M$ with a new $\mathcal{O}[[G_{\infty}]]$-action given by $\sigma * m := \iota(\sigma)m$ for $\sigma \in \mathcal{O}[[G_{\infty}]]$ and $m \in M$. In the following we mean by the ideal generated by $m^\#$, for any $m \in \mathcal{O}[[G_{\infty}]]$, the inverse ideal $(m)^\#$ of $(m)$.

**Lemma 3.3.** Assume that $E$ is the maximal real subfield of $\bar{E} = E(\zeta_{2p})$ where $\zeta_{2p}$ is a primitive $2p$-th root of unity. For all even $n$ we have:
\[
(\iota \circ t_n)(G_S) = \sum_{\psi \in H} G_{\psi - 1, \omega^n, S}(u^n(\gamma)^{-1} - 1) \cdot e_{\psi},
\]
\[
(\iota \circ t_n)(H_S) = \sum_{\psi \in H} H_{\psi - 1, \omega^n, S}(u^n(\gamma)^{-1} - 1) \cdot e_{\psi},
\]
where $u = \kappa(\gamma)$.
Proof. Under the assumptions of the lemma we first note that $E_\infty$ is the maximal real subfield of the field $E_\infty(\zeta_{2p})$ which contains all $p$-power roots of unity. So the actions of $t_n$ on $G_S$ and $H_S$ are defined for all even $n$. Now it suffices to observe that $(\iota \circ t_n)(\gamma - 1) = u^n \gamma^{-1} - 1$ and $(\iota \circ t_n)(e_\psi) = e_{\psi^{-1} \omega^n}$.

This lemma yields the following equality:

$$(\pi \circ \iota \circ t_n)G_S / H_S = \sum_{\chi \in G} \frac{G_{\chi^{-1} \omega^n, S}(u^n - 1)}{H_{\chi^{-1} \omega^n, S}(u^n - 1)} \cdot e_{\chi},$$

where $\pi : A \to \mathbb{Z}_p[G]$ is the projection mapping $\gamma - 1$ to zero, and $u = \kappa(\gamma)$. Therefore, we obtain:

**Corollary 3.4.** If we assume that $E$ is the maximal real subfield of $\bar{E} = E(\zeta_{2p})$ where $\zeta_{2p}$ is a primitive 2$p$-th root of unity, and that $n$ is even, then

$$(\pi \circ \iota \circ t_n)G_S / H_S = \Theta_{E/F}^S(1 - n)$$

**Remark 3.5.** We note that in the non-dyadic $L$-functions we have defined, the set $S$ can be replaced by $S_f$, since infinite primes have no influence on the definitions.

4 An Equivariant Main Conjecture in Iwasawa Theory

We recall that for the abelian extension $E/F$ of totally real number fields, $E_\infty$ (resp. $F_\infty$) is the cyclotomic $\mathbb{Z}_p$-extension of $E$ (resp. $F$), $H = \text{Gal}(E_\infty/F_\infty)$, and $G_\infty = \text{Gal}(E_\infty/F)$, which is abelian and hence of the form $G_\infty = H \times \Gamma$ for $\Gamma \simeq \mathbb{Z}_p$. We also recall that $d_\infty = e_\infty((\gamma - 1)e + (1 - e))$ is a non-zero divisor in the augmentation ideal of $A = \mathbb{Z}_p[[G_\infty]]$ so that $A/d_\infty A$ is a finitely generated $\mathbb{Z}_p$-free module, e.g. $d_\infty = \gamma - 1$.

**Conjecture 4.1.** (The Equivariant Main Conjecture). With notations as above, we have the following equality of ideals in $A$:

$$\text{Fitt}_A(Z^*_E) = (e_\infty G^*_S)$$

**Remark 4.2.** For any odd prime $p$, the formulation of the Equivariant Main Conjecture 4.1 is equivalent to the formulation of Ritter-Weiss in [35] (cf. [28], §2, (CPE2)). We note that Ritter-Weiss use the translation functor to define $\mathcal{Y}_\infty$, which equals to $\mathcal{Y}_\infty = \frac{\Delta \mathcal{H}}{\Delta \mathcal{H} \Delta \mathcal{G}}$. Here $\Delta(\mathcal{G}, \mathcal{H}) = \ker(Z_p[[\mathcal{G}]] \to Z_p[[G_\infty]])$, which is the same as $(\Delta \mathcal{H}) Z_p[[\mathcal{G}]]$ (cf. for example page 275 in [31]). As a result, $\mathcal{Y}_\infty = \frac{\Delta \mathcal{H}}{\Delta \mathcal{H} \Delta \mathcal{G}} = H_0(\mathcal{H}, \Delta \mathcal{G})$, which is the same as the definition used in this article.

Under the assumption $\mu = 0$ we prove that Conjecture 4.1 follows from the classical Main Conjecture in Iwasawa theory [47]. For some technical reasons, we need to apply the contravariant functors $E^i(-) := Ext^i_{A}(-, A)$ to $Z_{E}^*$, for $i = 0, 1$. We will see that $E^1(Z_{E}^*)$ is a finitely generated $A$-torsion module of projective dimension at most one, whose Fitting ideal is generated by the
modified equivariant $L$-function. First, we recall the definition and some basic properties of Fitting ideals (see [30] for more properties).

For a commutative ring $R$ with identity, the Fitting ideal $\text{Fitt}_R(M)$ of a finitely presented $R$-module $M$ is defined as follows: Given a presentation of $M$ as

$$ R^a \xrightarrow{h} R^b \rightarrow M \rightarrow 0, $$

let $A$ be the matrix associated to the map $h$. The (initial) Fitting ideal of $M$ is defined to be the ideal of $R$ generated by all $b$-minors of $A$ if $a \geq b$, and $(0)$ otherwise. Here are some properties:

1. $\text{Fitt}_R(M)$ is a finitely generated ideal of $R$ satisfying

$$ (\text{Ann}_R(M))^b \subseteq \text{Fitt}_R(M) \subseteq \text{Ann}_R(M), $$

where $\text{Ann}_R(M)$ is the annihilator ideal of $M$ and $b$ is an integer so that $M$ can be generated by $b$ elements as a $R$-module.

2. If $M \twoheadrightarrow M'$ is a surjective map of finitely presented $R$-modules, then

$$ \text{Fitt}_R(M) \subseteq \text{Fitt}_R(M'). $$

3. If $0 \rightarrow M' \rightarrow M \rightarrow M'' \rightarrow 0$ is an exact sequence of finitely presented $R$-modules, then

$$ \text{Fitt}_R(M') \cdot \text{Fitt}_R(M'') \subseteq \text{Fitt}_R(M). $$

Moreover, we have equality if the exact sequence splits, i.e.

$$ \text{Fitt}_R(M' \oplus M'') = \text{Fitt}_R(M') \cdot \text{Fitt}_R(M''). $$

4. If $M \cong R/a$ is a cyclic module, then

$$ \text{Fitt}_R(M) = \text{Ann}_R(M) = a. $$

More generally, if we apply the previous property to a direct sum of $n$ cyclic $R$-modules

$$ M \cong R/a_1 \oplus R/a_2 \oplus \cdots \oplus R/a_n, $$

then we obtain

$$ \text{Fitt}_R(M) = a_1 a_2 \cdots a_n. $$

5. Let $M$ be a finite $R$-module for a group ring $R$ of a finite abelian group with coefficients in $\mathbb{Z}_p$. If $M$ is cyclic as a $\mathbb{Z}_p$-module, then

$$ \text{Fitt}_R(M^\ast) = \text{Ann}_R(M^\ast) \cong \text{Ann}_R(M) = \text{Fitt}_R(M), $$

where $M^\ast = \text{Hom}(M, \mathbb{Q}_p/\mathbb{Z}_p)$ is the Pontryagin dual equipped with the covariant action.
We recall that \# denotes the inverse action defined in (3.4). In the next lemma we list some general properties of $E^i(M)$ for an $\mathcal{A}$-module $M$. For a proof see propositions 5.4.17, 5.5.6 and corollary 5.5.7 in [31], or [17].

**Lemma 4.3.** Let $M$ be an $\mathcal{A}$-module, let $\alpha(M)$ denote the adjoint of $M$ with the contravariant action, and let $M^\vee = Hom_{\mathbb{Z}_p}(M, \mathbb{Z}_p)$ be the dual with the contravariant $G_\infty$-action. Then

1. $E^i(M) = Ext^i_{\mathcal{A}}(M, \mathcal{A})$ as $\mathcal{A}$-modules for any $\mathcal{A}$-module $M$ and $i \geq 0$,

2. $E^1(M)\# \simeq \alpha(M)$ as $\mathcal{A}$-modules, provided $M$ is a finitely generated $\mathcal{A}$-torsion module,

3. $E^1(M)\# \simeq M^\vee$ as $\mathcal{A}$-modules, provided $M$ is a $\mathcal{A}$-torsion module with trivial $\mu$-invariant, i.e. $M$ is a finitely generated $\mathbb{Z}_p$-module.

We list some results obtained by applying the contravariant functors $E^i(-)$ for $i = 1, 2$, to some of the exact sequences arising from diagram (2.8). We first remark that $E^i(\mathcal{A}) = Hom_{\mathcal{A}}(\mathcal{A}, \mathcal{A})$ is a left exact functor, and that $E^i(\mathcal{A}) = 0$ for $i \geq 1$, since $pd_{\mathcal{A}}(\mathcal{A}) = 0$ (cf. Proposition 5.2.11 in [31]).

**Lemma 4.4.** The $\mathcal{A}$-module $E^1(\mathbb{Z}_\infty^f)^\#$ is of projective dimension at most one, and

$$Fitt_{\mathcal{A}}(E^1(\mathbb{Z}_\infty^f)^\#) = Fitt_{\mathcal{A}}(\mathbb{Z}_\infty^f)^\#.$$

**Proof.** This lemma is a consequence of Proposition 2 in [13] and Lemma 4.3. We give here a direct proof, since we need some of the methods in the proof of Lemma 4.5 below: We first apply $E^i(-)$ to the last column of diagram (2.8). We observe that $Hom_{\mathcal{A}}(\mathbb{Z}_\infty^f, \mathcal{A})$ is the set of all morphisms in $Hom_{\mathcal{A}}(\Delta G_\infty, \mathcal{A})$, whose restriction to $d_\infty \mathcal{A}$ vanishes. This observation and the choice of $d_\infty \mathcal{A}$ as a non-zero divisor imply that $E^0(\mathbb{Z}_\infty^f) = 0$. By part 1 of Lemma 4.3 $E^0(\mathcal{X}_\infty^f)$ is also trivial for the $\mathcal{A}$-torsion module $\mathcal{X}_\infty^f$. Hence by applying the contravariant functor $E^i(-)$ to the last row of diagram (2.8), we obtain $E^0(\mathbb{Z}_\infty^f) = 0$. On the other hand $E^i(\mathcal{A})$ is trivial for $i \geq 1$ as we mentioned before. Therefore, applying $E^i(-)$ to the middle column of diagram (2.10) leads to the exact sequence

$$0 \to E^0(\mathcal{A})^r + 1 \to E^0(\mathcal{A})^{r+1} \to E^1(\mathbb{Z}_\infty^f) \to 0.$$

Here $E^0(\Psi)$ is the transpose of $\Psi$. We now apply “$\#$” to the exact sequence above to obtain:

$$0 \to \mathcal{A}^{r+1} \to E^0(\Psi)^\# \to E^1(\mathbb{Z}_\infty^f)^\# \to 0,$$

which shows that the projective dimension of the $\mathcal{A}$-module $E^1(\mathbb{Z}_\infty^f)^\#$ is at most one. To complete the proof it is enough to note that the Fitting ideal of the $\mathcal{A}$-module $E^1(\mathbb{Z}_\infty^f)^\#$ is given by the determinant of $E^0(\Psi)^\#$, whereas the Fitting ideal of $\mathbb{Z}_\infty^f$ is given by the determinant of the map $\Psi$ defined in diagram (2.10).
Lemma 4.5. We have the following exact sequence of finitely generated \(k\)-torsion modules:

\[ 0 \to \mathbb{Z}_p \to (k/d_\infty k)^\# \to E^1(z_{\infty})^\# \to 0. \]

Proof. We saw in the proof of Lemma 4.4 that \(E^0(z_{\infty})\) is trivial. As a consequence of Proposition 5.2.11 in [31] we obtain that \(E^2(\mathbb{Z}_p)\) is trivial as well, since the projective dimension of the \(\Lambda\)-module \(\mathbb{Z}_p\) is one. By applying \(E^i(\cdot)\) to the exact sequence (2.9) we therefore obtain the exact sequence

\[ 0 \to E^1(\mathbb{Z}_p) \to E^1(k/d_\infty k) \to E^1(z_{\infty}) \to 0. \]

By part 2 of Lemma 4.3 together with the fact that an elementary module is isomorphic to the inverse module of its adjoint (cf. 1.3 in [16]) we have \(E^1(\mathbb{Z}_p)^\# \cong \mathbb{Z}_p\), where \(\mathbb{Z}_p\) has the trivial \(G_\infty\)-action. We also have the isomorphism \(E^1(k/d_\infty k)^\# \cong (k/d_\infty k)^\#\), since \(\text{pd}_A(k) = 0\). Therefore, by applying “\(\#\)” to the exact sequence above, we obtain:

\[ 0 \to \mathbb{Z}_p \to (k/d_\infty k)^\# \to E^1(z_{\infty})^\# \to 0. \]

Lemma 4.6. We have the following exact sequence of finitely generated \(k\)-torsion modules:

\[ 0 \to E^1(z_{\infty}) \to E^1(\mathbb{Z}_p) \to E^1(z_{\infty}) \to 0. \]

Proof. First we observe that \(E^2(z_{\infty})\) is trivial by applying \(E^i(\cdot)\) to the exact sequence (2.9) and by noting that the projective dimensions of \(\mathbb{Z}_p\) and \(k/d_\infty k\) are both one. Now we apply \(E^i(\cdot)\) to the last row of diagram (2.8) to obtain the exact sequence above. We note that the surjectivity of the last map in the diagram follows from the observation that \(E^2(z_{\infty}) = 0\), and that the injectivity of the first map in the diagram is a consequence of the observation that \(E^0(\mathbb{X}_{\infty}) = 0\) in the proof of lemma (4.4).

We combine Lemmas 4.5 and 4.6 to obtain the following theorem:

Theorem 4.7. We have the following exact sequence of finitely generated \(k\)-torsion modules:

\[ 0 \to \mathbb{Z}_p \to (k/d_\infty k)^\# \to E^1(\mathbb{Z}_p)^\# \to E^1(z_{\infty})^\# \to 0, \]

in which

\[ \text{pd}_A((k/d_\infty k)^\#) \leq 1 \quad \text{and} \quad \text{pd}_A(E^1(\mathbb{Z}_p)^\#) \leq 1. \]

The \(\Lambda\)-module \(E^1(\mathbb{X}_{\infty})^\#\) is isomorphic to the adjoint of \(\mathbb{X}_{\infty}\) by part 2 of Lemma 4.3 and so it is a finitely generated \(\mathbb{Z}_p\)-free module under the assumption \(\mu = 0\). Therefore we obtain from Theorem 4.7 the first statement in the following proposition:

Documenta Mathematica 18 (2013) 749–783
Proposition 4.8. Let $d_\infty \in \Delta G_\infty$ be a non-zero divisor so that $\mathcal{A}/d_\infty \mathcal{A}$ is a finitely generated $\mathbb{Z}_p$-free module. Then, under the assumption $\mu = 0$, the sequence

$$0 \to \mathbb{Z}_p \to (\mathcal{A}/d_\infty \mathcal{A})^\# \to E^1(\mathcal{Z}_\infty^f)^\# \to E^1(\mathcal{X}_\infty^f)^\# \to 0$$

is an exact sequence of finitely generated $\mathbb{Z}_p$-free modules. Moreover, if we consider this sequence as an exact sequence of $\mathbb{Z}_p[\Gamma]$-modules, then under the assumptions of $\mu = 0$ we have

$$pd_{\mathbb{Z}_p[\Gamma]}((\mathcal{A}/d_\infty \mathcal{A})^\#) = 0 \quad \text{and} \quad pd_{\mathbb{Z}_p[\Gamma]}(E^1(\mathcal{Z}_\infty^f)^\#) = 0.$$

Proof. We only need to prove the second part. We first remark that $\mathcal{A}/d_\infty \mathcal{A}$ and $E^1(\mathcal{Z}_\infty^f)$ are both $H$-cohomologically trivial by Proposition 2.2 in [32], since their projective dimensions are at most one as $\mathcal{A}$-modules. By a classical theorem of Nakayama $pd_{\mathbb{Z}_p[\Gamma]}(M) = 0$ if and only if $M$ is $\mathbb{Z}_p$-free and $H$-cohomologically trivial ($p$-adic version of Theorem 8 in [40], Chapter IX, §5). Therefore, both modules are of projective dimension zero as $\mathbb{Z}_p[\Gamma]$-modules.

To finish the proof that the Equivariant Main Conjecture follows from the classical Main Conjecture under the assumption $\mu = 0$, we first review the definition of the determinantal ideal, which plays a role similar to that of the characteristic ideal for some $\Lambda$-modules with an extra group action: For a commutative ring $R$ with identity, a finitely generated projective $R$-module $P$ and $f \in \text{End}_R(P)$, the determinant of $f$ is defined as

$$\text{det}_R(f \mid P) := \text{det}_R(f \oplus \text{id}_Q \mid P \oplus Q),$$

where $Q$ is a complement of $P$, i.e. $P \oplus Q$ is free. One can check that the definition is independent of $Q$ by using Schanuel’s lemma. By the same strategy, since $P \otimes_R R[X]$ is a finitely generated projective $R[X]$-module, the monic polynomial $\text{det}_R(X - f \mid P) \in R[X]$ is defined to be

$$\text{det}_R(X - f \mid P) := \text{det}_{R[X]}(\text{id}_P \otimes X - f \otimes 1 \mid P \otimes_R R[X]),$$

for any projective $R$-module $P$. One can see that these definitions are well-behaved under base-change, i.e.

$$\begin{align*}
\text{det}_R(f \mid P) &= \text{det}_R(f \otimes \text{id}_{R'} \mid P \otimes R') \\
\text{det}_R(X - f \mid P) &= \text{det}_{R'}(X - (f \otimes \text{id}_{R'}) \mid P \otimes R'),
\end{align*}$$

(4.1)

where $R'$ is any commutative $R$-algebra. We have the following general proposition:

Proposition 4.9. ([15], Proposition 7.2). Let $R$ be a commutative, semi-local, compact topological ring and let $\Gamma$ be a pro-cyclic group with topological generator $\gamma$. Let $M$ be a topological $R[[\Gamma]]$-module, which is projective and finitely generated as an $R$-module. Let

$$F(X) := \text{det}_R(X - m_\gamma \mid M),$$

Documenta Mathematica 18 (2013) 749–783.
where \( m_\gamma \) is the \( R[[\Gamma]] \)-module automorphism of \( M \) given by multiplication by \( \gamma \). Then the following holds:

1. \( M \) is finitely presented as an \( R[[\Gamma]] \)-module. If we let \( F(\gamma) \) be the image of \( F(X) \) via the \( R\)-algebra morphism \( R[X] \to R[[\Gamma]] \) sending \( X \) to \( \gamma \), we have an equality of \( R[[\Gamma]] \)-ideals

\[
\text{Fitt}_{R[[\Gamma]]}(M) = (F(\gamma)).
\]

2. If we view \( M_\gamma^R = \text{Hom}_R(M, R) \) as a topological \( R[[\Gamma]] \)-module with the covariant \( \Gamma \)-action, then

\[
\text{Fitt}_{R[[\Gamma]]}(M) = \text{Fitt}_{R[[\Gamma]]}(M_\gamma^R).
\]

3. If we view \( M^\vee = \text{Hom}_{\mathbb{Z}_p}(M, \mathbb{Z}_p) \) as a topological \( R[[\Gamma]] \)-module with the covariant \( \Gamma \)-action, where \( R = \mathbb{Z}_p[G] \) and \( G \) is a finite abelian group, then

\[
\text{Fitt}_{R[[\Gamma]]}(M) = \text{Fitt}_{R[[\Gamma]]}(M^\vee).
\]

By using Proposition 4.9 for the ring \( R = \mathbb{Z}_p[H] \) and the finitely generated \( R \)-modules \( M = E^1(\mathbb{Z}_\infty^f)^\# \) and \( M = (\mathbb{A}/d_\infty\mathbb{A})^\# \), which are projective by Proposition 4.8, we obtain:

**Lemma 4.10.** If \( m_\gamma \) denotes the \( R[[\Gamma]] \)-module automorphism of \( M \) given by multiplication by \( \gamma \), then

\[
\begin{align*}
\text{Fitt}_{\mathbb{A}}(E^1(\mathbb{Z}_\infty^f)^\#) &= (\det_{\mathbb{Z}_p[\Gamma]}((T + 1) - m_\gamma | E^1(\mathbb{Z}_\infty^f)^\#)), \\
\text{Fitt}_{\mathbb{A}}((\mathbb{A}/d_\infty\mathbb{A})^\#) &= (\det_{\mathbb{Z}_p[\Gamma]}((T + 1) - m_\gamma | (\mathbb{A}/d_\infty\mathbb{A})^\#)).
\end{align*}
\]

Let \( \mathcal{O} \) be the ring of integers obtained by adjoining all character values of the characters of \( H \) to \( \mathbb{Z}_p \), let \( \pi \) be a fixed uniformizer in \( \mathcal{O} \) and let \( Q(\mathcal{O}) \) denote the field of fractions of \( \mathcal{O} \). We consider \( \mathcal{O} \) and \( Q(\mathcal{O}) \) as \( \mathbb{A} \)-modules with trivial \( G_\infty \) action. We note that for the idempotent \( e \) attached to the trivial character of \( H \) we have \( H_\mathcal{S}(T) = T \cdot e + (1 - e) \) using the identification (2.2) (see (3.1) for the definition of \( H_\mathcal{S}(T) \)). Therefore, using Lemma 4.10 we obtain the following lemma:

**Lemma 4.11.** We have the following equalities of ideals in \( Q(\mathcal{O})[[H]][[\Gamma]] \):

\[
\begin{align*}
(H_\mathcal{S}(T)) &= (\det_{Q(\mathcal{O})[[H]]}((T + 1) - m_\gamma | Q(\mathcal{O}))), \\
(d^m_\mathcal{S}) &= (\det_{Q(\mathcal{O})[[H]]}((T + 1) - m_\gamma | (Q(\mathcal{O})[[H]][[\Gamma]]/d_\infty)^\#)).
\end{align*}
\]

**Remark 4.12.** Any character \( \chi \) of \( H \) can be extended to a \( Q(\mathcal{O})[[X]] \)-algebra homomorphism, for a variable \( X \), and to a \( Q(\mathcal{O}) \otimes_{\mathcal{O}} \mathcal{O}[[\Gamma]] \)-algebra homomorphism

\[
\begin{align*}
\chi : Q(\mathcal{O})[[X]] &\to Q(\mathcal{O})[[X]], \\
\chi : Q(\mathcal{O}) \otimes_{\mathcal{O}} \mathcal{O}[[G_\infty]] &\to Q(\mathcal{O}) \otimes_{\mathcal{O}} \mathcal{O}[[\Gamma]],
\end{align*}
\]

which maps \( h \to \chi(h) \) for \( h \in H \).
Lemma 4.13. We have the following equality of ideals in $O[[\Gamma]]$ under the assumption $\mu = 0$:

$$(\det_{Q(O)}((T + 1) - m_\gamma | e_\psi(\mathcal{X}_\infty^f \otimes_{Z_p} Q(O)))^\# = (\det_{Q(O)}((T + 1) - m_\gamma | e_\psi(E^1(\mathcal{X}_\infty^f)^\# \otimes_{Z_p} Q(O))))^\#.$$ 

Proof. Since the $R := O_\psi[[\Gamma]]$-torsion module $M_1 := e_\psi(\mathcal{X}_\infty^f \otimes_{Z_p} Q(O))$ is of projective dimension at most one (cf. for example Lemma 2.4 in [32]), the $R$-torsion module $M_2 := e_\psi(E^1(\mathcal{X}_\infty^f)^\# \otimes_{Z_p} Q(O))$ is also of projective dimension at most one. Hence by the first part of Proposition 4.9 it is enough to show that $\text{Fitt}_R(M_1)^{\#} = \text{Fitt}_R(M_2)$. But this is a direct consequence of the second part of Lemma 4.3 and Proposition 2 in [13].

The equality of fractional ideals

$$(G^*_\psi, S(T)) = (\det_{Q(O)}((T + 1) - m_\gamma | e_\psi(\mathcal{X}_\infty^f \otimes_{Z_p} Q(O))))$$

in $Q(O)[[T]]$ for any character $\psi$ of $H$ (see equality (1.2) and part 1 of Proposition 4.9). Here we note that the assumption $\mu = 0$ implies the vanishing of the $\mu$-invariants of $\mathcal{X}^f_{\infty, \psi}$ for all (even) characters $\psi$ of $H$. Hence, using Lemma 4.13, we obtain the following equality of fractional ideals in $Q(O)[[T]]$:

$$(\psi(G^*_\psi, S(T))^\#) = (G^*_\psi, S(T)^\#) = (\det_{Q(O)}((T + 1) - m_\gamma | e_\psi(E^1(\mathcal{X}_\infty^f)^\# \otimes_{Z_p} Q(O)))) = (\psi(\det_{Q(O)[\Gamma]}((T + 1) - m_\gamma | E^1(\mathcal{X}_\infty^f)^\# \otimes_{Z_p} Q(O))))).$$

Consequently the exact sequence

$$0 \to Z_p \to (\mathbb{A}/d_\infty \mathbb{A})^\# \to E^1(Z_\infty^f)^\# \to E^1(X_\infty^f)^\# \to 0$$

in Theorem 4.7, and Lemma 4.10 together with the base-change property of determinants (cf. (4.1)) imply the following equality of ideals in $O[[T]]$:

$$\psi((c_\infty G^*_\psi(T))^\#) = (\psi(\det_{Q(O)[\Gamma]}((T + 1) - m_\gamma | E^1(Z_\infty^f)^\# \otimes_{Z_p} Q(O))))).$$

Before completing the proof we recall that the $\mu$-invariant $\mu(F)$ of a power series $F \in O[[T]]$ is the largest exponent $\mu \geq 0$ such that $f \in (\pi^n)O[[T]]$. For $F \in \mathbb{A}$ we define the $\mu$-invariant of $F$ to be zero if $\mu(\chi(F)) = 0$ for any $p$-adic valued character $\chi$ of $H$.

Let $F := \det_{Z_p[H]}((T + 1) - m_\gamma | E^1(Z_\infty^f)^\#)$ and $G := (c_\infty G^*_\psi(T))^\#$ in $\mathbb{A}$ (cf. Lemma 3.2). Then

- $\mu(F) = 0$, since the determinantal polynomial $F \in Z_p[H][[T]]$ of the projective $Z_p[H]$-module $E^1(Z_\infty^f)^\#$ (cf. Proposition 4.8) is monic.
• $\mu(G) = 0$, since for any $p$-adic character $\psi$ of the group $H$, we have
$$\mu(\psi(G^*_S)) = \mu(g^*_S \cdot u_{\psi,S}) = 0, \quad \mu(\psi(H_S)) = 0, \quad \mu(\psi(d_\infty)) = 0.$$  

Here we note that the determinantal polynomials of the field $Q(O)$ and the projective $\mathbb{Z}_p[H]$-module $(h/d_\infty h)^\#$ (cf. Proposition 4.8), which are generated by $H_S$ and $d_\infty^\#$, respectively (cf. Lemma 4.11), are monic.

• $(\psi(F)) = (\psi(G))$, using equality (4.2).

In the terminology of [2], $R := \mathbb{Z}_p[H]$ is admissible for the abelian group $H$, i.e. $R$ is a finite product of strictly admissible rings $R_i$, which means that each $R_i$ is separated and complete in the $\text{rad}(R_i)$-adic topology and also $R_i/\text{rad}(R_i)$ is a skew field. Since the $\mu$-invariants of $F, G \in R[[T]]$ are both zero, Proposition 2.1 in [2] as an equivariant Weierstrass Preparation Theorem implies the existence of unique distinguished polynomials $f^*, g^* \in R[[T]]$ and units $u, v \in (R[[T]])^\times$ such that
$$F = u \cdot f^* \quad \text{and} \quad G = v \cdot g^*.$$  

We apply a $p$-adic character $\psi$ of $H$ to both sides, and note that $\psi(f^*)$ and $\psi(g^*)$ are both distinguished polynomials in $O[[T]]$, and that $\psi(u), \psi(v) \in O[[T]]^\times$ are units. Hence the equality $(\psi(F)) = (\psi(G))$ together with the uniqueness of the Weierstrass decomposition yields
$$\psi(f^*) = \psi(g^*)$$
for any $p$-adic character $\psi$ of $H$. Therefore, $f^* = g^*$ and $F = uv^{-1}G$. The equality $(F) = (G)$ now implies the following:

$$(\iota(c_\infty G^*_S)) = (\text{det}_{\mathbb{Z}_p[H]}((T + 1) - m_\gamma | E^1(\mathbb{Z}_S^\infty)^\#))$$

$$= \text{Fitt}_h(E^1(\mathbb{Z}_S^\infty)^\#) \quad \text{by Lemma 4.10} \quad (4.3)$$

$$= \text{Fitt}_h(\mathbb{Z}_S^\infty)^\# \quad \text{by Lemma 4.4.}$$

Consequently, the equality
$$\text{Fitt}_h(\mathbb{Z}_S^\infty) = (c_\infty G^*_S)$$
holds, and this completes the proof of the Equivariant Main Conjecture 4.1 under the assumptions of the classical Main Conjecture and of $\mu = 0$.

**Theorem 4.14.** The Equivariant Main Conjecture 4.1 follows from the classical Main Conjecture in Iwasawa theory under the assumption $\mu = 0$.

**Remark 4.15.** For any odd prime $p$, or for the prime 2 if $F$ is an absolutely abelian number field, the classical Main Conjecture holds, and hence the Equivariant Main Conjecture 4.1 is verified under the assumption $\mu = 0$.  

---

**Documenta Mathematica 18 (2013) 749–783**
We recall that the assumption $\mu = 0$ holds for any absolute abelian number field $E$, i.e. for any number field $E$ whose Galois group over $\mathbb{Q}$ is abelian, by the result of Ferrero and Washington in [10]. Hence

**Corollary 4.16.** If $E$ is an absolute abelian number field, the Equivariant Main Conjecture 4.1 holds unconditionally.

5 The Coates-Sinnott Conjecture as an application

Let $E/F$ be an abelian extension of number fields with Galois group $G$, let $n \geq 2$ be an integer, and let $p$ be an arbitrary prime. Let $S$ be a finite set of primes in $F$ containing the primes above $p$, the primes ramified in $E$ and the infinite primes, and let $S_f$ denote the set of all finite primes in $S$. Let

$$\Theta_{E/F}^S(s) = \sum_{\chi \in \hat{G}} L_{E/F}^S(s, \chi^{-1}) \cdot e_{\chi}$$

be the $G$-equivariant $S$-incomplete $L$-function associated to $E/F$. We recall that for an integer $n \geq 1$ by a result of Siegel [41]

$$\Theta_{E/F}^S(1 - n) \in \mathbb{Q}[G],$$


$$\text{Ann}_{\mathbb{Z}[G]}(H^0(E, \mathbb{Q}/\mathbb{Z}(n))) \cdot \Theta_{E/F}^S(1 - n) \subset \mathbb{Z}[G].$$

For $n \geq 1$ the $n$-th higher Stickelberger ideal is defined as follows:

$$\text{Stick}_{E/F}^S(n) := \text{Ann}_{\mathbb{Z}[G]}(H^0(E, \mathbb{Q}/\mathbb{Z}(n))) \cdot \Theta_{E/F}^S(1 - n) \subset \mathbb{Z}[G].$$

**Remark 5.1.** The classical theorem of Stickelberger states that

$$\text{Stick}_{E/F}^S(1) \subseteq \text{Ann}_{\mathbb{Z}[G]}(\text{Cl}(O_E)),$$

where $\text{Cl}(O_E)$ denotes the class group of $O_E$. Brumer conjectured that the same holds for any abelian extension $E/F$ of number fields.

The original formulation of the Coates-Sinnott Conjecture is as follows:

**Conjecture 5.2.** (The Coates-Sinnott Conjecture, $K$-theoretic version). Let $E/F$ be an abelian Galois extension of number fields with Galois group $G$, and let $n \geq 2$. Then

$$\text{Stick}_{E/F}^S(n) \subseteq \text{Ann}_{\mathbb{Z}[G]}(K_{2n-2}(O_E)).$$

As a consequence of the work of Voevodsky (cf. [45]) the Quillen-Lichtenbaum Conjecture holds, i.e. for all odd primes $p$ the étale Chern characters defined by Soulé [42]

$$ch_{i,n}^{(p)} : K_{2n-i}(O_F) \otimes \mathbb{Z}_p \to H_{et}^{2i}(O'_F, \mathbb{Z}_p(n))$$
are isomorphisms for \( i = 1, 2 \), and all \( n \geq 2 \). Here \( O_F' = O_F[1/p] \). The surjectivity of the Chern characters was proved by Soulé for even \( n \) (see [42]) and by Dwyer and Friedlander in general (see [7]).

For the prime 2 the situation is in general different. The deviation between \( K_{2n-i}(O_F) \otimes \mathbb{Z}_2 \) and \( H_{et}^i(O_F', \mathbb{Z}(n)) \) has been determined by Rognes and Weibel [37]. In [21] it was suggested to replace the \( K \)-groups \( K_{2n-i}(O_F) \) by the motivic cohomology groups \( H^*_{et}(E, \mathbb{Z}(n)) \), because the latter groups have the advantage that their \( p \)-parts are isomorphic to \( H^2_{et}(O_E', \mathbb{Z}_p(n)) \) for all primes \( p \) ([19], Theorem 2.4). This leads to the following motivic version of the Coates-Sinnott Conjecture:

**Conjecture 5.3.** (The Coates-Sinnott Conjecture, motivic version). Let \( E/F \) be an abelian Galois extension of number fields with Galois group \( G \), and let \( n \geq 2 \). Then

\[
\text{Stick}_{E/F}(n) \subseteq \text{Ann}_{\mathbb{Z}[G]}(H^2_{et}(E, \mathbb{Z}(n))).
\]

The explicit results of Rognes-Weibel show that the motivic version implies the \( K \)-theoretic version. Moreover, the validity of the motivic version is equivalent to the validity of the following \( p \)-adic version for all primes \( p \):

**Conjecture 5.4.** (The Coates-Sinnott Conjecture, \( p \)-adic version). Let \( E/F \) be an abelian Galois extension of number fields with Galois group \( G \), let \( p \) be prime, and let \( n \geq 2 \). Then

\[
\text{Ann}_{\mathbb{Z}_p[G]}(H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n))) \cdot \Theta^S_{E/F}(1-n) \subseteq \text{Ann}_{\mathbb{Z}_p[G]}(H^2_{et}(O_E', \mathbb{Z}_p(n))).
\]

We remark that by the functional equation of \( L \)-functions (see for example [1]) \( L_{E/F}(s, \chi) \) vanishes at negative integers \( 1 - n \) for \( n \geq 2 \), unless \( F \) is a totally real number field and \( \chi(-1) = (-1)^n \). Therefore only the following cases are of interest:

- \( E \) is a totally real number field and \( n \geq 2 \) is even.
- \( E \) is a totally complex number field, \( F \) is totally real and \( n \geq 2 \) is odd.

We will consider the first case, namely that \( E/F \) is an abelian extension of totally real fields with Galois group \( G \), and that \( n \geq 2 \) is even. We show that in this case the \( p \)-adic version of the Coates-Sinnott Conjecture follows from the Equivariant Main Conjecture for all primes \( p \) assuming \( \mu = 0 \). For odd primes this has been done by Nguyen Quang Do [28] and independently by Greither-Popescu [15] (see also [2] for a slightly weaker result).

We first assume without loss of generality for the proof of the Coates-Sinnott Conjecture that \( E \) is the maximal real subfield of \( E(\zeta_{2p}) \) (this assumption clearly holds for \( p = 2 \), and for odd primes \( p \) one can see for instance Lemma 6.14 in [15]). Here \( \zeta_{2p} \) is a primitive \( 2p \)-th root of unity. We recall the set up from Section 2: Let \( E_\infty \) (resp. \( F_\infty \)) be the cyclotomic \( \mathbb{Z}_p \)-extension of \( E \) (resp. \( F \))
with Galois group $\Gamma_E$ (resp. $\Gamma_F$) over $E$ (resp. over $F$). We denote by $G_\infty$ the Galois group of $E_\infty/F$, by $H$ the Galois group of $E_\infty/F_\infty$, and by $\Gamma = \langle \gamma \rangle$ the image of $\Gamma_F$ under the splitting map in (2.1). The following diagram illustrates the situation:

\[
\begin{array}{c}
E_\infty \\
\downarrow \Gamma_E \\
F_\infty \\
\downarrow \Gamma_F \\
E \\
\downarrow G \\
F
\end{array}
\]

Since $G_\infty$ is abelian, $G_\infty = H \times \Gamma$ and the completed group ring $\mathcal{A} = \mathbb{Z}_p[[G_\infty]]$ is identified with $\mathbb{Z}_p[H][[\Gamma]]$ under the identification (2.2). We let $d_\infty \in \Delta G_\infty$ be a non-zero divisor so that $\mathcal{A}/d_\infty \mathcal{A}$ is a finitely generated $\mathbb{Z}_p$-free module, e.g. $d_\infty = \gamma - 1$. By Theorem 4.7 we obtain an exact sequence of finitely generated $\mathcal{A}$-torsion modules

\[
0 \to \mathbb{Z}_p \to (\mathcal{A}/d_\infty \mathcal{A})^\# \to E^1(\mathbb{Z}_p^f)^\# \to E^1(X_p^f)^\# \to 0
\] (5.1)
in which the middle terms are of projective dimensions at most one. Using the equalities in (4.3) we also have

\[
\begin{align*}
\text{Fitt}_\mathcal{A}((\mathcal{A}/d_\infty \mathcal{A})^\#(n)) &= ((t \circ t_n)(d_\infty)) \\
\text{Fitt}_\mathcal{A}(E^1(\mathbb{Z}_p^f)^\#(n)) &= ((t \circ t_n)(c_\infty G_\infty^S)).
\end{align*}
\] (5.2)

We note that the sequence (5.1) is also an exact sequence of finitely generated $\mathcal{A}$-modules, where $\mathcal{A} = \mathbb{Z}_p[[\Gamma]]$. Moreover, the exact sequence (5.1) is an exact sequence of finitely generated $\mathbb{Z}_p$-free modules since we have assumed $\mu = 0$.

**Lemma 5.5.** Let $G_E^S$ be the Galois group of the maximal algebraic pro-$p$-extension of $E$ unramified outside the primes above $S_f$, over $E$. Under the assumption that $E$ is the maximal real subfield of $E(\zeta_{2p})$, we have the following:

1. $H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n)) \simeq \mathbb{Z}_p(n)_{\Gamma_E}$ for $n \geq 2$

2. $H^2(G_E^S, \mathbb{Z}_p(n)) \simeq (E^1(X_p^f)^\#(n))_{\Gamma_E}$ for even $n \geq 2$ (under the hypothesis $\mu = 0$)

**Proof.** 1. It is enough to take the $\Gamma_E$-invariants and $\Gamma_E$-coinvariants of the exact sequence $0 \to \mathbb{Z}_p(n) \to \mathbb{Q}_p(n) \to \mathbb{Q}_p(n)/\mathbb{Z}_p(n) \to 0$ to get a 6 term exact sequence in which $\mathbb{Q}_p(n)_{\Gamma_E} = \mathbb{Q}_p(n)^{G_E^S} = 0$ for $n \geq 2$. We note that
\[ H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n)) = \mathbb{Q}_p/\mathbb{Z}_p(n)^{\Gamma_E}. \]

2. Let \( E_\infty \) be the cyclotomic \( \mathbb{Z}_p \)-extension of \( E \), with Galois group \( \Gamma_E \). As before we denote by \( \Omega_E^{S_f} \) the maximal algebraic pro-\( p \)-extension of \( E \) unramified outside the primes above \( S_f \). We have the following isomorphisms:

\[ E^1(X_{\infty}^f)(\mathbb{Q}_p/\mathbb{Z}_p(n))_{\Gamma_E} \]

\[ \simeq \text{Hom}(X_{\infty}^f, \mathbb{Q}_p/\mathbb{Z}_p(n))_{\Gamma_E} \]

\[ \simeq \text{Hom}(X_{\infty}^f(-n), \mathbb{Q}_p/\mathbb{Z}_p) \quad \text{By Lemma 5.18 in [14]} \]

\[ \simeq \text{Hom}(X_{\infty}^f, \mathbb{Q}_p/\mathbb{Z}_p(n))_{\Gamma_E} \]

\[ \simeq \text{Hom}(\text{Gal}(\Omega_{E}^{S_f} / E_\infty), \mathbb{Q}_p/\mathbb{Z}_p(n))_{\Gamma_E}. \]

If we assume that

\[ \text{Hom}(X_{\infty}^f, \mathbb{Q}_p/\mathbb{Z}_p(n)) \simeq H^1(\text{Gal}(\Omega_{E}^{S_f} / E_\infty), \mathbb{Q}_p/\mathbb{Z}_p(n)), \quad (5.3) \]

then - using the fact that \( cd_p(\Gamma_E) = 1 \) - we can continue the isomorphisms above as follows:

\[ H^1(\text{Gal}(\Omega_{E}^{S_f} / E_\infty), \mathbb{Q}_p/\mathbb{Z}_p(n))_{\Gamma_E} \simeq H^1(\text{Gal}(\Omega_{E}^{S_f} / E), \mathbb{Q}_p/\mathbb{Z}_p(n)) \]

\[ \simeq H^2(G_{E}^{S_f}, \mathbb{Q}_p(n)) \]

where the last isomorphism follows from the finiteness of the groups \( H^1(G_{E}^{S_f}, \mathbb{Q}_p/\mathbb{Z}_p(n)) \) and \( H^2(G_{E}^{S_f}, \mathbb{Q}_p(n)) \) for even \( n \) and the totally real field \( E \) (cf. Corollary 2.5 in [20] and Proposition 2.3 in [44]).

Hence, to complete the proof it is enough to show that the claim (5.3) is true. Clearly

\[ \text{Hom}(X_{\infty}^f, \mathbb{Q}_p/\mathbb{Z}_p(n)) \simeq \text{Hom}(\text{Gal}(\Omega_{E}^{S_f} / E_\infty), \mathbb{Q}_p/\mathbb{Z}_p(n)) \]

since \( X_{\infty}^f \) is the abelianization of \( \text{Gal}(\Omega_{E}^{S_f} / E_\infty) \). Now we notice that the Galois group \( \text{Gal}(\Omega_{E}^{S_f} / E_\infty) \) acts trivially on \( \mathbb{Q}_p/\mathbb{Z}_p(n) \), since \( n \) is even, and therefore

\[ \text{Hom}(\text{Gal}(\Omega_{E}^{S_f} / E_\infty), \mathbb{Q}_p/\mathbb{Z}_p(n)) \simeq H^1(\text{Gal}(\Omega_{E}^{S_f} / E_\infty), \mathbb{Q}_p/\mathbb{Z}_p(n)) \]

At this point we recall the following lemma from Iwasawa theory (See Lemma 6.3 in [5], where a special case is proved, or [31], Chapter V, §3, Ex. 3):

**Lemma 5.6.** Let \( M \) be an \( \mathcal{O}[\![\Gamma]\!] \)-torsion module, where \( \mathcal{O} \) is a finite extension of \( \mathbb{Z}_p \), and let \( F(T) \) be the characteristic polynomial of \( M \). The following are equivalent:

1. \( M^\Gamma \) is finite.
2. $M_T$ is finite.

3. $F(0) \neq 0$

If these conditions hold, then

$$\frac{|M^T|}{|M_T|} = |F(0)|_v = p^{-f \cdot v(F(0))},$$

where $v$ is the normalized valuation, i.e. $v(\pi) = 1$ for a uniformizer $\pi \in \mathcal{O}$, and $f$ is the residue degree of $\pi$ over $p$.

As a consequence of Lemma 5.5 we see that $(E^1(X^f_\infty)^\#(n))_{\Gamma_E}$ and $Z_p(n)_{\Gamma_E}$ are both finite and so by Lemma 5.6 $(E^1(X^f_\infty)^\#(n))_{\Gamma_E}$ and $Z_p(n)_{\Gamma_E}$ are both trivial for even $n \geq 2$. We note that both $Z_p(n)$ and $E^1(X^f_\infty)^\#(n)$ have no non-trivial finite $\Lambda$-submodules. For $E^1(X^f_\infty)^\#(n)$ this follows from the fact that by Lemma 4.3 $E^1(X^f_\infty)^\#$ is isomorphic to the adjoint of $X^f_\infty$ and as such it does not have any non-trivial finite $\Lambda$-submodules (see [16], Section 1.3). Moreover, $((A/d_{\infty}A)^\#(n))_{\Gamma_E} = 0$, since for $n \geq 2$ we have $\kappa(\gamma)^n \neq 1$. Therefore, $((A/d_{\infty}A)^\#(n))_{\Gamma_E}$ is again finite by Lemma 5.6. As a result the $\Gamma_E$-coinvariants of $E^1(Z^f_\infty)^\#(n)$ are also finite and similarly $(E^1(Z^f_\infty)^\#(n))_{\Gamma_E} = 0$, for any even $n \geq 2$. Hence by taking the $\Gamma_E$-coinvariants of the exact sequence (5.1) we obtain the following exact sequence of finite $\mathcal{Z}_p[G]$-modules for any even $n \geq 2$:

$$0 \to H^0(E, \mathbb{Q}_p/Z_p(n)) \to ((A/d_{\infty}A)^\#(n))_{\Gamma_E} \to (E^1(Z^f_\infty)^\#(n))_{\Gamma_E} \to H^2(G^S_E, Z_p(n)) \to 0,$$

where the two middle $Z_p[G]$-modules are of projective dimension at most one as a consequence of the last part of Theorem 4.7 and the facts that $Z_p[G^S_E]_{\Gamma_E} = 0$ and $Z_p[G^S_E]_{\Gamma_E} = Z_p[G]$. Furthermore, following the equalities in (5.2), we have

$$\text{Fitt}_{Z_p[G]}(((A/d_{\infty}A)^\#(n))_{\Gamma_E}) = ((\pi \circ t_0 \circ t_n)(d_{\infty})), $$

$$\text{Fitt}_{Z_p[G]}((E^1(Z^f_\infty)^\#(n))_{\Gamma_E}) = ((\pi \circ t_0 \circ t_n)(c_{\infty}G^*_S)),$$

where $\pi : \mathcal{A} \to Z_p[G]$ is the projection mapping $\gamma - 1$ to 0.

Now we take advantage of the following Proposition due to Burns-Greither, which relates the fitting ideals of the modules of a 4-term exact sequence under some assumptions:

**Proposition 5.7. ([2], Lemma 5)** Let $R := Z_p[G]$ for a finite abelian group $G$ and a prime number $p$. Assume that we have an exact sequence of finite $R$-modules

$$0 \to A \to P \to P' \to A' \to 0,$$

**Documenta Mathematica 18 (2013) 749–783**
Further, assume that \( pd_{Z_p[G]} P \leq 1 \) and \( pd_{Z_p[G]} P' \leq 1 \). Then, we have
\[
\text{Fitt}_R(A^*) \cdot \text{Fitt}_R(P') = \text{Fitt}_R(A') \cdot \text{Fitt}_R(P),
\]
where the Pontryagin dual \( A^* := \text{Hom}(A, \mathbb{Q}_p/\mathbb{Z}_p) \) is endowed with the covariant \( G \)-action.

Applying Proposition 5.7 to the exact sequence (5.4) of finite \( Z_p[G] \)-modules yields the following equality:
\[
\text{Fitt}_{Z_p[G]}(H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n))^*) \text{Fitt}_{Z_p[G]}((E^1(\mathbb{Z}_\infty)^\#(n))_E)
= \text{Fitt}_{Z_p[G]}((\mathbb{A}/d_\infty \mathbb{A})^\#(n)_E) \text{Fitt}_{Z_p[G]}(H^2(G^{S_1}_{E}, \mathbb{Z}_p(n))).
\]

Property 5 of Fitting ideals (section 4) shows that the Fitting ideal of \( H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n))^* \) is the same as the annihilator ideal of \( H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n)) \). This, together with (5.2), yields the following equality of fractional ideals in \( Z_p[G] \):
\[
\text{Fitt}_{Z_p[G]}(H^2(G^{S_1}_{E}, \mathbb{Z}_p(n))) = \text{Ann}_{Z_p[G]}(H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n)))((\pi \circ t \circ t_n)(c_{\infty}/d_\infty \cdot G^S_{Z})).
\]

Finally we have \( d_\infty/c_\infty = (\gamma - 1)e + (1 - e) \), which can be identified by (2.2) with \( Te + 1 - e = H_S(T) \), and obtain the following theorem:

**Theorem 5.8.** We have the following equality of ideals of \( Z_p[G] \):
\[
\text{Fitt}_{Z_p[G]}(H^2(G^{S_1}_{E}, \mathbb{Z}_p(n))) = \text{Ann}_{Z_p[G]}(H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n)))((\pi \circ t \circ t_n)(G^S_{Z}/H^S_{Z})).
\]

Now let \( p \) be an odd prime. In this case we have an equality of the Galois groups \( G^S_{Z} \) and \( G^{S_1}_{E} \), as we noticed before. Furthermore, by Lemma 3.1 we have \( G^S_{Z} = G_S \) under the assumption \( \mu = 0 \). On the other hand for odd primes \( p \), since \( E \) is the maximal real subfield of \( E(\zeta_p) \), we have \( (\pi \circ t \circ t_n)(G^S_{Z}/H^S_{Z}) = \Theta_{E/F}^S(1 - n) \) for any even \( n \) by Corollary 3.4. Therefore from Theorem 5.8 we obtain
\[
\text{Fitt}_{Z_p[G]}(H^2_{\mathfrak{d}}(\mathcal{O}^S_{E}, \mathbb{Z}_p(n))) = \text{Ann}_{Z_p[G]}(H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n))) \cdot \Theta_{E/F}^S(n).
\]

This implies the \( p \)-adic version of the Coates-Sinnott Conjecture for odd primes. Here we note that \( H^2_{\mathfrak{d}}(\mathcal{O}^S_{E}, \mathbb{Z}_p(n)) \subseteq H^2_{\mathfrak{d}}(\mathcal{O}^S_{E}, \mathbb{Z}_p(n)) \).

For \( p = 2 \), we use the following lemma:

**Lemma 5.9.** Let \( E/F \) be an abelian extension of totally real fields with Galois group \( G \), and let \( n \geq 2 \) be an integer. Let \( \tau_1(F) = [F : \mathbb{Q}] \). Then we have the following exact sequence of \( \mathbb{A} \)-modules for \( A = \Lambda[H] \):
\[
0 \to (\mathbb{A}/2\mathbb{A})^{\tau_1(F)} \to \mathcal{X}^S_{\infty} \to \mathcal{X}^f_{\infty} \to 0.
\]
Proof. We have the following commutative diagram by class field theory (cf. [18]) for the finite sets \( S \) and \( S_f \) of primes in \( F \):

\[
\begin{array}{cccc}
D_E & \to & \hat{U}_E^S & \to \prod_{v \in S} \mathbb{Z}_p \mathbb{Z}_v \hat{E}_w & \to & Gal(M_E^S/H_E^S) & \to & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
D'_E & \to & \hat{U}_{E'}^S & \to \prod_{v \in S} \mathbb{Z}_p \mathbb{Z}_v \hat{E}_w & \to & Gal(M_{E'}^S/H_{E'}^S) & \to & 0,
\end{array}
\]

where \( D_E \) and \( D'_E \) are the kernels of the corresponding maps and are bounded by the Leopoldt defect \( \delta_E, \hat{U}_E^S \) (resp. \( \hat{U}_{E'}^S \)) is the \( p \)-adification of the \( S \)-unit (resp. \( S_f \)-unit) group of the ring of integers of \( E \), \( \hat{E}_w \) is the \( p \)-adic completion of the local field \( E_w, M_E^S \) (resp. \( M_{E'}^S \)) is the maximal abelian pro-\( p \)-extension of \( E \) unramified outside the primes in \( S \) (resp. \( S_f \)), and \( H_E^S \) (resp. \( H_{E'}^S \)) is the Hilbert \( S \)-class \((S_f\)-class) field of \( E \). Here the \( S_f \)-unit group means the group of the totally real \( S \)-units. Since \( \hat{E}_w \) is isomorphic to \( \mathbb{Z}/2\mathbb{Z} \) for a real prime \( w \), and \( G \) acts transitively on the set

\[ \{ \hat{E}_w \mid w : \text{infinite primes of } E \text{ lying above } v \} \]

for any infinite prime \( v \) of \( F \), we obtain the following exact sequence of \( \Lambda[H] \)-modules:

\[ 0 \to D_E \to D'_E \to \prod_{w \in S \setminus S_f} \mathbb{Z}/2 \to Gal(M_E^S/M_{E'}^S) \to 0, \]

where \( \prod_{w \in S \setminus S_f} \mathbb{Z}/2 \simeq (\mathbb{Z}/2^e)^{r_1(F)} \). Since the real primes are unramified in \( E_\infty/E \), we can write the exact sequence above for the unique intermediate fields \( E_n \) of \( E_\infty/E \) with \( G_n := Gal(E_n/E) \simeq \mathbb{Z}/2^e \mathbb{Z} \) for all \( n \geq 0 \), as follows:

\[ 0 \to D_{E_n} \to D'_{E_n} \to (\mathbb{Z}/2\mathbb{Z}[G_n])^{r_1(F)} \to Gal(M_{E_n}^S/M_{E_n}^{S_f}) \to 0. \]

Now the claim is that we have the isomorphism

\[ \lim_{\psi}(\mathbb{Z}/2\mathbb{Z}[G_n])^{r_1(F)} \simeq (\mathbb{A}/2\mathbb{A})^{r_1(F)} \]

for \( \mathbb{A} = \Lambda[H] \). Since \( \Lambda/2\Lambda \simeq \lim_{\psi} \mathbb{Z}/2\mathbb{Z}[T]/T^{2n} \), it suffices to show for a fixed real prime \( v \) of \( F \) that the inverse limits of \( \{ \prod_{w \mid v} \mathbb{Z}/2\mathbb{Z} \} \) and \( \{ \mathbb{Z}/2\mathbb{Z}[T]/T^{2n} \} \) are isomorphic. For this we inductively define an isomorphism

\[ f_n : \prod_{v_{n+1} \mid v_n} \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}[T]/T^{2n} \]

compatible with the norm maps as follows: Let \( f_0 \) be the identity and assume we have defined the isomorphisms \( f_m \) compatible with the norm maps for all \( m \leq n \). Let \( v_{n+1} \) and \( v'_{n+1} \) be the extensions of \( v_n \) to \( F_{n+1} \). We define

\[ f_{n+1} : \prod_{v_{n+1}} \mathbb{Z}/2\mathbb{Z} \cdot \prod_{v'_{n+1}} \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}[T]/T^{2^{n+1}}. \]
as follows:
\[ f_{n+1}(a_1, \cdots, a_{2^n}, b_1, \cdots, b_{2^n}) = f_n(a_1 + b_1, \cdots, a_{2^n} + b_{2^n}) + T^{2^n} f_n(a_1, \cdots, a_{2^n}). \]

Now we have the commutative diagram
\[
\begin{array}{ccc}
\prod_{v_n|v} \mathbb{Z}/2\mathbb{Z} & \xrightarrow{f_n} & \mathbb{Z}/2\mathbb{Z}[T]/T^{2^n} \\
\downarrow & & \downarrow \\
\prod_{v_{n+1}|v} \mathbb{Z}/2\mathbb{Z} & \xrightarrow{f_{n+1}} & \mathbb{Z}/2\mathbb{Z}[T]/T^{2^{n+1}}
\end{array}
\]
for any \( n \geq 0 \), and hence
\[
\lim_{\rightarrow n} \prod_{v_n|v} \mathbb{Z}/2\mathbb{Z} \simeq \Lambda/2\Lambda.
\]

This completes the proof of the claim. Now the exact sequence
\[
0 \rightarrow \text{Gal}(M_S^E/M_{E_f}^S) \rightarrow X^\infty_S \rightarrow X^\infty_f \rightarrow 0
\]
yields an exact sequence
\[
(\mathcal{A}/2\mathcal{A})^\ast_{\text{et}}(F) \rightarrow X^\infty_S \rightarrow X^\infty_f \rightarrow 0.
\]

We now take \( \Gamma_E \)-coinvariants and then the Pontryagin dual of the exact sequence of Lemma 5.9. The same calculation as in the proof of the second part of Lemma 5.5 leads to the exact sequence
\[
0 \rightarrow H^2(G_E^{S_f}, \mathbb{Z}_2(n)) \rightarrow H^2(G_E^S, \mathbb{Z}_2(n)) \rightarrow (\mathbb{Z}/2\mathbb{Z}[G])^\ast_{\text{et}}((\pi \circ \iota \circ t_n)(G_E^S)) \rightarrow 0
\]
of \( \mathbb{Z}_2[G] \)-modules. This yields
\[
2^{\ast_{\text{et}}(F)} \text{Ann}_{\mathbb{Z}_2[G]}(H^2(G_E^{S_f}, \mathbb{Z}_2(n))) \subseteq \text{Ann}_{\mathbb{Z}_2[G]}(H^2(G_E^S, \mathbb{Z}_2(n))).
\]

Consequently by Theorem 5.8 we obtain
\[
2^{\ast_{\text{et}}(F)} \text{Ann}_{\mathbb{Z}_2[G]}(H^0(E, \mathbb{Q}_p/\mathbb{Z}_p(n)))((\pi \circ \iota \circ t_n)(G_E^S)) \subseteq \text{Ann}_{\mathbb{Z}_2[G]}(H^2_{\text{et}}(O_E, \mathbb{Z}_2(n))).
\]
Under the assumption $\mu = 0$, we have $G_S = 2^{r_1(F)} \cdot G_S^*$ for the prime 2 (cf. Lemma 3.1). Consequently $\Theta^{G_S}_{E/F}(1 - n) = 2^{r_1(F)}((\pi \circ t \circ t_n)(G_S^*))$ (cf. 3.4) for any even integer $n \geq 2$ and as a result,$$
abla_{2'(\zeta_2)}(H^0(E, \mathbb{Q}_2/\mathbb{Z}_2(n))) \cdot \Theta^{G_S}_{E}(1 - n) \subseteq \nabla_{2'(\zeta_2)}(H^0_2(E, \mathbb{Z}_2(n)))$$Finally, we note that $H^2_2(O_E', \mathbb{Z}_2(n)) \subseteq H^2_2(O_E', \mathbb{Z}_2(n))$. Hence, the 2-adic version of the Coates-Sinnott Conjecture 5.4 holds. This finishes the proof of the following result:

**Theorem 5.10.** Let $E/F$ be an abelian extension of totally real number fields with Galois group $G$, and let $n \geq 2$ be an even integer. Then the motivic version - and therefore the original version - of the Coates-Sinnott Conjecture holds under the assumptions that $\mu = 0$ and that the 2-primary part of the classical Main Conjecture in Iwasawa theory is valid.

We note that both assumptions are true if $E$ is abelian over $\mathbb{Q}$, and therefore we obtain the following unconditional result:

**Corollary 5.11.** Let $E$ be a totally real absolute abelian field. For an abelian extension $E/F$ with Galois group $G$ and even $n \geq 2$, the Coates-Sinnott Conjecture 5.3 holds.

**References**


An Equivariant Main Conjecture


Reza Taleb
Department of Mathematics
McMaster University
Hamilton, Ontario
Canada L8S 4K1
talebr@math.mcmaster.ca
A GEOMETRIC QUANTIZATION OF THE KOSTANT–SEKIGUCHI CORRESPONDENCE FOR SCALAR TYPE UNITARY HIGHEST WEIGHT REPRESENTATIONS

JAN MÖLLERS

Received: February 10, 2013
Communicated by Patrick Delorme

Abstract. For any Hermitian Lie group $G$ of tube type we give a geometric quantization procedure of certain $K_C$-orbits in $p_C^*$ to obtain all scalar type highest weight representations. Here $K_C$ is the complexification of a maximal compact subgroup $K \subseteq G$ with corresponding Cartan decomposition $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ of the Lie algebra of $G$. We explicitly realize every such representation $\pi$ on a Fock space consisting of square integrable holomorphic functions on its associated variety $\text{Ass}(\pi) \subseteq p_C^*$. The associated variety $\text{Ass}(\pi)$ is the closure of a single nilpotent $K_C$-orbit $O_{K_C} \subseteq p_C^*$ which corresponds by the Kostant–Sekiguchi correspondence to a nilpotent coadjoint $G$-orbit $O^G \subseteq \mathfrak{g}^*$. The known Schrödinger model of $\pi$ is a realization on $L^2(O)$, where $O \subseteq O^G$ is a Lagrangian submanifold. We construct an intertwining operator from the Schrödinger model to the new Fock model, the generalized Segal–Bargmann transform, which gives a geometric quantization of the Kostant–Sekiguchi correspondence (a notion invented by Hilgert, Kobayashi, Ørsted and the author).

The main tool in our construction are multivariable $I$- and $K$-Bessel functions on Jordan algebras which appear in the measure of $O_{K_C}$, as reproducing kernel of the Fock space and as integral kernel of the Segal–Bargmann transform. As a corollary to our construction we also obtain the integral kernel of the unitary inversion operator in the Schrödinger model in terms of a multivariable $J$-Bessel function as well as explicit Whittaker vectors.

2010 Mathematics Subject Classification: Primary 22E45; Secondary 17C30, 30H20, 33C70, 44A15, 46E22.
Keywords and Phrases: unitary highest weight representation, orbit method, Schrödinger model, Fock model, Jordan algebra, Segal–Bargmann transform, Bessel function, Bessel operator, unitary inversion operator, Whittaker vectors, branching law.

Introduction

The unitary highest weight representations of a semisimple Lie group \( G \) were classified by Enright–Howe–Wallach \[13\] and Jakobsen \[29\] independently (see also \[14\]). Among them are the scalar type representations which were first completely described by Berezin \[6\] for the classical groups and by Wallach \[51\] in the general case and are parameterized by the so-called Wallach set (sometimes also referred to as the Berezin–Wallach set). There are various different realizations of these representations, e.g. on spaces of holomorphic functions on a bounded symmetric domain. Rossi–Vergne \[48\] gave a realization which can by the work of Hilgert–Neeb–Ørsted \[25, 26, 27\] be understood as the geometric quantization of a nilpotent coadjoint \( G \)-orbit. More precisely, their model lives on the space \( L^2(O) \) of square integrable functions on a Lagrangian submanifold \( O \subseteq O^G \) of a nilpotent coadjoint \( G \)-orbit \( O^G \subseteq g^\ast \) in the dual of the Lie algebra \( g \) of \( G \). We refer to this realization as the Schrödinger model.

The Kostant–Sekiguchi correspondence asserts a bijection between the set of nilpotent coadjoint \( G \)-orbits in \( g^\ast \) and the set of nilpotent \( K_C \)-orbits in \( p^\ast_C \), where \( K_C \) is the complexification of a maximal compact subgroup \( K \subseteq G \) of \( G \) with corresponding Cartan decomposition \( g = \mathfrak{t} + \mathfrak{p} \). Under this bijection \( O^G \) corresponds to a nilpotent \( K_C \)-orbit \( O^K_C \) in \( p^\ast_C \). Harris–Jakobsen \[19\] obtained a realization of each scalar type unitary highest weight representation of \( g \) on regular functions on the corresponding associated variety in \( p^\ast_C \) (see also \[30\]). However, their construction is not explicit and of purely algebraic nature. They neither give a geometric construction of the invariant inner product nor do they analytically describe the representation space for the group action.

However, for the special case of the even part of the metaplectic representation, which is a highest weight representation of scalar type of \( g = \mathfrak{sp}(k, \mathbb{R}) \), a geometric quantization of the corresponding \( K_C \)-orbit is well-known. It gives the even part \( \mathcal{F}_{\text{even}}(\mathbb{C}^k) \) of the classical Fock space, consisting of even holomorphic functions on \( \mathbb{C}^k \) which are square integrable with respect to the Gaussian measure \( e^{-|z|^2} \). In this case also the intertwining operator between the Schrödinger model on \( L^2_{\text{even}}(\mathbb{R}^k) \), the space of even \( L^2 \)-functions on \( \mathbb{R}^k \), and the Fock model \( \mathcal{F}_{\text{even}}(\mathbb{C}^k) \) is explicit. It is given by the classical Segal–Bargmann transform which is the unitary isomorphism \( \mathbb{B} : L^2_{\text{even}}(\mathbb{R}^k) \rightarrow \mathcal{F}_{\text{even}}(\mathbb{C}^k) \) given by

\[
\mathbb{B}\psi(z) = e^{-\frac{1}{2}z^2} \int_{\mathbb{R}^k} e^{2zx} e^{-x^2} \psi(x) \, dx.
\]  

(0.1)

For Hermitian groups of tube type, we construct in a completely analytic and geometric way analogues of the Fock space and the Segal–Bargmann transform.
for any unitary highest weight representation of scalar type. The generalized Fock space consists of holomorphic functions on $O^{KC}$ which are square integrable with respect to a certain measure. This establishes a geometric quantization of the nilpotent $K_C$-orbit $O^{KC}$. We further find explicitly the intertwining operator between this new Fock model and the Schrödinger model, the generalized Segal–Bargmann transform. This integral transform can be understood as a geometric quantization of the Kostant–Sekiguchi correspondence, a notion invented in [23]. We remark that the special case of the minimal nilpotent $K_C$-orbit was treated in [23] which was in fact the starting point of our investigations.

Although some of our proofs are obvious generalizations of the proofs in [23] for the rank 1 case (the minimal scalar type highest weight representation), the general case requires new techniques. Whereas in [23] it was possible to work with classical one-variable Bessel functions, we have to use multivariable $J$-, $I$- and $K$-Bessel functions on Jordan algebras in the general case. These were studied before in [8, 10, 11, 12, 16], partly in a different context. We systematically investigate them further, proving asymptotic expansions, growth estimates, invariance properties, differential equations, integral formulas and their restrictions to the $K_C$-orbits $O^{KC}$. We then use the $K$-Bessel functions in the construction of the measures on the orbits $O^{KC}$ and the $I$-Bessel functions as reproducing kernels of the Fock spaces and as integral kernels of the Segal–Bargmann transforms. Additional results not studied in [23] involve a detailed analysis of corresponding $K_C$-orbits and rings of differential operators on them (Section 4.5), the intertwining operator to the bounded symmetric domain model (Section 5.2), Whittaker vectors in the Schrödinger model (Section 6.2) and applications to branching laws (Section 7). We further believe that our more general construction extends the known realizations of unitary highest weight representations in a natural and organic way.

We now explain our results in more detail. Let $G$ be a simple connected Hermitean Lie group of tube type with trivial center. Then $G$ occurs as the identity component of the conformal group of a simple Euclidean Jordan algebra $V$. Its Lie algebra $\mathfrak{g}$ acts on $V$ by vector fields up to order 2. More precisely, we have a decomposition

$$\mathfrak{g} = \mathfrak{n} + \mathfrak{l} + \mathfrak{p},$$

(0.2)

where $\mathfrak{n} \cong V$ acts on $V$ by translations, $\mathfrak{l} = \mathfrak{str}(V)$ is the structure algebra of $V$ acting by linear transformations and $\mathfrak{p}$ acts by quadratic vector fields. There is a natural Cartan involution $\vartheta$ on $G$ given in terms of the inversion in the Jordan algebra. Denote by $K = G^\vartheta$ the corresponding maximal compact subgroup of $G$ and by $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ the associated Cartan decomposition. Then $\mathfrak{k}$ has a one-dimensional center $Z(\mathfrak{k}) = \mathbb{R}Z_0$. Let $K_C$ be a complexification of $K$, acting on $p_C$. Then $p_C$ decomposes into two irreducible $K_C$-modules $p^+$ and $p^-$ and $Z_0 \in Z(\mathfrak{k})$ can be chosen such that $p^\pm$ is the $\pm i$ eigenspace of $\text{ad}(Z_0)$
on \( p_C \). We have the well-known decomposition
\[
g_C = p^- + k_C + p^+.
\] (0.3)

There is a Cayley type transform \( C \in \text{Int}(g_C) \) (see (1.5) for the precise definition) which interchanges the decompositions (0.2) and (0.3) in the sense that
\[
C(p^-) = n_C, \quad C(k_C) = l_C, \quad C(p^+) = n_C.
\]

UNITARY HIGHEST WEIGHT REPRESENTATIONS

Let \((\pi, H)\) be an irreducible unitary representation of the universal cover \( \tilde{G} \) of \( G \) and denote by \((d\pi, H_{\tilde{K}})\) its underlying \((g, \tilde{K})\)-module, \( \tilde{K} \subseteq \tilde{G} \) being the universal cover of \( K \). We say that \((\pi, H)\) or \((d\pi, H_{\tilde{K}})\) is a highest weight representation if
\[
H_{p^+ + \tilde{K}} := \{ v \in H_K : d\pi(p^+)v = 0 \} \neq 0.
\]
The space \( H_{p^+ + \tilde{K}}^+ \) is an irreducible \( \mathfrak{k} \)-representation. If further \( \dim H_{p^+ + \tilde{K}}^+ = 1 \) we say that \((\pi, H)\) is of scalar type. In this case only the center \( Z(\mathfrak{k}) = \mathbb{R}Z_0 \) of \( \mathfrak{k} \) can act nontrivially on \( H_{p^+ + \tilde{K}}^+ \) and the scalar type highest weight representations of \( \tilde{G} \) are parameterized by this scalar. More precisely, we define the W-algebra set
\[
W = \left\{ 0, \frac{d}{2}, \ldots, (r-1)d\frac{d}{2} \right\} \cup \left( (r-1)d\frac{d}{2}, \infty \right),
\]
where \( d \) is a certain root multiplicity of \( g \) and \( r \) denotes the rank of the Hermitian symmetric space \( G/K \). Then for each \( \lambda \in W \) there is exactly one scalar type highest weight representation \( H \) of \( \tilde{G} \) in which \( Z_0 \) acts on \( H_{p^+ + \tilde{K}}^+ \) by the scalar \(-i\frac{\lambda}{\sqrt{2}}\).

NILPOTENT ORBITS

We identify \( g^* \) with \( g \) and \( p_C^* \) with \( p_C \) via the Killing form of \( g_C \). By the Kostant–Sekiguchi correspondence the nilpotent adjoint \( G \)-orbits in \( g \) are in bijection with the nilpotent \( K_C \)-orbits in \( p_C \). For a certain class of orbits we make this correspondence more precise. The analytic subgroup \( L \) of \( G \) with Lie algebra \( l \) acts on \( V \cong n \) via the adjoint representation. It has finitely many orbits, among them an open orbit \( \Omega \subseteq V \) (the orbit of the identity element in the Jordan algebra) which is a symmetric cone. Its boundary has a stratification into lower dimensional orbits \( O_0, \ldots, O_{r-1} \) with \( O_k \subseteq O_{k+1} \). Put \( O_r := \Omega \). These orbits in \( n \) generate nilpotent adjoint \( G \)-orbits
\[
O_k^G := G \cdot O_k \subseteq g, \quad k = 0, \ldots, r.
\]
Via the Kostant–Sekiguchi correspondence these $G$-orbits correspond to nilpotent $K_C$-orbits $O_k^{K_C}, \ldots, O_r^{K_C}$ in $p^+ \subseteq p_C$. Under the Cayley type transform $C \in \text{Int}(g_C)$ these $K_C$-orbits map to

$$X_k := C(O_k^{K_C}) \subseteq \mathfrak{p}_C \cong V_C, \quad k = 0, \ldots, r.$$ 

Since $C(t_C) = t_C$ each $X_k$ is an $L_C$ orbit in $V_C$, where $L_C$ is a complexification of $L$, acting linearly on $V_C$. Under the identifications described above $O_k \subseteq X_k$ is a totally real submanifold.

**The Schrödinger model**

Rossi–Vergne [18] showed that the scalar type unitary highest weight representation corresponding to the Wallach point $\lambda \in W$ can be realized on a space of $L^2$-functions on the orbit $O_k$ for $\lambda = k\frac{\lambda}{2}$, $k = 0, \ldots, r - 1$, and on the orbit $O_r = \Omega$ for $\lambda > (r - 1)\frac{\lambda}{2}$. For convenience we put $O_\lambda := O_k$ where either $\lambda = k\frac{\lambda}{2}$, $0 \leq k \leq r - 1$, or $k = r$ for $\lambda > (r - 1)\frac{\lambda}{2}$. For each $\lambda \in W$ there is a certain $L$-equivariant measure $d\mu_\lambda$ on $O_\lambda$ (see Section 1.2 for details). Then the natural irreducible unitary representation of $L \rtimes \text{exp}(\mathfrak{n})$ on $L^2(O_\lambda, d\mu_\lambda)$ extends to an irreducible unitary representation $\pi_\lambda$ of $\tilde{G}$ which is the scalar type highest weight representation belonging to the Wallach point $\lambda \in W$. The differential representation $d\pi_\lambda$ of $\mathfrak{g}$ in this realization was computed in [22] and is given by differential operators on $O_\lambda$ up to order 2 with polynomial coefficients. A special role is played by the second order vector-valued Bessel operator $\mathcal{B}_\lambda$ which gives the differential action of $\pi$ (see Section 1.3 for its definition). This operator was first introduced by Dib [10] and studied further in [22].

**Bessel functions on Jordan algebras**

For each $\lambda \in W$ we construct Bessel functions $J_\lambda(z, w)$ and $I_\lambda(z, w)$ for $z, w \in X_\lambda := L_C \cdot O_\lambda$ (see Sections 3.1 and 3.2). Both $J_\lambda(z, w)$ and $I_\lambda(z, w)$ can be extended to functions on $V_C \times X_\lambda$ which are holomorphic in $z$ and antiholomorphic in $w$. They solve the differential equations

$$(\mathcal{B}_\lambda)z J_\lambda(z, w) = -\overline{\pi} J_\lambda(z, w), \quad (\mathcal{B}_\lambda)z I_\lambda(z, w) = \overline{\pi} I_\lambda(z, w).$$

For $w = e$ the identity element and $\lambda > (r - 1)\frac{\lambda}{2}$ the corresponding one-variable functions $J_\lambda(x) = J_\lambda(x, e)$ and $I_\lambda(x) = I_\lambda(x, e)$ on $\Omega$ were first studied by Herz [21] for real symmetric matrices and later for the more general case in [10][11][12][15][16]. We further define a $K$-Bessel function $K_\lambda(x)$ on $O_\lambda$ for every $\lambda \in W$ which solves the same differential equation as $I_\lambda(x)$, but has a different asymptotic behaviour (see Section 3.3). These functions were introduced by Dib [10] and studied further in [8][15].
The Fock space

The closure of $\mathcal{X}_k \cong \mathcal{O}_k^{\mathbb{C}}$ is an affine algebraic variety and hence the space $\mathbb{C}[\mathcal{X}_k]$ of regular functions on $\mathcal{X}_k$ equals the space $\mathcal{P}(\mathcal{X}_k)$ of restrictions of holomorphic polynomials on $V_{\mathbb{C}}$ to $\mathcal{X}_k$. We construct measures on $\mathcal{X}_k$ in two steps. Fix $\lambda \in W$. First the $L_{\mathbb{C}}$-equivariant measure $d\mu_\lambda$ on $\mathcal{O}_\lambda \subseteq \mathcal{X}_\lambda$ gives rise to an $L_{\mathbb{C}}$-equivariant measure $d\nu_\lambda$ on $\mathcal{X}_\lambda$ for every $\lambda \in W$. In the second step we extend the function $\omega_\lambda(x) := K_\lambda \left( \frac{x}{2} \right)$, $x \in \mathcal{O}_\lambda$, uniquely to a positive density on $\mathcal{X}_\lambda = L_{\mathbb{C}} \cdot \mathcal{O}_\lambda$ which, under the identification of $\mathcal{X}_\lambda$ with a $K_{\mathbb{C}}$-orbit, is invariant under the action of $K$. Then every polynomial $F \in \mathcal{P}(\mathcal{X}_\lambda)$ is square integrable on $\mathcal{X}_\lambda$ with respect to the measure $\omega_\lambda d\nu_\lambda$ and we let $F_\lambda$ be the closure of $\mathcal{P}(\mathcal{X}_\lambda)$ with respect to the inner product $(F, G)_{\mathcal{F}_\lambda} := \text{const} \cdot \int_{\mathcal{X}_\lambda} F(z) \overline{G(z)} \omega_\lambda(z) d\nu_\lambda(z), \quad F, G \in \mathcal{P}(\mathcal{X}_\lambda)$.

**Theorem A** (see Propositions 4.4 & Theorems 4.15 & 5.7). For each $\lambda \in W$ the space $F_\lambda$ is a Hilbert space of holomorphic functions on $\mathcal{X}_\lambda$ with reproducing kernel $K_\lambda(z, w) = I_\lambda \left( \frac{z}{2}, \frac{w}{2} \right)$, $z, w \in \mathcal{X}_\lambda$.

We further have the following intrinsic description:

$\mathcal{F}_\lambda = \left\{ F|_{\mathcal{X}_\lambda} : F : V_{\mathbb{C}} \to \mathbb{C} \text{ holomorphic and } \int_{\mathcal{X}_\lambda} |F(z)|^2 \omega_\lambda(z) d\nu_\lambda(z) < \infty \right\}.$

In Section 4.2 we prove that the inner product on $\mathcal{F}_\lambda$ can also be expressed as a Fischer type inner product involving the Bessel operator $B_\lambda$.

Unitary action on the Fock space

Complexifying the differential operators $d\pi_\lambda(X), X \in \mathfrak{g}$, in the differential representation of the Schrödinger model one obtains a Lie algebra representation $d\pi^{\mathbb{C}}_\lambda$ of $\mathfrak{g}$ on holomorphic functions on $\mathcal{X}_\lambda$ by polynomial differential operators (see Section 2.2 for details). We define a $\mathfrak{g}$-module structure on $\mathcal{P}(\mathcal{X}_\lambda)$ by composing this action with the Cayley type transform $C \in \text{Int}(\mathfrak{g}_{\mathbb{C}})$:

$d\rho_\lambda(X) := d\pi^{\mathbb{C}}_\lambda(C(X)), \quad X \in \mathfrak{g}.$

**Theorem B** (see Propositions 4.9, 4.10 & 4.17 & Theorem 4.11). For each $\lambda \in W$ the $\mathfrak{g}$-module $(d\rho_\lambda, \mathcal{P}(\mathcal{X}_\lambda))$ is an irreducible $(\mathfrak{g}, \mathbb{K})$-module which is infinitesimally unitary with respect to the inner product on $\mathcal{F}_\lambda$. It integrates to an irreducible unitary representation $\rho_\lambda$ of $\mathbb{K}$ on $\mathcal{F}_\lambda$ with associated variety $\text{Ass}(\rho_\lambda) = \mathcal{O}_k^{\mathbb{C}}$, where $k \in \{0, \ldots, r\}$ is such that $\mathcal{X}_\lambda = \mathcal{X}_k$. 
The group action of $\tilde{K}$ in the Fock model is induced by the geometric action of $K_C$ on the orbits $X_k \cong O_k^{K_C}$ up to multiplication with a character (see Proposition 4.12). Therefore the $K$-type decomposition of $\rho_\lambda$ equals the decomposition of the space of polynomials on $O_k^{K_C} \subseteq p^+$ with respect to the natural $K_C$-action which is essentially the Hua–Kostant–Schmid decomposition (see Proposition 4.8).

In Section 4.5 we use the Fock model to give explicit generators for the rings $D(X_k)$ of differential operators on the singular affine algebraic varieties $X_k \subseteq V_C$, $k = 0, \ldots, r - 1$, in terms of the Bessel operator $B_\lambda$.

Although there are already known explicit realizations for highest weight representations on spaces of holomorphic functions our construction has a certain advantage. In the realization on holomorphic functions on the corresponding bounded symmetric domain $G/K$ the group action is quite explicit, but the representation space as well as the inner product are defined in a rather abstract way, at least for the singular highest weight representations. In our Fock model the representation space is by Theorem A explicitly given by holomorphic functions on $V_C$ which are square integrable with respect to a certain measure on the orbit $\lambda$. Although our group action is less explicit we still have a useful expression of the Lie algebra action in terms of the Bessel operators which allows e.g. the study of discrete branching laws as done in Theorem E.

**The Segal–Bargmann transform**

The unitary representations $(\pi_\lambda, L^2(O_\lambda, d\mu_\lambda))$ and $(\rho_\lambda, F_\lambda)$ are isomorphic and we explicitly find the integral kernel of the intertwining operator which resembles the classical Segal–Bargmann transform (0.1). For this let $\text{tr} \in V_C^*$ denote the Jordan trace of the complex Jordan algebra $V_C$ and recall the $I$-Bessel function $I_\lambda(z,w)$.

**Theorem C** (see Proposition 5.1 & Theorem 5.4). Let $\lambda \in W$, then for $\psi \in L^2(O_\lambda, d\mu_\lambda)$ the formula

$$B_\lambda \psi(z) = e^{-\frac{1}{2}\text{tr}(z)} \int_{O_\lambda} I_\lambda(z,x) e^{-\text{tr}(x)} \psi(x) d\mu_\lambda(x), \quad z \in X_\lambda,$$

defines a function $B_\lambda \psi \in F_\lambda$. This gives a unitary isomorphism $B_\lambda : L^2(O_\lambda, d\mu_\lambda) \rightarrow F_\lambda$ which intertwines the representations $\pi_\lambda$ and $\rho_\lambda$.

The $\ell$-finite vectors in the Fock model and in the bounded symmetric domain model have the same realization as polynomials on $V_C$ and therefore also the intertwiner between these two models is worth studying. In Theorem 5.10 we find an explicit formula for it in terms of its integral kernel. We remark that Brylinski–Kostant [7] construct a Fock space realization for minimal representations of non-Hermitian Lie groups as geometric quantization of the minimal nilpotent $K_C$-orbit in $p_C^*$ (see also [2]). Apart from the fact...
that their cases are disjoint to ours they do not construct an intertwining
operator between the known Schrödinger model and their Fock model such
as our generalized Segal–Bargmann transform. Further their measure on the
$K_C$-orbit is not positive whereas our measure is explicitly given in terms of
the $K$-Bessel function and hence positive. Moreover, the Lie algebra action in
their picture is given by pseudodifferential operators while in our model the Lie
algebra acts by differential operators up to second order.

Recently Achab [1] also constructed a Fock space realization for the minimal
representation of Hermitian groups. Her construction looks different from ours
and it would be interesting to find a connection between her model and our
model.

The unitary inversion operator

The explicit integral formula for the Segal–Bargmann transform can be used to
find the integral kernel of the unitary inversion operator $U_\lambda$ in the Schrödinger
model. This operator $U_\lambda$ is a unitary automorphism on $L^2(O_\lambda, d\mu_\lambda)$ of order
2 (see Section 6 for its precise definition). The action of $U_\lambda$ together with the
relatively simple action of a maximal parabolic subgroup generates the whole
group action in the Schrödinger model. To describe its integral kernel recall
the $J$-Bessel function $J_\lambda(z, w)$.

**Theorem D** (see Theorem 6.3). For each $\lambda \in \mathcal{W}$ the unitary inversion operator
$U_\lambda$ is given by

$$
U_\lambda \psi(x) = 2^{-r\lambda} \int_{O_\lambda} J_\lambda(x, y) \psi(y) \, d\mu_\lambda(y), \quad \psi \in C^\infty_c(O_\lambda).
$$

for the relative holomorphic discrete series, i.e. $\lambda$ is contained in the con-
tinuous part of the Wallach set and is ‘large enough’. (In fact Ding–Gross
proved an analogous version of Theorem D for all holomorphic discrete series
representations.) For $g = \mathfrak{su}(k, k)$ and $\lambda$ in the discrete part of the Wallach
For $g = \mathfrak{so}(2, n)$ and $\lambda$ the minimal non-zero discrete Wallach point Kobayashi–
Mano gave two different proofs for this result (see [33] Theorem 6.1.1] and [34]
Theorem 5.1.1]). The case of the minimal non-zero discrete Wallach point was
systematically treated in [23]. The proof we present for the general case is
quite simple, once the Bargmann transform is established, and works in the
same way for all scalar type unitary highest weight representations.

In Section 6.2 we use the integral kernel of $U_\lambda$ to construct explicit algebraic
and smooth Whittaker vectors in the Schrödinger model.
Application to branching laws

We demonstrate the use of our Fock model in one specific example, the branching $\mathfrak{so}(2, n) \rightarrow \mathfrak{so}(2, m) \oplus \mathfrak{so}(n - m)$ of the minimal unitary highest weight representation of $\mathfrak{so}(2, n)$. Note that the subalgebra $\mathfrak{h} := \mathfrak{so}(2, m) \oplus \mathfrak{so}(n - m)$ of $\mathfrak{g} = \mathfrak{so}(2, n)$ is symmetric and of holomorphic type, i.e. the corresponding involution of $\mathfrak{g}$ acts as the identity on the center of $k$. Kobayashi [32] proved that any highest weight representation of scalar type if restricted to a holomorphic type subalgebra is discretely decomposable and the decomposition is multiplicity free. He further determined the explicit decomposition for representations in the relative holomorphic discrete series. For arbitrary highest weight representation he gives a general algorithm to find the explicit branching law. We find the explicit decomposition for the non-zero discrete Wallach point by an easy computation in the Fock model. For $j = m, n$ we denote by $d_{\pi}^{\mathfrak{so}(2,j)}$ the unitary highest weight representation of $\mathfrak{so}(2, j)$ of scalar type with Wallach parameter $\lambda$ as constructed above. Further we let $\mathcal{H}^k(\mathbb{R}^{n-m})$ be the irreducible representation of $\mathfrak{so}(n - m)$ on the space of spherical harmonics on $\mathbb{R}^{n-m}$ of degree $k$.

**Theorem E** (see Theorem 7.2). The unitary highest weight representation of $\mathfrak{g} = \mathfrak{so}(2, n)$ belonging to the minimal non-zero discrete Wallach point $\lambda = \frac{n-2}{2}$ decomposes under restriction to the subalgebra $\mathfrak{h} = \mathfrak{so}(2, m) \oplus \mathfrak{so}(n - m)$ as follows:

$$d_{\pi}^{\mathfrak{so}(2,n)} = \bigoplus_{k=0}^{\infty} d_{\pi}^{\mathfrak{so}(2,m)} \boxtimes \mathcal{H}^k(\mathbb{R}^{n-m}).$$

We remark that for odd $n$ the representation $d_{\pi}^{\mathfrak{so}(2,n)}$ cannot be obtained via the theta correspondence and hence also the branching law cannot be obtained in this way.

For even $n$ this branching law was proved by Kobayashi–Ørsted in [35, Theorem A] and, as pointed out to the author by B. Ørsted, their proof actually carries over also to the case of odd $n$, although it is not explicitly stated in their paper. However, the techniques they use are more involved than the computation we do in Section 7 to prove Theorem E.

In the general case of a Hermitian Lie algebra $\mathfrak{g}$ Kobayashi–Oshima [36] recently classified all symmetric subalgebras of $\mathfrak{g}$ such that the restriction of the minimal scalar type unitary highest weight representation of $\mathfrak{g}$ is discretely decomposable. In [42] the explicit decompositions in these cases are determined.

**Organization of the paper**

This paper is organized as follows: In Section I we recall the necessary background of Euclidean Jordan algebras, their corresponding groups and nilpotent orbits. We also study polynomials and differential operators such as the Bessel operators on Jordan algebras. Section II describes three known
models for unitary highest weight representations of scalar type, the bounded symmetric domain model, the tube domain model and the Schrödinger model. We further investigate a natural complexification of the latter one. Section 3 deals with Bessel functions on Jordan algebras. We introduce $J$, $I$- and $K$-Bessel functions and discuss their properties. In Section 4 the Fock model for unitary highest weight representations of scalar type is constructed. We calculate the reproducing kernel of the Fock space and investigate rings of differential operators on the associated varieties. The Segal–Bargmann transform intertwining the Fock model and the Schrödinger model is introduced in Section 5. Here we also give a formula for the intertwiner between Fock model and bounded symmetric domain model. In Section 6 we use the Bargmann transform to obtain the integral kernel of the unitary inversion operator. We then describe Whittaker vectors in the Schrödinger model in terms of this integral kernel. Finally we use the Fock model in Section 7 to obtain in a very simple way the branching law for the smallest non-zero highest weight representation of $\mathfrak{so}(2, n)$ restricted to $\mathfrak{so}(2, m) \oplus \mathfrak{so}(n - m)$.

Acknowledgements: We thank T. Kobayashi and B. Ørsted for helpful discussions on the topic of this paper.

Notation: $\mathbb{N} = \{1, 2, 3, \ldots\}$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathbb{R}^x = \mathbb{R} \setminus \{0\}$, $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$, $\mathbb{C}^x = \mathbb{C} \setminus \{0\}$, $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.

$V^*$: algebraic dual of a vector space $V$, $V'$: topological dual of a topological vector space $V$.

1 Preliminaries

In this section we set up the notation and recall the construction of three known models for unitary highest weight representations of Hermitian Lie groups of tube type. For this we use the theory of Euclidean Jordan algebras which formalizes various aspects in a simple fashion. We mostly follow the notation of [15].

1.1 Simple Euclidean Jordan algebras

Let $V$ be a simple Euclidean Jordan algebra of dimension $n = \dim V$ with unit element $e$. For $x \in V$ we denote by $L(x) \in \text{End}(V)$ the multiplication by $x$ and by $P(x) = 2L(x)^2 - L(x^2)$ the quadratic representation. Its polarized version is given by

$$P(x, y) = L(x)L(y) + L(y)L(x) - L(xy), \quad x, y \in V.$$ 

For $x, y \in V$ we also define the box operator $x \square y \in \text{End}(V)$ by

$$x \square y := L(xy) + [L(x), L(y)].$$
and note the identity \((x \sqcap y)z = P(x, z)y\) for \(x, y, z \in V\).

The generic minimal polynomial of \(V\) provides the Jordan trace \(\text{tr} \in V^*\) and the Jordan determinant \(\Delta\), a homogeneous polynomial of degree \(r = \text{rk} V\), the rank of \(V\). Both can be expressed in terms of \(L(x)\) and \(P(x)\) (see [15 Proposition III.4.2]):

\[
\text{Tr} \, L(x) = \frac{n}{r} \text{tr}(x), \quad \text{Det} \, P(x) = \Delta(x)^{\frac{2n}{r}},
\]

for \(x \in V\) (\(\frac{n}{r}\) is always an integer). Here we use \(\text{Tr}\) and \(\text{Det}\) for the trace and determinant of a linear operator. Using the Jordan trace we define the trace form

\[
(x \mid y) := \text{tr}(xy), \quad x, y \in V,
\]

which is positive definite since \(V\) is Euclidean. Both \(L(x)\) and \(P(x)\) are symmetric operators with respect to the trace form (see [15 Proposition II.4.3]). For any orthonormal basis \((e_\alpha)_\alpha \subseteq V\) with respect to the trace form we have the identity (see [15 Exercise III.6])

\[
\sum_\alpha e_\alpha^2 = \frac{n}{r}.
\]  

We use the trace form to normalize the Lebesgue measure \(dx\) on \(V\). The trace form also defines a norm on \(V\) by

\[ |x| := \sqrt{(x \mid x)}, \quad x \in V. \]

Let \(c \in V\) be an idempotent, i.e. \(c^2 = c\). The only possible eigenvalues of the symmetric operator \(L(c)\) are \(1, \frac{1}{2}, 0\). Denote by \(V(c, 1)\), \(V(c, \frac{1}{2})\) and \(V(c, 0)\) the corresponding eigenspaces. The decomposition

\[ V = V(c, 1) \oplus V(c, \frac{1}{2}) \oplus V(c, 0) \]

is called Peirce decomposition. The subspaces \(V(c, 1)\) and \(V(c, 0)\) are themselves simple Euclidean Jordan algebras with identity elements \(c\) and \(e - c\), respectively. An idempotent \(c \in V\) is called primitive if it is non-zero and cannot be written as a nontrivial sum of two idempotents. Further, two idempotents \(c_1, c_2 \in V\) are called orthogonal if \(c_1c_2 = 0\). (This implies \((c_1|c_2) = 0\).) A set of primitive orthogonal idempotents which add up to the identity \(e\) is called a Jordan frame. The cardinality of every Jordan frame is equal to the rank \(r\) of \(V\). Further we say that an idempotent \(c \in V\) has rank \(k\) if it can be written as the sum of \(k\) primitive orthogonal idempotents. From now on we fix a Jordan frame \(c_1, \ldots, c_r\). The symmetric operators \(L(c_1), \ldots, L(c_r)\) commute pairwise and therefore we obtain the mutual Peirce decomposition

\[ V = \bigoplus_{1 \leq i \leq j \leq r} V_{ij}, \]
where
\[ V_{ii} = V(c_i, 1), \quad (1 \leq i \leq r), \quad V_{ij} = V(c_i, \frac{1}{2}) \cap V(c_j, \frac{1}{2}), \quad (1 \leq i < j \leq r). \]
The spaces $V_{ii}$ are one-dimensional and spanned by $c_i$ whereas the spaces $V_{ij}$, $1 \leq i < j \leq r$ have a common dimension $d$. In particular, this implies the dimension formula
\[ \frac{n}{r} = 1 + (r - 1)\frac{d}{2}. \]

1.2 The structure group and its orbits

The structure group $\text{Str}(V)$ of $V$ is defined to be the group of $g \in \text{GL}(V)$ such that
\[ P(gx) = gP(x)g^\# \quad x \in V, \]
where $g^\#$ denotes the adjoint with respect to the trace form. This is by [15] Exercise VIII.5] equivalent to the existence of a scalar $\chi(g) \in \mathbb{R}^\times$ such that
\[ \Delta(gx) = \chi(g)\Delta(x) \quad x \in V. \]
The map $\chi : \text{Str}(V) \rightarrow \mathbb{R}^\times$ defines a character of $\text{Str}(V)$ which is on the identity component $L := \text{Str}(V)_0$ given by (see [15] Proposition III.4.3)]
\[ \chi(g) = \text{Det}(g)^{\frac{1}{n}}, \quad g \in L. \]

Let $O(V)$ denote the orthogonal group of $V$ with respect to the trace form. Then $K^L := L \cap O(V)$ is a maximal compact subgroup of $L$, the corresponding Cartan involution being $\vartheta(g) = (g^{-1})^\#$. The group $K^L$ coincides with the identity component of the group of Jordan algebra automorphisms, i.e.
\[ K^L = \{ g \in L : g(x \cdot y) = gx \cdot gy \forall x, y \in V \}. \]

At the same time $K^L$ is the stabilizer subgroup of the unit element $e$, i.e.
\[ K^L = \{ g \in L : ge = e \}. \]

The Lie algebra $\mathfrak{l} = \mathfrak{str}(V)$ of $L$ has the Cartan decomposition (see [15] Proposition VIII.2.6)]
\[ \mathfrak{l} = \mathfrak{k}^L \oplus \mathfrak{p}^L, \]
where $\mathfrak{k}^L$ is the Lie algebra of $K^L$ consisting of all derivations of $V$ and $\mathfrak{p}^L$ is the space of multiplication operators:
\[ \mathfrak{k}^L = \{ D \in \mathfrak{gl}(V) : D(x \cdot y) = Dx \cdot y + x \cdot Dy, \forall x, y \in V \}, \]
\[ \mathfrak{p}^L = L(V) = \{ L(x) : x \in V \}. \]
The group $L$ acts on $V$ with finitely many orbits. The orbit $\Omega := L \cdot e$ is a symmetric cone and is isomorphic to the Riemannian symmetric space $L/K_L$. The Lebesgue measure $dx$ on $V$ clearly restricts to $\Omega$. The following integral formula due to J.-L. Clerc [8, Proposition 2.7] is used in Section 3.3 to calculate $K$-Bessel functions on boundary orbits:

**Lemma 1.1.** Let $c \in V$ be an idempotent of rank $k$ and let $\Omega_1$ and $\Omega_0$ denote the symmetric cones in the Euclidean subalgebras $V(c, 1)$ and $V(c, 0)$, respectively. Further we let $\Delta^0(x)$ be the determinant function of the Euclidean Jordan algebra $V(c, 0)$. Then the following integral formula holds:

$$\int_{\Omega} f(x) \, dx = 2^{-(r-k)d} \int_{\Omega_1} \int_{V(c, \frac{d}{2})} \int_{\Omega_0} f(\exp(c \Box x_\frac{d}{2})(x_1 + x_0)) \Delta^0(x_0)^{kd} \, dx_1 \, dx_0 \, dx_1.$$

The boundary $\partial \Omega$ of the symmetric cone $\Omega$ has a stratification into lower dimensional orbits. In fact

$$\Omega = \mathcal{O}_0 \cup \mathcal{O}_1 \cup \ldots \cup \mathcal{O}_r,$$

where

$$\mathcal{O}_j := L \cdot e_j \quad \text{with} \quad e_j = c_1 + \cdots + c_j.$$  

(We use the convention $e_0 = 0$ here.) Note that $\mathcal{O}_r = \Omega$. Each $x \in \mathcal{O}_j$ has a polar decomposition $x = ka$ with $k \in K_L$ and $a = \sum_{i=1}^{j} a_i c_i$, $a_1, \ldots, a_j > 0$ (see [15, Chapter VI.2]). Assuming $a_1 \geq \ldots \geq a_j > 0$ the element $a$ is unique.

We define complex powers of $x = k a \in \mathcal{O}_r$ by $x^s := k a^s$, $a^s = \sum_{i=1}^{j} a_i^s c_i$ for $s \in \mathbb{C}$. Note that for $s \in \mathbb{N}$ this definition agrees with the usual definition of powers by the Jordan algebra multiplication. Since every $k \in K_L$ is a Jordan algebra automorphism the identity $x^s \cdot x^t = x^{s+t}$ holds for $s, t \in \mathbb{C}$.

The Gamma function of $\Omega$ is for $\Re \lambda > (r - 1)\frac{d}{2}$ defined by the absolutely converging integral

$$\Gamma_{\Omega}(\lambda) := \int_{\Omega} e^{-\operatorname{tr}(x)} \Delta(x)^{\lambda - \frac{d}{2}} \, dx$$

and is extended meromorphically to the whole complex plane by the identity (see [15, Theorem VII.1.1])

$$\Gamma_{\Omega}(\lambda) = (2\pi)^{-\frac{d^2}{2}} \prod_{j=1}^{r} \Gamma \left( \lambda - (j - 1)\frac{d}{2} \right).$$

Using the Gamma function we define the Riesz distributions $R_\lambda \in S'(V)$ by

$$\langle R_\lambda, \varphi \rangle := \frac{2^r \lambda}{\Gamma_{\Omega}(\lambda)} \int_{\Omega} \varphi(x) \Delta(x)^{\lambda - \frac{d}{2}} \, dx, \quad \varphi \in S(V).$$
This integral converges for \( \lambda > (r-1)\frac{d}{2} \) and defines a tempered distribution \( R_\lambda \in \mathcal{S}'(V) \) which has an analytic continuation to all \( \lambda \in \mathbb{C} \) (see [15 Proposition VII.2.1 & Theorem VII.2.2]). The distribution \( R_\lambda \) is positive if and only if \( \lambda \) belongs to the Wallach set (sometimes referred to as the Berezin–Wallach set, see [15, Theorem VII.3.1])

\[
W = \left\{ 0, \frac{d}{2}, \ldots, (r-1)\frac{d}{2} \right\} \cup \left( (r-1)\frac{d}{2}, \infty \right).
\]

Note that the Wallach set consists of a discrete part \( W_{\text{disc}} \) and a continuous part \( W_{\text{cont}} \)

\[
W_{\text{disc}} = \left\{ 0, \frac{d}{2}, \ldots, (r-1)\frac{d}{2} \right\}, \quad W_{\text{cont}} = \left( (r-1)\frac{d}{2}, \infty \right).
\]

For \( \lambda \in W_{\text{cont}} \) in the continuous part the distribution \( R_\lambda \) is an \( L \)-equivariant measure \( d\mu_\lambda \) supported on \( \Omega \) which is absolutely continuous with respect to the Lebesgue measure \( dx \). Hence the boundary \( \partial \Omega = \overline{O}_{r-1} \) is a set of measure zero. For \( \lambda = k\frac{d}{2} \in W_{\text{disc}} \) in the discrete part the distribution \( R_\lambda \) is an \( L \)-equivariant measure \( d\mu_\lambda \) supported on \( O_k \) for which \( \partial O_k = \overline{O}_{k-1} \) is a set of measure zero (see [15, Proposition VII.2.3]). If we put

\[
O_\lambda := \begin{cases} 
O_k & \text{for } \lambda = k\frac{d}{2}, \ 0 \leq k \leq r-1, \\
\Omega & \text{for } \lambda > (r-1)\frac{d}{2},
\end{cases}
\]

then we obtain measure spaces \( (O_\lambda, d\mu_\lambda) \) for \( \lambda \in W \). The \( L \)-equivariant measures \( d\mu_\lambda \) transform by

\[
d\mu_\lambda(gx) = \chi(g)\lambda^\lambda \ d\mu(x), \quad g \in L.
\]

Note that with this normalization of the measures \( d\mu_\lambda \) the function \( \psi_0(x) = e^{-\text{tr}(x)} \) always has norm 1 in \( L^2(O_\lambda, d\mu_\lambda) \). In fact, for \( \lambda > (r-1)\frac{d}{2} \) we find

\[
\int_{O_\lambda} |\psi_0(x)|^2 \ d\mu_\lambda(x) = \frac{2\pi^\lambda}{\Gamma(\Omega(\lambda))} \int_{\Omega} e^{-2\text{tr}(x)} \Delta(x) \lambda^\lambda \ d\mu = 1,
\]

and the general case follows by analytic continuation.

In terms of the polar decomposition of each orbit \( O_k \) we have the following integral formula (cf. [22 Proposition 1.8])

\[
\int_{O_\lambda} f(x) \ d\mu_\lambda(x) = \text{const} \cdot \int_{R^k} \int_{t_1, \ldots, t_k} f(ke^{t})J_\lambda(t) \ dt \ dk, \quad (1.3)
\]

where we use the notation \( t = (t_1, \ldots, t_k) \),

\[
e^t = \sum_{i=1}^{k} e^{t_i} c_i, \quad J_\lambda(t) = e^{d\lambda} \sum_{j=1}^{k} t_j \prod_{1 \leq i < j \leq k} \sinh^d \left( \frac{t_i - t_j}{2} \right), \quad (1.4)
\]

and \( k \in \{0, \ldots, r\} \) such that \( O_\lambda = O_k \).
1.3 Conformal group and KKT algebra

For \( a \in V \) let \( n_a \) be the translation by \( a \), i.e.

\[ n_a(x) = x + a, \quad x \in V. \]

Denote by \( N := \{ n_a : a \in V \} \) the group of all translations. Further define the conformal inversion \( j \) by

\[ j(x) := -x^{-1}, \quad \text{for } x \in V \text{ invertible.} \]

The map \( j \) defines a rational transformation of \( V \) of order 2. The conformal group \( \text{Co}(V) \) of \( V \) is defined as the subgroup of all rational transformations of \( V \) generated by \( N \), \( \text{Str}(V) \) and \( j \):

\[ \text{Co}(V) := \langle N, \text{Str}(V), j \rangle_{\text{grp}}. \]

We let \( G := \text{Co}(V)_0 \) be its identity component. \( G \) is a connected simple Hermitian Lie group of tube type with trivial center and every such group occurs in this fashion (see Table 1 for a classification).

Conjugation with \( j \) defines a Cartan involution \( \vartheta \) of \( G \) by

\[ \vartheta(g) = j \circ g \circ j, \quad g \in G. \]

Its restriction to \( L \) agrees with the previously introduced Cartan involution \( \vartheta(g) = (g^{-1})^\# \), \( g \in L \). Let \( K := G^\theta \) be the corresponding maximal compact subgroup of \( K \).

The Lie algebra \( g \) of \( G \) is called Kantor–Koecher–Tits algebra. It acts on \( V \) by quadratic vector fields of the form

\[ X(z) = u + Tz - P(z)v, \quad z \in V, \]

with \( u, v \in V \) and \( T \in l \). For convenience we write \( X = (u, T, v) \in g \) for short. In this notation the Lie bracket of \( X_j = (u_j, T_j, v_j) \), \( j = 1, 2 \), is given by (see [15, Proposition X.5.8])

\[ [X_1, X_2] = (T_1 u_2 - T_2 u_1, [T_1, T_2] + 2(u_1 \square v_2) - 2(u_2 \square v_1), -T_1^\# v_2 + T_2^\# v_1). \]

This yields the grading

\[ g = n \oplus l \oplus \Pi, \]

where \( n \cong V \) is the Lie algebra of \( N \) and \( \Pi = \vartheta n \). Thus,

\[ n = \{(u, 0, 0) : u \in V\} \]

acts via constant vector fields, \( l \) via linear vector fields and

\[ \Pi = \{(0, 0, v) : v \in V\}. \]
by quadratic vector fields. Note that the abelian subalgebras \( \mathfrak{n} \) and \( \mathfrak{p} \) together 
generate \( \mathfrak{g} \) as a Lie algebra.

On \( \mathfrak{g} \) the Cartan involution \( \vartheta \) acts via

\[
\vartheta(u, D + L(a), v) = (-v, D - L(a), -u)
\]

and hence gives the Cartan decomposition \( \mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p} \) with

\[
\mathfrak{k} = \{(u, D, -u) : u \in V, D \in \mathfrak{t}^1\}, \quad \mathfrak{p} = \{(u, L(a), u) : u, a \in V\}.
\]

The Lie algebra \( \mathfrak{t} \) of \( K \) has a one-dimensional center (see [11 Lemma 1.6.2])

\[
Z(\mathfrak{t}) = \mathbb{R}Z_0, \quad Z_0 = -\frac{1}{2}(e, 0, -e),
\]

and hence the universal cover \( \tilde{K} \) of \( K \) is the direct product of \( \mathbb{R} \) with a compact
connected simply-connected semisimple group. Therefore the deck transformation group of the universal cover \( \tilde{G} \) of \( G \) is a finite extension of \( \mathbb{Z} \).

Under the action of \( K \) the space \( p \) decomposes into two irreducible components

\[
p = p^+ \oplus p^-.
\]

In fact, \( p^\pm \) is the \( \pm i \) eigenspace of \( \text{ad}(Z_0) \) on \( p \) and is explicitly given by

\[
p^\pm = \{(u, \pm 2iL(u), u) : u \in V\}.
\]

We set

\[
E := (e, 0, 0), \quad H := (0, 2 \text{id}, 0), \quad F := (0, 0, e).
\]

Then \( (E, F, H) \) forms an \( \mathfrak{sl}_2 \)-triple in \( \mathfrak{g} \). We define a Cayley type transform
\( C \in \text{Int}(\mathfrak{g}_C) \) by the formula

\[
C := \exp(-\frac{1}{2}i \text{ad}(E)) \exp(-i \text{ad}(F)). \tag{1.5}
\]

It is then routine to check the following formulas:

\[
C(a, 0, 0) = (\frac{1}{4}, iL(a), a), \quad C(0, L(a) + D, 0) = (i\frac{a}{4}, D, -ia), \quad C(0, 0, a) = (\frac{1}{4}, -iL(a), a). \tag{1.6-1.8}
\]

From these formulas it is easy to see that the transform \( C \) induces isomorphisms

\[
\mathfrak{k}_C \rightarrow \mathfrak{k}_C, \quad (u, D, -u) \mapsto D + 2iL(u),
\]

\[
\mathfrak{p}^+ \rightarrow \mathfrak{p}^-C, \quad (u, 2iL(u), u) \mapsto (0, 0, 4u),
\]

\[
\mathfrak{p}^- \rightarrow \mathfrak{n}_C, \quad (u, -2iL(u), u) \mapsto (u, 0, 0).
\]

Since \( \mathfrak{n} \) and \( \mathfrak{p} \) together generate \( \mathfrak{g} \) as a Lie algebra it is now immediate that
\( \mathfrak{p}^+ + \mathfrak{p}^- \) generates \( \mathfrak{g}_C \) as a Lie algebra. Hence the real form \( \{(a, L(b), a) : a, b \in V\} \) of \( \mathfrak{p}^+ + \mathfrak{p}^- \) generates \( \mathfrak{g} \).
Table 1: Simple Euclidean Jordan algebras, corresponding groups and structure constants

<table>
<thead>
<tr>
<th>V</th>
<th>(\mathfrak{so}(V))</th>
<th>(\mathfrak{str}(V))</th>
<th>n</th>
<th>r</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mathbb{R})</td>
<td>(\mathfrak{so}(2,\mathbb{R}))</td>
<td>(\mathbb{R})</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(\text{Sym}(k,\mathbb{R})) ((k \geq 2))</td>
<td>(\mathfrak{sp}(k,\mathbb{R}))</td>
<td>(\mathfrak{so}(k,\mathbb{R}) \oplus \mathbb{R})</td>
<td>(\frac{1}{2}k(k+1))</td>
<td>k</td>
<td>1</td>
</tr>
<tr>
<td>(\text{Herm}(k,\mathbb{C})) ((k \geq 2))</td>
<td>(\mathfrak{su}(k,\mathbb{C}))</td>
<td>(\mathfrak{su}(k,\mathbb{C}) \oplus \mathbb{R})</td>
<td>(k^2)</td>
<td>k</td>
<td>2</td>
</tr>
<tr>
<td>(\text{Herm}(k,\mathbb{H})) ((k \geq 2))</td>
<td>(\mathfrak{so}^*(4k))</td>
<td>(\mathfrak{su}^*(2k) \oplus \mathbb{R})</td>
<td>(k(2k-1))</td>
<td>k</td>
<td>4</td>
</tr>
<tr>
<td>(\mathbb{R}^{1,k-1}) ((k \geq 3))</td>
<td>(\mathfrak{so}(2,k))</td>
<td>(\mathfrak{so}(1,k-1) \oplus \mathbb{R})</td>
<td>k</td>
<td>2</td>
<td>(k-2)</td>
</tr>
<tr>
<td>(\text{Herm}(3,\mathbb{O}))</td>
<td>(\mathfrak{so}(7))</td>
<td>(\mathfrak{so}(7) \oplus \mathbb{R})</td>
<td>27</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

1.4 Complexifications

The complexification \(V_{\mathbb{C}}\) of \(V\) is a complex Jordan algebra. The Jordan trace of \(V_{\mathbb{C}}\) is given by the \(\mathbb{C}\)-linear extension of the Jordan trace \(\text{tr}\) of \(V\). The same holds for the trace form which is the \(\mathbb{C}\)-bilinear extension of the trace form of \(V\). By abuse of notation we use the same notation \(\text{tr}\) and \((\cdot|\cdot)\) for the complex objects. A norm on \(V_{\mathbb{C}}\) is given by

\[
|z| := \sqrt{(z|z)}, \quad z \in V_{\mathbb{C}}.
\]

\(V_{\mathbb{C}}\) can also be considered as a real Jordan algebra and we write \(W = V_{\mathbb{C}}\) for it. As real Jordan algebra \(W\) is simple since \(V\) is simple Euclidean. The trace form \((\cdot|\cdot)_W\) of \(W\) is of signature \((n,n)\) and explicitly given by (see 41, Lemma 1.2.3 (1))

\[
(z|w)_W = 2 ((\text{Re}(z)|\text{Re}(w)) - (\text{Im}(z)|\text{Im}(w))) = 2 \text{Re}(z|w), \quad z, w \in W.
\]

Note that \((\cdot|\cdot)_W\) is not \(\mathbb{C}\)-linear.

The identity component \(L_{\mathbb{C}} := \text{Str}(V_{\mathbb{C}})_0\) of the structure group of the complex Jordan algebra \(V_{\mathbb{C}}\) is a natural complexification of \(L\) (see 13, Proposition VIII.2.6)). As a real Lie group it is the same as \(\text{Str}(W)_0\), the identity component of the structure group of the real Jordan algebra \(W\). Let \(U \subseteq L_{\mathbb{C}}\) denote the analytic subgroup of \(L_{\mathbb{C}}\) with Lie algebra

\[
\mathfrak{u} := \mathfrak{t}^I + i\mathfrak{p}^I,
\]

then \(U\) is a maximal compact subgroup of \(L_{\mathbb{C}}\). The isomorphism \(C : \mathfrak{t}_{\mathbb{C}} \to \mathfrak{l}_{\mathbb{C}}\) introduced in Section 13 restricts to an isomorphism \(\mathfrak{t} \to \mathfrak{u}\) which integrates to a covering map \(\eta : K \to U \subseteq L_{\mathbb{C}}\) with differential

\[
d\eta(u, D, -u) = D + 2iL(u).
\]

Since \(u_{\mathbb{C}} = l_{\mathbb{C}}\) the subgroup \(U\) is totally real in \(L_{\mathbb{C}}\).

For \(g \in L_{\mathbb{C}}\) we denote by \(g^#\) the adjoint of \(g\) with respect to the \(\mathbb{C}\)-bilinear trace form \((\cdot|\cdot)\) of \(V_{\mathbb{C}}\). Since \((\cdot|\cdot)_W = 2 \text{Re}(\cdot|\cdot)\) this is also the adjoint with respect to the trace form \((\cdot|\cdot)_W\) of \(W\). Further put

\[
g^* := g^# , \quad g \in L_{\mathbb{C}}.
\]
Then $g \mapsto (g^{-1})^*$ defines a Cartan involution of $L_C$ with corresponding maximal compact subgroup $U$.

The determinant function $\Delta_W$ of the real Jordan algebra $W$ is a polynomial of degree $2r$. As in the Euclidean case an element $g \in \text{Str}(W)$ is characterized by the property that there exists a scalar $\chi_W(g) \in \mathbb{R}^\times$ such that

$$\Delta_W(gz) = \chi_W(g)\Delta_W(z), \quad z \in W.$$ 

This defines a real character $\chi_W : L_C \to \mathbb{R}_+^\times$.

The group $L_C$ has only one open orbit $L_C \cdot e$ (see [31, Theorem 4.2]). Its boundary again has a stratification by lower-dimensional orbits. More precisely we have

$$L_C \cdot e = X_0 \cup X_1 \cup \ldots \cup X_r,$$

where $X_k = L_C \cdot e_k$ (see [31, Theorem 4.2]). Note that $O_k \subseteq X_k$ is totally real and $X_k$ is closed under conjugation. The closure $\overline{X}_k$ of each $L_C$-orbit is an affine algebraic subvariety of $V_C$ (see [17, Theorem 2.9]). From [41, Section 1.5.2] it is easy to deduce the following formula for the dimension of $X_k$:

$$\dim_C X_k = \dim_{\mathbb{R}} O_k = k + k(2r - k - 1) \frac{d^2}{2}. \quad (1.9)$$

As in the real case every $z \in X_k$ has a polar decomposition $z = ua$ with $u \in U$ and $a = \sum_{i=1}^{k} a_i c_i$, $a_1 \geq \ldots \geq a_k > 0$, the element $a$ being unique with this property (see [15, Proposition X.3.2]).

For $\lambda \in W$ and $k \in \{0, \ldots, r\}$ such that $O_\lambda = O_k$ we similarly put $X_\lambda := X_k$.

**Proposition 1.2.** For every $\lambda \in W$ the measure $d\nu_\lambda$ on $X_\lambda$ defined by

$$\int_{X_\lambda} f(z) \, d\nu_\lambda(z) = \int_U \int_{O_\lambda} f(ua^2) \, d\mu_\lambda(x) \, du,$$ \hspace{1cm} (1.10)

is the unique (up to scalar multiples) $L_C$-equivariant measure on $X_\lambda$ which transforms by $\chi_\lambda^W$.

**Proof.** Except for the integral formula (1.10) this is essentially [22, Proposition 1.8]. By [22, Proposition 1.8] a $\chi_\lambda^W$-equivariant measure on $X_\lambda$ is given by

$$f \mapsto \int_U \int_{t_1 > \ldots > t_k} f(ue^t) J_\lambda(W)(t) \, dt \, du,$$

where $e^t$ is as in (1.4),

$$J_\lambda(W)(t) = e^{\frac{2}{k^2} \sum_{j=1}^{k} t_j} \prod_{1 \leq i < j \leq k} \sinh \left( \frac{t_i - t_j}{2} \right) \cosh \left( \frac{t_i - t_j}{2} \right).$$
and \( k \in \{0, \ldots, r\} \) such that \( X_k = X_\lambda \). Using \( \sinh x \cdot \cosh x = \frac{1}{2} \sinh 2x \) we find
\[
J^{W}_\lambda (t) = 2^{-k(k-1)\frac{d}{2}} e^{\frac{d}{2} \sum_{j=1}^{2r} t_j} \prod_{1 \leq i < j \leq k} \sinh^d (t_i - t_j) = 2^{-k(k-1)\frac{d}{2}} J_\lambda (2t)
\]
with \( J_\lambda (s) \) as in (1.3). Now note that \( K^L \subseteq U \) and \( (ke^t)^{\frac{d}{2}} = ke^{\frac{d}{2}t} \) for \( k \in K^L \) and hence
\[
\int_U \int_{t_1 > \ldots > t_k} f(ue^t)J^{W}_\lambda (t) \, dt \, du = \text{const} \cdot \int_U \int_{K^L} \int_{t_1 > \ldots > t_k} f(uke^t)J_\lambda (2t) \, dt \, dk \, du = \text{const} \cdot \int_U \int_{K^L} \int_{s_1 > \ldots > s_k} f(uke^{\frac{d}{2}s})J_\lambda (s) \, ds \, dk \, du = \text{const} \cdot \int_U \int_{O_\lambda} f(ux^{\frac{d}{2}}) \, d\mu_\lambda (x) \, du,
\]
where we have used the integral formula (1.3) for the last equality.

1.5 Nilpotent orbits and the Kostant–Sekiguchi correspondence

Nilpotent coadjoint \( G \)-orbits

We identify \( g^* \) with \( g \) by means of the Killing form and view coadjoint \( G \)-orbits on \( g^* \) as adjoint \( G \)-orbits on \( g \). Further we use the embedding
\[
V \hookrightarrow g, \ u \mapsto (u, 0, 0)
\]
to identify the \( L \)-orbits \( O_k \subseteq V, \ k = 0, \ldots, r, \) with \( L \)-orbits in \( n \cong V \). Since \( n \) is nilpotent, the \( G \)-orbits
\[
O^G_k := G \cdot (e_k, 0, 0)
\]
are nilpotent adjoint orbits in \( g \). Since further \( O_k = L \cdot (e_k, 0, 0) \) we clearly have \( O_k \subseteq O^G_k \). As usual we endow \( O^G_k \) with the Kirillov–Kostant–Souriau symplectic form. The following result was proved in [22, Theorem 2.9 (4)] for the minimal orbit \( O_1 \). For the general case the author could not find the statement in the existing literature.

Proposition 1.3. \( O_k \subseteq O^G_k \) is a Lagrangian subvariety.

Proof. Since \( O_k \subseteq n \) and \( n \) is abelian the symplectic form vanishes on \( O \) and it remains to show that \( \dim O^G_k = 2 \dim O_k \). By (1.9) we have
\[
\dim O_k = k + k(2r - k - 1) \frac{d}{2}.
\]
To determine \( \dim O^G_k \) we note that
\[
0 = [(u, T, v), (e_k, 0, 0)] = (Te_k, -2e_k \Box v, 0)
\]
The Cayley type transform

Observe that

\[ \dim \mathcal{O}_k^G = \dim \mathcal{O}_k + \dim (V/V(e_k,0)) = \dim \mathcal{O}_k + \dim V(e_k,1) + \dim V(e_k,\frac{1}{2}) = \dim \mathcal{O}_k + (k + k(k - 1)\frac{1}{2}) + k(r - k)d = 2 \dim \mathcal{O}_k. \]

\[ \Box \]

Nilpotent \( K_C \)-orbits

Via the Kostant–Sekiguchi correspondence the nilpotent adjoint \( G \)-orbits \( \mathcal{O}_k^G \subseteq g^* \) correspond to nilpotent \( K_C \)-orbits \( \mathcal{O}_k^{K_C} \subseteq p_C^* \). Identifying \( p^* \) with \( p \) by means of the Killing form we view \( \mathcal{O}_k^{K_C} \) as \( K_C \)-orbits in \( p_C = p^+ + p^- \).

Following [17] we let \((E_k, F_k, H_k)\) be the strictly normal \( \mathfrak{sl}_2 \)-triple in \( g \) given by

\[ E_k := (e_k,0,0), \quad H_k := (0,2L(e_k),0), \quad F_k := (0,0,e_k). \]

We form a new \( \mathfrak{sl}_2 \)-triple \( (E_k^d, F_k^d, H_k^d) \) by putting

\[ E_k^d := \frac{1}{2}(E_k + F_k + iH_k) = \frac{1}{2}(e_k,2iL(e_k),e_k), \]
\[ H_k^d := i(E_k - F_k) = i(e_k,0,-e_k), \]
\[ F_k^d := \frac{1}{2}(E_k + F_k - iH_k) = \frac{1}{2}(e_k,-2iL(e_k),e_k). \]

Then \( E_k^d \in p^+ \) and the \( K_C \) orbit \( \mathcal{O}_k^{K_C} \) corresponding to \( \mathcal{O}_k^G \) is given by

\[ \mathcal{O}_k^{K_C} = K_C \cdot E_k^d. \]

Since \( p^+ \) is \( K_C \)-stable we have \( \mathcal{O}_k^{K_C} \subseteq p^+ \).

If we use the embedding \( V_C \hookrightarrow g_C, u \mapsto (0,0,u) \) to identify the \( L_C \)-orbits \( X_k \subseteq V_C, k = 0, \ldots, r \), with \( L_C \)-orbits in \( \mathfrak{p}_C \cong V_C \) we obtain the following result:

**Proposition 1.4.** The Cayley type transform \( C \in \text{Int}(g_C) \) is a bijection from the \( K_C \)-orbit \( \mathcal{O}_k^{K_C} \subseteq p^+ \) onto the \( L_C \)-orbit \( X_k \subseteq \mathfrak{p}_C \) for every \( k \in \{0, \ldots, r\} \).

**Proof.** Observe that \( C(E_k^d) = (0,0,4e_k) \). Since the Lie algebra homomorphism \( C \) is a bijection from \( k_C \) onto \( l_C \) it gives a bijection between the orbits \( \mathcal{O}_k^{K_C} = K_C \cdot E_k^d \) and \( X_k = L_C \cdot (0,0,4e_k) \) which shows the claim. \( \Box \)

1.6 Differential operators

**General notation**

For a map \( f : X \to X \) on a (real or complex) vector space \( X \) we write \( Df(x) : X \to X \) for its (real or complex) Jacobian at the point \( x \in X \). For a
complex-valued function \( g : X \to \mathbb{C} \) we denote the directional derivative in the direction of \( a \in X \) by \( \partial_a g \):

\[
\partial_a g(x) = \left. \frac{d}{dt} \right|_{t=0} g(x + ta).
\]

**Bessel operators - the real case**

Let \( J \) be a real semisimple Jordan algebra with trace form \((-|-)\). The gradient with respect to \((-|-)\) will be denoted by \( \partial_x \). If \( (e_\alpha)_\alpha \subseteq J \) is a basis of \( J \) with dual basis \( (\hat{e}_\alpha)_\alpha \subseteq J \) with respect to the trace form then the gradient is in coordinates \( x = \sum_\alpha x_\alpha e_\alpha \in J \) expressed as

\[
\partial_x = \sum_\alpha \frac{\partial}{\partial x_\alpha} \hat{e}_\alpha.
\]

In terms of the directional derivative the gradient is characterized by the identity \( (a \partial_x)_J = \partial_a \) for \( a \in J \). The Bessel operator \( B^J_\lambda \) with parameter \( \lambda \in \mathbb{C} \) is a vector-valued second order differential operator on \( J \) given by

\[
B^J_\lambda := P \left( \frac{\partial}{\partial x} \right) x + \lambda \frac{\partial}{\partial x}.
\]

In coordinates it is given by

\[
B^J_\lambda = \sum_{\alpha,\beta} \frac{\partial^2}{\partial x_\alpha \partial x_\beta} P(\hat{e}_\alpha, \hat{e}_\beta)x + \lambda \sum_\alpha \frac{\partial}{\partial x_\alpha} \hat{e}_\alpha.
\]

Denote by \( \ell \) the left-action of \( \text{Str}(J) \) on functions on \( J \) given by \( \ell(g)f(x) = f(g^{-1}x) \). Then the Bessel operator satisfies the following equivariance property:

\[
\ell(g)B^J_\lambda \ell(g^{-1}) = g^# B^J_\lambda, \quad g \in \text{Str}(J),
\]

where \( g^# \) denotes the adjoint with respect to the trace form of \( J \). We further have the following product rule:

\[
B^J_\lambda [f(x)g(x)] = B^J_\lambda f(x) \cdot g(x) + 2P \left( \frac{\partial f}{\partial x} \cdot \frac{\partial g}{\partial x} \right) x + f(x) \cdot B^J_\lambda g(x).
\]

For either \( J = V \) or \( J = W = V_{\mathbb{C}} \) the operator \( B^J_\lambda \) is tangential to the orbit \( \text{Str}(J)_0 \cdot e_k \) if \( \lambda = k \frac{d}{dx} \), \( 0 \leq k \leq r - 1 \), and hence defines a differential operator on this orbit (see \[11 \] Theorem 1.7.5). On \( K^L \)-invariant (resp. \( U \)-invariant) functions on \( V \) (resp. \( W \)) we have the following formulas:

**Proposition 1.5.** \([13 \] Theorem XV.2.7\) Let \( f \in C^\infty(V) \) be \( K^L \)-invariant. Then with \( F(a_1, \ldots, a_r) = f(a) \), \( a = \sum_{i=1}^r a_i c_i \), we have

\[
B^V_\lambda f(a) = \sum_{i=1}^r (B^V_\lambda)^i F(a_1, \ldots, a_r) c_i,
\]
where

$$(B^W_\lambda)^i = a_i \frac{\partial^2}{\partial a_i^2} + \left( \lambda - (r - 1) \frac{d}{2} \right) \frac{\partial}{\partial a_i} + \frac{d}{2} \sum_{j \neq i} \frac{1}{a_i - a_j} \left( \frac{\partial}{\partial a_i} - \frac{\partial}{\partial a_j} \right).$$

(2) \((43)\) Let \(f \in C^\infty(W)\) be \(U\)-invariant. Then with \(F(a_1, \ldots, a_r) = f(a)\), we have

$$B^W_\lambda f(a) = \sum_{i=1}^r (B^W_\lambda)^i F(a_1, \ldots, a_r)c_i,$$

where

$$(B^W_\lambda)^i = \frac{1}{4} \left( a_i \frac{\partial^2}{\partial a_i^2} + (2\lambda - 1 - (r - 1)d) \frac{\partial}{\partial a_i} \right)$$

$$+ \frac{d}{2} \sum_{j \neq i} \left( \frac{1}{a_i - a_j} + \frac{1}{a_i + a_j} \right) \left( \frac{\partial}{\partial a_i} - \frac{\partial}{\partial a_j} \right).$$

For \(J = V\) we will write \(B_\lambda := B^V_\lambda\) for short.

**Bessel operators - the complex case**

We also need the Bessel operator of the complex Jordan algebra \(V_C\). It is a holomorphic vector-valued differential operator on \(V_C\) which is defined in the same way as the Bessel operator of \(V\). More precisely, if \((e_\alpha)_\alpha \subseteq V_C\) is a \(C\)-basis of \(V_C\) with dual basis \((\hat{e}_\alpha)_\alpha \subseteq V_C\) with respect to the trace form \((\cdot, \cdot)\) (which is \(C\)-bilinear) we define the gradient of \(V_C\) by

$$\frac{\partial}{\partial z} = \sum_{\alpha} \frac{\partial}{\partial z_{\alpha}}, \quad \text{where} \quad \frac{\partial}{\partial z_{\alpha}} = \frac{1}{2} \left( \frac{\partial}{\partial x_{\alpha}} - i \frac{\partial}{\partial y_{\alpha}} \right)$$

is the Wirtinger derivative and \(z = \sum_{\alpha} z_\alpha e_\alpha\) with \(z_\alpha = x_\alpha + iy_\alpha\). The Bessel operator \(B^V_\lambda\) of \(V_C\) is then defined by

$$B^V_\lambda = P \left( \frac{\partial}{\partial z} \right) z + \lambda \frac{\partial}{\partial z} = \sum_{\alpha,\beta} \frac{\partial^2}{\partial z_{\alpha} \partial z_{\beta}} P(\hat{e}_\alpha, \hat{e}_\beta)z + \lambda \sum_{\alpha} \frac{\partial}{\partial z_{\alpha}} \hat{e}_\alpha.$$

For a holomorphic function \(f\) on \(V_C\) we have

$$\left. \left( B^V_\lambda f \right) \right|_V = B^V_\lambda (f|_V)$$

and hence \(B^V_\lambda\) is a natural complexification of the Bessel operator \(B_\lambda\) of \(V\). Therefore, by abuse of notation, we also abbreviate \(B_\lambda = B^V_\lambda\).
1.7 Bounded symmetric domains of tube type

Let $T_Ω := V + iΩ ⊆ V_ℂ$ be the tube domain associated with the symmetric cone $Ω$. Each rational transformation $g ∈ G$ extends to a holomorphic automorphism of $T_Ω$. This establishes an isomorphism between $G$ and the identity component $\text{Aut}(T_Ω)_0$ of the group of all holomorphic automorphisms of $T_Ω$ (see [15 Theorem X.5.6]). Under this isomorphism the maximal compact subgroup $K ⊆ G$ corresponds to the stabilizer subgroup of the element $ie ∈ T_Ω$ so that $T_Ω ∼= G/K$. We identify an element $g ∈ G$ with its holomorphic extension to $T_Ω$.

Let $p : T_Ω → V_ℂ, p(z) := (z - ie)(z + ie)^{-1}$, and define $D := p(T_Ω)$. Then $p$ restricts to a holomorphic isomorphism $p : T_Ω ∼→ D$ with inverse the Cayley transform (see [15 Theorem X.4.3])

$$c : D ∼→ T_Ω, c(w) = i(e + w)(e - w)^{-1}. \quad (1.13)$$

The open set $D ⊆ V_ℂ$ is a bounded symmetric domain of tube type and every such domain occurs in this fashion. Clearly $G$ is also isomorphic to the identity component $\text{Aut}(D)_0$ of the group of all holomorphic automorphisms of $D$ via the conjugation map

$$α : G → \text{Aut}(D)_0, g ↦ p ∘ g ∘ p^{-1} = c^{-1} ∘ g ∘ c. \quad (1.14)$$

The stabilizer subgroup of the origin $0 ∈ D$ corresponds to the maximal compact subgroup $K ⊆ G$ and also $D ∼= G/K$.

Denote by $Σ ⊆ ∂D$ the Shilov boundary of $D$. The group $U$ acts transitively on $Σ$ and the stabilizer of $e ∈ Σ$ is equal to $K^L ⊆ U$ (see [15 Proposition X.3.1 & Theorem X.4.6]). Hence $Σ ∼= U/K^L$ is a compact symmetric space. Denote by $dσ$ the normalized $U$-invariant measure on $Σ ∼= U/K^L$. We write $L^2(Σ, dσ)$ and denote the corresponding $L^2$-inner product by $⟨−, −⟩_Σ$.

1.8 Polynomials

In this Section we recall known properties for the space of polynomials on $V_ℂ$.

Principal minors and decompositions

Let $P(V_ℂ)$ denote the space of holomorphic polynomials on $V_ℂ$. The space $P(V_ℂ)$ carries a natural representation $ℓ$ of $L_ℂ$ given by

$$ℓ(g)p(z) = p(g^{-1}z), \quad g ∈ L_ℂ, p ∈ P(V_ℂ).$$

Since $L$ and $U$ are both real forms of the complex group $L_ℂ$ the decompositions of $P(V_ℂ)$ into irreducible $L$- and $U$-representations are the same. This decomposition can be described as follows:

$\textbf{Documenta Mathematica 18 (2013) 785–855}$
For $j \in \{1, \ldots, r\}$ we denote by $P_j$ the orthogonal projection $V \to V(e_j, 1)$. Let $\Delta_{V(e_j, 1)}$ be the Jordan determinant of the Euclidean Jordan algebra $V(e_j, 1)$. We define a polynomial $\Delta_j$ on $V$ by the formula

$$\Delta_j(x) := \Delta_{V(e_j, 1)}(P_j x), \quad x \in V,$$

and extend it to a holomorphic polynomial on $V_\mathbb{C}$. The polynomials $\Delta_1, \ldots, \Delta_r$ are called principal minors of $V$. Note that $\Delta_r = \Delta$ is the Jordan determinant of $V$. For $m \in \mathbb{N}_0$ we write $m \geq 0$ if $m_1 \geq \ldots \geq m_r \geq 0$. If $m \geq 0$ we define a polynomial $\Delta_m$ on $V_\mathbb{C}$ by

$$\Delta_m(z) := \Delta_1(z)^{m_1-m_2} \ldots \Delta_{r-1}(z)^{m_{r-1}-m_r} \Delta_r(z)^{m_r}, \quad z \in V_\mathbb{C}.$$

The polynomials $\Delta_m(z)$ are called generalized power functions. Define $\mathcal{P}_m(V_\mathbb{C})$ to be the subspace of $\mathcal{P}(V_\mathbb{C})$ spanned by $\ell(g)\Delta_m$ for $g \in L$. (Equivalently one can let $g \in U$ or $g \in L_\mathbb{C}$.) We write $d_m := \dim \mathcal{P}_m(V_\mathbb{C})$ for its dimension.

**Theorem 1.6** (Hua–Kostant–Schmid, see e.g. [15, Theorem XI.2.4]). For each $m \geq 0$ the space $\mathcal{P}_m(V_\mathbb{C})$ is an irreducible $L$-module and the space $\mathcal{P}(V_\mathbb{C})$ decomposes into irreducible $L$-modules as

$$\mathcal{P}(V_\mathbb{C}) = \bigoplus_{m \geq 0} \mathcal{P}_m(V_\mathbb{C}).$$

The same spaces $\mathcal{P}_m(V_\mathbb{C})$ occur in the decomposition of $L^2(\Sigma)$ into irreducible $U$-representations. To make this precise let $\mathbb{Z}_+^r$ be the set of all $m \in \mathbb{Z}^r$ with $m_1 \geq \ldots \geq m_r$. For $m \in \mathbb{Z}_+^r$ we define another tuple $m' := (m_1 - m_r, \ldots, m_r - m_r, 0)$. Then $m' \geq 0$ and we define

$$\mathcal{P}_m(\Sigma) := \{\Delta^{m'} p|\Sigma : p \in \mathcal{P}_m(V_\mathbb{C})\}.$$

If $m \geq 0$ then $\mathcal{P}_m(\Sigma)$ coincides with the space of restrictions of polynomials in $\mathcal{P}_m(V_\mathbb{C})$ to $\Sigma$.

**Theorem 1.7** (Cartan–Helgason, see e.g. [15, Theorem XII.2.2]). For each $m \in \mathbb{Z}_+^r$ the space $\mathcal{P}_m(\Sigma)$ is an irreducible unitary $U$-representation and the space $L^2(\Sigma)$ decomposes into the direct Hilbert space sum

$$L^2(\Sigma) = \bigoplus_{m \in \mathbb{Z}_+^r} \mathcal{P}_m(\Sigma).$$

We now study the restriction of polynomials to the orbits $X_k \subseteq V_\mathbb{C}$ and $O_k \subseteq V$. Note that for $m \geq 0$ and $k \in \{0, \ldots, r-1\}$ the condition $m_{k+1} = 0$ means $m_{k+1} = \ldots = m_r = 0$. For convenience we also use this notation for $k = r$. In this case $m_{k+1} = 0$ should mean no restriction on $m$.

**Proposition 1.8** ([24, Proposition 1.7 (vi) (a)]). For $x \in X_k$ we have $\Delta_{k+1}(x) = \ldots = \Delta_r(x) = 0$. In particular, $\Delta_m$ vanishes on $X_k$ iff $m_{k+1} \neq 0$, $m \geq 0$. 
For $0 \leq k \leq r$ let $\mathcal{P}(X_k)$ and $\mathcal{P}_m(X_k)$ be the images of $\mathcal{P}(V_C)$ and $\mathcal{P}_m(V_C)$ under the restriction map $C^\infty(V_C) \to C^\infty(X_k)$. Similar we define $\mathcal{P}(O_k)$ and $\mathcal{P}_m(O_k)$ to be the images of $\mathcal{P}(V_C)$ and $\mathcal{P}_m(V_C)$ under the restriction map $C^\infty(V_C) \to C^\infty(O_k)$.

**Corollary 1.9.** Let $k \in \{0, \ldots, r\}$ and $m \geq 0$. Then $\mathcal{P}_m(X_k)$ (resp. $\mathcal{P}_m(O_k)$) is non-trivial if and only if $m k + 1 = 0$. In this case the restriction from $V_C$ to $X_k$ (resp. $O_k$) induces an isomorphism $\mathcal{P}_m(V_C) \cong \mathcal{P}_m(X_k)$ (resp. $\mathcal{P}_m(V_C) \cong \mathcal{P}_m(O_k)$) of $L$-modules (resp. $L_C$-modules). In particular we have the following decompositions:

$$
\mathcal{P}(X_k) = \bigoplus_{m \geq 0, m k + 1 = 0} \mathcal{P}_m(X_k), \quad \mathcal{P}(O_k) = \bigoplus_{m \geq 0, m k + 1 = 0} \mathcal{P}_m(O_k).
$$

For $\lambda \in \mathbb{C}$ and $m \geq 0$ we define the Pochhammer symbol $(\lambda)_m$ by

$$(\lambda)_m := \prod_{i=1}^r \left( \lambda - (i-1)\frac{d}{2} \right)^{m_i},$$

where $(a)_n = a(a+1) \cdots (a+n-1)$ denotes the classical Pochhammer symbol for $a \in \mathbb{C}$, $n \in \mathbb{N}_0$.

**Lemma 1.10 ([15, Lemma XI.2.3]).** For $p \in \mathcal{P}_m(V_C)$ and $\lambda \in \mathcal{W}$ we have

$$
\int_{O_{\lambda}} e^{-\langle z|y \rangle} p(x) \, d\mu_{\lambda}(x) = 2^r \lambda (\lambda)_m \Delta(y)^{-\lambda} p(y^{-1}), \quad y \in \Omega.
$$

**Spherical polynomials**

The representations $\mathcal{P}_m(V_C)$ of $L$ (resp. $U$) are $K^L$-spherical. The $K^L$-spherical vectors in $\mathcal{P}_m(V_C)$ are spanned by (see [15, Proposition XI.3.1])

$$
\Phi_m(z) = \int_{K^L} \Delta_m(kz) \, dk, \quad z \in V_C.
$$

The $L^2(\Sigma)$-norm of these functions are given by (see [15, Proposition XI.4.1 (i)])

$$
\|\Phi_m\|_2^2 = \frac{1}{d_m}, \quad m \geq 0. \quad (1.15)
$$

By [15, Corollary XI.3.4] there exists a unique polynomial $\Phi_m(z, w)$ on $V_C \times V_C$ holomorphic in $z$ and antiholomorphic in $w$ such that

$$
\Phi_m(gz, w) = \Phi_m(z, g^*w), \quad \forall g \in L_C,
$$

$$
\Phi_m(x, x) = \Phi_m(x^2), \quad \forall x \in V.
$$

**Lemma 1.11.** (1) For all $z \in V_C$ we have

$$
\Phi_m(z, e) = \Phi_m(z).
$$
(2) For $x \in V$ and $y \in \overline{\Omega}$ we have
\[ \Phi_m(x, y) = \Phi_m(P(y^\frac{1}{2})x). \]

(3) For $z, w \in V_C$ we have
\[ \Phi_m(z, w) = \Phi_m(w, z). \]

Proof. (1) First let $x \in \Omega$. Recall the complex powers $x^s$, $s \in \mathbb{C}$, defined in Section 1.2. Then
\[ \Phi_m(x, e) = \Phi_m(P(x^\frac{1}{2})x^\frac{1}{2}, e) = \Phi_m(x^\frac{1}{2}, P(x^\frac{1}{2})e) = \Phi_m(x^\frac{1}{2}, x^\frac{1}{2}) = \Phi_m(x). \]
Since both sides are holomorphic in $x$ and $\Omega \subseteq V_C$ is totally real, this also holds for $x \in V_C$.

(2) Let $x \in V$ and $y \in \Omega$. Then by (1) we obtain
\[ \Phi_m(x, y) = \Phi_m(x, P(y^\frac{1}{2})e) = \Phi_m(P(y^\frac{1}{2})x, e) = \Phi_m(P(y^\frac{1}{2})x). \]
Now both sides are continuous in $y \in \Omega$ and the claim follows.

(3) First let $x, y \in \Omega$. Then by [15, Lemma XIV.1.2] there exists $k \in K^L$ such that $P(y^\frac{1}{2})x = kP(x^\frac{1}{2})y$. Using (2) and the $K^L$-invariance of $\Phi_m$ we find that
\[ \Phi_m(x, y) = \Phi_m(P(y^\frac{1}{2})x) = \Phi_m(P(x^\frac{1}{2})y) = \Phi_m(y, x). \]
Since $\Phi_m(x, y) \in \mathbb{R}$ for $x, y \in V$, we obtain
\[ \Phi_m(z, w) = \Phi_m(w, z) \quad \forall z, w \in \Omega. \]
Both sides are holomorphic in $z$ and antiholomorphic in $w$ and $\Omega \subseteq V_C$ is totally real. Hence, the formula also holds for $z, w \in V_C$.

Lemma 1.12. If $m \geq 0$ with $m_{k+1} \neq 0$, then $\Phi_m(z, w) = 0$ for all $z \in V_C$ and $w \in \overline{V_C}$.

Proof. Write $w = he_k$ with $g \in L_C$. Then
\[ \Phi_m(z, w) = \Phi_m(z, ge_k) = \Phi_m(g^*z, e_k). \]
Therefore, it suffices to show that $\Phi_m(-, e_k) = 0$ as a polynomial on $V_C$. Since $\Phi_m(z, w)$ is holomorphic in the first variable it suffices to show that $\Phi_m(-, e_k) = 0$ as a polynomial on $V$ and since $\Omega \subseteq V$ is open it is enough to prove $\Phi_m(x, e_k) = 0$ for $x \in \Omega$. But for $x \in \Omega$ we have by Lemma 1.11(2) and (3)
\[ \Phi_m(x, e_k) = \Phi_m(P(x^\frac{1}{2})e_k). \]
Now write $x^\sharp = ge$ with $g \in L$, then

$$P(x^\sharp)e_k = P(ge)e_k = gP(e)g^#e_k = gg^#e_k$$

and hence $P(x^\sharp)e_k \in O_k = L \cdot e_k$. But by Corollary [13]\ref{cor:1.3} we have $\Phi_m|O_k = 0$ since $m_{k+1} \neq 0$ and the proof is complete.

**Example 1.13.** On the rank 1 orbit $X_1$ the spherical polynomials $\Phi_m(z, w)$ are non-zero if and only if $m = (m, 0, \ldots, 0)$, $m \in \mathbb{N}_0$, and in this case

$$\Phi_m(z, w) = \left(\frac{\pi}{m}\right)^m \frac{d_m}{d_m m!} (z|\overline{w})^m, \quad z, w \in X_1.$$

(This can e.g. be derived from the expansion of $\text{tr}(x^k)$ into the polynomials $\Phi_m(x)$, see [15] Chapter XI.5.)

**The Fischer inner product**

We equip $\mathcal{P}(V_\mathbb{C})$ with the *Fischer inner product*

$$[p, q] := p \left( \frac{\partial}{\partial \overline{z}} \right) \overline{q(z)} \bigg|_{z=0}, \quad p, q \in \mathcal{P}(V_\mathbb{C}),$$

where $\overline{q(z)} := q(z)$, $z \in V_\mathbb{C}$. The action of $U$ on $\mathcal{P}(V_\mathbb{C})$ is unitary with respect to this inner product and hence the irreducible constituents $\mathcal{P}_m(V_\mathbb{C})$ are pairwise orthogonal. Since $U$ also acts unitarily on $\mathcal{P}(V_\mathbb{C})$ with respect to the inner product on $L^2(\Sigma)$ the two inner products are proportional on each irreducible constituent $\mathcal{P}_m(V_\mathbb{C})$ (see [15] Corollary XI.4.2):

$$[p, q] = \left(\frac{\pi}{m}\right)^m \langle p, q \rangle_{\Sigma}, \quad p, q \in \mathcal{P}_m(V_\mathbb{C}).$$

Denote by $K^m(z, w)$ the reproducing kernel of $\mathcal{P}_m(V_\mathbb{C})$ with respect to the Fischer inner product. By [15] Propositions XI.3.3 & XI.4.1 we have

$$K^m(z, w) = \frac{d_m}{\phi_m} \Phi_m(z, w), \quad z, w \in V_\mathbb{C}.$$

Since by [15] Proposition XI.1.1 the completion of $\mathcal{P}(V_\mathbb{C})$ with respect to the Fischer inner product is a Hilbert space with reproducing kernel $K(z, w) = e^{(z|\overline{w})}$ we obtain the following expansion:

$$e^{(z|\overline{w})} = \sum_{m \geq 0} \frac{d_m}{\phi_m} \Phi_m(z, w), \quad z, w \in V_\mathbb{C}.$$

**Lemma 1.14.** For $m, n \geq 0$ and $p \in \mathcal{P}_m(V_\mathbb{C})$, $q \in \mathcal{P}_n(V_\mathbb{C})$:

$$\int_U p(uz)q(\overline{uw}) \, du = \Phi_m(z, w) \langle p, q \rangle_{\Sigma} \quad \forall z, w \in V_\mathbb{C}.$$
Proof. Using the reproducing kernel property and the Schur orthogonality relations we obtain
\[
\int_U p(u)q(u) du = \int_U [p, K^m(-,uz)]q, K^n(-,uw) du
= \int_U [\ell(u^{-1})p, K^m(-,z)]\ell(u^{-1})q, K^n(-,w) du
= \frac{1}{d_m}[p, q][K^m(-,z), K^n(-,w)].
\]

With (1.16) and
\[
[K^m(-,z), K^n(-,w)] = \delta_{mn}K^{m}(w,z) = \delta_{mn}d_m \Phi_m(z, w)
\]
the claimed identity follows.

LAGUERRE FUNCTIONS

For \( m \geq 0 \) the polynomial \( \Phi_m(e + x) \) is \( K^L \)-invariant of degree \( |m| \) and can hence be written as a linear combination of the \( K^L \)-invariant polynomials \( \Phi_n(x) \) for \( |n| \leq |m| \). Following [15, Chapter XV.4] we define the generalized binomial coefficients \( \binom{m}{n} \) by the formula
\[
\Phi_m(e + x) = \sum_{|n| \leq |m|} \binom{m}{n} \Phi_n(x), \quad x \in V.
\]

Using the generalized binomial coefficients we define the generalized Laguerre polynomials \( L^\lambda_m(x) \) by
\[
L^\lambda_m(x) := (\lambda)_m \sum_{|n| \leq |m|} \binom{m}{n} \frac{1}{(\lambda)_n} \Phi_n(-x), \quad x \in V, \quad (1.18)
\]
and the generalized Laguerre functions \( \ell^\lambda_m(x) \) by
\[
\ell^\lambda_m(x) := e^{-\text{tr}(x)}L^\lambda_m(2x), \quad x \in V. \quad (1.19)
\]

Both \( L^\lambda_m(x) \) and \( \ell^\lambda_m(x) \) are \( K^L \)-invariant. Note that for \( \lambda > (r - 1)\frac{d}{2} \) we have \( (\lambda)_n \neq 0 \) for all \( n \geq 0 \) and hence \( L^\lambda_m(x) \) and \( \ell^\lambda_m(x) \) are defined on \( V \). For \( x \in \mathcal{O}_k, k = 0, \ldots, r - 1 \), we further have \( \Phi_n(x) = 0 \) if \( n_{k+1} \neq 0 \). Therefore the sum in (1.15) reduces to a sum over \( n \) with \( n_{k+1} = \ldots = n_r = 0 \). For such \( n \) we have \( (\lambda)_n \neq 0 \) for \( \lambda = k\frac{d}{2} \) and hence the expression (1.15) is defined. This gives Laguerre polynomials \( L^\lambda_m(x) \) and Laguerre functions \( \ell^\lambda_m(x) \) on \( \mathcal{O}_\lambda \) for each \( \lambda \in W \).

Properties of the generalized Laguerre functions have been studied in [4].
2 Three different realizations of unitary highest weight representations of scalar type

We describe three known realizations of highest weight representations of scalar type: the \textit{bounded symmetric domain model}, the \textit{tube domain model} and the \textit{Schrödinger model}. We further discuss a complexification of the Schrödinger model which we use in Section 4.3 to construct yet another model, the Fock model.

2.1 The Schrödinger model

The highest weight representation of the universal cover \( \tilde{G} \) belonging to the Wallach point \( \lambda \in \mathcal{W} \) can be realized on \( L^2(O_\lambda, d\mu_\lambda) \). We sketch the construction here (see e.g. [41] for details). First, there is a Lie algebra representation \( d\pi_\lambda \) of \( g \) on \( C^\infty(O_\lambda) \) for every \( \lambda \in \mathcal{W} \) given by

\[
d\pi_\lambda(u, 0, 0) = i(u|x),
d\pi_\lambda(0, T, 0) = \partial_T x + \frac{r\lambda}{2n} \text{Tr}(T^#),
d\pi_\lambda(0, 0, v) = i(v|B_\lambda).
\]

Note that for \( \lambda = k\frac{d}{d} \), \( k = 0, \ldots, r-1 \), the Bessel operator \( B_\lambda \) is tangential to the orbit \( O_k \) and hence defines a differential operator acting on \( C^\infty(O_k) \) (see Section 1.6). The representation \( d\pi_\lambda \) is further infinitesimally unitary with respect to the \( L^2 \)-inner product on \( L^2(O_\lambda, d\mu_\lambda) \).

The subrepresentation of \( (d\pi_\lambda, C^\infty(O_\lambda)) \) generated by the function

\[
\psi_0(x) = e^{-\text{tr}(x)}, \quad x \in O_\lambda,
\]

defines a \((g, \tilde{K})\)-module \( (d\pi_\lambda, W_\lambda) \) whose underlying vector space turns out to be (see e.g. [15] Proposition XIII.3.2)]

\[
W_\lambda = \mathcal{P}(O_\lambda)e^{-\text{tr}(x)}.
\]

The \((g, \tilde{K})\)-module \( (d\pi_\lambda, W_\lambda) \) integrates to an irreducible unitary representation \((\pi_\lambda, L^2(O_\lambda, d\mu_\lambda))\) of \( \tilde{G} \). The minimal \( \tilde{K} \)-type is spanned by the function \( \psi_0 \) which is \( \tilde{K} \)-equivariant:

\[
\pi_\lambda(k)\psi_0 = \xi_\lambda(k)\psi_0, \quad k \in \tilde{K},
\]

where \( \xi_\lambda: \tilde{K} \to \mathbb{T} \) is the character of \( \tilde{K} \) with differential

\[
d\xi_\lambda(u, T, -u) = i\lambda \text{tr}(u).
\]

This implies that the representation \( \pi_\lambda \) descends to a finite cover of \( G \) if and only if \( \lambda \in \mathcal{Q} \) (which holds in particular for \( \lambda \in \mathcal{W}_{\text{disc}} \)). For \( \lambda > \frac{2r}{d} - 1 \) the representation \( \pi_\lambda \) belongs to the relative holomorphic discrete series of \( \tilde{G} \).
2.2 The tube domain model

For $\lambda \in \mathbb{C}$ consider the function

$$T_\Omega \times T_\Omega \to \mathbb{C}, (z, w) \mapsto \Delta \left(\frac{z - \overline{w}}{2i}\right)^{-\lambda}.$$ 

It is of positive type if and only if $\lambda \in \mathcal{W}$ (see [15, Proposition XIII.1.2 & Theorem XIII.2.4]). Denote by $H^2_\lambda(T_\Omega)$ the Hilbert space of holomorphic functions on $T_\Omega$ which has the function $\Delta \left(\frac{z - \overline{w}}{2i}\right)^{-\lambda}$ as reproducing kernel. For $\lambda > 1 + (r - 1)d$ this space coincides with the space of holomorphic functions $F$ on $T_\Omega$ such that

$$\int_{T_\Omega} |F(z)|^2 \Delta(y)^{\lambda - \frac{2r}{n}} \, dx \, dy < \infty,$$

where $z = x + iy \in T_\Omega$ (see [15, Chapter XIII.1]). For every $\lambda \in \mathcal{W}$ there is an irreducible unitary representation $\pi^T_\lambda$ of $\tilde{G}$ on $H^2_\lambda(T_\Omega)$. Note that $\tilde{G}$ acts on $T_\Omega$ by composition of the action of $\tilde{G} = \text{Aut}(T_\Omega)_0$ on $T_\Omega$ with the covering map $\tilde{G} \to G$. Then the representation $\pi^T_\lambda$ is given by

$$\pi^T_\lambda(g) F(z) = \mu^T_\lambda(g^{-1}, z) F(g^{-1}z), \quad g \in \tilde{G}, z \in T_\Omega,$$

where the cocycle $\mu^T_\lambda(g, z)$ is given by

$$\mu^T_\lambda(g, z) = \text{Det}(Dg(z))^{\frac{\lambda}{2}}, \quad g \in \tilde{G}, z \in T_\Omega,$$

the powers being well-defined on the universal cover $\tilde{G}$. The representations $H^2_\lambda(T_\Omega)$ and $L^2(\mathcal{O}_\lambda, d\mu_\lambda)$ are isomorphic, the intertwining operator being the Laplace transform (see [15, Theorems XIII.1.1 & XIII.3.4])

$$\mathcal{L}_\lambda : L^2(\mathcal{O}_\lambda, d\mu_\lambda) \to H^2_\lambda(T_\Omega), \mathcal{L}_\lambda \psi(z) := \int_{\mathcal{O}_\lambda} e^{i(z|x)} \psi(x) \, d\mu_\lambda(x). \quad (2.1)$$

The Laplace transform $\mathcal{L}_\lambda$ is a unitary isomorphism (up to a scalar) intertwining the group actions.

2.3 The bounded symmetric domain model

The polynomial $h(x) := \Delta(e - x^2)$, $x \in V$, is $K^L$-invariant and therefore by [15, Corollary XI.3.4] there exists a unique polynomial $h(z, w)$ holomorphic in $z \in V_C$ and antiholomorphic in $w \in V_C$ such that

$$h(gz, w) = h(z, g^*w), \quad \forall g \in L_C,$$

$$h(x, x) = h(x), \quad \forall x \in V.$$

We also write $h(z) := h(z, z)$ for $z \in V_C$. Consider the powers $h(z, w)^{-\lambda}$ for $\lambda \in \mathbb{C}$ as functions on $D \times D$. Then $h(z, w)^{-\lambda}$ is positive definite on $D \times D$. 

Documenta Mathematica 18 (2013) 785–855
if and only if $\lambda \in \mathcal{W}$ (see [15 Theorem XIII.2.7]). The corresponding Hilbert space of holomorphic functions on $D$ with reproducing kernel $h(z, w)^{-\lambda}$ will be denoted by $\mathcal{H}_{\lambda}^2(D)$. For $\lambda > 1 + (r - 1)d$ this space coincides with the space of holomorphic functions $f$ on $D$ such that

$$\int_D |f(w)|^2 h(w)^{\lambda - \frac{2m}{n}} du dv < \infty,$$

where $w = u + iv \in D$ (see [15 Proposition XIII.1.4]). For each $\lambda \in \mathcal{W}$ there is an irreducible unitary representation $\pi_{\lambda}^D$ of $\tilde{G}$ on $\mathcal{H}_{\lambda}^2(D)$. To give an explicit formula recall the isomorphism $\alpha : G \rightarrow \text{Aut}(D)_0$ defined in (1.14) and view it as a covering map $\alpha : \tilde{G} \rightarrow \text{Aut}(D)_0$. Then the representation $\pi_{\lambda}^D$ is given by

$$\pi_{\lambda}^D(g)f(w) = \mu_{\lambda}^D(g, w)f(\alpha(g)^{-1}w), \quad g \in \tilde{G}, w \in D,$$

where the cocycle $\mu_{\lambda}^D(g, w)$ is given by

$$\mu_{\lambda}^D(g, w) = \text{Det}(D(\alpha(g))(w))^{\frac{1}{2n}}, \quad g \in \tilde{G}, w \in D,$$

the powers being well-defined on the universal cover $\tilde{G}$.

The representations $\mathcal{H}_{\lambda}^2(D)$ and $\mathcal{H}_{\lambda}^2(T_{\Omega})$ are isomorphic, the intertwining operator being given by

$$\gamma_{\lambda} : \mathcal{H}_{\lambda}^2(T_{\Omega}) \rightarrow \mathcal{H}_{\lambda}^2(D), \quad \gamma_{\lambda}F(w) := \Delta(e - w)^{-\lambda}F(c(w)),$$

where $c(w)$ is the Cayley transform defined in (1.13) (see [15 Proposition XIII.1.3 & Theorem XIII.3.4]). The operator $\gamma_{\lambda}$ is unitary (up to a scalar) and intertwines the group actions.

### 2.4 $\mathfrak{t}$-type decompositions

In the bounded symmetric domain model the $\tilde{K}$-type decomposition is very explicit. Let $\lambda \in \mathcal{W}$ and $0 \leq k \leq r$ such that $O_{\lambda} = O_k$. Then $\mathcal{H}_{\lambda}^2(D)$ decomposes into the direct Hilbert space sum

$$\mathcal{H}_{\lambda}^2(D) = \bigoplus_{m \geq 0, m_{k+1} = 0} \mathcal{P}_m(V_{\mathcal{C}}),$$

each summand $\mathcal{P}_m(V_{\mathcal{C}})$ is irreducible under the action $\pi_{\lambda}^D(\tilde{K})$ and on it the norm is given by (see [15 Theorem XIII.2.7])

$$\|p\|_{\mathcal{H}_{\lambda}^2(D)}^2 = \frac{(\frac{\lambda}{m})}{(\lambda)^m_m} \|p\|_{\mathcal{P}_m(V_{\mathcal{C}})}^2, \quad p \in \mathcal{P}_m(V_{\mathcal{C}}). \quad (2.2)$$

Further, the $\mathfrak{t}^k$-spherical vector in each $\tilde{K}$-type $\mathcal{P}_m(V_{\mathcal{C}})$ is the spherical polynomial $\Phi_m(z)$ which has norm

$$\|\Phi_m\|_{\mathcal{H}_{\lambda}^2(D)}^2 = \frac{(\frac{\lambda}{m})}{d_m(\lambda)^m_m}.$$
Correspondingly the underlying \((\mathfrak{g}, \tilde{K})\)-module \(W^\lambda\) of the Schrödinger model \(L^2(O_\lambda, d\mu_\lambda)\) decomposes into \(\tilde{K}\)-types

\[
W^\lambda = \bigoplus_{m \geq 0, m_{k+1} = 0} W^\lambda_m,
\]

where \(W^\lambda_m = \mathcal{L}_\lambda^{-1} \circ \gamma^{-1}_\lambda(P_m(V_\mathbb{C}))\). The \(\mathfrak{k}\)-spherical vector in \(W^\lambda\) is given by the Laguerre function \(\ell^\lambda_m(x)\) defined in Section 1.8 (see [15, Proposition XV.4.2]) which has norm (see [15, Corollary XV.4.3 (i)])

\[
\|\ell^\lambda_m\|^2_{L^2(O_\lambda, d\mu_\lambda)} = \left(\frac{2}{m}\right)_{m(\lambda)m} \sum_{i=0}^{m} \alpha_i.
\]

### 2.5 Complexification of the Schrödinger model

The infinitesimal action \(d\pi_\lambda\) in the Schrödinger model is given by second order differential operators on \(O_\lambda\) with polynomial coefficients. Hence the action can be extended to an action \(d\pi^C_\lambda\) of \(\mathfrak{g}\) on \(C^\infty(X_\lambda)\) by holomorphic differential operators. More precisely, let \(D\) be a differential operator on \(V\) with polynomial coefficients. Choose some basis \(e_1, \ldots, e_n\) of \(V\) and write \(x = \sum_{j=1}^{n} x_j e_j \in V\). Then \(D\) is in coordinates given by

\[
D = \sum_{\alpha \in \mathbb{N}^n_0} c_\alpha(x) \frac{\partial^{\left|\alpha\right|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}
\]

with polynomials \(c_\alpha(x)\) of which only finitely many are non-zero. The coefficients extend uniquely to holomorphic polynomials \(c_\alpha(z)\) on \(V_\mathbb{C}\). We define the complexification \(D^C\) of \(D\) to a holomorphic differential operator on \(V_\mathbb{C}\) by

\[
D^C := \sum_{\alpha \in \mathbb{N}^n_0} c_\alpha(z) \frac{\partial^{\left|\alpha\right|}}{\partial z_1^{\alpha_1} \cdots \partial z_n^{\alpha_n}}, \quad \text{where} \quad \frac{\partial}{\partial z_j} = \frac{1}{2} \left( \frac{\partial}{\partial x_j} - i \frac{\partial}{\partial y_j} \right)
\]

denotes the Wirtinger derivative for \(z_j = x_j + iy_j\) in coordinates \(z = \sum_{j=1}^{n} z_j e_j \in V_\mathbb{C}\). The operator \(D^C\) is a complexification of \(D\) in the sense that for a holomorphic function \(f\) on \(V_\mathbb{C}\) we have

\[
(D^C f)|_V = D(f|_V).
\]

Applying this procedure to \(d\pi_\lambda(X)\) for \(X \in \mathfrak{g}\) we put

\[
d\pi^C_\lambda(X) := d\pi_\lambda(X)^C.
\]

It remains to show that these holomorphic differential operators are actually tangential to the orbit \(X_\lambda\) and hence define an action \(d\pi^C_\lambda\) of \(\mathfrak{g}\) on \(C^\infty(X_\lambda)\). For this we use the Schrödinger model of certain unipotent representations of the complex group \(G_\mathbb{C} = \text{Co}(V_\mathbb{C})\) (the conformal group of the complex Jordan algebra \(V_\mathbb{C}\)), viewed as a real Lie group.
In [22] Proposition 2.14 the authors construct a representation $d\tau_\lambda$ of $g_C$, viewed as a real Lie algebra, on $C^\infty(X_\lambda)$ for each $\lambda \in W$. It is explicitly given by

\[
\begin{align*}
    d\tau_\lambda(u,0,0) &= i(u|z)_W, \\
    d\tau_\lambda(0,T,0) &= \partial_T z + \frac{r\lambda}{2n} \text{Tr}_W(T^#), \\
    d\tau_\lambda(0,0,v) &= i(v|B_W)_W.
\end{align*}
\]

By $\text{Tr}_W$ we mean the real trace of an operator on the real vector space $W$. Note that $d\tau_\lambda$ does not act via holomorphic differential operators, but via real differential operators up to second order on $X_\lambda$. Further, it is shown in [22] Theorem 1.12 that the representation $d\tau_\lambda$ is infinitesimally unitary with respect to the inner product of $L^2(X_\lambda, d\nu_\lambda)$.

We have the following result relating $d\tau_\lambda$ to the complexification $d\pi_C^\lambda$:

**Proposition 2.1.** For $X \in g$ we have

\[
    d\pi_C^\lambda(X) = \frac{1}{2} \left( d\tau_\lambda(X) - i d\tau_\lambda(iX) \right).
\]  

In particular, for every $X = (u,T,v) \in g$ and all $F,G \in C^\infty(X_\lambda)$ we have

\[
    \int_{X_\lambda} d\pi_C^\lambda(u,T,v) F(z) \cdot G(z) \, d\nu_\lambda(z) = \int_{X_\lambda} F(z) \cdot d\pi_C^\lambda(u,-T,v) G(z) \, d\nu_\lambda(z).
\]

**Remark 2.2.** The formula (2.4) can be understood as an analog of the Wirtinger derivative

\[
    \frac{\partial}{\partial z} = \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right).
\]

**Proof of Proposition 2.1.** First note that since $n$ and $\Pi$ together generate $g$ as a Lie algebra it suffices to show (2.4) for $X \in n$ and $X \in \Pi$. Further note that for $z \in V_C$ we have

\[
    (a|z) = \frac{1}{2} \langle (a|z)W - i(ia|z)W \rangle, \quad a \in V.
\]  

Then (2.4) is immediate for $X \in n$. Now let $X = (0,0,a) \in \Pi$. Let $(e_\alpha)$ be any orthonormal basis of $V$ with respect to the trace form $-|-$). Write $x = \sum_\alpha x_\alpha e_\alpha$. We view $W = V_C$ as a real Jordan algebra. The vectors

\[
    f_\alpha := \frac{1}{\sqrt{2}} e_\alpha \quad \text{and} \quad g_\alpha := \frac{1}{\sqrt{2}} ie_\alpha
\]

constitute an $\mathbb{R}$-basis of $W$. Its dual basis with respect to the trace form $-|-$ is given by $(f_\alpha := f_\alpha)_A \cup (g_\alpha := -g_\alpha)_A$. We write $z = \sum_\alpha z_\alpha e_\alpha$. We write $z = \sum_\alpha z_\alpha e_\alpha = \sum_\alpha (a_\alpha f_\alpha + b_\alpha g_\alpha)$ with $a_\alpha, b_\alpha \in \mathbb{R}$ and $z_\alpha = \frac{1}{\sqrt{2}} (a_\alpha + ib_\alpha)$. Hence,

\[
    \frac{\partial}{\partial a_\alpha} = \frac{1}{\sqrt{2}} \frac{\partial}{\partial x_\alpha} \quad \text{and} \quad \frac{\partial}{\partial b_\alpha} = \frac{1}{\sqrt{2}} \frac{\partial}{\partial y_\alpha}.
\]
\[ \frac{1}{2} \left( d\tau_\lambda(X) - i d\tau_\lambda(iX) \right) = \frac{i}{2} \left( (a|B^W_\lambda) - i(i|B^W_\lambda) \right) \]

\[ = i \sum_{\alpha,\beta} \left( \frac{\partial^2}{\partial a_\alpha \partial a_\beta}(a|P(\hat{f}_\alpha, \hat{f}_\beta)z) + 2 \frac{\partial^2}{\partial a_\alpha \partial b_\beta}(a|P(\hat{f}_\alpha, \hat{g}_\beta)z) \right) + i\lambda \sum_\alpha \left( \frac{\partial}{\partial a_\alpha} \hat{f}_\alpha + \frac{\partial}{\partial b_\alpha} \hat{g}_\alpha \right) \]

\[ = \frac{i}{4} \sum_{\alpha,\beta} \left( \frac{\partial^2}{\partial x_\alpha \partial x_\beta}(a|P(e_\alpha, e_\beta)z) - 2i \frac{\partial^2}{\partial x_\alpha \partial y_\beta}(a|P(e_\alpha, e_\beta)z) \right) + i\lambda \sum_\alpha \left( \frac{\partial}{\partial x_\alpha} e_\alpha - i \frac{\partial}{\partial y_\alpha} e_\alpha \right) \]

\[ = i \sum_{\alpha,\beta} \frac{\partial^2}{\partial z_\alpha \partial z_\beta}(a|P(e_\alpha, e_\beta)z) + i\lambda \sum_\alpha \frac{\partial}{\partial z_\alpha} e_\alpha = d\pi_C^\lambda(X). \]

This shows (2.4) for \( X \in \bar{n} \) and hence it follows for all \( X \in \mathfrak{g} \).

The stated integral formula now follows from (2.4) using the fact that \( d\tau_\lambda(X) \) is given by skew-adjoint real differential operators operators on \( L^2(X_\lambda, d\nu_\lambda) \) with real coefficients if \( X = (0, T, 0) \in \mathfrak{g} \) and purely imaginary coefficients if \( X = (u, 0, v) \in \mathfrak{g} \).

Since \( d\tau_\lambda \) restricts to an action on \( C^\infty(X_\lambda) \) by differential operators, the same is true for \( d\pi_C^\lambda \) by the previous proposition. Therefore, \( d\pi_C^\lambda \) is a representation of \( \mathfrak{g} \) on \( C^\infty(X_\lambda) \) by holomorphic polynomial differential operators of order at most 2.

3 Bessel functions on Jordan algebras

In this section we study \( J \), \( I \)- and \( K \)-Bessel functions on symmetric cones and their boundary orbits. These functions play a fundamental role in the study of Schrödinger and Fock models and the intertwining operators between them.

3.1 J-Bessel function

For \( \lambda \in \mathbb{C} \) with \( (\lambda)_m \neq 0 \) for all \( m \geq 0 \) and \( z, w \in V_\mathbb{C} \) we put

\[ \mathcal{J}_\lambda(z, w) := \sum_{m \geq 0} (-1)^m \frac{d_m}{(\lambda)_m} \Phi_m(z, w). \]

(This notation agrees with the one used in [15 Chapter XV.2].) One problem is that for a discrete Wallach point \( \lambda = k^2 \), \( k = 0, \ldots, r-1 \), we have \( (\lambda)_m = 0 \)
for all \( \mathbf{m} \geq 0 \) with \( m_{k+1} \neq 0 \). However, by Lemma [1.12] we find that for \( z \in \overline{X}_k \) or \( w \in \overline{X}_k \):

\[
\mathcal{J}_\lambda(z, w) = \sum_{m \geq 0, m_{k+1} = 0} (-1)^{|m|} \frac{d_m}{(\pi)^m(\lambda)_m} \Phi_m(z, w).
\]

In this expression the coefficients are non-singular at \( \lambda = k \frac{d}{2} \) since \( (\lambda)_m \neq 0 \) for all \( \mathbf{m} \geq 0 \) with \( m_{k+1} = 0 \). Therefore we obtain for each \( \lambda \in \mathcal{W} \) a \( J \)-Bessel function \( \mathcal{J}_\lambda(z, w) \) on \( V_C \times \overline{X}_\lambda \). It only remains to show convergence of the defining series.

**Lemma 3.1.** For \( \lambda \in \mathcal{W} \) the series for \( \mathcal{J}_\lambda(z, w) \) converges absolutely for all \( z \in V_C \) and \( w \in \overline{X}_\lambda \) and the following estimate holds:

\[
|\mathcal{J}_\lambda(z, w)| \leq C(1 + |z| \cdot |w|)^{\frac{n}{2}} e^{2r\sqrt{|z| \cdot |w|}} \quad \forall z \in V_C, w \in \overline{X}_\lambda
\]

for some constant \( C > 0 \) which depends only on the structure constants of \( V \) and \( \lambda \).

**Proof.** Let \( z, w \in V_C \). Then by [15] Proposition X.3.2] there exist \( u \in U \) and \( a = \sum_{j=1}^k a_j c_j \) with \( a_1 \geq \ldots \geq a_r \geq 0 \) such that \( w = ua = uP(a^{\frac{d}{2}})e \). With Lemma [1.11] (2) we find

\[
\Phi_m(z, w) = \Phi_m(z, uP(a^{\frac{d}{2}})e) = \Phi_m(P(a^{\frac{d}{2}})u^{-1}z, c) = \Phi_m(P(a^{\frac{d}{2}})u^{-1}z).
\]

Now suppose further that \( w \in \overline{X}_k \), \( 0 \leq k \leq r \), and \( m_{k+1} = 0 \). Then \( a_{k+1} = \ldots = a_r = 0 \) and hence \( P(a^{\frac{d}{2}}) \) projects onto \( V(\epsilon_k, 1) \subseteq \overline{X}_k \). Thus we find that \( P(a^{\frac{d}{2}})u^{-1}z \in \overline{X}_k \) and again by [15] Proposition X.3.2] we can write

\[
P(a^{\frac{d}{2}})u^{-1}z = u'b,
\]

where \( u' \in U \) and \( b = \sum_{j=1}^k b_j c_j \), \( b_1 \geq \ldots \geq b_k \geq 0 \). Now, by [15] Theorem XII.1.1 (i) we obtain

\[
|\Phi_m(z, w)| = |\Phi_m(u'b)| \leq b_1^{m_1} \cdots b_k^{m_k}.
\]

We further have the following obvious inequalities (assuming \( m_{k+1} = 0 \) for \( \lambda = k \frac{d}{2} \))

\[
d_m \leq \dim P_m(V_C) = \binom{n + |m| - 1}{n - 1} \leq C_1(1 + |m|)^{n-1}
\]

\[
\leq C_1(1 + m_1)^{n-1} \cdots (1 + m_k)^{n-1},
\]

\[
\frac{\Gamma}{(\pi)^m(\lambda)_m} = \prod_{j=1}^r \frac{\Gamma_n - (j - 1) \frac{d}{2}}{\Gamma_n - (m_j - 1)} \geq \prod_{j=1}^r (1) = m!,
\]

\[
\prod_{j=1}^k (\lambda - (j - 1) \frac{d}{2}) \geq \prod_{j=1}^r (\lambda - (k - 1) \frac{d}{2})m_j.
\]
We abbreviate \( \lambda' := \lambda - (k - 1) \frac{d}{d} \geq 0 \). Putting things together gives

\[
|J_{\lambda}(z, w)| \leq \sum_{m \geq 0, m_{k+1}=0} \frac{d_m}{(\mathcal{A})_{m}(\lambda)m} |\Phi_m(z, w)|
\]

\[
\leq C_1 \sum_{m \in \mathbb{N}_0} \frac{(1 + m_1)^{n-1} \cdots (1 + m_k)^{n-1}}{m!(\lambda_{1})_{m_1} \cdots (\lambda_{k})_{m_k}} b_1^{m_1} \cdots b_k^{m_k}
\]

\[
= C_1 \prod_{j=1}^k \left( \sum_{m=0}^{\infty} \frac{(1 + m)^{n-1}}{m!(\lambda')_{m}} b_j^m \right).
\]

Now note that \( mb^m = (b \frac{d}{db}) b^m \), so

\[
\sum_{m=0}^{\infty} \frac{(1 + m)^{n-1}}{m!(\lambda')_{m}} b^m = \left( 1 + b \frac{d}{db} \right)^{n-1} \sum_{m=0}^{\infty} \frac{1}{m!(\lambda')_{m}} b^m
\]

\[
= \left( 1 + b \frac{d}{db} \right)^{n-1} _0F_1(\lambda'; b),
\]

where \(_0F_1(\beta; z)\) denotes the hypergeometric function. For \(_0F_1(\beta; z)\) we have the obvious identity

\[
\frac{d}{dz} _0F_1(\beta; z) = \frac{\beta}{\beta + 1} _0F_1(\beta + 1; z)
\]

and hence \( (1 + b \frac{d}{db})^{n-1} _0F_1(\lambda'; b) \) is a linear combination of functions of the type

\[
b^k _0F_1(\lambda' + k; b), \quad k = 0, \ldots, n - 1.
\]

Now by \([3]\) equations (4.5.2) & (4.8.5) the asymptotic behaviour of the hypergeometric function as \( z \to \infty \) can be estimated by

\[
|_0F_1(\beta; z)| \lesssim |z|^{1-\frac{\beta}{\beta}} e^{2|z|^\frac{1}{2}}
\]

and hence we obtain

\[
\left| \left( 1 + b \frac{d}{db} \right)^{n-1} _0F_1(\lambda'; b) \right| \leq C_2 (1 + |b|)^{\frac{2n-2\lambda'}{2}-\frac{1}{2}} e^{2\sqrt{b}} \quad \forall b \in [0, \infty),
\]

the constant \( C_2 > 0 \) only depending on \( n \) and \( \lambda' \). Inserting this into the estimate for \( J_{\lambda}(z, w) \) gives

\[
|J_{\lambda}(z, w)| \leq C_1 C_2^k \prod_{j=1}^k \left( (1 + |b_j|)^{\frac{2n-2\lambda'}{2}-\frac{1}{2}} e^{2\sqrt{b_j}} \right).
\]
Now note that
\[ |b| = |a'| = |P(a^\frac{1}{2})u^{-1}z| \leq \|P(a^\frac{1}{2})\| \cdot |u^{-1}z| = \|P(a^\frac{1}{2})\| \cdot |z|, \]
where \(\|P(a^\frac{1}{2})\|^2\) is the largest eigenvalue of \(P(a^\frac{1}{2}) P(a^\frac{1}{2}) = P(a^\frac{1}{2})^2\). Since \(P(a^\frac{1}{2})^2\) acts on \(V_{ij}\) by \(a_i a_j\), its largest eigenvalue is \(a_i^2\). Hence
\[ |b| \leq a_1|z| \leq |a| \cdot |z| = |w| \cdot |z|. \]

Altogether we finally obtain
\[
|J_\lambda(z, w)| \leq C_1 C_2^k \left( \prod_{j=1}^k (1 + |b_j|) \right) \frac{2n-2 \lambda'}{k} e^{2(\sqrt{b_1} + \cdots + \sqrt{b_r})} \\
\leq C_1 C_2^k (1 + b_1 + \cdots + b_r)^{\frac{2n-2 \lambda'}{k}} e^{2\sqrt{b_1 + \cdots + b_r}} \\
\leq C_1 C_2^k \frac{e^{2n-2 \lambda'} k}{k} (1 + |b|)^{\frac{2n-2 \lambda'}{k}} e^{2\sqrt{|b|}} \\
\leq C (1 + |z| |w|)^{\frac{2n-2 \lambda'}{k}} e^{2\sqrt{|z| |w|}}
\]
with \(C = C_1 C_2^k \frac{e^{2n-2 \lambda'}}{k} > 0\) which gives the claim. \(\square\)

The estimate obtained in Lemma \[6.1\] is not sharp, but suffices for most of our purposes. Recently, Nakahama \[44\] obtained a sharper estimate which we use in Section \[6.2\] to find explicit Whittaker vectors:

**Proposition 3.2** (\[44\] Corollary 1.2). For \(\lambda \in \mathcal{W}\) and \(k \in \mathbb{N}_0\) with \(\Re \lambda + k > \frac{2n}{r} - 1\) there exists a constant \(C_{\lambda, k} > 0\) such that
\[ |J_\lambda(z^2, e)| \leq C_{\lambda, k} (1 + |z|^k) e^{2|\text{Im} z|}, \]
where \(|z| = \sum_{j=1}^r |a_j|\), \(z = u \sum_{j=1}^r a_j c_j\), \(u \in U\), \(a_j \in \mathbb{R}\).

**Example 3.3.** On the rank 1 orbit we have, using Lemma \[6.1\] and Example \[6.2\]
\[ J_\lambda(z, w) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m! (\lambda + m)_{m}^m} (z|w|)^m = \Gamma(\lambda) \tilde{J}_{\lambda-1}(2 \sqrt{|z| |w|}), \]
where \(\tilde{J}_{\alpha}(z) = (\frac{z}{2})^{-\alpha} J_{\alpha}(z)\) is the classical renormalized \(J\)-Bessel function which is an even entire function on \(\mathbb{C}\).

The following proposition is clear with the results of Section \[6.8\]

**Proposition 3.4.** The \(J\)-Bessel function \(J_\lambda(z, w)\) has the following properties:
(1) \(J_\lambda(z, w) = \overline{J_\lambda(w, z)}\) for \(z \in V_C\), \(w \in \overline{\mathcal{W}}\),
To prove the differential equation for \( \mathcal{J}_\lambda(z,w) \) we use the same method as Faraut–Koranyi [15] Theorem XV.2.6 for the one-variable \( J \)-Bessel function \( J(x,e) \). The first step is to calculate the Laplace transform as defined in (2.1).

**Lemma 3.5.** Fix \( w \in \mathcal{A}_X \) and consider the function \( \mathcal{J}_\lambda(-,w) \) on \( V \) given by \( x \mapsto \mathcal{J}_\lambda(x,w) \). The Laplace transform of \( \mathcal{J}_\lambda(-,w) \) is given by
\[
(L_\lambda \mathcal{J}_\lambda(-,w))(z) = 2^{r+1} \Delta(-iz)^{-\lambda} e^{-i(z^{-1}|w|)}, \quad z \in T_\Omega.
\]

**Proof.** Let \( 0 \leq k \leq r \) be such that \( \mathcal{O}_\lambda = \mathcal{O}_k \). The we find, using Lemma 1.10 and (1.17):
\[
(L_\lambda \mathcal{J}_\lambda(-,w))(z) = \int_{\mathcal{O}_\lambda} e^{iz|z|} \mathcal{J}_\lambda(x,w) \, d\mu_\lambda(x)
= \sum_{m \geq 0, m_{k+1}=0} (-1)^{|m|} \frac{d_m}{(z^{\lambda})^{|m|}} \int_{\mathcal{O}_\lambda} e^{iz|z|} \Phi_m(x,w) \, d\mu_\lambda(x)
= 2^{r+1} \sum_{m \geq 0} (-1)^{|m|} \frac{d_m}{(z^{\lambda})^{|m|}} \Delta(-iz)^{-\lambda} \Phi_m(i z^{-1}, w) = 2^{r+1} \Delta(-iz)^{-\lambda} e^{-i(z^{-1}|w|)}. \quad \square
\]

**Proposition 3.6.** For \( \lambda \in \mathcal{W} \) the function \( \mathcal{J}_\lambda(z,w) \) solves the following differential equation:
\[
(B_\lambda)z \mathcal{J}_\lambda(z,w) = -w \mathcal{J}_\lambda(z,w), \quad z \in V_\mathcal{C}, \ w \in \mathcal{A}_\mathcal{C}.
\]

**Proof.** First note that it suffices to show the differential equation for \( \lambda > (r - 1)\frac{1}{2} \), then the general case follows by analytic continuation. Further, it suffices to show the differential equation for \( z \in \Omega \) as \( \mathcal{J}_\lambda(z,w) \) is holomorphic in \( z \in V_\mathcal{C} \). Since the Laplace transform \( L_\lambda \) is injective on functions on \( \Omega \) the differential equation is equivalent to the identity
\[
L_\lambda(B_\lambda \mathcal{J}_\lambda(-,w)) = -w L_\lambda \mathcal{J}_\lambda(-,w).
\]

Let \( z \in T_\Omega \), then using the symmetry of the Bessel operator \( B_\lambda \) we find
\[
(L_\lambda(B_\lambda \mathcal{J}_\lambda(-,w))) (z) = \int_{\mathcal{O}_\lambda} e^{iz|z|} (B_\lambda)z \mathcal{J}_\lambda(x,w) \, d\mu_\lambda(x)
= \int_{\mathcal{O}_\lambda} (B_\lambda)z e^{iz|z|} \mathcal{J}_\lambda(x,w) \, d\mu_\lambda(x) = \int_{\mathcal{O}_\lambda} (P(z)x + iz) e^{iz|z|} \mathcal{J}_\lambda(x,w) \, d\mu_\lambda(x)
= i \left( P(z) \frac{\partial}{\partial z} + \lambda z \right) (L_\lambda \mathcal{J}_\lambda(-,w))(z).
\]

Now by Lemma 3.6 we have \( (L_\lambda \mathcal{J}_\lambda(-,w))(z) = 2^{r+1} \Delta(-iz)^{-\lambda} e^{-i(z^{-1}|w|)} \). Using
\[ D(x^{-1}) = -P(x)^{-1} \] and \( \partial_y \Delta(x) = \Delta(x)(y|x^{-1}) \) we find
\[
i \left( P(z) \frac{\partial}{\partial z} + \lambda z \right) \left[ \Delta(-iz)^{-\lambda} e^{-i(z^{-1}|w)} \right]
\begin{align*}
&= i \left( -\lambda P(z)z^{-1} + iP(z)P(z)^{-1}w + \lambda z \right) \Delta(-iz)^{-\lambda} e^{-i(z^{-1}|w)} \\
&= -\overline{w} \Delta(-iz)^{-\lambda} e^{-i(z^{-1}|w)}
\end{align*}

and (3.1) follows. \( \square \)

### 3.2 \( I \)-Bessel function

For \( \lambda \in \mathcal{W} \), \( z, w \in V_{\mathbb{C}} \) and \( z \in X_\lambda \) or \( w \in X_\lambda \) we put
\[
\mathcal{I}_\lambda(z, w) := \mathcal{J}_\lambda(-z, w) = \mathcal{J}_\lambda(z, -w)
\begin{align*}
&= \sum_{m \geq 0} \frac{d_m}{(\frac{z}{w})^m} \Phi_m(z, w).
\end{align*}

By definition the \( I \)-Bessel function \( \mathcal{I}_\lambda(z, w) \) also satisfies the estimates in Proposition 3.1 and has the same properties as in Proposition 3.4.

**Example 3.7.** By Example 3.3 it is immediate that on the rank 1 orbit we have
\[
\mathcal{I}_\lambda(z, w) = \Gamma(\lambda) \tilde{I}_{\lambda-1}(2\sqrt{z|w})), \quad z, w \in X_1,
\]
where \( \tilde{I}_\lambda(z) = \left( \frac{z}{w} \right)^{-\alpha} I_\lambda(z) \) is the classical renormalized \( I \)-Bessel function which is an even entire function on \( \mathbb{C} \).

**Lemma 3.8.** For \( \lambda \in \mathcal{W} \) and \( y \in \Omega \), \( z \in V_{\mathbb{C}} \)
\[
\int_{\Omega_\lambda} e^{-(z|y)} \mathcal{I}_\lambda(x, z) \, d\mu_\lambda(x) = 2^{r_\lambda} \Delta(y)^{-\lambda} e^{-(y^{-1}|w)}.
\]

**Proof.** This follows immediately from Lemma 5.5. \( \square \)

**Proposition 3.9.** For \( \lambda \in \mathcal{W} \) the function \( \mathcal{I}_\lambda(z, w) \) solves the following differential equation:
\[
(B_{\lambda, z}) \mathcal{I}_\lambda(z, w) = \overline{w} \mathcal{I}_\lambda(z, w), \quad z, w \in X_\lambda.
\]

**Proof.** Since \( \mathcal{I}_\lambda(z, w) = \mathcal{J}_\lambda(z, -w) \) this is equivalent to Proposition 3.6. \( \square \)
K-Bessel function

For $\lambda \in \mathbb{C}$ and $x \in \Omega$ we put

$$K_\lambda(x) := \int_{\Omega} e^{-\text{tr}(u^{-1} - (x|u))} \Delta(u)^{\lambda - \frac{2d}{r}} \, du = \int_{\Omega} e^{-\text{tr}(v - (x|v))} \Delta(v)^{-\lambda} \, dv.$$  

Note that our normalization of the parameter $\lambda$ differs from the one used in [8 and 15, Chapter XVI.3]. By [15, Proposition XVI.3.1] these integrals converge for all $\lambda \in \mathbb{C}$ and $x \in \Omega$. Since the integrand is positive on $\Omega$ we have $K_\lambda(x) > 0$ for $x \in \Omega$. To extend the $K$-Bessel function also to lower rank orbits we need the following result due to J.-L. Clerc [8, Théorème 4.1]:

**Proposition 3.10.** Let $c \in V$ be an idempotent of rank $k$. Let $\Omega_1$ and $\Omega_0$ be the symmetric cones in the Euclidean Jordan algebras $V(c,1)$ and $V(c,0)$, respectively. Further, let $K^{1}_{\lambda}$ be the $K$-Bessel function of $\Omega_1$, $\Gamma_{\Omega_0}$ the Gamma function of $\Omega_0$ and $n_0$ and $r_0$ the dimension and rank of $V(c,0)$. Then for $x_1 \in \Omega_1$

$$K_\lambda(x_1) = (2\pi)^k (r-k)^{\frac{d}{2}} \Gamma_{\Omega_0} \left( \frac{n_0}{r_0} + \frac{k}{2} - \lambda \right) K^{1}_\lambda(x_1). \quad (3.2)$$

This shows that for $\lambda$ near $k d^2$ the Bessel function $K_\lambda(x)$ is defined for $x \in \Omega_k$ and hence we obtain Bessel functions $K_\lambda$ on $\Omega_{\lambda}$ for $\lambda \in \mathcal{W}$. Note that by (3.2) the function $K_\lambda$ is positive on $\Omega_{\lambda}$.

**Example 3.11.** (1) For $V = \mathbb{R}$ we have by [18, formula 3.471 (9)]

$$K_\lambda(x) = 2K_{\lambda-1}(2\sqrt{x}), \quad x \in \Omega = \mathbb{R}_+,$$

where $\tilde{K}_\alpha(z) = (\frac{z}{2})^{-\alpha} K_{\alpha}(z)$ is the classical renormalized $K$-Bessel function.

(2) In the general case the Bessel function $K_\lambda$ is by Proposition 3.10 on the rank 1 orbit $\Omega_1$ given by

$$K_\lambda(x) = \text{const} \cdot K^{1}_\lambda(|x|c_1) = \text{const} \cdot \tilde{K}_{\lambda-1}(2\sqrt{|x|}), \quad x \in \Omega_1.$$

**Lemma 3.12.** For $\lambda \in \mathcal{W}$ and $m \geq 0$ we have

$$\int_{\Omega_{\lambda}} p(x)K_\lambda(x) \, d\mu_\lambda(x) = 2^\lambda \Gamma_{\Omega} \left( \frac{n}{r} \right) (\frac{\lambda}{n}) m(\lambda) p(e), \quad p \in \mathcal{P}_m(V_{\mathbb{C}}).$$

In particular, for every $N > 0$ we have

$$\int_{\Omega_{\lambda}} (1 + |x|^N) K_\lambda(x) \, d\mu_\lambda(x) < \infty.$$
Proof. It suffices to show the claim for $\lambda > (r - 1)\frac{d}{2}$. The general case then follows by analytic continuation. For $\lambda > (r - 1)\frac{d}{2}$ we have, using Lemma 1.10

\[
\int_{0}^{\lambda} p(x)K_{\lambda}(x)\,d\mu_{\lambda}(x) = \frac{2^{r\lambda}}{\Gamma_\Omega(\lambda)} \int_{\Omega} p(x)K_{\lambda}(x)\Delta(x)^{\lambda - \frac{n}{2}}\,dx
\]

\[
= \frac{2^{r\lambda}}{\Gamma_\Omega(\lambda)} \int_{\Omega} \int_{\Omega} p(x)e^{-\text{tr}(u^{-1})-(x|u)}\Delta(u)^{\lambda - \frac{n}{2}}\Delta(x)^{\lambda - \frac{n}{2}}\,dx\,du
\]

\[
= 2^{r\lambda}(\lambda)m \int_{\Omega} e^{-\text{tr}(u^{-1})}p(u^{-1})\Delta(u)^{-\frac{n}{2}}\,du
\]

and under the coordinate change $v = u^{-1}$, $dv = \Delta(u)^{-\frac{n}{2}}\,du$, this is

\[
= 2^{r\lambda}(\lambda)m \int_{\Omega} e^{-\text{tr}(v)}p(v)\,dv = 2^{r\lambda}(\lambda)m\Gamma_\Omega(\frac{\Omega}{2})(\frac{\Omega}{m})mp(e),
\]

where we have again used Lemma 1.10 for the last equality. This shows the desired integral formula.

For the second claim we observe that by the previous calculations every polynomial can be integrated against the positive measure $K_{\lambda}(x)\,d\mu_{\lambda}(x)$ since $\mathcal{P}(V_\mathcal{C}) = \bigoplus_{m \geq 0} \mathcal{P}_m(V_\mathcal{C})$ and hence the claim follows.

**Proposition 3.13.** The function $K_{\lambda}(x)$ solves the following differential equation:

\[B_{\lambda}K_{\lambda}(x) = eK_{\lambda}(x).\]

Proof. Differentiating under the integral we obtain

\[
B_{\lambda}K_{\lambda}(x) = \int_{\Omega} (P(-u)x + \lambda(-u))e^{-\text{tr}(u^{-1})-(x|u)}\Delta(u)^{\lambda - \frac{n}{2}}\,du
\]

\[
= \int_{\Omega} -\left( P(u)\frac{\partial}{\partial u} + \lambda u \right)e^{-(x|u)}\cdot e^{-\text{tr}(u^{-1})}\Delta(u)^{\lambda - \frac{n}{2}}\,du
\]

\[
= \sum_{\alpha} \int_{\Omega} -\left( P(u)e_{\alpha}\frac{\partial}{\partial u_{\alpha}} + \lambda u \right)e^{-(x|u)}\cdot e^{-\text{tr}(u^{-1})}\Delta(u)^{\lambda - \frac{n}{2}}\,du
\]

\[
= \sum_{\alpha} \int_{\Omega} e^{-(x|u)}\cdot \left( \frac{\partial}{\partial u_{\alpha}}P(u)e_{\alpha} - \lambda u \right) [e^{-\text{tr}(u^{-1})}\Delta(u)^{\lambda - \frac{n}{2}}]\,du.
\]
Using $\partial_y P(x) = 2P(x, y)$, $D(x^{-1}) = -P(x)^{-1}$ and $\partial_y \Delta(x) = \Delta(x)(y|x^{-1})$ we obtain

$$\sum \int e^{-x|u} \left( 2P(e_\alpha, u)e_\alpha + P(u)e_\alpha \text{tr}(P(u)^{-1}e_\alpha) ight)$$

$$+ (\lambda - \frac{2n}{r})P(u)e_\alpha - \lambda u \right) e^{-\text{tr}(u^{-1})}\Delta(u)^{\lambda - \frac{2n}{r}} du$$

$$= \int e^{-x|u} \left( 2 \left( \sum e_\alpha \right)^2 u + P(u)P(u)^{-1} e^2 ight)$$

$$+ (\lambda - \frac{2n}{r})P(u)u^{-1} - \lambda u \right) e^{-\text{tr}(u^{-1})}\Delta(u)^{\lambda - \frac{2n}{r}} du.$$ 

By (1.1) this is

$$= \int e^{-x|u} \left( \frac{2n}{r} u + e + (\lambda - \frac{2n}{r})u - \lambda u \right) e^{-\text{tr}(u^{-1})}\Delta(u)^{\lambda - \frac{2n}{r}} du$$

$$= e\mathcal{K}_\lambda(x).$$

Now let $\lambda \in \mathcal{W}$. For $z \in \mathcal{X}_\lambda$, $z = ua$ with $u \in U$, $a = \sum_{i=1}^r t_i c_i$, $t_i \geq 0$, we put

$$\omega_\lambda(z) := \mathcal{K}_\lambda \left( \left( \frac{a}{2} \right)^2 \right).$$

We note that $\omega_\lambda$ is positive on $\mathcal{X}_\lambda$.

**Proposition 3.14.** The function $\omega_\lambda(z)$ solves the following differential equation:

$$B_\lambda \omega_\lambda(z) = \frac{3}{4} \omega_\lambda(z), \qquad z \in \mathcal{X}_\lambda.$$ 

**Proof.** Recall the operators $B_\lambda^Y$ and $B_\lambda^W$ acting on functions of $r$ variables (see Proposition 1.5). Let $F(a_1, \ldots, a_r) = \mathcal{K}_\lambda(a)$, $a = \sum_{i=1}^r a_i c_i$, then $F$ solves the system $(B_\lambda^Y)^i F = F$, $i = 1, \ldots, r$ by Propositions 1.5 and 3.13. Put

$\text{Documenta Mathematica} 18$ (2013) 785–855
$$G(a_1,\ldots,a_r) := \omega_\lambda(a) = F((\frac{\omega}{2})^2,\ldots, (\frac{\omega}{2})^2).$$ Then

$$(B^W_\lambda)^t G(a_1,\ldots,a_r)$$

$$= \frac{1}{4} \left( a_i \frac{\partial^2}{\partial a_i^2} + (2\lambda - 1 - (r - 1)d) \frac{\partial}{\partial a_i} \right)$$

$$+ \frac{d}{2} \sum_{j \neq i} \left( \frac{1}{a_i - a_j} + \frac{1}{a_i + a_j} \right) \left( a_i \frac{\partial}{\partial a_i} - a_j \frac{\partial}{\partial a_j} \right) G(a_1,\ldots,a_r)$$

$$= \frac{1}{4} \left( a_i \left( \frac{a_i^2 - 2d}{4} \right) + \frac{d}{2} \frac{\partial}{\partial a_i} \right)$$

$$+ \frac{d}{2} \sum_{j \neq i} \left( \frac{1}{a_i - a_j} + \frac{1}{a_i + a_j} \right) \left( \frac{a_i^2}{2} \frac{\partial}{\partial a_i} - \frac{a_j^2}{2} \frac{\partial}{\partial a_j} \right) F\left( \left( \frac{a_i}{2} \right)^2,\ldots, \left( \frac{a_r}{2} \right)^2 \right)$$

$$= \frac{a_i}{4} \left( \frac{a_i}{2} \right)^2 \frac{\partial^2}{\partial a_i^2} + \left( \lambda - (r - 1)d \right) \frac{\partial}{\partial a_i}$$

$$+ \frac{d}{2} \sum_{j \neq i} \left( \frac{1}{(\frac{a_i}{2})^2 - (\frac{a_j}{2})^2} \right) \left( \frac{a_i^2}{2} \frac{\partial}{\partial a_i} - \frac{a_j^2}{2} \frac{\partial}{\partial a_j} \right) F\left( \left( \frac{a_i}{2} \right)^2,\ldots, \left( \frac{a_r}{2} \right)^2 \right)$$

$$= \frac{a_i}{4} (B^W_\lambda)^t F\left( \left( \frac{a_i}{2} \right)^2,\ldots, \left( \frac{a_r}{2} \right)^2 \right) = \frac{a_i}{4} \frac{\partial}{\partial a_i}$$

Hence we obtain $B^W_\lambda \omega_\lambda(z) = \frac{\tau}{4} \omega_\lambda(z)$ for $z = a$. Since $\omega_\lambda(z)$ is further U-invariant, we obtain with (1.11) for $z = ua$ with $u \in U$ and $a = \sum_{i=1}^{r} a_i c_i$:

$B^W_\lambda \omega_\lambda(z) = (u^{-1})^# B^W_\lambda \omega_\lambda(a) = \tau^a \omega_\lambda(a) = \frac{\tau}{4} \omega_\lambda(z)$

since $\tau^# = u^{-1}$ for $u \in U$. Finally we use Proposition 2.11 to find for every $a \in V$:

$$i(a|B_\lambda) \omega_\lambda(z) = d \tau^a_\lambda (0,0,a) \omega_\lambda(z)$$

$$= \frac{1}{2} \left( d \tau_\lambda (0,0,a) \omega_\lambda(z) - i d \tau_\lambda (0,0,i a) \omega_\lambda(z) \right)$$

$$= \frac{1}{2} i \left( (a|B^W_\lambda)_W \omega_\lambda(z) - i (i a|B^W_\lambda)_W \omega_\lambda(z) \right)$$

$$= \frac{1}{2} i \left( (a|\frac{\tau}{4})_W - i (i a|\frac{\tau}{4})_W \right) \omega_\lambda(z)$$

$$= i(a|\frac{\tau}{4}) \omega_\lambda(z).$$

Since this holds for any $a \in V$ we find $B_\lambda \omega_\lambda(z) = \frac{\tau}{4} \omega_\lambda(z)$ and the proof is complete. □
Example 3.15. On the rank 1 orbit the function $\omega_\lambda$ takes by Example 3.11 (2) the form

$$\omega_\lambda(z) = \omega_\lambda(|z|c_1) = K_\lambda \left( \frac{|z|^2}{4} c_1 \right) = \text{const} \cdot \tilde{K}_{\lambda-1}(|z|).$$

4 A Fock model for unitary highest weight representations of scalar type

In this section we construct a Fock space $F_\lambda = F(X_\lambda, \omega_\lambda d\nu_\lambda)$ of holomorphic functions on the orbit $X_\lambda$ for every $\lambda \in W$, calculate its reproducing kernel and find a realization on $F_\lambda$ of the unitary highest weight representation corresponding to the Wallach point $\lambda$.

4.1 Construction of the Fock space

Let $\lambda \in W$. Recall the positive function $\omega_\lambda \in C^\infty(X_\lambda)$ from Section 3.3. We endow the space $P(X_\lambda)$ of polynomials on $X_\lambda$ with the $L^2$-inner product of $L^2(X_\lambda, \omega_\lambda d\nu_\lambda)$:

$$\langle F,G \rangle_{F_\lambda} := \frac{1}{c_\lambda} \int_{X_\lambda} F(z) \overline{G(z)} \omega_\lambda(z) d\nu_\lambda(z), \quad F,G \in P(X_\lambda) \quad (4.1)$$

with $c_\lambda = 2^{\nu_\lambda} \Gamma(\frac{n-1}{2})$. This turns $P(X_\lambda)$ into a pre-Hilbert space. Its completion $F_\lambda := F(X_\lambda, \omega_\lambda d\nu_\lambda)$ will be called the Fock space on $X_\lambda$.

It remains to show that the integral in (4.1) converges. Using the polarization principle the following lemma suffices:

**Lemma 4.1.** For $F \in P(V_C)$ we have

$$\int_{X_\lambda} |F(z)|^2 \omega_\lambda(z) d\nu_\lambda(z) < \infty.$$

**Proof.** Using the integral formula (1.10) we obtain

$$\int_{X_\lambda} |F(z)|^2 \omega_\lambda(z) d\nu_\lambda(z) = \int_U \int_{\partial X} |F(uxz^2)|^2 K_\lambda(x) d\mu_\lambda(x) du.$$

Now put

$$p(x) := \int_U |F(ux)|^2 du.$$

Clearly $p$ is a polynomial on $V$, so there are constants $C_1 > 0$ and $N \in \mathbb{N}$ such that $|p(x)| \leq C_1 (1 + |x|)^N$. Now, every $x \in \overset{\circ}{V}$ has a decomposition $x = ka$ with $k \in K^L$ and $a = \sum_{j=1}^n a_j c_j$, $a_j \geq 0$. In this decomposition the square root
is of the form $x^{\frac{1}{2}} = ka^{\frac{1}{2}}$ with $a^{\frac{1}{2}} = \sum_{j=1}^{r} a_{j}^{\frac{1}{2}} c_{j}$. Since the norm $|\cdot|$ on $V$ is $K^r$-invariant we obtain

\[ |p(x^{\frac{1}{2}})| \leq C_1 \left(1 + (a_1 + \cdots + a_r)^{\frac{1}{2}}\right)^{N} \leq C_1 \left(1 + \sqrt{r}(a_1^2 + \cdots + a_r^2)^{\frac{1}{2}}\right)^{N} \]

\[ \leq C_1 \sqrt{r} \left(1 + |x|^\frac{1}{2}\right)^{N} \leq C_1 \sqrt{r} \left(\sqrt{2}(1 + |x|^\frac{1}{2})\right) = C_2 (1 + |x|)^\frac{2}{r} \]

with $C_2 = C_1 \sqrt{2^{\frac{2}{r}}}$. Hence, we find

\[ \int_{\mathcal{X}_{\lambda}} |F(z)|^2 \omega_{\lambda}(z) \, d\nu_{\lambda}(z) = \int_{\mathcal{O}_{\lambda}} p(x) \mathcal{K}_{\lambda}(x) \, d\mu_{\lambda}(x) \]

\[ \leq C_2 \int_{\mathcal{O}_{\lambda}} (1 + |x|)^\frac{2}{r} \mathcal{K}_{\lambda}(x) \, d\mu_{\lambda}(x). \]

The latter integral is finite by Lemma 5.12 and the proof is complete. $\square$

We explicitly calculate the norms on the finite-dimensional subspaces $\mathcal{P}_m(\mathcal{X}_{\lambda})$.

**Proposition 4.2.** Let $m, n \geq 0$ and $F \in \mathcal{P}_m(\mathcal{X}_{\lambda})$, $G \in \mathcal{P}_n(\mathcal{X}_{\lambda})$. Then

\[ \langle F, G \rangle_{\mathcal{X}_{\lambda}} = 4^{\text{im}\left(\frac{m}{2}\right)} \text{im}(\lambda)^m \langle F, G \rangle_{\Sigma} \]

In particular the subspaces $\mathcal{P}_m(\mathcal{X}_{\lambda}) \subseteq \mathcal{P}(\mathcal{X}_{\lambda})$ are pairwise orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{X}_{\lambda}}$ and for $F = \Phi_m$ we have

\[ \|\Phi_m\|_{\mathcal{X}_{\lambda}}^2 = \frac{4^{\text{im}\left(\frac{m}{2}\right)} \text{im}(\lambda)^m}{d_m}. \]

**Proof.** Using the integral formula (1.10), Lemma 1.14 and Lemma 5.12 we obtain

\[ \langle F, G \rangle_{\mathcal{X}_{\lambda}} = \frac{1}{c_{\lambda}} \int_{\mathcal{X}_{\lambda}} F(z) \overline{G(z)} \omega_{\lambda}(z) \, d\nu_{\lambda}(z) \]

\[ = \frac{1}{c_{\lambda}} \int_{\mathcal{O}_{\lambda}} \int_{U} F(ux^{\frac{1}{2}}) \overline{G(ux^{\frac{1}{2}})} \omega_{\lambda}(x^{\frac{1}{2}}) \, d\mu_{\lambda}(x) \, du \]

\[ = \frac{1}{c_{\lambda}} \langle F,G \rangle_{\Sigma} \int_{\mathcal{O}_{\lambda}} \Phi_m(x) \mathcal{K}_{\lambda}(\frac{x^{\frac{1}{2}}}{2}) \, d\mu_{\lambda}(x) \]

\[ = 4^{r}\lambda+|m| \frac{1}{c_{\lambda}} \langle F,G \rangle_{\Sigma} \int_{\mathcal{O}_{\lambda}} \Phi_m(y) \mathcal{K}_{\lambda}(y) \, d\mu_{\lambda}(y) \]

\[ = 4^{\text{im}\left(\frac{m}{2}\right)} \text{im}(\lambda)^m \langle F, G \rangle_{\Sigma}. \]

Since $\|\Phi_m\|_{\Sigma}^2 = \frac{1}{d_m}$ by (1.15) this finishes the proof. $\square$

**Remark 4.3.** Comparing the norm on $\mathcal{F}_{\lambda}$ with the norm on the space $\mathcal{H}_{\lambda}^2(D)$ gives by (2.2)

\[ \|F\|_{\mathcal{F}_{\lambda}}^2 = 4^{\text{im}(\lambda)} \|F\|_{\mathcal{H}_{\lambda}^2(D)}^2 \quad \forall F \in \mathcal{P}_m(V_{\Sigma}). \]
If we denote by $O(\mathcal{X}_\lambda)$ the space of holomorphic functions on the complex manifold $\mathcal{X}_\lambda$, we obtain the following result:

**Proposition 4.4.** $O(\mathcal{X}_\lambda) \cap L^2(\mathcal{X}_\lambda, \omega_\lambda \, d\nu_\lambda)$ is a closed subspace of $L^2(\mathcal{X}_\lambda, \omega_\lambda \, d\nu_\lambda)$ and the point evaluation $O(\mathcal{X}_\lambda) \cap L^2(\mathcal{X}_\lambda, \omega_\lambda \, d\nu_\lambda) \to \mathbb{C}$, $F \mapsto F(z)$, is continuous for every $z \in \mathcal{X}_\lambda$. In particular, $\mathcal{F}_\lambda \subseteq O(\mathcal{X}_\lambda) \cap L^2(\mathcal{X}_\lambda, \omega_\lambda \, d\nu_\lambda)$ and the point evaluation $\mathcal{F}_\lambda \to \mathbb{C}$, $F \mapsto F(z)$, is continuous for every $z \in \mathcal{X}_\lambda$.

**Proof.** This is a local statement and hence, we may transfer it with a chart map to an open domain $U \subseteq \mathbb{C}^k$. Here the measure $\omega_\lambda \, d\nu_\lambda$ is absolutely continuous with respect to the Lebesgue measure $dz$ and hence it suffices to show that $O(C^k) \cap L^2(C^k, dz)$ is a closed subspace with continuous point evaluations. This is done e.g. in [20, Proposition 3.1 and Corollary 3.2].

The particular choice of the density $\omega_\lambda$ yields the following result:

**Proposition 4.5.** The adjoint of $\mathcal{B}_\lambda$ on $\mathcal{F}_\lambda$ is $\frac{z}{4}$.

**Proof.** Let $F$ and $G$ be holomorphic functions on $\mathcal{X}_\lambda$. Then by Proposition [2.1] we know that

$$\int_{\mathcal{X}_\lambda} \mathcal{B}_\lambda F(z) \overline{G(z)} \omega_\lambda(z) \, d\nu_\lambda(z) = \int_{\mathcal{X}_\lambda} F(z) \mathcal{B}_\lambda \left[ \overline{G(z)} \omega_\lambda(z) \right] \, d\nu_\lambda(z).$$

The function $\overline{G(z)}$ is antiholomorphic and hence

$$= \int_{\mathcal{X}_\lambda} F(z) \overline{G(z)} \mathcal{B}_\lambda \omega_\lambda(z) \, d\nu_\lambda(z).$$

By Proposition [5.14] we have $\mathcal{B}_\lambda \omega_\lambda(z) = \frac{z}{4} \omega_\lambda(z)$ and therefore

$$= \int_{\mathcal{X}_\lambda} F(z) \frac{z}{4} \overline{G(z)} \omega_\lambda(z) \, d\nu_\lambda(z).$$

### 4.2 The Bessel–Fischer inner product

We introduce another inner product on the space $\mathcal{P}(\mathcal{X}_\lambda)$ of polynomials, the Bessel–Fischer inner product. For two polynomials $p$ and $q$ it is defined by

$$[p, q]_\lambda := p(B_\lambda \overline{\mathcal{P}(z)})_{z=0},$$

where $\mathcal{P}(z) = \overline{q(z)}$ is obtained by conjugating the coefficients of the polynomial $q$. A priori it is not even clear that this sesquilinear form is positive definite.

**Theorem 4.6.** For $p, q \in \mathcal{P}(\mathcal{X}_\lambda)$ we have

$$[p, q]_\lambda = \langle p, q \rangle_{\mathcal{F}_\lambda}.$$  \hspace{1cm} (4.2)
The proof is similar to the proof of [3, Proposition 3.8]

Proof. First note that for all $p, q \in \mathcal{P}(\lambda)$

\[
[(a|\frac{\lambda}{z})p, q]_\lambda = [p, (\overline{D}_\lambda)q]_\lambda, \quad \text{for } a \in \mathbb{C},
\]

\[
\langle (a|\frac{\lambda}{z})p, q \rangle_{\mathcal{F}_\lambda} = \langle p, (\overline{D}_\lambda)q \rangle_{\mathcal{F}_\lambda}, \quad \text{for } a \in \mathbb{C}.
\]

In fact, the second equation follows from Proposition 4.3. The first equation is immediate since the components $(a|\lambda)$, $a \in \mathbb{C}$, of the Bessel operator form a commuting family of differential operators on $X_\lambda$. Therefore, $(a|\lambda)p(B_\lambda)q(4z) = 4p(B_\lambda)(\overline{D}_\lambda)q(4z)$ and the claim follows.

To prove (4.2) we proceed by induction on $\deg(q)$. First, if $p = q = 1$, the constant polynomial with value 1, it is clear that $[p, q]_\lambda = 1$ and by Proposition 4.3 we also have $\langle p, q \rangle_{\mathcal{F}_\lambda} = 1$. Thus, (4.2) holds for $\deg(p) = \deg(q) = 0$. If now $\deg(p)$ is arbitrary and $\deg(q) = 0$ then $(\overline{D}_\lambda)q = 0$ and hence

\[
[(a|\frac{\lambda}{z})p, q]_\lambda = [p, (\overline{D}_\lambda)q]_\lambda = 0
\]

and

\[
\langle (a|\frac{\lambda}{z})p, q \rangle_{\mathcal{F}_\lambda} = \langle p, (\overline{D}_\lambda)q \rangle_{\mathcal{F}_\lambda} = 0.
\]

Therefore, (4.2) holds if $\deg(q) = 0$. We note that (4.2) also holds if $\deg(p) = 0$ and $\deg(q)$ is arbitrary. In fact,

\[
[p, q]_\lambda = p(0)q(0) = \overline{[q, p]}_\lambda \quad \text{and} \quad \langle p, q \rangle_{\mathcal{F}_\lambda} = \overline{\langle q, p \rangle}_{\mathcal{F}_\lambda},
\]

and (4.2) follows from the previous considerations. Now assume (4.2) holds for $\deg(q) \leq k$. For $\deg(q) \leq k + 1$ we then have $\deg((\overline{D}_\lambda)q) \leq k$ and hence, by the assumption

\[
[(a|\frac{\lambda}{z})p, q]_\lambda = [p, (\overline{D}_\lambda)q]_\lambda = \langle p, (\overline{D}_\lambda)q \rangle_{\mathcal{F}_\lambda} = \langle (a|\frac{\lambda}{z})p, q \rangle_{\mathcal{F}_\lambda}.
\]

This shows (4.2) for $\deg(q) \leq k + 1$ and $p(0) = 0$, i.e. without constant term. But for constant $p$, i.e. $\deg(p) = 0$, we have already seen that (4.2) holds and therefore the proof is complete. $\square$

4.3 Unitary action on the Fock space

In Section 2.3 we verified that for every $\lambda \in \mathcal{W}$ the complexification $d\pi_X^\lambda$ of the action $d\pi_X$ defines a Lie algebra representation of $\mathfrak{g}$ on $C^\infty(X_\lambda)$ by holomorphic polynomial differential operators in $z$. It is clear that this action preserves the subspace $\mathcal{P}(\lambda)$ of holomorphic polynomials. Using this we now construct an action of $\mathfrak{g}$ on $\mathcal{P}(\lambda)$ by composing the action $d\pi_X^\lambda$ with the Cayley type transform $C \in \text{Int}(\mathfrak{g})$ introduced in Section 1.3.

Definition 4.7. Let $\lambda \in \mathcal{W}$. On $\mathcal{P}(\lambda)$ we define a $\mathfrak{g}$-action $d\rho_\lambda$ by

\[
d\rho_\lambda(X) := d\pi_X^\lambda(C(X)), \quad X \in \mathfrak{g}.
\]
Proposition 4.8. Let $\lambda \in \mathcal{W}$ and $k \in \{0, \ldots, r\}$ such that $\mathcal{X}_\lambda = \mathcal{X}_k$. The $\mathfrak{k}$-type decomposition of $(d\rho_\lambda, \mathcal{P}(\mathcal{X}_\lambda))$ is given by

$$\mathcal{P}(\mathcal{X}_\lambda) = \bigoplus_{m \geq 0, m_{k+1} = 0} \mathcal{P}_m(\mathcal{X}_\lambda)$$

and in every $\mathfrak{k}$-type $\mathcal{P}_m(\mathcal{X}_\lambda)$ the space of $\mathfrak{k}^1$-fixed vectors is one-dimensional and spanned by the polynomial $\Phi_m(z)$. In particular $(d\rho_\lambda, \mathcal{P}(\mathcal{X}_\lambda))$ is an admissible $(\mathfrak{g}, \mathfrak{k})$-module.

Proof. The Cayley type transform $C : \mathfrak{g}_\mathbb{C} \to \mathfrak{g}_\mathbb{C}$ maps $\mathfrak{t}_\mathbb{C}$ to $\mathfrak{l}_\mathbb{C}$ and $\mathfrak{t}^1$ to itself. The action of $\mathfrak{l}_\mathbb{C}$ in the complexified Schrödinger model $d\pi_C^L$ is induced by the action of $\mathcal{L}_\mathbb{C}$ on the orbit $\mathcal{X}_\lambda = \mathcal{L}_\mathbb{C} \cdot e_k$ up to multiplication by a character. Hence the $\mathfrak{k}$-type decomposition of $\mathcal{P}(\mathcal{X}_\lambda)$ is the same as the decomposition into $\mathcal{L}_\mathbb{C}$-representations under the natural action $\ell$ of $\mathcal{L}_\mathbb{C}$. Therefore the claimed decomposition is clear by Corollary 4.10. The action of $\mathfrak{t}^1$ is induced by the natural action of $\mathcal{K}^L$ and hence the unique (up to scalar) $\mathfrak{k}^1$-invariant vector in the $\mathfrak{k}$-type $\mathcal{P}_m(\mathcal{X}_\lambda)$ is the unique (up to scalar) $\mathcal{K}^L$-invariant vector in $\mathcal{P}_m(\mathcal{X}_\lambda)$ under the natural action $\ell$. This finishes the proof. □

Proposition 4.9. For each $\lambda \in \mathcal{W}$ the representation $(d\rho_\lambda, \mathcal{P}(\mathcal{X}_\lambda))$ is an irreducible $(\mathfrak{g}, \mathfrak{k})$-module.

Proof. Let $m, n \geq 0$ be arbitrary and denote by $\mathcal{U}(\mathfrak{g})$ the universal enveloping algebra of $\mathfrak{g}$. Then it suffices to show that there exists an element of $d\rho_\lambda(\mathcal{U}(\mathfrak{g}))$ which maps $\Phi_m$ to $\Phi_n$. Using Theorem 4.10 we have

$$0 \neq \|\Phi_m\|_{\mathcal{X}_\lambda}^2 = \Phi_m(\mathcal{B}_\lambda) \Phi_m(4z)|_{z=0} = 4^{|m|} \Phi_m(\mathcal{B}_\lambda) \Phi_m(z)|_{z=0}.$$

Since $\mathcal{B}_\lambda$ is a differential operator of Euler degree $-1$ the polynomial $\Phi_m(\mathcal{B}_\lambda) \Phi_m(z)$ is constant and by the previous observation it is non-zero. Note that $d\rho_\lambda(\mathcal{U}(\mathfrak{g})) = d\pi_C^L(\mathcal{U}(\mathfrak{g}))$ contains multiplication by arbitrary polynomials and differential operators which are polynomials in the Bessel operator $\mathcal{B}_\lambda$. Hence, the operator $\Phi_m(z) \Phi_m(\mathcal{B}_\lambda) \in d\rho_\lambda(\mathcal{U}(\mathfrak{g}))$ maps $\Phi_m(z)$ to a non-zero multiple of $\Phi_n(z)$ and the claim follows. □

Proposition 4.10. The $(\mathfrak{g}, \mathfrak{k})$-module $(d\rho_\lambda, \mathcal{P}(\mathcal{X}_\lambda))$ is infinitesimally unitary with respect to the $L^2$-inner product $\langle -, - \rangle_{\mathcal{X}_\lambda}$.

Proof. As remarked in Section 4.2 the subspace $\{(a, L(b), a) : a, b \in \mathcal{V}\} \subseteq \mathfrak{g}$ generates $\mathfrak{g}$ as a Lie algebra and therefore it suffices to show that the operators

$$d\rho_\lambda(a, 0, a) = \frac{1}{2} (a | 4\mathcal{B}_\lambda + z) \quad \text{and} \quad d\rho_\lambda(0, 2L(a), 0) = \frac{1}{2} (a | 4\mathcal{B}_\lambda - z)$$

are skew-adjoint on $\mathcal{P}(\mathcal{X}_\lambda)$ with respect to the $L^2$-inner product $\langle -, - \rangle_{\mathcal{X}_\lambda}$. But this is clear by Proposition 4.12. □
Theorem 4.11. The \((\mathfrak{g}, \mathfrak{t})\)-module \((d\rho_\lambda, \mathcal{P}(X_\lambda))\) integrates to an irreducible unitary representation \(\rho_\lambda\) of the universal cover \(\tilde{G}\) of \(G\) on \(F_\lambda\). This representation factors to a finite cover of \(G\) if and only if \(\lambda \in \mathbb{Q}\). In particular it factors to a finite cover of \(G\) if \(\lambda \in W_{\text{disc}}\).

Proof. By the previous results it only remains to check in which cases the minimal \(k\)-type \(P_0(X_\lambda) = C_1\) integrates to a finite cover. The \(k\)-action on \(1\) is given by
\[
d\rho_\lambda(a, D, -a) 1 = d\pi^\mathbb{C}(0, D + 2iL(a), 0) 1 = \frac{r_\lambda}{2n} \text{Tr}(2iL(a)) 1 = i\lambda \text{tr}(a) 1.
\]
Therefore, the center \(Z(\mathfrak{k}) = \mathbb{R}(e, 0, -e)\) acts by
\[
d\rho_\lambda(e, 0, -e) 1 = ir\lambda 1.
\]
In \(K\) we have \(e^{\pi(e, 0, -e)} = 1\) and hence, the claim follows.

In the Fock model the action of the maximal compact subgroup is quite explicit. For this recall the group homomorphism \(\eta: \tilde{K} \to U \subseteq L_C\) with differential \(d\eta(u, D, -u) = D + 2iL(u)\) defined in Section 1.4 and the character \(\xi_\lambda: \tilde{K} \to \mathbb{T}\) of \(\tilde{K}\) with differential \(d\xi_\lambda(u, T, -u) = i\lambda \text{tr}(u)\) defined in Section 2.1.

Proposition 4.12. For \(k \in \tilde{K}\) we have
\[
\rho_\lambda(k) F(z) = \xi_\lambda(k) F(\eta(k)^# z), \quad z \in X_\lambda.
\]
Proof. The action of \(X = (u, D, -u) \in \mathfrak{t}\) is given by
\[
d\rho_\lambda(X) = d\pi^\mathbb{C}(0, d\eta(X), 0)
\]
\[
= \partial_{d\eta(X)^#} + \frac{r_\lambda}{2n} \text{Tr}(d\eta(X)^#) = \partial_{d\eta(X)^#} + i\lambda \text{tr}(u)
\]
which implies the claim.

4.4 The reproducing kernel

We now calculate the reproducing kernel \(K_\lambda(z, w)\) of the Hilbert space \(F_\lambda\). For this we first calculate the reproducing kernels \(K^m_\lambda(z, w)\) on the finite-dimensional subspaces \(P_m(X_\lambda)\) endowed with the inner product of \(F_\lambda\).

Lemma 4.13. Let \(\lambda \in \mathcal{W}\) and \(m \geq 0\). If \(\lambda = k\frac{d}{2}, k = 0, \ldots, r - 1\), we additionally assume that \(m_{k+1} = \ldots = m_r = 0\). Then the following invariance property holds:
\[
K^m_\lambda(gz, w) = K^m_\lambda(z, g^*w), \quad z, w \in X_\lambda, g \in L_C.
\]
Proof. Using Proposition 4.12 we find that for \( k \in \tilde{K} \) and \( F \in \mathcal{P}_m(X_\lambda) \) we have

\[
\rho_\lambda(k)F(z) = \langle \rho_\lambda(k)F, \mathcal{K}_\lambda^m(-,z) \rangle_{\mathcal{F}_\lambda} = \langle F, \rho_\lambda^{-1}(\mathcal{K}_\lambda^m(-,z)) \rangle_{\mathcal{F}_\lambda} \\
= \xi_\lambda(k^{-1}) \mathcal{K}_\lambda^m((\eta(k)^{-1})^\#-z, z)_{\mathcal{F}_\lambda} \\
= \xi_\lambda(k) \mathcal{K}_\lambda^m((\eta(k)^{-1})^\#-z, z)_{\mathcal{F}_\lambda}
\]

and on the other hand

\[
\rho_\lambda(k)F(z) = \xi_\lambda(k)F(\eta(k)^\#z) = \xi_\lambda(k)\mathcal{K}_\lambda^m(-,\eta(k)^\#z)_{\mathcal{F}_\lambda}.
\]

Since \( \eta : \tilde{K} \to U \) is surjective and \( u^{-1} = u^* = \pi^\# \) as well as \( \pi \in U \) for \( u \in U \) we obtain

\[
\mathcal{K}_\lambda^m(uz, w) = \mathcal{K}_\lambda^m(z, u^*w), \quad z, w \in X_\lambda, u \in U.
\]

Now both sides are holomorphic in \( u \in L_C \) and \( U \subseteq L_C \) is totally real. Hence the claim follows.

PROPOSITION 4.14. Let \( \lambda \in \mathcal{W} \) and \( m \geq 0 \). If \( \lambda = k \frac{d}{2}, k = 0, \ldots, r - 1, \) we additionally assume that \( m_{k+1} = \ldots = m_r = 0 \). Then

\[
\mathcal{K}_\lambda^m(z, w) = \frac{d_m}{(\lambda)_m} \Phi_m \left( \frac{z}{2}, \frac{w}{2} \right), \quad z, w \in X_\lambda. \quad (4.3)
\]

Proof. First let \( \lambda > (r - 1) \frac{d}{2} \). By Lemma 4.13 the function \( \mathcal{K}_\lambda^m(-, e) \in \mathcal{P}_m(X_\lambda) \) is \( K^L \)-invariant and hence there is a constant \( c_\lambda^m \) such that \( \mathcal{K}_\lambda^m(z, e) = c_\lambda^m \Phi_m(z) \). Since

\[
1 = \Phi_m(e) = \langle \Phi_m, \mathcal{K}_\lambda^m(-, e) \rangle_{\mathcal{F}_\lambda} = c_\lambda^m \| \Phi_m \|_{\mathcal{F}_\lambda}^2,
\]

formula (4.3) now follows with Proposition 4.12. Now, for \( \lambda > (r - 1) \frac{d}{2} \) this implies the following identity:

\[
p(z) = [p, \mathcal{K}_\lambda^m(-,z)]_{\lambda} = \frac{d_m}{(\lambda)_m} \cdot p(B_\lambda)_w \Phi_m(w, \overline{z}) \big|_{w=0} \quad (4.4)
\]

for \( z \in V_C \) and \( p \in \mathcal{P}_m(V_C) \). The right hand side is clearly meromorphic in \( \lambda \) with possible poles at the points where \( (\lambda)_m = 0 \). For \( \lambda = k \frac{d}{2}, k \in \{0, \ldots, r - 1\}, \) and \( m \geq 0 \) with \( m_{k+1} = \ldots = m_r = 0 \) we have \( (\lambda)_m \neq 0 \) and hence the identity (4.4) holds for such \( \lambda \) and \( m \) by analytic continuation. Since the reproducing kernel of \( \mathcal{P}_m(X_\lambda) \) is uniquely determined by (4.4) the claim now follows also for \( \lambda = k \frac{d}{2} \).

THEOREM 4.15. The reproducing kernel \( \mathcal{K}_\lambda(z, w) \) of the Hilbert space \( \mathcal{F}_\lambda \) is given by

\[
\mathcal{K}_\lambda(z, w) = \mathcal{I}_\lambda(\frac{z}{2}, \frac{w}{2}), \quad z, w \in X_\lambda.
\]
Proof. Let $0 \leq k \leq r$ be such that $X_\lambda = X_k$. By the previous result $\mathbb{K}_m^\lambda(z,w)$ is the reproducing kernel for the subspace $P_m(X_\lambda)$. Further we know from Proposition [12] that the spaces $P_m(X_\lambda)$ are pairwise orthogonal. Therefore, by [15] Proposition 1.1.8, the sum

$$\sum_{m \geq 0, m_k+1 = 0} \mathbb{K}_m^\lambda(z,w) = \sum_{m \geq 0, m_k+1 = 0} \frac{d_m}{(\mathbb{K}_m^\lambda(z,w))} = \mathcal{I}_\lambda \left( \frac{z}{2}, \frac{w}{2} \right)$$

converges pointwise to the reproducing kernel $\mathbb{K}_\lambda(z,w)$ of the direct Hilbert sum of all subspaces $P_m(X_\lambda)$ with $m \geq 0, m_k+1 = 0$. But this direct Hilbert sum is by definition $\mathcal{F}_\lambda$ and the proof is complete.

The following consequence is a standard result for reproducing kernel spaces and can e.g. be found in [15], page 9.

**Corollary 4.16.** For every $F \in \mathcal{F}_\lambda$ and every $z \in X_\lambda$ we have

$$|F(z)| \leq \mathcal{I}_\lambda \left( \frac{z}{2}, \frac{z}{2} \right)^{\frac{1}{2}} \|F\|_{\mathcal{F}_\lambda}.$$ 

### 4.5 Rings of differential operators and associated varieties

We recall the definition of the associated variety of an admissible representation for the example $(\rho_\lambda, \mathcal{F}_\lambda)$. Let $(U_k(g))_{k \in \mathbb{N}_0}$ denote the usual filtration of the universal enveloping algebra $U(g)$ and form the corresponding graded algebra $gr U(g)$ which is by the Poincare–Birkhoff–Witt theorem naturally isomorphic to the symmetric algebra $S(g)$. The underlying $(g, \mathcal{K})$-module $X^\lambda = P(X_\lambda)$ of the representation $\rho_\lambda$ carries an action of $U(g)$. For $m \in \mathbb{N}_0$ further let $X^\lambda_m$ be the subspace of polynomials in $P(X_\lambda)$ of degree $\leq m$. Then $X^\lambda_0 = \mathbb{C}1$ is $\mathcal{K}$-invariant and generates $X^\lambda$ as a $U(g)$-module. We further have

$$d \rho_\lambda(U_k(g))X^\lambda_m = X^\lambda_{k+m}, \quad k, m \in \mathbb{N}_0,$$

i.e. the filtrations $(U_k(g))_k$ and $(X^\lambda_m)_m$ are compatible. Thus the corresponding graded space

$$gr X^\lambda = \bigoplus_{m=0}^{\infty} X^\lambda_m / X^\lambda_{m-1}$$

is a module over $gr U(g)$. Consider the annihilator ideal

$$J_\lambda := \text{Ann}_{gr U(g)}(gr X^\lambda) \subseteq gr U(g) \cong S(g) \cong \mathbb{C}[g]^*_g.$$ 

Then the associated variety $\mathcal{V}(\rho_\lambda)$ of $\rho_\lambda$ is by definition the affine subvariety of $g^*_g$ consisting of the common zeros of $J_\lambda$. Since $\mathfrak{t}_g$ lives in degree 1 in $gr U(g)$, but leaves each $X^\lambda_m$ invariant, every element in $\mathcal{V}(\rho_\lambda)$ vanishes on $\mathfrak{t}_g$ and we can view $\mathcal{V}(\rho_\lambda)$ as a subset of $\mathfrak{p}_g^*$.

Via the Killing form we identify $\mathfrak{p}_g^*$ with $\mathfrak{p}_g$.

---

*Documenta Mathematica 18 (2013) 785–855*
and view $\mathcal{V}(\rho_\lambda)$ as a subset of $p^*_\mathbb{C}$. Then $\mathcal{V}(\rho_\lambda)$ is a $K_{\mathbb{C}}$-stable closed subvariety of $p^*_\mathbb{C}$ consisting of nilpotent elements and hence the union of finitely many nilpotent $K_{\mathbb{C}}$-orbits (see [30] Corollary 5.23]).

Recall from Section 4.19 the $K_{\mathbb{C}}$-orbits $O^k_{\mathbb{C}}$ which are isomorphic to the $L_{\mathbb{C}}$-orbits $X_k$ via the Cayley type transform $C \in \text{Int}(g_{\mathbb{C}})$. The following result is due to A. Joseph [30] Theorem 7.14):

**Proposition 4.17.** Let $\lambda \in \mathcal{W}$ and $0 \leq k \leq r$ such that $X_\lambda = X_k$. Then $\mathcal{V}(\rho_\lambda) = \overline{O^k_{\mathbb{C}}}$.

**Corollary 4.18.** For $\lambda \in \mathcal{W}$ let $k \in \{0, \ldots, r\}$ such that $X_\lambda = X_k$. Then the Gelfand–Kirillov dimension of $\rho_\lambda$ is $k + k(2r - k - 1)\frac{d}{2}$.

**Proof.** The Gelfand–Kirillov dimension of an irreducible unitary representation equals the dimension of its associated variety in $p^*_\mathbb{C}$ (combine [30] Corollary 4.7 and [30] Theorem 8.4]). Therefore the result follows from [30].

For an algebraic variety $X$ over $\mathbb{C}$ denote by $\mathbb{C}[X]$ the ring of regular functions. Further let $\mathbb{D}(X)$ be the ring of algebraic differential operators on $X$. This subring of $\text{End}_\mathbb{C}(\mathbb{C}[X])$ can be defined inductively as follows: Let $\mathbb{D}_0(X) := \mathbb{C}[X]$ be the ring of multiplication operators and for $m \in \mathbb{N}$ put

$$\mathbb{D}_m(X) := \{ D \in \text{End}_\mathbb{C}(\mathbb{C}[X]) : [D, f] \in \mathbb{D}_{m-1}(X) \forall f \in \mathbb{C}[X]\}.$$ 

Then $\mathbb{D}(X) = \bigcup_{m \in \mathbb{N}} \mathbb{D}_m(X)$. Since the varieties $X_k$ are affine it follows that $\mathbb{C}[\overline{X_k}] = \mathcal{P}(X_k)$. The representation $d\rho_\lambda$ acts on $\mathcal{P}(X_\lambda)$ by differential operators and hence it induces a map

$$d\rho_\lambda : \mathcal{U}(\mathfrak{g}) \to \mathbb{D}(X_\lambda).$$

The following result is a qualitative version of [30] Theorem 4.5):

**Theorem 4.19.** For $\lambda = k\frac{d}{2} \in \mathcal{W}_{\text{disc}}$ the map $d\rho_\lambda : \mathcal{U}(\mathfrak{g}) \to \mathbb{D}(X_\lambda)$ is surjective and induces an isomorphism $\mathcal{U}(\mathfrak{g})/J_k \cong \mathbb{D}(X_\lambda)$, where $J_k = \text{Ann}_{\mathcal{U}(\mathfrak{g})}(X_\lambda)$.

**Proof.** By [30] Theorem 4.5] the map $d\rho_\lambda$ is surjective onto the space of $\mathbb{C}$-endomorphisms of $\mathcal{P}(X_\lambda) = \mathbb{C}[\overline{X_k}]$ which are locally finite under the diagonal action of $p^-$. Now $p^-$ acts by multiplication with coordinate functions and hence the condition for $D \in \text{End}_\mathbb{C}(\mathbb{C}[\overline{X_k}])$ to be locally finite under the action of $p^-$ is equivalent to the existence of $N \in \mathbb{N}$ such that the iterated commutator $[[[\ldots[D, f_1(x)], \ldots, f_{N-1}(x)], f_N(x)] = 0$ for all $f_1, \ldots, f_N \in \mathbb{C}[\overline{X_k}]$. This again is equivalent to $D \in \mathbb{D}(X_\lambda)$ and the proof is complete.

**Remark 4.20.** A quantitative version of Theorem 4.19 was obtained by Levasseur–Smith–Stafford [38] for the minimal orbit $X_1$, by Levasseur–Stafford [31] for classical $\mathfrak{g}$ and finally by Joseph [30] for the general case. However, their version is less explicit and does not provide a geometric construction of the unitary structure.
Corollary 4.21. For $k = 1, \ldots, r - 1$ the ring $\mathbb{D}(\mathfrak{A}_k)$ of algebraic differential operators on the affine variety $\mathfrak{A}_k$ is generated by the multiplications with regular functions in $\mathbb{C}[\mathfrak{A}_k]$ and the Bessel operators $(u|\mathcal{B}_k)$, $u \in \mathcal{V}_C$, for $\lambda = k\frac{d}{d}$.  

Proof. By Theorem 4.19 the ring $\mathbb{D}(\mathfrak{A}_k)$ is generated by the constants and $d\rho_\lambda(g)$. Note that in the decomposition $\mathfrak{g}_\mathbb{C} = \mathfrak{p}^- \oplus \mathfrak{t}_\mathbb{C} \oplus \mathfrak{p}^+$ the subalgebra $\mathfrak{t}_\mathbb{C}$ is generated by $\mathfrak{p}^+$ and $\mathfrak{p}^-$ and hence $\mathbb{D}(\mathfrak{A}_k)$ is generated by the constants, $d\rho_\lambda(\mathfrak{p}^+) = \{(u|\mathcal{B}_k) : u \in \mathcal{V}_C\}$ and $d\rho_\lambda(\mathfrak{p}^-) = \{(v|z) : v \in \mathcal{V}_C\}$. □

Remark 4.22. Neither Theorem 4.19 nor Corollary 4.21 can hold for $\lambda \in \mathcal{W}_{\text{cont}}$ resp. $k = r$ since in this case $\mathfrak{A}_\lambda = \mathfrak{A}_r = \mathfrak{g}_\mathbb{C}$ and $\mathbb{D}(\mathcal{V}_C) = \mathbb{C}[x, \frac{1}{2}]$ is a Weyl algebra (see [39] Lemma IV.1.5)).

5 The Segal–Bargmann transform

For every $\lambda \in \mathcal{W}$ we explicitly construct an intertwining operator, the Segal–Bargmann transform, between the Schrödinger model $(\pi_\lambda, L^2(\mathcal{O}_\lambda, d\mu_\lambda))$ and the Fock model $(\rho_\lambda, \mathcal{F}_\lambda)$ in terms of its integral kernel.

5.1 Construction of the Segal–Bargmann transform

For each $\lambda \in \mathcal{W}$ the Segal–Bargmann transform $\mathcal{B}_\lambda$ is defined for $\psi \in L^2(\mathcal{O}_\lambda, d\mu_\lambda)$ by

$$\mathcal{B}_\lambda \psi(z) := e^{-\frac{1}{2} \text{tr}(z)} \int_{\mathcal{O}_\lambda} \mathcal{I}_\lambda(z, x)e^{-\text{tr}(x)} \psi(x) d\mu_\lambda(x), \quad z \in \mathcal{V}_C.$$  

5.1 Proposition 5.1. For $\psi \in L^2(\mathcal{O}_\lambda, d\mu_\lambda)$ the integral in (5.1) converges uniformly on bounded subsets and defines a function $\mathcal{B}_\lambda \psi \in \mathcal{O}(\mathcal{V}_C)$. The Segal–Bargmann transform $\mathcal{B}_\lambda$ is a continuous linear operator $L^2(\mathcal{O}_\lambda, d\mu_\lambda) \to \mathcal{O}(\mathcal{V}_C)$.

Proof. Since the kernel function $e^{-\frac{1}{2} \text{tr}(z)} \mathcal{I}_\lambda(z, x)e^{-\text{tr}(x)}$ is analytic in $z$, it suffices to show that its $L^2$-norm in $x$ has a uniform bound for $|z| \leq R$, $R > 0$. By Lemma 5.2 there exists $C > 0$ such that

$$|\mathcal{I}_\lambda(z, x)| \leq C(1 + |z|, |x|)^{(2r - 1)} e^{2R \sqrt{|x|}}.$$  

Then for $x \in \mathcal{O}_\lambda$, $z \in \mathcal{V}_C$ with $|z| \leq R$, we find

$$|e^{-\frac{1}{2} \text{tr}(z)} \mathcal{I}_\lambda(z, x)e^{-\text{tr}(x)}| \leq C'(1 + R : |x|)^{(2r - 1)} e^{2R \sqrt{|x|}} e^{-\text{tr}(x)} \leq C'(1 + R : |x|)^{(2r - 1)} e^{2R \sqrt{|x|} - |x|}$$  

with $C' = C \max_{|z| \leq R} |e^{-\frac{1}{2} \text{tr}(z)}|$. This is certainly $L^2$ as a function of $x \in \mathcal{O}_\lambda$ with norm independent of $z$ and the claim follows. □
Next, we show that $\mathcal{B}_\lambda$ intertwines the action $d\pi_\lambda$ on $L^2(\mathcal{O}_\lambda, d\mu_\lambda)$ with the action $d\rho_\lambda$ on $\mathcal{F}_\lambda$. Recall the Cayley type transform $C \in \text{Int}(\mathfrak{g}_C^*)$ introduced in Section 1.3.

**Proposition 5.2.** The following intertwining identity holds on $L^2(\mathcal{O}_\lambda, d\mu_\lambda)^\infty$:

$$\mathcal{B}_\lambda \circ d\pi_\lambda(X) = d\pi_\lambda^C(C(X)) \circ \mathcal{B}_\lambda, \quad X \in \mathfrak{g}. \quad (5.2)$$

**Proof.** As remarked in Section 1.3, the subspace $\{(a, L(b), a) : a, b \in V\}$ generates $\mathfrak{g}$ as a Lie algebra. Hence it suffices to prove (5.2) for the elements $(a, +2iL(a), a), a \in V$. We show (5.2) for $X = (a, -2iL(a), a), a \in V$, the proof for $(a, +2iL(a), a)$ works similarly. For $X = (a, -2iL(a), a)$ we have $C(X) = (a, 0, 0)$ and hence

$$(d\pi_\lambda^C(C(X))) \circ \mathcal{B}_\lambda \psi(z) = i(a|z)e^{-\frac{i}{2}tr(z)}\int_{\mathcal{O}_\lambda} I_\lambda(z, x)e^{-tr(x)}\psi(x) d\mu_\lambda(x).$$

By Proposition 5.1 we have $z I_\lambda(z, x) = (\mathcal{B}_\lambda)_z I_\lambda(z, x)$. Further the Bessel operator $\mathcal{B}_\lambda$ is symmetric on $L^2(\mathcal{O}_\lambda, d\mu_\lambda)$ and we obtain

$$= ie^{-\frac{i}{2}tr(z)}\int_{\mathcal{O}_\lambda} (a|\mathcal{B}_\lambda)_z I_\lambda(z, x)e^{-tr(x)}\psi(x) d\mu_\lambda(x)$$

$$= ie^{-\frac{i}{2}tr(z)}\int_{\mathcal{O}_\lambda} I_\lambda(z, x)(a|\mathcal{B}_\lambda)\left[e^{-tr(x)}\psi(x)\right] d\mu_\lambda(x).$$

By the product rule (1.12) for the Bessel operator we obtain

$$\mathcal{B}_\lambda\left[e^{-tr(x)}\psi(x)\right]$$

$$= \mathcal{B}_\lambda e^{-tr(x)} \cdot \psi(x) + 2P (\frac{\partial}{\partial x} e^{-tr(x)}, \frac{\partial}{\partial x} x) + e^{-tr(x)}, \mathcal{B}_\lambda\psi(x)$$

$$= (x - \lambda e)e^{-tr(x)}\psi(x) - 2e^{-tr(x)}x \frac{\partial}{\partial x} + e^{-tr(x)} \mathcal{B}_\lambda\psi(x)$$

since

$$\mathcal{B}_\lambda e^{-tr(x)} = (P(-e)x - \lambda e)e^{-tr(x)} = (x - \lambda e)e^{-tr(x)}$$

and

$$\frac{\partial}{\partial x} e^{-tr(x)} = -e \cdot e^{-tr(x)}.$$ 

Hence we have

$$(a|\mathcal{B}_\lambda)\left[e^{-tr(x)}\psi(x)\right] = e^{-tr(x)}[(a|x)\psi(x) - \lambda \text{tr}(a)\psi(x) - 2\partial_{a} \psi(x) + \mathcal{B}_\lambda\psi(x)].$$

Inserting this into our calculation above we find

$$(d\pi_\lambda^C(C(X))) \circ \mathcal{B}_\lambda \psi(z)$$

$$= \mathcal{B}_\lambda \circ \left(i(a|x) - 2i \left[\partial_{L(a)}x + \frac{r\lambda}{2ni} \text{Tr}(L(a))\right] + i(a|\mathcal{B}_\lambda)\right)\psi(z)$$

$$= (\mathcal{B}_\lambda \circ d\pi_\lambda(X))\psi(z). \quad \square$$

Documenta Mathematica 18 (2013) 785–855
To conclude that $B_\lambda$ is an isomorphism $L^2(O_\lambda, d\mu_\lambda) \to \mathcal{F}_\lambda$, we show that it maps the underlying $(g, \mathfrak{t})$-module $W^\lambda \subseteq L^2(O_\lambda, d\mu_\lambda)$ to the $(g, \mathfrak{t})$-module $\mathcal{P}(X_\lambda)$. In order to do so we show that the function

$$\Psi_0 := B_\lambda \psi_0 \in \mathcal{O}(V_C)$$

is $L_C$-invariant. In fact, the function $\psi_0$ is $\tilde{K}$-equivariant by the character $\xi_\lambda$ (see Section 5.1). By Proposition 5.2 this has to be true for $\Psi_0$. But by Proposition 5.4 this implies that $\Psi_0$ is invariant under $\eta(K)^\# = U$. Now $U$ is a real form of $L_C$ and the action of $L_C$ on $\Psi_0$ is holomorphic, whence $\Psi_0$ has to be $L_C$-invariant. Therefore it has to be constant on every $L_C$-orbit. Since $\Psi_0$ is holomorphic on $V_C$ and $V_C$ decomposes into finitely many $L_C$-orbits, it follows that $\Psi_0$ is constant on $V_C$. It remains to show that $\Psi_0$ is non-zero.

**Proposition 5.3.** $\Psi_0(0) = 1$ and hence $\Psi_0 = B_\lambda \psi_0 = 1$.

**Proof.** Since $L_\lambda(0, x) = 1$ we have

$$\Psi_0(0) = \int_{O_\lambda} e^{-2 \text{tr}(x)} d\mu_\lambda(x) = \|\psi_0\|^2_{L^2(O_\lambda, d\mu_\lambda)} = 1$$

as shown in Section 5.2.

**Theorem 5.4.** $B_\lambda$ is a unitary isomorphism $L^2(O_\lambda, d\mu_\lambda) \to \mathcal{F}_\lambda$ intertwining the actions $\pi_\lambda$ and $\rho_\lambda$.

**Proof.** Since $W^\lambda = L^2(O_\lambda, d\mu_\lambda)_\mathfrak{t}$ is generated by $\psi_0$, $\mathcal{P}(X_\lambda) = (\mathcal{F}_\lambda)_\mathfrak{t}$ is generated by $\Psi_0 = 1$, $B_\lambda \psi_0 = \Psi_0$ and $B_\lambda$ intertwines the actions $d\pi_\lambda$ and $d\rho_\lambda$, it has to map the irreducible $(g, \mathfrak{t})$-module $W^\lambda$ into the irreducible $(g, \mathfrak{t})$-module $\mathcal{P}(X_\lambda)$ and is therefore, thanks to Schur’s Lemma, an isomorphism between the underlying $(g, \mathfrak{t})$-modules. It only remains to show that $B_\lambda$ is isometric between $W^\lambda$ and $\mathcal{P}(X_\lambda)$, then the statement follows since $W^\lambda \subseteq L^2(O_\lambda, d\mu_\lambda)$ and $\mathcal{P}(X_\lambda) \subseteq \mathcal{F}_\lambda$ are dense.

Both $W^\lambda$ and $\mathcal{P}(X_\lambda)$ are irreducible infinitesimally unitary $(g, \mathfrak{t})$-modules and hence $B_\lambda$ is a scalar multiple of a unitary operator. Since further

$$\|B_\lambda \psi_0\|_{\mathcal{F}_\lambda} = \|1\|_{\mathcal{F}_\lambda} = 1 = \|\psi_0\|_{L^2(O_\lambda, d\mu_\lambda)}$$

the operator $B_\lambda$ itself has to be unitary.

**Remark 5.5.** The Segal–Bargmann transform can also be obtained via a restriction principle (see [28] [40] for other instances of this principle). The formula $R_\lambda F(x) = e^{-\frac{1}{2} \text{tr}(x)} F(x)$ defines an operator $\mathcal{P}(X_\lambda) \to L^2(O_\lambda, d\mu_\lambda)$ and hence we obtain a densely defined unbounded operator $R_\lambda : \mathcal{F}_\lambda \to L^2(O_\lambda, d\mu_\lambda)$. We consider its adjoint $R_\lambda^* : L^2(O_\lambda, d\mu_\lambda) \to \mathcal{F}_\lambda$ as a densely defined unbounded operator. One can show that the Segal–Bargmann transform appears as the unitary part in the polar decomposition of the operator $R_\lambda^*$:

$$R_\lambda^* = B_\lambda \circ \sqrt{R_\lambda R_\lambda^*}.$$
For the case of the minimal discrete Wallach point this is done in [23, Proposition 5.5].

**Corollary 5.6.** The inverse Segal–Bargmann transform is given by

$$\mathbb{B}_\lambda^{-1} F(x) = e^{-\text{tr}(x)} \int_{X_\lambda} I_\lambda (x, z) e^{-\frac{1}{2} \text{tr}(z) F(z) \omega_\lambda (z)} \, d\nu_\lambda (z), \quad F \in \mathcal{F}_\lambda.$$  

**Proof.** Since the Segal–Bargmann transform is a unitary operator we have

$$\langle \mathbb{B}_\lambda^{-1} F, \psi \rangle_{L^2(O_\lambda, d\mu_\lambda)} = \langle F, \mathbb{B}_\lambda \psi \rangle_{\mathcal{F}_\lambda} = \frac{1}{c_\lambda} \int_{X_\lambda} F(z) \overline{\mathbb{B}_\lambda \psi (z) \omega_\lambda (z)} \, d\nu_\lambda (z)$$

which implies the claim.  

We now use the Segal–Bargmann transform to obtain an intrinsic description of the Fock space.

**Theorem 5.7.**

$$\mathcal{F}_\lambda = \left\{ F|_{X_\lambda} : F \in \mathcal{O}(V_\mathbb{C}), \int_{X_\lambda} |F(z)|^2 \omega_\lambda (z) \, d\nu_\lambda (z) < \infty \right\}.$$  

**Proof.** The Segal–Bargmann transform is an isomorphism $\mathbb{B}_\lambda : L^2(O_\lambda, d\mu_\lambda) \to \mathcal{F}_\lambda$ and hence by Proposition 5.1 every function in $\mathcal{F}_\lambda$ extends to a holomorphic function on $V_\mathbb{C}$. This shows the inclusion $\subseteq$.

For the other inclusion let $F \in \mathcal{O}(V_\mathbb{C})$ such that

$$\int_{X_\lambda} |F(z)|^2 \omega_\lambda (z) \, d\nu_\lambda (z) < \infty.$$  

We expand $F$ into a power series which we can arrange as

$$F = \sum_{m \geq 0} p_m \quad \text{with} \quad p_m \in \mathcal{P}_m(V_\mathbb{C})$$  

by Theorem 1.6. Since $p_m|_{X_\lambda} = 0$ for $m_{k+1} \neq 0$ by Corollary 1.9 we may assume $F = \sum_{m \geq 0, m_{k+1} = 0} p_m$ for the study of $F|_{X_\lambda}$. This series converges uniformly on compact subsets. We show that this series also converges in $L^2(X_\lambda, \omega_\lambda \, d\nu_\lambda)$. For $R > 0$ let

$$X_\lambda^R := \{ z \in X_\lambda : |z| \leq R \},$$

$$O_\lambda^R := \{ x \in O_\lambda : |x| \leq R \}.$$
Note that $X^R_\lambda$ and $O^R_\lambda$ are compact and hence integration over these sets commutes with taking the limit $F = \sum_{m \geq 0, m_{k+1} = 0} p_m$. With (1.10) we find

$$\infty > \int_{X^R_\lambda} |F(z)|^2 \omega_\lambda(z) \, d\nu_\lambda(z) = \lim_{R \to \infty} \int_{X^R_\lambda} |F(z)|^2 \omega_\lambda(z) \, d\nu_\lambda(z)$$

$$= \lim_{R \to \infty} \sum_{m,n \geq 0, m_{k+1} = n_{k+1} = 0} \int_{O^R_\lambda} \int_{U} \frac{p_m(x^\perp) p_n(x^\perp)}{p_m(x^\perp) \omega_\lambda(x^\perp) \, d\mu_\lambda(x)}.$$

By Theorem 1.7 and Lemma 1.14 the integrals over $U$ for $m \neq n$ vanish and we obtain

$$= \lim_{R \to \infty} \sum_{m \geq 0, m_{k+1} = 0} \int_{X^R_\lambda} |p_m(z)|^2 \omega_\lambda(z) \, d\nu_\lambda(z) = \sum_{m \geq 0, m_{k+1} = 0} \|p_m\|^2_{L^2(X^R_\lambda, \omega_\lambda \, d\nu_\lambda)}$$

which shows convergence in $L^2(X^R_\lambda, \omega_\lambda \, d\nu_\lambda)$. Since $F^R_\lambda$ is the closure of the space of polynomials with respect to the norm of $L^2(X^R_\lambda, \omega_\lambda \, d\nu_\lambda)$ it is clear that $F|_{X^R_\lambda} \in F^R_\lambda$ which shows the other inclusion. 

This intrinsic description leads to the following conjecture which was proved in [23, Theorem 2.26] for the minimal orbit, i.e. $\lambda \in \mathcal{W}$ with $X^R_\lambda = X^R_1$:

**Conjecture 5.8.**

$$F^R_\lambda = \left\{ F \in O(X^R_\lambda) : \int_{X^R_\lambda} |F(z)|^2 \omega_\lambda(z) \, d\nu_\lambda(z) < \infty \right\}.$$

Recall the Laguerre functions $\ell^\lambda_m$ introduced in Section 1.8

**Proposition 5.9.** Let $\lambda \in \mathcal{W}$ and $k \in \{0, \ldots, r\}$ such that $O^R_\lambda = O_k$. Then for every $m \geq 0$ with $m_{k+1} = 0$ we have

$$B^l_\lambda \ell^\lambda_m = \frac{(-1)^{|m|}}{2^{|m|}} \Phi_m.$$

**Proof.** $\ell^\lambda_m$ is the unique (up to scalar multiples) $\ell^l$-invariant vector in the $\ell^l$-type $W^\lambda_m \subseteq L^2(O^R_\lambda, \omega_\lambda \, d\mu_\lambda)$ whereas $\Phi_m$ is the unique (up to scalar multiples) $\ell^l$-invariant vector in the $\ell^l$-type $P_m(X^R_\lambda) \subseteq F^R_\lambda$. Hence

$$B^l_\lambda \ell^\lambda_m = \text{const.} \Phi_m.$$

To find the constant we evaluate $B^{-1}_\lambda \Phi_m(x)$ and $\ell^\lambda_m(x)$ at $x = 0$. First observe that

$$\ell^\lambda_m(0) = L^\lambda_m(0) = (\lambda)_m.$$
Next we calculate $B^{-1}_m(0)$ using Corollary 5.6:

$$B^{-1}_m(0) = \frac{1}{c_\lambda} \int_{X_\lambda} e^{-\frac{1}{2} \text{tr}(z)} \Phi_m(z) \omega_\lambda(z) d\nu_\lambda(z).$$

By (1.17) the function $e^{\text{tr}(z)}$ has the following expansion:

$$e^{\text{tr}(z)} = \sum_{m \geq 0} \frac{d_m}{(P)^m} \Phi_m(z).$$

Inserting this yields

$$B^{-1}_m(0) = \sum_{n \geq 0} (-1)^{|n|} \frac{d_n}{2^{|n|} (\frac{2}{\tau})_n} \frac{1}{c_\lambda} \int_{X_\lambda} \Phi_m(z) \Phi_n(z) \omega_\lambda(z) d\nu_\lambda(z)$$

$$= \sum_{n \geq 0} (-1)^{|n|} \frac{d_n}{2^{|n|} (\frac{2}{\tau})_n} (\Phi_m, \Phi_n)_{F_\lambda}$$

$$= (-1)^{|m|} \frac{1}{2^{|m|} (\frac{2}{\tau})_m} \cdot \frac{4^{|m|} (\frac{2}{\tau})_m (\lambda)_m}{d_m}$$

$$= (-1)^{|m|/2^{|m|} (\lambda)_m},$$

where we have used Proposition 4.2.

\[ \square \]

5.2 Intertwiner to the bounded symmetric domain model

The Fock space $F_\lambda$ and the bounded symmetric domain model $H^2_\lambda(D)$ have (as vector spaces) the same underlying $(g, \tilde{K})$-module, namely

$$\bigoplus_{m \geq 0, m_{k+1}=0} P_m(V_\mathbb{C}),$$

where $k \in \{0, \ldots, r\}$ is such that $A_\lambda = X_k$. However, inner product and group action differ. In Remark 4.3 we already compared the norms in the two models. To gain a better understanding of the relation between the two models we investigate the intertwiner between them.

For this we compose the inverse Segal–Bargmann transform $B_\lambda^{-1}$ with the Laplace transform $L_\lambda$ and the pullback of the Cayley transform $\gamma_\lambda$ to obtain an intertwining operator between the Fock space picture and the realization on functions on the bounded symmetric domain. (For the definition of $L_\lambda$ and $\gamma_\lambda$ see Sections 2.2 and 2.3.) Let

$$A_\lambda := \gamma_\lambda \circ L_\lambda \circ B_\lambda^{-1} : F_\lambda \to H^2_\lambda(D).$$

The operator $A_\lambda$ intertwines the actions $\rho_\lambda$ and $\pi^D_\lambda$.

**Theorem 5.10.** For $F \in F_\lambda$ we have

$$A_\lambda F(z) = \frac{1}{c_\lambda} \int_{X_\lambda} e^{-\frac{1}{2} \text{tr}(w)} F(w) \omega_\lambda(w) d\nu_\lambda(w), \quad z \in D.$$
Proof. Using Lemma 3.8 we obtain
\[ A_{\lambda}F(z) = \gamma_{\lambda} \circ L_{\lambda} \circ \mathbb{B}_{\lambda}^{-1}F(z) = \Delta(e - z)^{-\lambda}L_{\lambda} \circ \mathbb{B}_{\lambda}^{-1}F(e(z)) \]
\[ = \Delta(e - z)^{-\lambda} \int_{\Omega_{\lambda}} e^{(e(z(u))} \mathbb{B}_{\lambda}^{-1}F(u) \, d\mu_{\lambda}(u) \]
\[ = \frac{\Delta(e - z)^{-\lambda}}{c_{\lambda}} \int_{\Omega_{\lambda}} \int_{X_{\lambda}} e^{(e(z(u))} e^{-\frac{1}{2}tr(w)} F(w) \]
\[ \omega_{\lambda}(w) \, d\nu_{\lambda}(w) \]
\[ = \frac{\Delta(e - z)^{-\lambda}}{c_{\lambda}} \int_{X_{\lambda}} \Delta(e - ic(z))^{-\lambda}e^{(e-ic(z))^{-1}w} e^{-\frac{1}{2}tr(w)} F(w) \]
\[ \omega_{\lambda}(w) \, d\nu_{\lambda}(w). \]

Now
\[ e - ic(z) = e + (e + z)(e - z)^{-1} = 2(e - z)^{-1} \]
and therefore
\[ \Delta(e - ic(z))^{-\lambda} = 2^{-\lambda} \Delta(e - z)^{\lambda}. \]

This also implies that
\[ ((e - ic(z))^{-1}w) = \frac{1}{2} tr(w) - \frac{1}{2}(z(w)) \]
so that altogether we obtain
\[ A_{\lambda}F(z) = \frac{1}{c_{\lambda}} \int_{X_{\lambda}} e^{-\frac{1}{2}(z(w))} F(w) \omega_{\lambda}(w) \, d\nu_{\lambda}(w). \]

Corollary 5.11. The operator \( A_{\lambda} \) acts on \( \mathcal{P}_{m}(V_{\mathbb{C}}) \) by the scalar \( \frac{(-1)^{m}(\lambda)_{m}}{2^{m}(\lambda)_{m}} \).

Proof. Using the expansion (1.17) and Proposition 4.14 we obtain
\[ e^{(z(w))} = \sum_{n \geq 0} (\lambda)_{n} K_{\lambda}^{n}(z, w). \]

Hence for \( F \in \mathcal{P}_{m}(V_{\mathbb{C}}) \) we find
\[ A_{\lambda}F(z) = \sum_{n \geq 0} (\lambda)_{n} (F, K_{\lambda}^{n}(z, -\frac{1}{2}z)) \mathcal{F}_{\lambda} = (\lambda)_{m} F(-\frac{1}{2}z) = \frac{(-1)^{m}(\lambda)_{m}}{2^{m}(\lambda)_{m}} F(z). \]
Using the Segal–Bargmann transform we now calculate the integral kernel of the unitary inversion operator $U_\lambda$. The unitary inversion operator is essentially given by the action $\pi_\lambda(\tilde{j})$ of the inversion element (see [22 Section 3.3])

$$\tilde{j} = \exp(G(\frac{\pi}{2}(e,0,-e)) \in \tilde{G}.$$ 

More precisely, we put

$$U_\lambda := e^{-ir_\lambda^2 \pi_\lambda(\tilde{j})}.$$ 

The operator $U_\lambda$ is unitary on $L^2(O_\lambda, d\mu_\lambda)$ of order 2, i.e. $U_\lambda^2 = \text{id}$ (see [22 Proposition 3.17 (1)]). We also study Whittaker vectors in the Schrödinger model. They can be derived from the explicit expression of the integral kernel of $U_\lambda$.

6.1 The integral kernel of $U_\lambda$

Recall the underlying $(g,k)$-module $W_\lambda = P(O_\lambda) e^{-\text{tr}(x)}$ of the representation $(\pi_\lambda, L^2(O_\lambda, d\mu_\lambda))$.

**Proposition 6.1.** For every $\psi \in W_\lambda$ the integral

$$T_\lambda \psi(z) := 2^{-r_\lambda} \int_{O_\lambda} J_\lambda(z,x) \psi(x) \, d\mu_\lambda(x), \quad z \in V_C,$$

converges uniformly on compact subsets of $V_C$. This defines a linear operator $T_\lambda : W^\lambda \to C(O_\lambda)$.

**Proof.** Let $\psi(x) = p(x)e^{-\text{tr}(x)} \in P(O_\lambda)e^{-\text{tr}(x)} = W^\lambda$. Then $|p(x)| \leq C_1(1+|x|)^N$ for some constants $C_1, N > 0$. Further, by Lemma 3.1 we have

$$|J_\lambda(z,x)| \leq C_2(1+|z|\cdot |x|)^\frac{2(n-1)}{4} e^{2r\sqrt{|z|\cdot |x|}} \quad \forall x \in O_\lambda, z \in V_C$$

for a constant $C_2 > 0$. Hence, for $|z| \leq R, R \geq 1$, we obtain

$$\int_{O_\lambda} |J_\lambda(z,x)\psi(x)| \, d\mu_\lambda(x) \leq C_1 C_2 \int_{O_\lambda} (1+R|x|)^N |x|^{\frac{2(n-1)}{2}} e^{2r\sqrt{|z|\cdot |x|}} \, d\mu_\lambda(x) < \infty$$

and the proof is complete. 

**Proposition 6.2.** The operator $T_\lambda$ is a unitary isomorphism $T_\lambda : W^\lambda \to W^\lambda$ with $T_\lambda \psi_0 = \psi_0$ which intertwines the $g$-action $d\pi_\lambda$ with the $g$-action $d\pi_\lambda \circ \text{Ad}(j)$. 

Documenta Mathematica 18 (2013) 785–855
**Proof.** By Proposition 6.4 the operator $\mathcal{T}_\lambda$ intertwines the Bessel operator $B_\lambda$ with the coordinate multiplication $-x$ and vice versa. Therefore we obtain

$$U_\lambda \circ d\pi_\lambda(X) = d\pi_\lambda(Ad(j)X) \circ U_\lambda$$

for $X \in \mathfrak{n}\oplus\mathfrak{p}$. Since $\mathfrak{n}$ and $\mathfrak{p}$ together generate $\mathfrak{g}$ as a Lie algebra the intertwining formula (6.1) holds for all $X \in \mathfrak{g}$. Further, for $x \in \mathcal{O}_\lambda \subseteq \mathcal{A}_\lambda$ we find

$$T_\lambda \psi_0(x) = 2^{-r_\lambda} \int_{\mathcal{O}_\lambda} \mathcal{J}_\lambda(x, y)e^{-tr(y)} \, d\mu_\lambda(y) = \int_{\mathcal{O}_\lambda} \mathcal{I}_\lambda(-2x, y)e^{-2tr(y)} \, d\mu(y)$$

$$= e^{-tr(x)} \mathcal{B}_\lambda \psi_0(-2x) = e^{-tr(x)} \psi_0(x).$$

Since $W^\lambda = d\pi_\lambda(\mathcal{U}(\mathfrak{g}))\psi_0$, it follows that $T_\lambda W^\lambda \subseteq d\pi_\lambda(\mathcal{U}(\mathfrak{g}))T_\lambda \psi_0 = W^\lambda$. Now, since invariant Hermitian forms on the irreducible infinitesimally unitary $(\mathfrak{g}, \mathfrak{t})$-module $W^\lambda$ are unique up to a scalar, we find that $T_\lambda$ is a unitary isomorphism.

**Theorem 6.3.** $U_\lambda = T_\lambda$.

**Proof.** By the previous proposition $U_\lambda \circ T_\lambda^{-1}$ extends to a unitary isomorphism $L^2(\mathcal{O}_\lambda, d\mu_\lambda) \rightarrow L^2(\mathcal{O}_\lambda, d\mu_\lambda)$ which commutes with the $\mathfrak{g}$-action $d\pi_\lambda$. Therefore, by Schur’s Lemma, $U_\lambda$ is a scalar multiple of $T_\lambda$. Since $U_\lambda \psi_0 = \psi_0 = T_\lambda \psi_0$ this gives the claim.

**Remark 6.4.** The group $G$ is generated by the conformal inversion $j$ and the maximal parabolic subgroup $P := L^0 \ltimes N$, where

$$L^0 = G \cap \text{Str}(V).$$

Write $\tilde{P}$ for the preimage of $P$ under the covering map $\tilde{G} \rightarrow G$. The restriction of $\pi_\lambda$ to $\tilde{P}$ factors to $P$ and is in the Schrödinger model quite simple. In fact

$$\pi_\lambda(n_a)\psi(x) = e^{i(x|a)}\psi(x), \quad n_a \in N,$$

$$\pi_\lambda(g)\psi(x) = \chi(g^\#)\tilde{\psi}(g^\#x), \quad g \in L^0,$$

and by Mackey theory this representation is even irreducible on $L^2(\mathcal{O}_\lambda, d\mu_\lambda)$. Therefore, together with the action of $\tilde{j}$ in the Schrödinger model (see Theorem 6.3) this describes the complete group action $\pi_\lambda$.

For the case of the rank 1 orbit $\mathcal{O}_1$ Theorem 6.3 was shown in [23] Theorem 4.3. Earlier contributions to special cases are [33] for the case $\mathfrak{g} = so(2,k)$ and $\lambda$ the minimal discrete Wallach point and [37] for the case $\mathfrak{g} = sl(2,\mathbb{R})$ and arbitrary $\lambda \in \mathcal{W}$.

**Remark 6.5.** Since the functions $x \mapsto \mathcal{J}_\lambda(x, y)$, $y \in \mathcal{O}_\lambda$, are eigenfunctions of the Bessel operator, the inversion formula for $U_\lambda$ gives an expansion of any function $\psi \in L^2(\mathcal{O}_\lambda, d\mu_\lambda)$ into eigenfunctions of the Bessel operator:

$$\psi(x) = 2^{-r_\lambda} \int_{\mathcal{O}_\lambda} \mathcal{J}_\lambda(x, y)U_\lambda\psi(y) \, d\mu_\lambda(y), \quad \psi \in L^2(\mathcal{O}_\lambda, d\mu_\lambda).$$

*Documenta Mathematica* 18 (2013) 785–855
Compared to the action of $\tilde{\gamma}$ in the Schrödinger model its action in the Fock model is extremely simple. Define $(-1)^* \in F_\lambda$ by $(-1)^* F(z) = F(-z)$.

**Proposition 6.6.** $\mathbb{B}_\lambda \circ U_\lambda = (-1)^* \circ \mathbb{B}_\lambda$.

**Proof.** We have $d\rho_\lambda(t(e,0,-e)) = d\pi^\mathbb{C}_\lambda(2it(0,1,0)) = 2it(\partial_z + \frac{iz^2}{2})$. Therefore, we obtain

$$\rho_\lambda(e^{it(e,0,-e)}) F(z) = e^{itr} F(e^{2it}z).$$

For $t = \frac{1}{2}$ we obtain the action of $\tilde{\gamma}$ which is given by $e^{itr} \tilde{\gamma}(-1)^*$. \hfill $\square$

This now allows us to compute the action of the unitary inversion operator on the Laguerre functions $\ell^i_m$ introduced in Section 6.2.

**Proposition 6.7.** $U_\lambda \ell^i_m = (-1)^{im} \ell^i_m$.

**Proof.** Since $\mathbb{B}_\lambda \ell^i_m = \text{const} \cdot \Phi_m$ and $\Phi_m(-z) = (-1)^{im} \Phi_m(z)$ the claim follows with Proposition 6.6. \hfill $\square$

### 6.2 Whittaker vectors

The integral kernel of the unitary inversion operator provides us with explicit Whittaker vectors for the representations $\pi_\lambda$ on $L^2(O_\lambda, d\mu_\lambda)$ and hence with explicit embeddings into Whittaker models. For this let $L^2(O_\lambda, d\mu_\lambda)^{\infty}$ denote the space of smooth vectors of the representation $\pi_\lambda$ endowed with the usual Fréchet topology. Since $\pi_\lambda$ leaves $L^2(O_\lambda, d\mu_\lambda)^{\infty}$ invariant the representation extends to the space $L^2(O_\lambda, d\mu_\lambda)^{-\infty} = (L^2(O_\lambda, d\mu_\lambda)^{\infty})'$ of distribution vectors by

$$\langle \pi_\lambda(g) u, \varphi \rangle := \langle u, \bar{\pi_\lambda(g^{-1})}\varphi \rangle, \quad g \in \tilde{G},$$

for $u \in L^2(O_\lambda, d\mu_\lambda)^{-\infty}$ and $\varphi \in L^2(O_\lambda, d\mu_\lambda)^{\infty}$. Here we use the complex conjugate $\overline{U}$ of an operator $U$ on $L^2(O_\lambda, d\mu_\lambda)^{\infty}$ which is defined by $\overline{U}\varphi := \overline{U\varphi}$ for $\varphi \in L^2(O_\lambda, d\mu_\lambda)^{\infty}$. Further recall that $W^\lambda = P(O_\lambda)e^{-\text{tr}(x)} \subseteq L^2(O_\lambda, d\mu_\lambda)$ is the underlying $(\mathfrak{g}, \ell)$-module of $\pi_\lambda$. It does not carry a representation of the group $\tilde{G}$, but the representation $d\pi_\lambda$ restricts to $W^\lambda$. By duality $d\pi_\lambda$ extends to the algebraic dual $(W^\lambda)^*$:

$$(d\pi_\lambda(X) u, \varphi) := -\langle u, d\pi_\lambda(X) \varphi \rangle, \quad X \in \mathfrak{g},$$

for $u \in (W^\lambda)^*$ and $\varphi \in W^\lambda$. We have the following inclusions:

$$W^\lambda \subseteq L^2(O_\lambda, d\mu_\lambda)^{\infty} \subseteq L^2(O_\lambda, d\mu_\lambda) \subseteq L^2(O_\lambda, d\mu_\lambda)^{-\infty} \subseteq (W^\lambda)^*.$$  

Note that the unitary inversion operator $U_\lambda$ leaves each of these spaces invariant.
**Definition 6.8.** A distribution \( u \in (W^\lambda)^* \) is an algebraic \( n \)-Whittaker vector of weight \( \eta \in n_C^* \) (resp. an algebraic \( \mathfrak{n} \)-Whittaker vector of weight \( \eta \in \mathfrak{n}_C^* \)) if
\[
d\pi_\lambda(X)u = \eta(X)u, \quad \text{for all } X \in n \text{ (resp. } X \in \mathfrak{n}).
\]
If moreover \( u \in L^2(O_\lambda, d\mu_\lambda)^\sim \) we call \( u \) a smooth Whittaker vector.

For \( z \in V_C \) let \( \eta_z \in n_C^* \) and \( \bar{\eta}_z \in \mathfrak{n}_C^* \) be defined by
\[
\eta_z(u, 0, 0) := i(u|z), \quad u \in n_C, \\
\bar{\eta}_z(0, 0, v) := -i(v|z), \quad v \in \mathfrak{n}_C.
\]
Now fix \( \lambda \in W \). In [40, Theorem 1] it is shown that there can only be non-trivial algebraic \( n \)-Whittaker vectors (resp. \( \mathfrak{n} \)-Whittaker vectors) of weight \( \eta \in n_C^* \) (resp. \( \eta \in \mathfrak{n}_C^* \)) for \( d\pi_\lambda \) if \( \eta \) is contained in the associated variety of \( d\pi_\lambda \) in \( g_C^* \). The intersection of the associated variety of \( d\pi_\lambda \) with \( n_C^* \) (resp. \( \mathfrak{n}_C^* \)) is equal to \( X_\lambda \) (after identifying \( n_C^* \) and \( \mathfrak{n}_C^* \) with \( V_C \)). Hence the existence of a non-trivial algebraic \( n \)-Whittaker vector (resp. \( \mathfrak{n} \)-Whittaker vector) of weight \( \eta_z \) (resp. \( \bar{\eta}_z \)) implies \( z \in X_\lambda \). We prove the converse of this statement:

**Proposition 6.9.** Let \( z \in X_\lambda \).

1. The Dirac delta distribution \( \delta_{\lambda,z} \) at \( z \) is contained in \( (W^\lambda)^* \) and defines an algebraic \( n \)-Whittaker vector of weight \( \eta_z \).

2. The distribution
\[
\phi_{\lambda,z}(x) := \mathcal{J}_\lambda(z, x) d\mu_\lambda(x), \quad x \in O_\lambda,
\]
is contained in \( (W^\lambda)^* \) and defines an algebraic \( \mathfrak{n} \)-Whittaker vector of weight \( \bar{\eta}_z \).

3. The algebraic Whittaker vectors \( \delta_{\lambda,z} \) and \( \phi_{\lambda,z} \) are related by
\[
\phi_{\lambda,z} = 2^{r_\lambda} U_\lambda \delta_{\lambda,z}.
\]

**Proof.**

1. Since \( W^\lambda = \mathcal{P}(O_\lambda)e^{-t(x)} \) and \( \mathcal{P}(O_\lambda) \cong \mathbb{C}[X_\lambda] \) the point evaluation of every \( \varphi \in W^\lambda \) at \( z \in X_\lambda \) is well-defined. Now \( d\pi_\lambda(u, 0, 0) = i(u|x) \) and the Whittaker property follows.

2. By Proposition 6.1 the function \( \phi_{\lambda,z} \) belongs to \( (W^\lambda)^* \). Since \( d\pi_\lambda(0, 0, v) = i(v|B_\lambda) \) the Whittaker property follows from Proposition 3.6.

3. By Theorem 6.3 we have for \( \varphi \in W^\lambda \)
\[
\langle \phi_{\lambda,z}, \varphi \rangle = \int_{O_\lambda} \mathcal{J}_\lambda(z, x) \varphi(x) d\mu_\lambda(x) = 2^{r_\lambda} U_\lambda \varphi(z)
\]
\[
= 2^{r_\lambda} \langle \delta_{\lambda,z}, U_\lambda \varphi \rangle = 2^{r_\lambda} \langle U_\lambda \delta_{\lambda,z}, \varphi \rangle
\]
since \( U_\lambda = U_\lambda \). \( \square \)
An interesting question is to determine the set of smooth Whittaker vectors for $\pi_\lambda$. We find that the algebraic $n$-Whittaker vector $\delta_{\lambda,z}$ (resp. the algebraic $\Pi$-Whittaker vector $\phi_{\lambda,z}$) is smooth if $z \in \overline{O_\lambda}$. For this we first prove estimates for $\delta_{\lambda,z}$ and $\phi_{\lambda,z}$ using the estimate for $J_\lambda(z,w)$ by Nakahama \cite{Nakahama} stated in Proposition \ref{prop:estimates}.

**Lemma 6.10.** For each $\lambda \in W$ and $x \in \overline{O_\lambda}$ there exist elements $X, Y \in U(\mathfrak{g})$ such that for all $\varphi \in W^\lambda$ we have

$$
|\langle \delta_{\lambda,x}, \varphi \rangle| \leq \|d\pi_\lambda(X)\varphi\|_{L^2(\omega_\lambda, d\mu_\lambda)}, \quad (6.2)
$$

$$
|\langle \phi_{\lambda,x}, \varphi \rangle| \leq \|d\pi_\lambda(Y)\varphi\|_{L^2(\omega_\lambda, d\mu_\lambda)}. \quad (6.3)
$$

**Proof.** We first prove (6.3). From Proposition \ref{prop:estimates} it follows that there exist constants $C_1 > 0$ and $k \in \mathbb{N}_0$ such that

$$
|J_\lambda(x, y)| \leq C_1 (1 + |y|^2)^k \quad \forall y \in \overline{O_\lambda}.
$$

Now there exists an $N \in \mathbb{N}_0$ with

$$
C_2 := \|(1 + |\cdot|^2)^{k-N}\|_{L^2(\omega_\lambda, d\mu_\lambda)} < \infty.
$$

Hence we obtain for $\varphi \in W^\lambda$, using Hölder’s inequality:

$$
|\langle \phi_{\lambda,x}, \varphi \rangle| \leq C_1 \int_{O_\lambda} (1 + |y|^2)^k |\varphi(y)| \, d\mu_\lambda(y)
$$

$$
= C_1 \int_{O_\lambda} (1 + |y|^2)^{k-N} \cdot (1 + |y|^2)^N |\varphi(y)| \, d\mu_\lambda(y)
$$

$$
\leq C_1 C_2 \|(1 + |\cdot|^2)^N \varphi\|_{L^2(\omega_\lambda, d\mu_\lambda)}.
$$

Since $d\pi_\lambda(U(\mathfrak{g}))$ contains all polynomials there exists $X \in U(\mathfrak{g})$ such that

$$
d\pi_\lambda(X)\varphi(y) = C_1 C_2 (1 + |y|^2)^N \varphi(y)
$$

which shows (6.3). Now by Proposition \ref{prop:estimates} (3) we have

$$
\delta_{\lambda,x} = 2^{-r\lambda} U_\lambda \phi_{\lambda,x}
$$

and we obtain

$$
|\langle \delta_{\lambda,x}, \varphi \rangle| = 2^{-r\lambda} |\langle U_\lambda \phi_{\lambda,x}, \varphi \rangle| = 2^{-r\lambda} |\langle \phi_{\lambda,x}, U_\lambda \varphi \rangle|
$$

$$
\leq 2^{-r\lambda} \|d\pi_\lambda(X)U_\lambda \varphi\|_{L^2(\omega_\lambda, d\mu_\lambda)}
$$

$$
= 2^{-r\lambda} \|d\pi_\lambda(X)\pi_\lambda(j)\varphi\|_{L^2(\omega_\lambda, d\mu_\lambda)}
$$

$$
= 2^{-r\lambda} \|\pi_\lambda(\text{Ad}(j^{-1})X)\varphi\|_{L^2(\omega_\lambda, d\mu_\lambda)}
$$

$$
= 2^{-r\lambda} \|d\pi_\lambda(\text{Ad}(j^{-1})X)\varphi\|_{L^2(\omega_\lambda, d\mu_\lambda)}.
$$

Hence (6.2) follows with $Y = 2^{-r\lambda} \text{Ad}(j^{-1})X$. \hfill \Box
Theorem 6.11. Let $\lambda \in W$. Then for every $x \in \mathcal{O}_{\lambda}$ the algebraic Whittaker vectors $\delta_{\lambda,x}$ and $\phi_{\lambda,x}$ are smooth, i.e., $\delta_{\lambda,x}, \phi_{\lambda,x} \in L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^{-\infty}$.

Proof. Recall that for $X \in U(\mathfrak{g})$ the map

$$L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda}) \to [0, \infty), \varphi \mapsto \|d\pi_{\lambda}(X)\varphi\|_{L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})}$$

is a continuous seminorm on $L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^\infty$. Since $W^\lambda \subseteq L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^\infty$ is dense the claim now follows from (6.2) and (6.3). 

Remark 6.12. For $x \in \mathcal{O}_{\lambda}$ it is easy to see that $\delta_{\lambda,x}, \phi_{\lambda,x} \in L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^{-\infty}$. In fact, by the Sobolev embedding theorem it follows that there is a continuous linear embedding $L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^\infty \hookrightarrow C^\infty(\mathcal{O}_{\lambda})$ and hence it is immediate that $\delta_{\lambda,x} \in L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^{-\infty}$. Further $\phi_{\lambda,x} = 2^s \mathcal{U}_1 \delta_{\lambda,x} \in L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^{-\infty}$ as $\mathcal{U}_1$ is an automorphism of $L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^{-\infty}$. However, it is a priori not clear that every function $\varphi \in L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^\infty$ extends to $\mathcal{O}_{\lambda}$ and that $\delta_{\lambda,x}$ is defined on $L^2(\mathcal{O}_{\lambda}, d\mu_{\lambda})^\infty$ for $x \in \partial \mathcal{O}_{\lambda}$. In [37] Theorems 5.15, 5.18 & 5.19 the statements of Theorem 6.11 are shown for the special case $V = \mathbb{R}$, i.e. $\mathfrak{g} \cong \mathfrak{sl}(2, \mathbb{R})$.

Remark 6.13. Note that since $\mathcal{U}_1(0, y) = 1$ for all $y \in V$ we have $\phi_{\lambda,0} = 1 d\mu_{\lambda}$, the constant function on $\mathcal{O}_{\lambda}$ with value 1, and hence

$$\mathcal{U}_1 \delta_{\lambda,0} = 2^s \mathcal{U}_1 \delta_{\lambda,0} d\mu_{\lambda}.$$

This formula resembles the fact that the Euclidean Fourier transform maps the Dirac delta distribution at the origin to a constant function.

7 Application to branching problems

For $\mathfrak{g} = \mathfrak{so}(2, n)$ we consider the representation $d\pi_{\lambda}^{\mathfrak{so}(2, n)}$ belonging to the minimal non-zero discrete Wallach point $\lambda = \frac{d}{2} = \frac{n-2}{2}$. We consider the restriction of $d\pi_{\lambda}^{\mathfrak{so}(2, n)}$ to the symmetric subalgebra $\mathfrak{h} = \mathfrak{so}(2, m) \oplus \mathfrak{so}(n-m)$ of $\mathfrak{g}$. Note that the pair $(\mathfrak{g}, \mathfrak{h})$ is a symmetric pair of holomorphic type and hence the restriction of the lowest weight representation $d\pi_{\lambda}^{\mathfrak{so}(2, n)}$ to $\mathfrak{h}$ is discretely decomposable with multiplicity one and the representations of $\mathfrak{h}$ appearing in the decomposition will again be lowest weight representations (see [32]).

Consider the Jordan algebra $V = \mathbb{R}^{1, n-1}$ which is the vector space $\mathbb{R} \times \mathbb{R}^{n-1}$ endowed with the multiplication

$$(x_1, x') \cdot (y_1, y') := (x_1 y_1 + x' \cdot y', x_1 y' + y_1 x')$$

for $x_1, y_1 \in \mathbb{R}, x', y' \in \mathbb{R}^{n-1}$, where $x' \cdot y'$ denotes the standard inner product on $\mathbb{R}^{n-1}$. We have $\det(V) = \mathfrak{so}(n-1)$, acting on the last $n-1$ variables, and $\text{str}(V) = \mathfrak{so}(1, n-1) \oplus \mathbb{R}$. The conformal Lie algebra of $V$ is given by $\mathfrak{g} = \mathfrak{so}(2, n)$. The subalgebra $\mathfrak{so}(2, m), 1 \leq m \leq n$, can be viewed as the conformal Lie algebra of the Jordan subalgebra $U = \{x \in V : x_{m+1} = \ldots = x_n = 0\} \cong \mathbb{R}^{1, m-1}$. Then $\mathfrak{so}(2, m)$ consists of the elements $(u, L(a) + D, v)$ with $u, a, v \in U$.
and $D \in \mathfrak{so}(m - 1)$ acting on the coordinates $x_2, \ldots, x_m$. The centralizer of $\mathfrak{so}(2, m)$ in $\mathfrak{g}$ is $\mathfrak{so}(n - m)$ acting linearly on the coordinates $x_{m+1}, \ldots, x_n$. Hence $\mathfrak{so}(n - m) \subseteq \mathfrak{t}^* = \mathfrak{so}(n - 1)$. Finally let $\mathfrak{h} = \mathfrak{so}(2, m) \oplus \mathfrak{so}(n - m) \subseteq \mathfrak{g}$ be the corresponding subalgebra of $\mathfrak{g}$.

To study the restriction $d\pi^\mathfrak{so}(2, n)|_{\mathfrak{h}}$ we use the Fock model $(d\rho^\mathfrak{so}(2, n), \mathcal{F}_\lambda)$ of the representation $d\pi^\mathfrak{so}(2, n)$. In this case

$$
\mathcal{X}_1^\mathfrak{so}(2, n) = \{ z \in \mathbb{C}^n : z_1^2 - z_2^2 - \cdots - z_n^2 = 0 \} \setminus \{ 0 \},
$$

$$
\mathcal{X}_2^\mathfrak{so}(2, n) = \{ z \in \mathbb{C}^n : z_1^2 - z_2^2 - \cdots - z_n^2 \neq 0 \},
$$

realized in the ambient space $V_{\mathbb{C}} = \mathbb{C}^n$, and hence

$$
\mathcal{P}(\mathcal{X}_1^\mathfrak{so}(2, n)) = \mathbb{C}[Z_1, \ldots, Z_n]/(Z_1^2 - Z_2^2 - \cdots - Z_n^2),
$$

$$
\mathcal{P}(\mathcal{X}_2^\mathfrak{so}(2, n)) = \mathbb{C}[Z_1, \ldots, Z_n].
$$

For arbitrary $\lambda \in \mathcal{W}$ the representation $d\rho^\mathfrak{so}(2, n)$ of $\mathfrak{g}_{\mathbb{C}}$ acts by

$$
d\rho_\lambda(a, -2iL(a), a) = 2i \sum_{j=1}^n a_j z_j, \quad d\rho_\lambda(a, 0, -a) = 2i \left( \partial_{a_2} + \frac{\lambda}{2} \text{tr}(a) \right),
$$

$$
d\rho_\lambda(a, 2iL(a), a) = -2i \sum_{j=1}^n a_j B^{n,j}_\lambda, \quad d\rho_\lambda(0, D, 0) = -\partial_{D^2},
$$

where

$$
B^{n,j}_\lambda = \varepsilon_j z_j \Box^n - 2(E^n + \lambda) \frac{\partial}{\partial z_j}, \quad \Box^n = \sum_{j=1}^n \varepsilon_j \left( \begin{array}{c} \partial^2 \\ \partial z_j \end{array} \right),
$$

$$
E^n = \sum_{j=1}^n \varepsilon_j \frac{\partial}{\partial z_j}, \quad \varepsilon_j = \begin{cases} +1 & \text{for } j = 1, \\ -1 & \text{for } 2 \leq j \leq n. \end{cases}
$$

We first consider the action of the compact part $\mathfrak{so}(n - m)$ which acts naturally on the coordinates $z_{m+1}, \ldots, z_n$. Note that

$$
\mathbb{C}[Z_1, \ldots, Z_n] = \mathbb{C}[Z_1, \ldots, Z_m] \otimes \mathbb{C}[Z_{m+1}, \ldots, Z_n].
$$

Every polynomial in $Z_{m+1}, \ldots, Z_n$ can be written as a sum of spherical harmonics multiplied with powers of $(Z_{m+1}^2 + \cdots + Z_n^2)$. In $\mathcal{P}(\mathcal{X}_1^{\mathfrak{so}(2, n)})$ we have $Z_{m+1}^2 + \cdots + Z_n^2 = Z_1^2 - Z_2^2 - \cdots - Z_n^2$ and hence

$$
\mathcal{P}(\mathcal{X}_1^{\mathfrak{so}(2, n)}) \cong \mathbb{C}[Z_1, \ldots, Z_n]/(Z_1^2 - Z_2^2 - \cdots - Z_n^2)
$$

$$
\cong \bigoplus_{k=0}^\infty \mathbb{C}[Z_1, \ldots, Z_m] \otimes \mathcal{H}^k(\mathbb{C}^{n-m}),
$$

which gives the decomposition into irreducible $\mathfrak{so}(n - m)$-representations.

Next we have to examine the action of $\mathfrak{so}(2, m)$ on each of the summands. For this we use the notation $V_{\mathbb{C}} = \mathbb{C}^n = \mathbb{C}^m \oplus \mathbb{C}^{n-m} = U_{\mathbb{C}} \oplus U_{\mathbb{C}}^\perp$.
Lemma 7.1. For \( k \in \mathbb{N}_0 \) and \( X \in \mathfrak{so}(2, m) \) we have
\[
d\lambda^{(2,n)}(X)|_{\mathbb{C}[Z_1,\ldots,Z_m] \otimes \mathcal{H}^k(\mathbb{C}^{n-m})} = d\lambda^{(2,m)}(X) \otimes \text{id}.
\]

Proof. On \( \mathbb{C}[Z_1,\ldots,Z_m] \otimes \mathcal{H}^k(\mathbb{C}^{n-m}) \) we check four exhaustive cases separately:

1. \( X = (0,D,0), D \in \mathfrak{so}(m - 1) \). The action is given by
\[
d\lambda^{(2,n)}(X) = -\partial_D z = d\lambda^{(2,m)}(X) \otimes \text{id}
\]

since \( D \) only acts on the coordinates \( z_2,\ldots,z_m \).

2. \( X = (a,-2iL(a), a), a \in U \). Since \( a \in U \subseteq \mathbb{C}^m \) it is immediate that
\[
d\lambda^{(2,n)}(X) = 2i \sum_{j=1}^{m} az_j = d\lambda^{(2,m)}(X) \otimes \text{id}.
\]

3. \( X = (a,0,-a), a \in U \). Assume first that \( a_1 = 0 \). Then \( L(a) \) annihilates \( \mathbb{C}^{n-m} \) and hence
\[
d\lambda^{(2,n)}(X) = 2i\partial_a z = d\lambda^{(2,m)}(X) \otimes \text{id}.
\]

For \( a = e \) we consider \( X = (e,0,-e) \) which acts by
\[
d\lambda^{(2,n)}(X) = 2i(E^n + \lambda) = 2i(E^n + \lambda + k) = d\lambda^{(2,m)}(X) \otimes \text{id}
\]

since \( E^{n-m} = \sum_{j=m+1}^{n} z_j \frac{\partial}{\partial z_j} \) acts on \( \mathcal{H}^k(\mathbb{C}^{n-m}) \) by the scalar \( k \).

4. \( X = (a,2iL(a), a), a \in U \). It suffices to check the claimed formula for \( a = \varepsilon_j, j = 1,\ldots,m \). We then have
\[
d\lambda^{(2,n)}(X) = -2i \left( \varepsilon_j z_j \Box^{n-m} - 2(E^n + \lambda) \frac{\partial}{\partial z_j} \right)
\]
\[
= -2i \left( \varepsilon_j z_j \Box^{n-m} - 2(E^n + \lambda + k) \frac{\partial}{\partial z_j} \right)
\]
\[
= d\lambda^{(2,m)}(X) \otimes \text{id}
\]

since

\[
\Box^{n}[f(z_1,\ldots,z_m)g(z_{m+1},\ldots,z_n)]
\]
\[
= \Box^{m} f(z_1,\ldots,z_m) \cdot g(z_{m+1},\ldots,z_n)
\]
\[
+ f(z_1,\ldots,z_m) \cdot \Delta^{n-m} g(z_{m+1},\ldots,z_n),
\]

where \( \Delta^{n-m} \) denotes the Laplacian on \( \mathbb{C}^{n-m} \) and \( E^{n-m} \) the Euler operator on \( \mathbb{C}^{n-m} \) which acts on \( \mathcal{H}^k(\mathbb{C}^{n-m}) \) by the scalar \( k \). \( \square \)
This shows the following theorem:

**Theorem 7.2.** For $\lambda = \frac{n-2}{2}$ the minimal non-zero discrete Wallach point for $\mathfrak{so}(2,n)$ we have

$$d_{\lambda}^{{\mathfrak{so}}(2,n)} = \bigoplus_{k=0}^{\infty} d_{\lambda+k}^{{\mathfrak{so}}(2,m)} \otimes \mathcal{H}^k(\mathbb{R}^{n-m}).$$

For $m < n$ we have $\lambda + k = \frac{2\lambda+2k-2}{2} > \frac{m-2}{2}$ for $k \in \mathbb{N}_0$, and hence the representations $d_{\lambda+k}^{{\mathfrak{so}}(2,m)}$ belong to Wallach points in the continuous part of the Wallach set. For $2k > 2m - n$ the representation $d_{\lambda+k}^{{\mathfrak{so}}(2,m)}$ are holomorphic discrete series. In particular, if $2m \geq n$ then there occur representations in the branching law which are not holomorphic discrete series.

**References**


[34] ———, The Schrödinger model for the minimal representation of the indefinite orthogonal group O(p,q), Memoirs of the AMS 213 (2011), no. 1000.


The Magnitude of Metric Spaces

Tom Leinster

Received: August 15, 2012
Communicated by Stefan Schwede

Abstract. Magnitude is a real-valued invariant of metric spaces, analogous to Euler characteristic of topological spaces and cardinality of sets. The definition of magnitude is a special case of a general categorical definition that clarifies the analogies between cardinality-like invariants in mathematics. Although this motivation is a world away from geometric measure, magnitude, when applied to subsets of $\mathbb{R}^n$, turns out to be intimately related to invariants such as volume, surface area, perimeter and dimension. We describe several aspects of this relationship, providing evidence for a conjecture (first stated in joint work with Willerton) that magnitude encodes all the most important invariants of classical integral geometry.

2010 Mathematics Subject Classification: 51F99 (primary), 18D20, 18F99, 28A75, 49Q20, 52A20, 52A38, 53C65 (secondary).
Keywords and Phrases: metric space, magnitude, enriched category, Möbius inversion, Euler characteristic of a category, finite metric space, convex set, integral geometry, valuation, intrinsic volume, fractal dimension, positive definite space, space of negative type.

Contents

Introduction 858

1 Enriched categories 861
  1.1 The magnitude of a matrix ......................... 861
  1.2 Background on enriched categories ................. 863
  1.3 The magnitude of an enriched category .......... 864
  1.4 Properties ........................................ 867

Documenta Mathematica 18 (2013) 857–905
INTRODUCTION

Many mathematical objects carry a canonical notion of size. Sets have cardinality, vector spaces have dimension, topological spaces have Euler characteristic, and probability spaces have entropy. This work adds a new item to the list: metric spaces have magnitude.

Already, several cardinality-like invariants are tied together by the notion of the Euler characteristic of a category [21]. This is a rational-valued invariant of finite categories. A network of theorems describes the close relationships between this invariant and established cardinality-like invariants, including the cardinality of sets and of groupoids [1], the Euler characteristic of topological spaces and of posets, and even the Euler characteristic of orbifolds. (That Euler characteristic deserves to be considered an analogue of cardinality was first made clear by Schanuel [41, 42].) These results attest that for categories, Euler characteristic is the fundamental notion of size.

Here we go further. Categories are a special case of the more general concept of enriched category. Much of ordinary category theory generalizes to the enriched setting, and this is true, in particular, of the Euler characteristic of categories. Rebaptizing Euler characteristic as ‘magnitude’ to avoid a potential ambiguity later, this gives a canonical definition of the magnitude of an enriched category. Metric spaces, as well as categories, are examples of enriched categories:

\[\text{(categories)} \subset \text{(enriched categories)} \supset \text{(metric spaces)}\]

[19, 20]. The analogy between categories and metric spaces can be understood in broad terms immediately. A category has objects; a metric space has points. For any two objects there is a set (the maps between them); for any two points there is a real number (the distance between them). For any three objects there is an operation of composition; for any three points there is a triangle inequality.
Having generalized the definition of magnitude (or Euler characteristic) from ordinary to enriched categories, we specialize it to metric spaces. This gives our invariant. The fundamental role of the Euler characteristic of categories strongly suggests that the magnitude of metric spaces should play a fundamental role too. Our faith is rewarded by a series of theorems showing that magnitude is intimately related to the classical invariants of integral geometry: dimension, perimeter, surface area, volume, ..., This is despite the fact that no concept of measure or integration goes into the definition of magnitude; they arise spontaneously from the general categorical definition.

This, then, is part of the appeal of magnitude. It is motivated in abstract terms that seem to have nothing to do with geometry or measure, and defined in the very wide generality of enriched categories—yet when specialized to the context of metric spaces, it turns out to have a great deal to say about geometry and measure.

There is a further surprise. While the author’s motivation was category-theoretic, magnitude had already arisen in work on the quantification of biological diversity. In 1994, Solow and Polasky [44] carried out a probabilistic analysis of the benefits of high biodiversity, and isolated a particular quantity that they called the ‘effective number of species’. It is the same as our magnitude. This is no coincidence: the theoretical problem of how to maximize diversity can be solved using the concept of magnitude and some of the results presented here [23]. Indeed, under suitable circumstances, magnitude can be interpreted as maximum diversity, a cousin to maximum entropy.

Our first step is to define the magnitude of an enriched category (Section 1). This puts the notion of the magnitude of a metric space into a wide mathematical context, showing how analogous theories can be built in parts of mathematics far away from metric geometry. The reader interested only in geometry can, however, avoid these general considerations without logical harm, and begin at Section 2.

A topological space is not guaranteed to have a well-defined Euler characteristic unless it satisfies some finiteness condition. Similarly, the magnitude of an enriched category is defined under an assumption of finiteness; specializing to metric spaces, the definition of magnitude is just for finite spaces (Section 2). The magnitude of a finite metric space can be thought of as the ‘effective number of points’. It deserves study partly because of its intrinsic interest, partly because of its applications to the measurement of diversity, and partly because it is used in the theory of magnitude of infinite metric spaces.

While categorical arguments do not (yet) furnish a definition of the magnitude of an infinite space, several methods for passing from finite to infinite immediately suggest themselves. Meckes [31] has shown that they are largely equivalent. Using the most elementary such method, coupled with some Fourier analysis, we produce evidence for the following conjectural principle:
magnitude encodes all the most important invariants of integral geometry

(Section 3). The most basic instance of this principle is the fact that a line segment of length \( t \) has magnitude \( 1 + t/2 \), enabling one to recover length from magnitude. Less basic is the notion of the magnitude dimension of a space \( A \), defined as the growth of the function \( t \mapsto |tA| \); here \( tA \) is \( A \) scaled up by a factor of \( t \), and \(|tA|\) is its magnitude. We show, for example, that a subset of \( \mathbb{R}^N \) with nonzero Lebesgue measure has magnitude dimension \( N \). Magnitude dimension also appears to behave sensibly for fractals: for instance, Theorem 11 of [27] implies that the magnitude dimension of the ternary Cantor set is the same as its Hausdorff dimension (namely, \( \log_3 2 \)).

It seems, moreover, that for any convex subset \( A \) of Euclidean space, all of the intrinsic volumes of \( A \) can be recovered from the function \( t \mapsto |tA| \). This was first conjectured in [27], and appears below as Conjecture 3.5.10. In two dimensions, the conjectured formula is

\[
|tA| = \frac{1}{2\pi} \text{area}(A) \cdot t^2 + \frac{1}{4} \text{perimeter}(A) \cdot t + \chi(A).
\]

This resembles the theorem of Willerton [51] that for a compact homogeneous Riemannian 2-manifold \( A \),

\[
|tA| = \frac{1}{2\pi} \text{area}(A) \cdot t^2 + \chi(A) + O(t^{-2})
\]
as \( t \to \infty \). This in turn resembles the celebrated tube formula of Weyl.

Review sections provide the necessary background on both enriched categories and integral geometry. No expertise in category theory or integral geometry is needed to read this paper.

Related Work The basic ideas of this paper were first written up in a 2008 internet posting [22]. Several papers have already built on this. Leinster and Willerton [27] studied the large-scale asymptotics of the magnitude of subsets of Euclidean space, and stated the conjecture just mentioned. The precise form of that conjecture was motivated by numerical evidence and heuristic arguments found by Willerton [50]. Leinster [23] established magnitude as maximum diversity. Meckes [31] proved, inter alia, the equivalence of several definitions of the magnitude of compact metric spaces, and by using more subtle analytical methods than are used here, extended some of the results of Section 3 below. The magnitude of spheres is especially well understood [27, 51, 31].

In the literature on quantifying biodiversity, magnitude appears not only in the paper of Solow and Polasky [44], but also in later papers such as [38]. The approach to biodiversity measurement taken in [26] arose from the theory in the present paper. This relationship is explored further in [23] and Section 2 of [31].

Geometry as the study of metric structures is developed in the books of Blumenthal [4] and Gromov [11], among others; representatives of the theory of finite metric spaces are [4] and papers of Dress and collaborators [2, 6]. We
will make contact with the theory of spaces of negative type, which goes back
to Menger [32] and Schoenberg [43]. This connection has been exploited by
Meckes [31]. It is notable that the complete bipartite graph $K_{3,2}$ appears as a
minimal example in both [2] and Example 2.2.7 below.

**Notation** Given $N \in \mathbb{N} = \{0, 1, 2, \ldots\}$, we write $\mathbb{R}^N$ for real $N$-dimensional
space as a set, topological space or vector space—but with no implied choice of
metric except when $N = 1$. The metric on a metric space $A$ is denoted by $d$ or
$d_A$. We write $\# X$ for the cardinality of a finite set $X$. When $\mathcal{C}$ is a category,
$C \in \mathcal{C}$ means that $C$ is an object of $\mathcal{C}$.

**Acknowledgements** I have had countless useful conversations on magni-
tude with Mark Meckes and Simon Willerton. Their insights have played a
very important role, and I thank them for it. I also thank John Baez, Neal
Bez, Paul Blackwell, Yemon Choi, Christina Cobbold, David Corfield, Alastair
Craw, Jacques Distler, Joe Fu, Anton Geraschenko, Martin Hyland, David Jor-
dan, André Joyal, Joachim Kock, Christian Korff, Urs Schreiber, Josh Shadlen,
Ivan Smith, David Speyer, and Terry Tao. Two web resources have been crucial
to the progress of this work: The $n$-Category Café\footnote{http://golem.ph.utexas.edu/category} and MathOverflow.\footnote{http://mathoverflow.net} Parts
of this work were carried out at the Centre de Recerca Matemàtica (Barcelona)
and the School of Mathematics and Statistics at the University of Sheffield. I
thank them for their hospitality. This work was supported by an EPSRC Ad-
vanced Research Fellowship.

1 Enriched categories

This section describes the conceptual origins of the notion of magnitude.
We define the magnitude of an enriched category, in two steps. First we assign
a number to every matrix; then we assign a matrix to every enriched category.
We pause in between to recall some basic aspects of enriched category theory:
the definitions, and how a metric space can be viewed as an enriched category.

1.1 The magnitude of a matrix

A rig (or semiring) is a ring without negatives: a set $k$ equipped with a com-
mutative monoid structure $(+, 0)$ and a monoid structure $(\cdot, 1)$, the latter dis-
tributing over the former. For us, rig will mean commutative rig: one whose
multiplication is commutative.

It will be convenient to use matrices whose rows and columns are indexed by
abstract finite sets. Thus, for finite sets $I$ and $J$, an $I \times J$ matrix over a rig $k$
is a function $I \times J \to k$. The usual operations can be performed, e.g. an $H \times I$
matrix can be multiplied by an $I \times J$ matrix to give an $H \times J$ matrix. The
identity matrix is the Kronecker $\delta$. An $I \times J$ matrix $\zeta$ has a $J \times I$ transpose $\zeta^\ast$.

Given a finite set $I$, we write $u_I \in k^I$ for the column vector with $u_I(i) = 1$ for all $i \in I$.

**Definition 1.1.1** Let $\zeta$ be an $I \times J$ matrix over a rig $k$. A *weighting* on $\zeta$ is a column vector $w \in k^J$ such that $\zeta w = u_I$. A *coweighting* on $\zeta$ is a row vector $v \in k^I$ such that $v \zeta = u_J^\ast$.

A matrix may admit zero, one, or many (co)weightings, but their freedom is subject to the following constraint.

**Lemma 1.1.2** Let $\zeta$ be an $I \times J$ matrix over a rig, let $w$ be a weighting on $\zeta$, and let $v$ be a coweighting on $\zeta$. Then

$$\sum_{j \in J} w(j) = \sum_{i \in I} v(i).$$

**Proof** $\sum_{j} w(j) = u_J^\ast w = v \zeta w = vu_I = \sum_{i} v(i)$.\[\Box\]

We refer to the entries $w(j) \in k$ of a weighting $w$ as *weights*, and similarly *coweights*. The lemma implies that if a matrix $\zeta$ has both a weighting and a coweighting, then the total weight is independent of the weighting chosen. This makes the following definition possible.

**Definition 1.1.3** A matrix $\zeta$ over a rig $k$ has *magnitude* if it admits at least one weighting and at least one coweighting. Its *magnitude* is then

$$|\zeta| = \sum_{j} w(j) = \sum_{i} v(i) \in k$$

for any weighting $w$ and coweighting $v$ on $\zeta$.

We will be concerned with square matrices $\zeta$. If $\zeta$ is invertible then there are a unique weighting and a unique coweighting. (Conversely, if $k$ is a field then a unique weighting or coweighting implies invertibility.) The weights are then the sums of the rows of $\zeta^{-1}$, and the coweights are the sums of the columns.

Lemma 1.1.2 is obvious in this case, and there is an easy formula for the magnitude:

**Lemma 1.1.4** Let $\zeta$ be an invertible $I \times I$ matrix over a rig. Then $\zeta$ has a unique weighting $w$, given by $w(j) = \sum_i \zeta^{-1}(j,i)$ ($j \in I$), and a unique coweighting given by the dual formula. Moreover,

$$|\zeta| = \sum_{i,j \in I} \zeta^{-1}(j,i).$$

\[\Box\]

Often our matrix $\zeta$ will be symmetric, in which case weightings and coweightings are essentially the same.

Documenta Mathematica 18 (2013) 857–905
1.2 Background on enriched categories

Here we review two standard notions: monoidal category, and category enriched in a monoidal category.

A monoidal category is a category \( \mathcal{V} \) equipped with an associative binary operation \( \otimes \) (which is formally a functor \( \mathcal{V} \times \mathcal{V} \to \mathcal{V} \)) and a unit object \( \mathbb{1} \in \mathcal{V} \). The associativity and unit axioms are only required to hold up to suitably coherent isomorphism; see [28] for details.

Examples 1.2.1

i. \( \mathcal{V} \) is the category \( \text{Set} \) of sets, \( \otimes \) is cartesian product \( \times \), and \( \mathbb{1} \) is a one-element set \( \{\ast\} \).

ii. \( \mathcal{V} \) is the category \( \text{Vect} \) of vector spaces over some field \( K \), the product \( \otimes \) is the usual tensor product \( \otimes_K \), and \( \mathbb{1} = K \).

iii. A poset can be viewed as a category in which each hom-set has at most one element. In particular, consider the poset \( ([0, \infty], \geq) \) of nonnegative reals together with infinity. The objects of the resulting category are the elements of \([0, \infty]\), there is one map \( x \to y \) when \( x \geq y \), and there are none otherwise. This is a monoidal category with \( \otimes = + \) and \( \mathbb{1} = 0 \).

iv. Let \( 2 \) be the category of Boolean truth values [19]: there are two objects, \( f \) (‘false’) and \( t \) (‘true’), and a single non-identity map, \( f \to t \). Taking \( \otimes \) to be conjunction and \( \mathbb{1} = t \) makes \( 2 \) monoidal. Then \( 2 \) is a monoidal subcategory of \( \text{Set} \), identifying \( f \) with \( \emptyset \) and \( t \) with \( \{\ast\} \). It is also a monoidal subcategory of \([0, \infty]\), identifying \( f \) with \( \infty \) and \( t \) with \( 0 \).

Let \( \mathcal{V} = (\mathcal{V}, \otimes, \mathbb{1}) \) be a monoidal category. The definition of category enriched in \( \mathcal{V} \), or \( \mathcal{V} \)-category, is obtained from the definition of ordinary category by asking that the hom-sets are no longer sets but objects of \( \mathcal{V} \). Thus, a (small) \( \mathcal{V} \)-category \( A \) consists of a set \( \text{ob} A \) of objects, an object \( \text{Hom}(a, b) \) of \( \mathcal{V} \), for each \( a, b \in \text{ob} A \), and operations of composition and identity satisfying appropriate axioms [16]. The operation of composition consists of a map

\[
\text{Hom}(a, b) \otimes \text{Hom}(b, c) \to \text{Hom}(a, c)
\]

in \( \mathcal{V} \) for each \( a, b, c \in \text{ob} A \), while the identities are provided by a map \( \mathbb{1} \to \text{Hom}(a, a) \) for each \( a \in \text{ob} A \).

There is an accompanying notion of enriched functor. Given \( \mathcal{V} \)-categories \( A \) and \( A' \), a \( \mathcal{V} \)-functor \( F : A \to A' \) consists of a function \( \text{ob} A \to \text{ob} A' \), written \( a \mapsto F(a) \), together with a map

\[
F(a) \to \text{Hom}(F(a), F(b))
\]

in \( \mathcal{V} \) for each \( a, b \in \text{ob} A \), satisfying suitable axioms [16]. We write \( \mathcal{V} \text{-Cat} \) for the category of \( \mathcal{V} \)-categories and \( \mathcal{V} \)-functors.

Examples 1.2.2

i. Let \( \mathcal{V} = \text{Set} \). Then \( \mathcal{V} \text{-Cat} \) is the category \( \text{Cat} \) of (small) categories and functors.
ii. Let $\mathcal{V} = \textbf{Vect}$. Then $\mathcal{V}$-Cat is the category of linear categories or algebroids: categories equipped with a vector space structure on each hom-set, such that composition is bilinear.

iii. Let $\mathcal{V} = [0, \infty]$. Then, as observed by Lawvere \cite{19, 20}, a $\mathcal{V}$-category is a generalized metric space. That is, a $\mathcal{V}$-category consists of a set $A$ of objects or points together with, for each $a,b \in A$, a real number $\text{Hom}(a,b) = d(a,b) \in [0, \infty]$, satisfying the axioms

$$d(a,b) + d(b,c) \geq d(a,c), \quad d(a,a) = 0$$

$(a,b,c \in A)$. Such spaces are more general than classical metric spaces in three ways: $\infty$ is permitted as a distance, the separation axiom $d(a,b) = 0 \Rightarrow a = b$ is dropped, and, most significantly, the symmetry axiom $d(a,b) = d(b,a)$ is dropped.

A $\mathcal{V}$-functor $f : A \to A'$ between generalized metric spaces $A$ and $A'$ is a distance-decreasing map: one satisfying $d(a,b) \geq d(f(a),f(b))$ for all $a,b \in A$. Hence $[0, \infty]$-Cat is the category $\textbf{MS}$ of generalized metric spaces and distance-decreasing maps. The isomorphisms in $\textbf{MS}$ are the isometries.

iv. Let $\mathcal{V} = 2$. A $\mathcal{V}$-category is a set equipped with a preorder (a reflexive transitive relation), which up to equivalence of $\mathcal{V}$-categories is the same thing as a poset.

The embedding $2 \hookrightarrow \textbf{Set}$ of monoidal categories induces an embedding $2$-Cat $\hookrightarrow \textbf{Set}$-Cat; this is the embedding $\textbf{Poset} \hookrightarrow \textbf{Cat}$ of Example 1.2.1(iii). Similarly, the embedding $2 \hookrightarrow [0, \infty]$ induces an embedding $\textbf{Poset} \hookrightarrow \textbf{MS}$: as observed in \cite{19}, a poset can be understood as a non-symmetric metric space whose distances are all 0 or $\infty$.

1.3 The magnitude of an enriched category

Here we meet the definition on which the rest of this work is built. Having already defined the magnitude of a matrix, we now assign a matrix to each enriched category. To do this, we assume some further structure on the base category $\mathcal{V}$. In fact, we assume that we have a notion of size for objects of $\mathcal{V}$. This, then, will lead to a notion of size for categories enriched in $\mathcal{V}$.

Let $\mathcal{V}$ be a monoidal category. We will suppose given a rig $k$ and a monoid homomorphism

$$|\cdot| : (\text{ob } \mathcal{V}/ \cong, \circ, 1) \to (k, \cdot, 1).$$

(This is, deliberately, the same symbol as for magnitude; no confusion should arise.) The domain here is the monoid of isomorphism classes of objects of $\mathcal{V}$.

Examples 1.3.1 i. When $\mathcal{V}$ is the monoidal category $\textbf{FinSet}$ of finite sets, we take $k = \mathbb{Q}$ and $|X| = \#X$. 
The Magnitude of Metric Spaces

ii. When $\mathcal{V}$ is the monoidal category $\text{FDVect}$ of finite-dimensional vector spaces, we take $k = \mathbb{Q}$ and $|X| = \dim X$.

iii. When $\mathcal{V} = [0, \infty]$, we take $k = \mathbb{R}$ and $|x| = e^{-x}$. (If $|\cdot|$ is to be measurable then the only possibility is $|x| = C^x$ for some constant $C \geq 0$.)

iv. When $\mathcal{V} = 2$, we take $k = \mathbb{Z}$, $|f| = 0$ and $|t| = 1$. This is a restriction of the functions $|\cdot|$ of (i) and (iii) along the embeddings $2 \hookrightarrow \text{FinSet}$ and $2 \hookrightarrow [0, \infty]$ of Example 1.2.1(iv).

Write $\mathcal{V}$-cat (with a small ‘c’) for the category whose objects are the $\mathcal{V}$-categories with finite object-sets and whose maps are the $\mathcal{V}$-functors between them.

Definition 1.3.2 Let $A \in \mathcal{V}$-cat.

i. The similarity matrix of $A$ is the ob $A \times \text{ob } A$ matrix $\zeta_A$ over $k$ defined by $\zeta_A(a, b) = |\text{Hom}(a, b)|$ ($a, b \in A$).

ii. A (co)weighting on $A$ is a (co)weighting on $\zeta_A$.

iii. $A$ has magnitude if $\zeta_A$ does; its magnitude is then $|A| = |\zeta_A|$.

iv. $A$ has Möbius inversion if $\zeta_A$ is invertible; its Möbius matrix is then $\mu_A = \zeta^{-1}_A$.

Magnitude is, then, a partially-defined function $|\cdot|: \mathcal{V}$-cat $\rightarrow k$.

Examples 1.3.3 i. When $\mathcal{V} = \text{FinSet}$, we obtain a notion of the magnitude $|A| \in \mathbb{Q}$ of any (suitable) finite category $A$. This is also called the Euler characteristic of $A$ and written as $\chi(A)$ [21]. There are theorems relating it to the Euler characteristics of topological spaces, of graphs, of posets and of orbifolds, the cardinality of sets, and the order of groups.

Very many finite categories have Möbius inversion, or are equivalent to some category with Möbius inversion; all such categories have Euler characteristic. (This includes all finite posets, groupoids, monoids, categories containing no nontrivial idempotents, and categories admitting an epimono factorization system.) The Möbius matrix $\mu_A$ is a generalization of Rota’s Möbius function for posets [40], which in turn generalizes the classical Möbius function on integers. For details, see [21], and for further material on Euler characteristic and Möbius inversion for categories, see Berger and Leinster [3], Fiore, Lück and Sauer [7, 8], Jacobsen and Møller [14], Leinster [25], and Noguchi [33, 34, 35, 36].

---

3I thank Mark Meckes for pointing out that the more obvious hypothesis of continuity can be weakened to measurability [9]. In fact it suffices to assume that $|\cdot|$ is bounded on some set of positive measure [18].
Similarly, taking \( \mathcal{V} = \text{FDVect} \) gives an invariant \( \chi(A) = |A| \in \mathbb{Q} \) of linear categories \( A \) with finitely many objects and finite-dimensional hom-spaces.

Taking \( \mathcal{V} = [0, \infty] \) gives the notion of the magnitude \( |A| \in \mathbb{R} \) of a (generalized) finite metric space \( A \). This is the main subject of this paper.

Taking \( \mathcal{V} = 2 \) gives a notion of the magnitude \( |A| \in \mathbb{Z} \) of a finite poset \( A \). Under the name of Euler characteristic, this goes back to Rota [40]; see [46] for a modern account. It is always defined. Indeed, every poset has Möbius inversion, and the Möbius matrix is the Möbius function of Rota mentioned in (i).

We have noted that a poset can be viewed as a category, or alternatively as a generalized metric space. The three notions of magnitude are compatible: the magnitude of a poset is the same as that of the corresponding category or generalized metric space.

Let \( \mathcal{V} \) be a category of topological spaces in which every object has a well-defined Euler characteristic (e.g. finite CW-complexes). Taking \( |X| \) to be the Euler characteristic of a space \( X \), we obtain a notion of the magnitude or Euler characteristic of a topologically-enriched category.

The definition of the magnitude of a \( \mathcal{V} \)-category \( A \) is independent of the composition and identities in \( A \), so could equally well be made in the generality of \( \mathcal{V} \)-graphs. (A \( \mathcal{V} \)-graph \( G \) is a set \( \text{ob} G \) of objects together with, for each \( a, b \in \text{ob} G \), an object \( \text{Hom}(a, b) \) of \( \mathcal{V} \).) However, it is not clear that it is fruitful to do so. Two theorems on the magnitude or Euler characteristic of ordinary categories illuminate the general situation.

The first, Proposition 2.10 of [21], concerns directed graphs. The Euler characteristic of a category \( A \) is not in general equal to the Euler characteristic of its underlying graph \( U(A) \). But the functor \( U \) has a left adjoint \( F \), assigning to a graph \( G \) the category \( F(G) \) whose objects are the vertices and whose maps are the paths in \( G \). If \( G \) is finite and circuit-free then \( F(G) \) is finite, and the theorem is that \( \chi(F(G)) = \chi(G) \). So the Euler characteristics of categories and graphs are closely related, but not in the most obvious way.

The second theorem, Proposition 1.5 of [25], concerns the classifying space \( \lvert NA \rvert \) of a category \( A \) (the geometric realization of its simplicial nerve \( NA \)). Under a suitable finiteness condition, the topological space \( \lvert NA \rvert \) has a well-defined Euler characteristic, and it is a theorem that \( \chi(\lvert NA \rvert) = \chi(A) \). It follows that if two categories have the same underlying graph but different compositions then their classifying spaces, although not usually homotopy equivalent, have the same Euler characteristic. So if we wish the Euler characteristic of a category to be defined in such a way that it is equal to the Euler characteristic of its classifying space, it is destined to be independent of composition.

(At least, this is the case under the finiteness condition concerned. This states that the simplicial nerve of \( A \) contains only finitely many nondegenerate simplices, as in Proposition 2.11 of [21].) In a different setting, with a different
definition of Euler characteristic, Fiore, Lück and Sauer have found a pair of finite categories whose underlying graphs are the same but whose Euler characteristics are different: see [7], after Theorem 1.14.)

1.4 Properties

Much of ordinary category theory generalizes smoothly to enriched categories. This includes many of the properties of the Euler characteristic of categories [21]. We list some of those properties now, using the symbols $\mathcal{V}$, $k$ and $|\cdot|$ as in the previous section.

There are notions of adjunction and equivalence between $\mathcal{V}$-categories [16], generalizing the case $\mathcal{V} = \text{Set}$ of ordinary categories. We write $\simeq$ for equivalence of $\mathcal{V}$-categories.

**Proposition 1.4.1** Let $A, B \in \mathcal{V}\text{-cat}$.

i. If there exist adjoint $\mathcal{V}$-functors $A \rightleftarrows B$, and $A$ and $B$ have magnitude, then $|A| = |B|$.

ii. If $A \simeq B$, and $A$ and $B$ have magnitude, then $|A| = |B|$.

iii. If $A \simeq B$ and $n \cdot 1 \in k$ has a multiplicative inverse for all positive integers $n$, then $A$ has magnitude if and only if $B$ does.

**Proof** Part (i) has the same proof as Proposition 2.4(a) of [21], and part (ii) follows immediately. Part (iii) has the same proof as Lemma 1.12 of [21].

For example, take a generalized metric space $A$ and adjoin a new point at distance zero from some existing point. Then the new space $A'$ is equivalent to $A$. By Proposition 1.4.1, if $A$ has magnitude then $A'$ does too, and $|A| = |A'|$.

On the other hand, the proposition is trivial for classical metric spaces $A, B$: if there is an adjunction between $A$ and $B$ (and in particular if $A \simeq B$) then in fact $A$ and $B$ are isometric.

So far, we have not used the multiplicativity of the function $|\cdot|$ on objects of $\mathcal{V}$. We now show that it implies a multiplicativity property of the function $|\cdot|$ on $\mathcal{V}$-categories.

Assume that the monoidal category $\mathcal{V}$ is symmetric, that is, equipped with an isomorphism $X \otimes Y \to Y \otimes X$ for each pair $X, Y$ of objects, satisfying axioms [28]. There is a product on $\mathcal{V}\text{-Cat}$, also denoted by $\otimes$, defined as follows. Let $A, B \in \mathcal{V}\text{-Cat}$. Then $A \otimes B$ is the $\mathcal{V}$-category whose object-set is $\text{ob} A \times \text{ob} B$ and whose hom-objects are given by

$$\text{Hom}((a, b), (a', b')) = \text{Hom}(a, a') \otimes \text{Hom}(b, b').$$

Composition is defined with the aid of the symmetry [16]. The unit for this product is the one-object $\mathcal{V}$-category $I$ whose single hom-object is $1 \in \mathcal{V}$.
Examples 1.4.2  

i. When $\mathcal{V} = \textbf{Set}$, this is the ordinary product $\times$ of categories.

ii. There is a one-parameter family of products on metric spaces. For $1 \leq p \leq \infty$ and metric spaces $A$ and $B$, let $A \otimes_p B$ be the metric space whose point-set is the product of the point-sets of $A$ and $B$, and whose distances are given by

$$d((a,b),(a',b')) = \begin{cases} 
(d(a,a')^p + d(b,b')^p)^{1/p} & \text{if } p < \infty \\
\max\{d(a,a'),d(b,b')\} & \text{if } p = \infty.
\end{cases}$$

Then the tensor product $\otimes$ defined above is $\otimes_1$.

Proposition 1.4.3 Let $A,B \in \mathcal{V}\text{-}\textbf{cat}$. If $A$ and $B$ have magnitude then so does $A \otimes B$, with

$$|A \otimes B| = |A||B|.$$  

Furthermore, the unit $\mathcal{V}$-category $I$ has magnitude 1.

Proof As for Proposition 2.6 of [21].

Magnitude is therefore a partially-defined monoid homomorphism

$$|\cdot|: (\mathcal{V}\text{-}\textbf{cat}/\simeq, \otimes, I) \rightarrow (k, \cdot, 1).$$

Under mild assumptions, coproducts of $\mathcal{V}$-categories exist and interact well with magnitude. Indeed, assume that $\mathcal{V}$ has an initial object $0$, with $X \otimes 0 \simeq 0 \simeq 0 \otimes X$ for all $X \in \mathcal{V}$. Then for any two $\mathcal{V}$-categories $A$ and $B$, the coproduct $A + B$ in $\mathcal{V}\text{-}\textbf{Cat}$ exists. It is constructed by taking the disjoint union of $A$ and $B$ and setting $\text{Hom}(a,b) = \text{Hom}(b,a) = 0$ whenever $a \in A$ and $b \in B$. There is also an initial $\mathcal{V}$-category $\emptyset$, with no objects.

When $\mathcal{V} = [0,\infty]$, the coproduct of metric spaces $A$ and $B$ is their distant union, the disjoint union of $A$ and $B$ with $d(a,b) = d(b,a) = \infty$ whenever $a \in A$ and $b \in B$.

Assume also that $|0| = 0$, where the 0 on the left-hand side is the initial object of $\mathcal{V}$. This assumption and the previous ones hold in all of our examples.

Proposition 1.4.4 Let $A,B \in \mathcal{V}\text{-}\textbf{cat}$. If $A$ and $B$ have magnitude then so does $A + B$, with

$$|A + B| = |A| + |B|.$$  

Furthermore, the initial $\mathcal{V}$-category $\emptyset$ has magnitude 0.

Proof As for Proposition 2.6 of [21].

It might seem unsatisfactory that not every $\mathcal{V}$-category with finite object-set has magnitude. This can be resolved as follows.

Given $A \in \mathcal{V}\text{-}\textbf{cat}$, there is an evident notion of (co)weighting on $A$ with values in a prescribed $k$-algebra. As in Lemma 1.1.2, the total weight is always equal
The Magnitude of Metric Spaces

869

The Magnitude of Metric Spaces

869

Let \( R(A) \) be the free \( k \)-algebra containing a weighting \( w \) and a coweighting \( v \) for \( A \). Then \( \sum_a w(a) = \sum_a v(a) = [A], \) say. This is always defined, and we may call \([A] \in R(A)\) the formal magnitude of \( A \).

A homomorphism \( \phi \) from \( R(A) \) to another \( k \)-algebra \( S \) amounts to a weighting and a coweighting for \( A \) in \( S \); moreover, \( \phi([A]) \in S \) is independent of the homomorphism \( \phi \) chosen. In particular, \( A \) has magnitude in the original sense if and only if there exists a \( k \)-algebra homomorphism \( \phi: R(A) \to k \). In that case, \( |A| = \phi([A]) \) for any such \( \phi \).

This may lead to a more conceptually satisfactory theory, but at a price: the magnitudes of different categories lie in different rigs, complicating results such as those of the present section. In any case, we say no more about this approach.

2 Finite metric spaces

The definition of the magnitude of a finite metric space is a special case of the definition for enriched categories. Its most basic properties are special cases of general results. But metric spaces have many features not possessed by enriched categories in general. By exploiting them, we uncover a geometrically rich theory.

A crucial feature of metric spaces is that they can be rescaled. When handed a space, we gain more information about it by considering the magnitudes of its rescaled brothers and sisters than by taking it in isolation. This information is encapsulated in the so-called magnitude function of the space.

For some spaces, the magnitude function exhibits wild behaviour: singularities, negative magnitude, and so on. But for geometrically orthodox spaces such as subsets of Euclidean space, it turns out to be rather tame. This is because they belong to the important class of ‘positive definite’ spaces. Positive definiteness will play a central role when we come to extend the definition of magnitude from finite to infinite spaces. It is explored thoroughly in the paper of Meckes [31], who also describes its relationship with the classical notion of negative type.

The term metric space will be used in its standard sense, except that \( \infty \) is permitted as a distance. Many of our theorems do hold for the generalized metric spaces of Example 1.2.2(iii), with the same proofs; but to avoid cluttering the exposition, we leave it to the reader to discern which.

Throughout, we use matrices whose rows and columns are indexed by abstract finite sets (as in Section 1.1). The identity matrix is denoted by \( \delta \).

2.1 The magnitude of a finite metric space

We begin by restating the definitions from Section 1, without reference to enriched categories. Let \( A \) be a finite metric space. Its similarity matrix \( \zeta_A \) in \( \mathbb{R}^{A \times A} \) is defined by \( \zeta_A(a,b) = e^{-d(a,b)} \) (\( a, b \in A \)). A weighting on \( A \) is a function \( w: A \to \mathbb{R} \) such that \( \sum_b \zeta_A(a,b)w(b) = 1 \) for all \( a \in A \). The space \( A \) has magnitude if it admits at least one weighting; its magnitude is then \( |A| = \sum_a w(a) \) for any weighting \( w \), and is independent of the weighting chosen.
A finite metric space $A$ has Möbius inversion if $\zeta_A$ is invertible. Its Möbius matrix is then $\mu_A = \zeta_A^{-1}$. In that case, there is a unique weighting $w$ given by $w(a) = \sum_b \mu(a, b)$, and $|A| = \sum_{a,b} \mu_A(a, b)$ (Lemma 1.1.4). A generic real square matrix is invertible, which suggests that most finite metric spaces should have Möbius inversion. Proposition 2.2.6(i) makes this precise.

Here are some elementary examples.

**Examples 2.1.1**

i. The empty space has magnitude 0, and the one-point space has magnitude 1.

ii. Let $A$ be the space consisting of two points distance $d$ apart. Then

$$\zeta_A = \begin{pmatrix} 1 & e^{-d} \\ e^{-d} & 1 \end{pmatrix}.$$ 

This is invertible, so $A$ has Möbius inversion and its magnitude is the sum of all four entries of $\mu_A = \zeta_A^{-1}$:

$$|A| = 1 + \tanh(d/2)$$

(Fig. 1). This can be interpreted as follows. When $d$ is small, $A$ closely resembles a 1-point space; correspondingly, the magnitude is little more than 1. As $d$ grows, the points acquire increasingly separate identities and the magnitude increases. In the extreme, when $d = \infty$, the two points are entirely separate and the magnitude is 2.

iii. A metric space $A$ is discrete [20] if $d(a, b) = \infty$ for all $a \neq b$ in $A$. Let $A$ be a finite discrete space. Then $\zeta_A$ is the identity matrix $\delta$, each point has weight 1, and $|A| = \#A$.

The definition of the magnitude of a metric space first appeared in a paper of Solow and Polasky [44], although with almost no mathematical development. They called it the ‘effective number of species’, since the points of their spaces represented biological species and the distances represented inter-species differences (e.g. genetic). We can view the magnitude of a metric space as the...
'effective number of points'. Solow and Polasky also considered the magnitude of correlation matrices, making connections with the statistical concept of effective sample size.

Three-point spaces have magnitude; the formula follows from the proof of Proposition 2.4.15. Meckes [31, Theorem 3.6] has shown that four-point spaces have magnitude. But spaces with five or more points need not have magnitude (Example 2.2.7).

We now describe two classes of space for which the magnitude exists and is given by an explicit formula.

**Definition 2.1.2** A finite metric space $A$ is **scattered** if $d(a, b) > \log((\#A) - 1)$ for all distinct points $a$ and $b$. (Vacuously, the empty space and one-point space are scattered.)

**Proposition 2.1.3** A scattered space has magnitude. Indeed, any scattered space $A$ has Möbius inversion, with Möbius matrix given by the infinite sum

$$\mu_A(a, b) = \sum_{k=0}^{\infty} \sum_{a_a \neq \cdots \neq a_k = b} (-1)^k \zeta_A(a_0, a_1) \cdots \zeta_A(a_{k-1}, a_k).$$

The inner sum is over all $a_0, \ldots, a_k \in A$ such that $a_0 = a$, $a_k = b$, and $a_j \neq a_i$ whenever $1 \leq j \leq k$. That a scattered space has magnitude was also proved in [27, Theorem 2], by a different method that does not produce a formula for the Möbius matrix.

**Proof** Write $n = \#A$. For $a, b \in A$ and $k \geq 0$, put

$$\mu_{A,k}(a, b) = \sum_{a_{a_0 \neq \cdots \neq a_k = b}} \zeta_A(a_0, a_1) \cdots \zeta_A(a_{k-1}, a_k).$$

(In particular, $\mu_{A,0}$ is the identity matrix.) Write $\varepsilon = \min_{a \neq b} d(a, b)$. Then

$$\mu_{A,k+1}(a, b) = \sum_{b' : b' \neq b} \sum_{a_{a_0 \neq \cdots \neq a_k = b'}} \zeta_A(a_0, a_1) \cdots \zeta_A(a_{k-1}, b') \zeta_A(b', b)$$

$$\leq \sum_{b' : b' \neq b} \sum_{a_{a_0 \neq \cdots \neq a_k = b'}} \zeta_A(a_0, a_1) \cdots \zeta_A(a_{k-1}, b') e^{-\varepsilon}$$

$$= e^{-\varepsilon} \sum_{b' : b' \neq b} \mu_{A,k}(a, b').$$

The last sum is over $(n-1)$ terms, so by induction, $\mu_{A,k}(a, b) \leq ((n-1)e^{-\varepsilon})^k$ for all $a, b \in A$ and $k \geq 0$. But $A$ is scattered, so $(n-1)e^{-\varepsilon} < 1$, so the sum $\sum_{k=0}^{\infty} (-1)^k \mu_{A,k}(a, b)$ converges for all $a, b \in A$. A telescoping sum argument finishes the proof. \qed

**Definition 2.1.4** A metric space is **homogeneous** if its isometry group acts transitively on points.
Proposition 2.1.5 (Speyer [45]) Every homogeneous finite metric space has magnitude. Indeed, if $A$ is a homogeneous space with $n \geq 1$ points then

$$|A| = \frac{n^2}{\sum_{a,b} e^{-d(a,b)}} = \frac{n}{\sum_a e^{-d(x,a)}}$$

for any $x \in A$. There is a weighting $w$ on $A$ given by $w(a) = |A|/n$ for all $a \in A$.

**Proof** By homogeneity, the sum $S = \sum_a \zeta_A(x,a)$ is independent of $x \in A$. Hence there is a weighting $w$ given by $w(a) = 1/S$ for all $a \in A$. □

**Example 2.1.6** For any (undirected) graph $G$ and $t \in (0, \infty]$, there is a metric space $tG$ whose points are the vertices and whose distances are minimal path-lengths, a single edge having length $t$. Write $K_n$ for the complete graph on $n$ vertices. Then

$$|tK_n| = \frac{n}{1 + (n-1)e^{-t}}.$$

In general, $e^{-d(a,b)}$ can be interpreted as the similarity or closeness of the points $a, b \in A$ [26, 44]. Proposition 2.1.5 states that the magnitude of a homogeneous space is the reciprocal mean similarity.

**Example 2.1.7** A subspace can have greater magnitude than the whole space. Let $K_{n,m}$ be the graph with vertices $a_1, \ldots, a_n, b_1, \ldots, b_m$ and one edge between $a_i$ and $b_j$ for each $i$ and $j$. If $n$ is large then the mean similarity between two points of $tK_{n,n}$ is approximately $\frac{1}{2}(e^{-t} + e^{-2t})$ (Fig. 2). On the other hand, $tK_{n,n}$ has a subspace $2tK_n = \{a_1, \ldots, a_n\}$ in which the mean similarity is approximately $e^{-2t}$. Since $e^{-t} > e^{-2t}$, the mean similarity between points of $tK_{n,n}$ is greater than that of its subspace $2tK_n$; hence $|tK_{n,n}| < |2tK_n|$. In fact, it can be shown using Proposition 2.1.5 that $|tK_{n,n}| < |2tK_n|$ whenever $n > e^t + 1$. 
2.2 Magnitude functions

In physical situations, distance depends on the choice of unit of length; making a different choice rescales the metric by a constant factor. In the definition of $|x|$ as $e^{-x}$ (Example 1.3.1(iii)), the constant $e^{-1}$ was chosen without justification; choosing a different constant between 0 and 1 also amounts to rescaling the metric. For both these reasons, every metric space should be seen as a member of the one-parameter family of spaces obtained by rescaling it.

**Definition 2.2.1** Let $A$ be a metric space and $t \in (0, \infty)$. Then $tA$ denotes the metric space with the same points as $A$ and $d_{tA}(a, b) = td_A(a, b)$ ($a, b \in A$).

Most familiar invariants of metric spaces behave in a predictable way when the space is rescaled. This is true, for example, of topological invariants, diameter, and Hausdorff measure of any dimension. But magnitude does not behave predictably under rescaling. Graphing $|tA|$ against $t$ therefore gives more information about $A$ than is given by $|A|$ alone.

**Definition 2.2.2** Let $A$ be a finite metric space. The magnitude function of $A$ is the partially-defined function $t \mapsto |tA|$, defined for all $t \in (0, \infty)$ such that $tA$ has magnitude.

**Examples 2.2.3**

i. Let $A$ be the space consisting of two points distance $d$ apart. By Example 2.1.1(ii), the magnitude function of $A$ is defined everywhere and given by $t \mapsto 1 + \tanh(dt/2)$.

ii. Let $A = \{a_1, \ldots, a_n\}$ be a nonempty homogeneous space, and write $E_i = d(a_1, a_i)$. By Proposition 2.1.5, the magnitude function of $A$ is

$$t \mapsto n \sum_{i=1}^{n} e^{-E_i t}.$$  

In the terminology of statistical mechanics, the denominator is the partition function for the energies $E_i$ at inverse temperature $t$.

iii. Let $R$ be a finite commutative ring. For $a \in R$, write $\nu(a) = \min\{k \in \mathbb{N} : a^{k+1} = 0\} \in \mathbb{N} \cup \{\infty\}$.

There is a metric $d$ on $R$ given by $d(a, b) = \nu(b - a)$, and the resulting metric space $A_R$ is homogeneous. Write $q = e^{-t}$, and $\text{Nil}(R)$ for the ideal of nilpotent elements. By Proposition 2.1.5, $A_R$ has magnitude function

$$t \mapsto |tA_R| = \#R \sum_{a \in \text{Nil}(R)} q^{\nu(a)} = \#R/(1 - q) \sum_{k=0}^{\infty} \#\{a \in R : a^{k+1} = 0\} \cdot q^k$$

where the last expression is an element of the field $\mathbb{Q}(q)$ of formal Laurent series.

---

\[4\]I thank Simon Willerton for suggesting that some such relationship should exist.
To establish the basic properties of magnitude functions, we need some auxiliary definitions and a lemma. A vector \( v \in \mathbb{R}^I \) is positive if \( v(i) > 0 \) for all \( i \in I \), and nonnegative if \( v(i) \geq 0 \) for all \( i \in I \). Recall the definition of distance-decreasing map from Example 1.2.2(iii).

**Definition 2.2.4** A metric space \( A \) is an expansion of a metric space \( B \) if there exists a distance-decreasing surjection \( A \to B \).

**Lemma 2.2.5** Let \( A \) and \( B \) be finite metric spaces, each admitting a nonnegative weighting. If \( A \) is an expansion of \( B \) then \(|A| \geq |B|\).

**Proof** Take a distance-decreasing surjection \( f : A \to B \). Choose a right inverse function \( g : B \to A \) (not necessarily distance-decreasing). Then \( \zeta_B(f(a), b) \geq \zeta_A(a, g(b)) \) for all \( a \in A \) and \( b \in B \). Let \( w_A \) and \( w_B \) be nonnegative weightings on \( A \) and \( B \) respectively. Then

\[
|A| = \sum_{a,b} w_A(a) \zeta_B(f(a), b) w_B(b) \geq \sum_{a,b} w_A(a) \zeta_A(a, g(b)) w_B(b) = |B|,
\]

as required. \( \Box \)

**Proposition 2.2.6** Let \( A \) be a finite metric space. Then:

i. \( tA \) has Möbius inversion (and therefore magnitude) for all but finitely many \( t > 0 \).

ii. The magnitude function of \( A \) is analytic at all \( t > 0 \) such that \( tA \) has Möbius inversion.

iii. For \( t \gg 0 \), there is a unique, positive, weighting on \( tA \).

iv. For \( t \gg 0 \), the magnitude function of \( A \) is increasing.

v. \(|tA| \to \#A \) as \( t \to \infty \).

**Proof** We use the space \( \mathbb{R}^{A \times A} \) of real \( A \times A \) matrices, and its open subset \( \text{GL}(A) \) of invertible matrices. We also use the notions of weighting on, and magnitude of, a matrix (Section 1.1). For \( \zeta \in \text{GL}(A) \), the unique weighting \( w_\zeta \) on \( \zeta \) and the magnitude of \( \zeta \) are given by

\[
w_\zeta(a) = \sum_{b \in A} \zeta^{-1}(a, b) = \sum_{b \in A} (\text{adj} \, \zeta)(a, b) / \det \zeta, \quad |\zeta| = \sum_{a \in A} w_\zeta(a) \quad (1)
\]

\((a \in A)\), where \( \text{adj} \) denotes the adjugate.

For (i), first note that \( \zeta_{tA} \to \delta \in \text{GL}(A) \) as \( t \to \infty \); hence \( \zeta_{tA} \) is invertible for \( t \gg 0 \). The matrix \( \zeta_{tA} = (e^{-td(a,b)}) \) is defined for all \( t \in \mathbb{C} \), and \( \det \zeta_{tA} \) is analytic in \( t \). But \( \det \zeta_{tA} \neq 0 \) for real \( t > 0 \), so by analyticity, \( \det \zeta_{tA} \) has only finitely many zeros in \((0, \infty)\).

Part (ii) follows from equations (1).
For (iii), each of the functions $\zeta \mapsto w_\zeta(a)$ ($a \in A$) is continuous on $\text{GL}(A)$ by (1). But $w_\delta(a) = 1$ for all $a \in A$, so there is a neighbourhood $U$ of $\delta$ in $\text{GL}(A)$ such that $w_\zeta(a) > 0$ for all $\zeta \in U$ and $a \in A$. Since $\zeta_{tA} \to \delta$ as $t \to \infty$, we have $\zeta_{tA} \in U$ for all $t \gg 0$.

Part (iv) follows from part (iii) and Lemma 2.2.5.

For (v), $\lim_{t \to \infty} |tA| = |\lim_{t \to \infty} \zeta_{tA}| = |\delta| = \#A$. \hfill \Box

Part (i) implies that magnitude functions have only finitely many singularities. Proposition 2.4.17 will provide an explicit lower bound for parts (iii) and (iv).

Part (v) also appeared as Theorem 3 of [27].

Many natural conjectures about magnitude are disproved by the following example. Later we will see that subspaces of Euclidean space are less prone to surprising behaviour.\footnote{Our approach to general metric spaces bears the undeniable imprint of early exposure to Euclidean geometry. We just love spaces sharing a common feature with $\mathbb{R}^n$. (Gromov [11], page xvi.)}

**Example 2.2.7** Fig. 3 shows the magnitude function of the space $K_{3,2}$ defined in Example 2.1.7. It is given by

$$|tK_{3,2}| = \frac{5 - 7e^{-t}}{(1 + e^{-t})(1 - 2e^{-2t})} \quad (t \neq \log \sqrt{2});$$

the magnitude of $(\log \sqrt{2})K_{3,2}$ is undefined. (One can compute this directly or use Proposition 2.3.13.) Several features of the graph are apparent. At some scales, the magnitude is negative; at others, it is greater than the number of points. There are also intervals on which the magnitude function is strictly decreasing. Furthermore, this example shows that a space with magnitude can have a subspace without magnitude: for $(\log \sqrt{2})K_{3,2}$ is a subspace of $(\log \sqrt{2})K_{3,3}$, which, being homogeneous, has magnitude (Proposition 2.1.5).
The graph $K_{3,2}$ is also a well-known counterexample in the theory of spaces of negative type \[10\]. The connection is explained, in broad terms, by the remarks in Section 2.4.

The first example of a finite metric space with undefined magnitude was found by Tao \[48\], and had 6 points. The first examples of $n$-point spaces with magnitude outside the interval $[0,n]$ were found by the author and Simon Willerton, and were again 6-point spaces.

**Example 2.2.8** This is an example of a space $A$ for which $\lim_{t\to 0} |tA| \neq 1$, due to Willerton (personal communication, 2009). Let $A$ be the graph $K_{3,3}$ (Fig. 2) with three new edges adjoined: one from $b_i$ to $b_j$ whenever $1 \leq i < j \leq 3$. Then $|tA| = 6/(1 + 4e^{-t}) \to 6/5$ as $t \to 0$.

### 2.3 New spaces from old

For each way of constructing a new metric space from old, we can ask: is the magnitude of the new space determined by the magnitudes of the old ones? Here we give a positive answer for four constructions: unions of a particular type, tensor products, fibrations, and constant-distance gluing.

**Unions**

Let $X$ be a metric space with subspaces $A$ and $B$. The magnitude of $A \cup B$ is not in general determined by the magnitudes of $A$, $B$ and $A \cap B$: consider one-point spaces. In this respect, magnitude of metric spaces is unlike cardinality of sets, for which there is the inclusion-exclusion formula. We do, however, have an inclusion-exclusion formula for magnitude when the union is of a special type.

**Definition 2.3.1** Let $X$ be a metric space and $A, B \subseteq X$. Then $A$ projects to $B$ if for all $a \in A$ there exists $\pi(a) \in A \cap B$ such that for all $b \in B$,

$$d(a,b) = d(a,\pi(a)) + d(\pi(a),b).$$

In this situation, $d(a,\pi(a)) = \inf_{b \in B} d(a,b)$. If all distances in $X$ are finite then $\pi(a)$ is unique for $a$.

**Proposition 2.3.2** Let $X$ be a finite metric space and $A, B \subseteq X$. Suppose that $A$ projects to $B$ and $B$ projects to $A$. If $A$, $B$ and $A \cap B$ have magnitude then so does $A \cup B$, with

$$|A \cup B| = |A| + |B| - |A \cap B|.$$ 

Indeed, if $w_A$, $w_B$ and $w_{A\cap B}$ are weightings on $A$, $B$ and $A \cap B$ respectively then there is a weighting $w$ on $A \cup B$ defined by

$$w(x) = \begin{cases} 
  w_A(x) & \text{if } x \in A \setminus B \\
  w_B(x) & \text{if } x \in B \setminus A \\
  w_A(x) + w_B(x) - w_{A\cap B}(x) & \text{if } x \in A \cap B.
\end{cases}$$
Proof Let $a \in A \setminus B$. Choose a point $\pi(a)$ as in Definition 2.3.1. Then
\[
\sum_{x \in A \cup B} \zeta(a, x)w(x) = \sum_{a' \in A} \zeta(a, a')w_A(a') + \sum_{b \in B} \zeta(a, b)w_B(b) - \sum_{c \in A \cap B} \zeta(a, c)w_{A \cap B}(c)
\]
\[
= 1 + \zeta(a, \pi(a))\left\{\sum_{b \in B} \zeta(\pi(a), b)w_B(b) - \sum_{c \in A \cap B} \zeta(\pi(a), c)w_{A \cap B}(c)\right\} = 1.
\]

Similar arguments apply when we start with a point of $B \setminus A$ or $A \cap B$. This proves that $w$ is a weighting, and the result follows. 

It can similarly be shown that if $A$, $B$ and $A \cap B$ all have Möbius inversion then so does $A \cup B$. The proof is left to the reader; we just need the following special case.

Corollary 2.3.3 Let $X$ be a finite metric space and $A, B \subseteq X$. Suppose that $A \cap B$ is a singleton $\{c\}$, that for all $a \in A$ and $b \in B$,
\[
d(a, b) = d(a, c) + d(c, b),
\]
and that $A$ and $B$ have magnitude. Then $A \cup B$ has magnitude $|A| + |B| - 1$. Moreover, if $A$ and $B$ have Möbius inversion then so does $A \cup B$, with
\[
\mu_{A \cup B}(x, y) = \begin{cases} 
\mu_A(x, y) & \text{if } x, y \in A \text{ and } (x, y) \neq (c, c) \\
\mu_B(x, y) & \text{if } x, y \in B \text{ and } (x, y) \neq (c, c) \\
\mu_A(c, c) + \mu_B(c, c) - 1 & \text{if } (x, y) = (c, c) \\
0 & \text{otherwise.}
\end{cases}
\]

Proof The first statement follows from Proposition 2.3.2, and the second is easily checked. 

Corollary 2.3.4 Every finite subspace of $\mathbb{R}$ has Möbius inversion. If $A = \{a_0 < \cdots < a_n\} \subseteq \mathbb{R}$ then, writing $d_i = a_i - a_{i-1}$,
\[
|A| = 1 + \sum_{i=1}^n \tanh \frac{d_i}{2}.
\]
The weighting $w$ on $A$ is given by
\[
w(a_i) = \frac{1}{2} \left(\tanh \frac{d_i}{2} + \tanh \frac{d_{i+1}}{2}\right)
\]
$(0 \leq i \leq n)$, where by convention $d_0 = d_{n+1} = \infty$ and $\tanh \infty = 1$.

Proof This follows by induction from Example 2.1.1(ii), Proposition 2.3.2 and Corollary 2.3.3. (An alternative proof is given in [27, Theorem 4].) 

Documenta Mathematica 18 (2013) 857–905
Thus, in a finite subspace of $\mathbb{R}$, the weight of a point depends only on the distances to its neighbours. This is reminiscent of the Ising model in statistical mechanics [5], although whether there is a substantial connection remains to be seen.

**Example 2.3.5** The magnitude function is not a complete invariant of finite metric spaces. Indeed, let $X = \{0,1,2,3\} \subseteq \mathbb{R}$. Let $Y$ be the four-vertex Y-shaped graph, viewed as a metric space as in Example 2.1.6. I claim that $X$ and $Y$ have the same magnitude function, even though they are not isometric. For put $A = \{0,1,2\} \subseteq \mathbb{R}$ and $B = \{0,1\} \subseteq \mathbb{R}$. Both $tX$ and $tY$ can be expressed as unions, satisfying the hypotheses of Corollary 2.3.3, of isometric copies of $tA$ and $tB$. Hence $|tX| = |tA| + |tB| - 1 = |tY|$ for all $t > 0$.

**Tensor products**

Recall from Example 1.4.2(ii) the definition of the tensor product of metric spaces. Proposition 1.4.3 implies (and it is easy to prove directly):

**Proposition 2.3.6** If $A$ and $B$ are finite metric spaces with magnitude then $A \otimes B$ has magnitude, given by $|A \otimes B| = |A||B|$. □

**Example 2.3.7** Let $q$ be a prime power, and denote by $\mathbb{F}_q$ the field of $q$ elements metrized by $d(a,b) = 1$ whenever $a \neq b$. Then for $N \in \mathbb{N}$, the metric tensor product $\mathbb{F}_q^\otimes N$ is the set $\mathbb{F}_q^N$ with the Hamming metric. Its magnitude function is

$$t \mapsto |t\mathbb{F}_q|^N = \left( \frac{q}{1 + (q - 1)e^{-t}} \right)^N$$

by Example 2.1.6 and Proposition 2.3.6.

More generally, a linear code is a vector subspace $C$ of $\mathbb{F}_q^N$ [29]. Its (single-variable) weight enumerator is the polynomial $W_C(x) = \sum_{i=0}^{N} A_i(C)x^i \in \mathbb{Z}[x]$, where $A_i(C)$ is the number of elements of $C$ whose Hamming distance from 0 is $i$. Since $C$ is homogeneous, Proposition 2.1.5 implies that its magnitude function is

$$t \mapsto (#C)/W_C(e^{-t}).$$

The magnitude function of a linear code therefore carries the same, important, information as its weight enumerator.

Similarly, if $A$ and $B$ are finite metric spaces with magnitude then their co-product or distant union $A+B$ (Section 1.4) has magnitude $|A+B| = |A| + |B|$.

**Fibrations**

A fundamental property of the Euler characteristic of topological spaces is its behaviour with respect to fibrations. If a space $A$ is fibred over a connected base $B$, with fibre $F$, then under suitable hypotheses, $\chi(A) = \chi(B)\chi(F)$. 
The Magnitude of Metric Spaces

An analogous formula holds for the Euler characteristic of a fibred category (Proposition 2.8 of [21] and, in a different context, Theorem 7.7 of [7]). Apparently no general notion of fibration of enriched categories has yet been formulated. Nevertheless, we define here a notion of fibration of metric spaces sharing common features with the categorical and topological notions, and we prove an analogous theorem on magnitude.

**Definition 2.3.8** Let $A$ and $B$ be metric spaces. A (metric) fibration from $A$ to $B$ is a distance-decreasing map $p: A \to B$ with the following property (Fig. 4): for all $a \in A$ and $b' \in B$ with $d(p(a), b') < \infty$, there exists $a' \in p^{-1}(b')$ such that for all $a' \in p^{-1}(b')$, $d(a, a') = d(p(a), b') + d(a', b')$.

**Example 2.3.9** Let $C_t$ be the circle of circumference $t$, metrized non-symmetrically by taking $d(a, b)$ to be the length of the anticlockwise arc from $a$ to $b$. (This is a generalized metric space in the sense of Example 1.2.2(iii).) Let $k$ be a positive integer. Then the $k$-fold covering $C_{kt} \to C_t$, locally an isometry, is a fibration.

**Lemma 2.3.10** Let $p: A \to B$ be a fibration of metric spaces. Let $b, b' \in B$ with $d(b, b') < \infty$. Then the fibres $p^{-1}(b)$ and $p^{-1}(b')$ are isometric.

**Proof** Equation (2) and finiteness of $d(b, b')$ imply that $a_{b'}$ is unique for $a \in p^{-1}(b)$, so we may define a function $\gamma_{b, b'}: p^{-1}(b) \to p^{-1}(b')$ by $\gamma_{b, b'}(a) = a_{b'}$. It is distance-decreasing: for if $a, c \in p^{-1}(b)$ then
\[
d(b, b') + d(\gamma_{b, b'}(a), \gamma_{b, b'}(c)) = d(a, \gamma_{b, b'}(c)) \\
\leq d(a, c) + d(c, \gamma_{b, b'}(c)) = d(a, c) + d(b, b'),
\]
giving $d(\gamma_{b, b'}(a), \gamma_{b, b'}(c)) \leq d(a, c)$ by finiteness of $d(b, b')$.

There is a distance-decreasing map $\gamma_{b', b}: p^{-1}(b') \to p^{-1}(b)$ defined in the same way. It is readily shown that $\gamma_{b, b'}$ and $\gamma_{b', b}$ are mutually inverse; hence they are isometries. 

Figure 4: Metric fibration

---

Documenta Mathematica 18 (2013) 857–905
Let $B$ be a nonempty metric space all of whose distances are finite, and let $p: A \to B$ be a fibration. The fibre of $p$ is any of the spaces $p^{-1}(b)$ ($b \in B$); it is well-defined up to isometry.

Theorem 2.3.11 Let $p: A \to B$ be a fibration of finite metric spaces. Suppose that $B$ is nonempty with $d(b, b') < \infty$ for all $b, b' \in B$, and that $B$ and the fibre $F$ of $p$ both have magnitude. Then $A$ has magnitude, given by $|A| = |B||F|$.

Proof Choose a weighting $w_B$ on $B$. Choose, for each $b \in B$, a weighting $w_b$ on the space $p^{-1}(b)$. For $a \in A$, put $w_A(a) = w_{p(a)}(a)w_B(p(a))$. It is straightforward to check that $w_A$ is a weighting, and the theorem follows. □

Examples 2.3.12 i. A trivial example of a fibration is a product-projection $B \odot F \to B$. In that case, Theorem 2.3.11 reduces to Proposition 2.3.6.

ii. Let $B$ be a finite metric space in which the triangle inequality holds strictly for every triple of distinct points. Let $F$ be a finite metric space of small diameter:

$$\text{diam}(F) \leq \min\{d(b, b') + d(b', b'') - d(b, b'') : b, b', b'' \in B, b \neq b' \neq b''\}.$$ 

Choose for each $b, b' \in B$ an isometry $\gamma_{b, b'}: F \to F$, in such a way that $\gamma_{b, b}$ is the identity and $\gamma_{b', b} = \gamma_{b, b'}^{-1}$. Then the set $A = B \times F$ can be metrized by putting

$$d((b, c), (b', c')) = d(b, b') + d(\gamma_{b, b'}(c), c')$$

$(b, b' \in B$, $c, c' \in F)$. The projection $A \to B$ is a fibration (but not a product-projection unless $\gamma_{b', b'} \circ \gamma_{b, b'} = \gamma_{b', b}$ for all $b, b', b''$). So if $B$ and $F$ have magnitude, $|A| = |B||F|$.

Arguments similar to Lemma 2.3.10 show that a fibration over $B$ amounts to a family $(A_b)_{b \in B}$ of metric spaces together with a distance-decreasing map $\gamma_{b, b'}: A_b \to A_{b'}$ for each $b, b' \in B$ such that $d(b, b') < \infty$, satisfying the following three conditions. First, $\gamma_{b, b}$ is the identity for all $b \in B$. Second, $\gamma_{b', b} = \gamma_{b, b'}^{-1}$. Third,

$$\sup_{a \in A_b} d(\gamma_{b', b''} \gamma_{b, b'}(a), \gamma_{b, b'}(a)) \leq d(b, b') + d(b', b'') - d(b, b'')$$

for all $b, b', b'' \in B$ such that $d(b, b'), d(b', b'') < \infty$.

Constant-distance gluing

Given metric spaces $A$ and $B$ and a real number $D \geq \max\{\text{diam} A, \text{diam} B\}/2$, there is a metric space $A +_D B$ defined as follows. As a set, it is the disjoint union of $A$ and $B$. The metric restricted to $A$ is the original metric on $A$; similarly for $B$; and $d(a, b) = d(a, b) = D$ for all $a \in A$ and $b \in B$. 

Documenta Mathematica 18 (2013) 857–905
Proposition 2.3.13 Let $A$ and $B$ be finite metric spaces, and take $D$ as above. Suppose that $A$ and $B$ have magnitude, with $|A||B| \neq e^{2D}$. Then $A + D B$ has magnitude

$$|A| + |B| - 2e^{-D}|A||B|$$

$$1 - e^{-2D}|A||B|.$$ 

Proof Given weightings $w_A$ on $A$ and $w_B$ on $B$, there is a weighting $w$ on $A + D B$ defined by

$$w(a) = \frac{1 - e^{-D}|B|}{1 - e^{-2D}|A||B|} w_A(a), \quad w(b) = \frac{1 - e^{-D}|A|}{1 - e^{-2D}|A||B|} w_B(b)$$

$(a \in A, b \in B)$. The result follows. □

This provides an easy way to compute the magnitude functions in Examples 2.2.7 and 2.2.8.

2.4 Positive definite spaces

We saw in Example 2.2.7 that the magnitude of a finite metric space may be undefined, or smaller than the magnitude of one of its subspaces, or even negative. We now introduce a class of spaces for which no such behaviour occurs. Very many spaces of interest—including all subsets of Euclidean space—belong to this class. It has been studied in greater depth by Meckes [31].

Definition 2.4.1 A finite metric space $A$ is positive definite if the matrix $\zeta_A$ is positive definite.

We emphasize that positive definiteness of a matrix is meant in the strict sense.

Lemma 2.4.2

i. A positive definite space has Möbius inversion.

ii. The tensor product of positive definite spaces is positive definite.

iii. A subspace of a positive definite space is positive definite.

Proof Parts (i) and (iii) are elementary. For (ii), $\zeta_A \otimes \zeta_B$ is the Kronecker product $\zeta_A \otimes \zeta_B$, and the Kronecker product of positive definite matrices is positive definite. □

In particular, a positive definite space has magnitude and a unique weighting.

Proposition 2.4.3 Let $A$ be a positive definite finite metric space. Then

$$|A| = \sup_{v \neq 0} \frac{\left(\sum_{a \in A} v(a)\right)^2}{v^* \zeta_A v}$$

where the supremum is over $v \in \mathbb{R}^A \setminus \{0\}$ and $v^*$ denotes the transpose of $v$. A vector $v$ attains the supremum if and only if it is a nonzero scalar multiple of the unique weighting on $A$. 
Proof Since \( \zeta_A \) is positive definite, we have the Cauchy–Schwarz inequality:
\[
(v^* \zeta_A v) \cdot (w^* \zeta_A w) \geq (v^* \zeta_A w)^2
\]
for all \( v, w \in \mathbb{R}^A \), with equality if and only if one of \( v \) and \( w \) is a scalar multiple of the other. Taking \( w \) to be the unique weighting on \( A \) gives the result.

**Corollary 2.4.4** If \( A \) is a positive definite finite metric space and \( B \subseteq A \), then \(|B| \leq |A|\).

**Corollary 2.4.5** A nonempty positive definite finite metric space has magnitude \( \geq 1 \).

For any finite metric space \( A \), the set \( \text{Sing}(A) = \{ t \in (0, \infty) : \zeta_{tA} \text{ is singular} \} \) is finite (Proposition 2.2.6(i)). When \( \text{Sing}(A) = \emptyset \), put \( \sup(\text{Sing}(A)) = 0 \).

**Proposition 2.4.6** Let \( A \) be a finite metric space. Then \( tA \) is positive definite for all \( t > 0 \). In particular, \( tA \) is positive definite for all \( t > \sup(\text{Sing}(A)) \).

Proof Write \( \lambda_{\min}(\xi) \) for the minimum eigenvalue of a real symmetric \( A \times A \) matrix \( \xi \). Then \( \lambda_{\min}(\xi) \) is continuous in \( \xi \). Also \( \lambda_{\min}(\xi) > 0 \) if and only if \( \xi \) is positive definite, and if \( \lambda_{\min}(\xi) = 0 \) then \( \xi \) is singular.

Now \( \zeta_{tA} \to \delta \) as \( t \to \infty \), and \( \lambda_{\min}(\delta) = 1 \), so \( \lambda_{\min}(\zeta_{tA}) > 0 \) for all \( t > 0 \). On the other hand, \( \lambda_{\min}(\zeta_{tA}) \) is continuous and nonzero for \( t > \sup(\text{Sing}(A)) \). Hence \( \lambda_{\min}(\zeta_{tA}) > 0 \) for all \( t > \sup(\text{Sing}(A)) \).

It follows that a space with Möbius inversion at all scales also satisfies an apparently stronger condition.

**Definition 2.4.7** A finite metric space \( A \) is stably positive definite if \( tA \) is positive definite for all \( t > 0 \).

**Corollary 2.4.8** Let \( A \) be a finite metric space. Then \( tA \) has Möbius inversion for all \( t > 0 \) if and only if \( A \) is stably positive definite.

**Example 2.4.9** Let \( A \) be the space of Example 2.2.8. It is readily shown that \( tA \) has a unique weighting for all \( t > 0 \). By the remarks after Definition 1.1.3, \( tA \) has Möbius inversion for all \( t > 0 \), so \( A \) is stably positive definite. Hence magnitude is not continuous with respect to the Gromov–Hausdorff metric even when restricted to stably positive definite finite spaces. (Theorem 2.6 of [31] implies that it is, however, lower semicontinuous.)

Meckes [31, Theorem 3.3] has shown that a finite metric space is stably positive definite if and only if it is of negative type. By definition, a finite metric space \( A \) is of negative type if \( \sum_{a,b} v(a) d(a,b) v(b) \leq 0 \) for all \( v \in \mathbb{R}^A \) such that \( \sum_a v(a) = 0 \). A general metric space \( A \) is of negative type if every finite subspace is of negative type, or equivalently if \((A, \sqrt{d_A})\) embeds isometrically into some Hilbert space [43]. Many of the most commonly encountered spaces...
are of negative type, including those that we prove below to be stably positive definite; see [31, Theorem 3.6] for a list. But whereas the classical results on negative type typically rely on embedding theorems, we are able to prove our results directly.

Lemma 2.2.5 gave additional hypotheses on finite metric spaces $A$ and $B$ guaranteeing that if $A$ is an expansion of $B$ then $|A| \geq |B|$. Some additional hypotheses are needed, since not every magnitude function is increasing (Example 2.2.7). The following will also do.

**Lemma 2.4.10** Let $A$ and $B$ be finite metric spaces. Suppose that $A$ is positive definite and $B$ admits a nonnegative weighting. If $A$ is an expansion of $B$ then $|A| \geq |B|$.

**Proof** First consider a distance-decreasing bijection $f: A \rightarrow B$. Choose a nonnegative weighting $w_B$ on $B$. Without loss of generality, $f$ is the identity as a map of sets; thus, $\zeta_A(a, a') \leq \zeta_B(a, a')$ for all points $a, a'$. Hence

$$|A| \geq \frac{(\sum w_B(a))^2}{w_B^* \zeta_A w_B} \geq \frac{(\sum w_B(a))^2}{w_B^* \zeta_B w_B} = |B|,$$

by Proposition 2.4.3.

Now consider the general case of a distance-decreasing surjection from $A$ to $B$. We may choose a subspace $A' \subseteq A$ and a distance-decreasing bijection $A' \rightarrow B$. The space $A'$ is positive definite, so $|A'| \geq |B|$ by the previous argument; but also $|A| \geq |A'|$ by Corollary 2.4.4. □

A positive definite space cannot have negative magnitude, but the following example shows that it can have magnitude greater than the number of points.

**Example 2.4.11** Take the space $K_{3,2}$ of Example 2.2.7. It is easily shown that $\text{Sing}(K_{3,2}) = \{\log \sqrt{2}\}$. Choose $u > \log \sqrt{2}$ such that $|uK_{3,2}| > 5$ (say, $u = 0.35$); then $A = uK_{3,2}$ is positive definite by Proposition 2.4.6, and $|A| > \#A$. This example also shows that a positive definite expansion of a positive definite space may have smaller magnitude: for if $s > 1$ then $sA$ is an expansion of $A$, but $|sA| < |A|$ (Fig. 3).

A different positivity condition is sometimes useful: the existence of a nonnegative weighting.

**Lemma 2.4.12** Let $A$ be a finite metric space admitting a nonnegative weighting. Then $0 \leq |A| \leq \#A$.

**Proof** Choose a nonnegative weighting $w$ on $A$. For all $a \in A$ we have $0 \leq w(a) \leq (\zeta_A w)(a) = 1$, so $0 \leq w(a) \leq 1$. Summing, $0 \leq |A| \leq \#A$. □

We now list some sufficient conditions for a space to be positive definite, or have a positive weighting, or both.
Proposition 2.4.13 Every finite subspace of $\mathbb{R}$ is positive definite with positive weighting.

Proof Let us temporarily say that a finite metric space $A$ is good if it has M"obius inversion and for all $v \in \mathbb{R}^A$, 
\[ v^*\mu_A v \geq \max_{a \in A} v(a)^2. \]

I claim that if $A \cup B$ is a union of the type in Corollary 2.3.3 and $A$ and $B$ are both good, then $A \cup B$ is good. Indeed, let $v \in \mathbb{R}^{A \cup B}$. By Corollary 2.3.3,
\[ v^*\mu_{A \cup B} v = v^*\mu_A v |_A + v^*\mu_B v |_B - v(c)^2 \]
where $v |_A$ is the restriction of $v$ to $A$. Now let $x \in A \cup B$. Without loss of generality, $x \in A$. Since $A$ is good, $v^*\mu_A v |_A \geq v(x)^2$. Since $B$ is good, $v^*\mu_B v |_B \geq v(c)^2$. Hence $v^*\mu_{A \cup B} v \geq v(x)^2$, proving the claim.

Every metric space with 0, 1 or 2 points is good. Every finite subset of $\mathbb{R}$ with 3 or more points can be expressed nontrivially as a union of the type in Corollary 2.3.3. It follows by induction that every finite subset of $\mathbb{R}$ is good and therefore positive definite.

Positivity of the weighting is immediate from Corollary 2.3.4.

□

For $N \in \mathbb{N}$ and $1 \leq p \leq \infty$, write $\ell_p^N = \mathbb{R}^{\otimes_p N}$, where $\otimes_p$ is as defined in Example 1.4.2(ii). Thus, $\ell_p^N$ is $\mathbb{R}^N$ with the metric induced by the $p$-norm, $\|x\|_p = (\sum_r |x_r|_p)^{1/p}$.

Theorem 2.4.14 Every finite subspace of $\ell_1^N$ is positive definite.

Proof Let $A$ be a finite subspace of $\ell_1^N$. Write $pr_1, \ldots, pr_N : \ell_1^N \to \mathbb{R}$ for the projections. Each space $pr_i A$ is positive definite by Proposition 2.4.13, so $\prod_{i=1}^N pr_i A \subseteq \ell_1^N$ is positive definite by Lemma 2.4.2(ii), so $A$ is positive definite by Lemma 2.4.2(iii).

□

We prove the same result for Euclidean space in the next section. In the category of metric spaces and distance-decreasing maps (Example 1.2.2(iii)), the categorical product $\times$ is $\otimes_{\infty}$. The class of positive definite spaces is not closed under $\times$. For if it were then, by an argument similar to the proof of Theorem 2.4.14, every finite subspace of $\ell_\infty^N$ would be positive definite. But in fact, every finite metric space embeds isometrically into $\ell_\infty^N$ for some $N$ ([43], p.535), whereas not every finite metric space is positive definite. Comprehensive results on (non-)preservation of positive definiteness by the products $\otimes_p$ have been proved by Meckes [31, Section 3.2].

Proposition 2.4.15 Every space with 3 or fewer points is positive definite with positive weighting.
The Magnitude of Metric Spaces

885

Proof The proposition is trivial for spaces with 2 or fewer points. Now take a 3-point space \( A = \{a_1, a_2, a_3\} \), writing \( Z_{ij} = \zeta(a_i, a_j) \). We use Sylvester’s criterion: a symmetric real \( n \times n \) matrix is positive definite if and only if the upper-left \( m \times m \) submatrix has positive determinant whenever \( 1 \leq m \leq n \). This holds for \( Z \) when \( m = 1 \) or \( m = 2 \), and

\[
\det Z = (1 - Z_{12})(1 - Z_{23})(1 - Z_{31}) + (1 - Z_{12})(Z_{12} - Z_{13}Z_{32}) + (1 - Z_{23})(Z_{23} - Z_{21}Z_{13}) + (1 - Z_{31})(Z_{31} - Z_{32}Z_{21})
\]

which is positive by the triangle inequality. The unique weighting is \( v / \det Z \), where

\[
v_1 = (1 - Z_{12})(1 - Z_{23})(1 - Z_{31}) + (1 - Z_{23})(Z_{23} - Z_{21}Z_{13}) > 0
\]

and similarly \( v_2 \) and \( v_3 \). □

Meckes [31, Theorem 3.6] has shown that 4-point spaces are also positive definite. By Example 2.2.7, his result is optimal.

Example 2.4.16 The weighting on a 4-point space may have negative components, as may the weighting on a finite subspace of \( \ell^N_1 \). Indeed, using Proposition 2.3.2 one can show that in the space \( \{ (0, 0), (t, 0), (0, t), (-t, 0) \} \subseteq \ell^2_1 \), the weight at \( (0, 0) \) is negative whenever \( t < \log 2 \).

Every finite metric space, when scaled up sufficiently, becomes positive definite with positive weighting (Propositions 2.2.6 and 2.4.6). The following result provides an alternative, quantitative proof, using the notion of scattered space (Definition 2.1.2).

Proposition 2.4.17 Every scattered space is positive definite with positive weighting.

Proof Let \( A \) be a scattered space with \( n \geq 2 \) points. Positive definiteness follows from a version of the Levy–Desplanques theorem (Theorem 6.1.10 of [13]), but since the argument is simple, we repeat it here. Let \( v \in \mathbb{R}^A \). Then

\[
v^* \zeta_A v = \sum_a v(a)^2 + \sum_{a \neq b} v(a) \zeta_A(a, b) v(b) \geq \sum_a v(a)^2 - \frac{1}{n - 1} \sum_{a \neq b} |v(a)||v(b)|
\]

\[
= \frac{1}{2(n - 1)} \sum_{a \neq b} (|v(a)| - |v(b)|)^2 \geq 0.
\]

The inequality \( \zeta_A(a, b) < 1/(n - 1) \) \((a \neq b)\) is strict, so if \( v^* \zeta_A v = 0 \) then \( v = 0 \). To show that the unique weighting \( w_A \) on \( A \) is positive, we use the proof of Proposition 2.1.3. There we showed that \( A \) has Möbius inversion and that the Möbius matrix is a sum \( \mu_A = \sum_{k=0}^{\infty} (-1)^k \mu_{A,k} \), where the matrices \( \mu_{A,k} \) satisfy

\[
\mu_{A,k+1}(a, b) < \frac{1}{n - 1} \sum_{b' : b' \neq b} \mu_{A,k}(a, b')
\]
for all $a,b$. Hence $w_A = \sum_{k=0}^{\infty} (-1)^k w_{A,k}$, where $w_{A,k}(a) = \sum_b \mu_{A,k}(a,b)$. Summing (3) over all $b \in A$ gives

$$w_{A,k+1}(a) < \frac{1}{n-1} \sum_{b,b' : b' \neq b} \mu_{A,k}(a,b') = w_{A,k}(a)$$

($a \in A$). Hence $w_A(a) = \sum_{k=0}^{\infty} (-1)^k w_{A,k}(a) > 0$ for all $a \in A$. □

A metric space $A$ is ultrametric if $\max\{d(a,b), d(b,c)\} \geq d(a,c)$ for all $a,b,c \in A$.

**Proposition 2.4.18** Every finite ultrametric space is positive definite with positive weighting.

Positive definiteness was proved by Varga and Nabben [49], and positivity of the weighting (rather indirectly) by Pavoine, Ollier and Pontier [38]. Another proof of positive definiteness is given by Meckes [31, Theorem 3.6]. Both parts of the following proof are different from those cited.

**Proof** Let $\Omega$ be the set of symmetric matrices $Z$ over $[0, \infty)$ such that $Z_{ik} \geq \min\{Z_{ij}, Z_{jk}\}$ for all $i,j,k$ and $Z_{ii} > \max_{j \neq k} Z_{jk}$ for all $i$. (For a $1 \times 1$ matrix, this maximum is to be interpreted as 0.) We show by induction that every matrix in $\Omega$ is positive definite and that its unique weighting (Definition 1.1.1) is positive. The proposition will follow immediately.

The result is trivial for $0 \times 0$ and $1 \times 1$ matrices. Now let $Z \in \Omega$ be an $n \times n$ matrix with $n \geq 2$. Put $z = \min_{i,j} Z_{ij}$. There is an equivalence relation $\sim$ on $\{1, \ldots, n\}$ defined by $i \sim j$ if and only if $Z_{ij} > z$.

It is not the case that $i \sim j$ for all $i,j$. Hence we may partition $\{1, \ldots, n\}$ into two nonempty subsets that are each a union of equivalence classes: say $\{1, \ldots, m\}$ and $\{m+1, \ldots, n\}$. We have $Z_{ij} = z$ whenever $i \leq m < j$, so $Z$ is a block sum

$$Z = \begin{pmatrix} Z' & zU_{n-m}^m \\ zU_{n-m}^m & Z'' \end{pmatrix}$$

where $U_k^\ell$ denotes the $k \times \ell$ matrix all of whose entries are 1. Since $Z' \in \Omega$ and $Z'_{ij} = Z_{ij} \geq z$ for all $i,j \leq m$, we have $Y' = Z' - zU_m^m \in \Omega$. Similarly, $Y'' = Z'' - zU_{n-m}^{n-m} \in \Omega$, and

$$Z = zU_n^n + \begin{pmatrix} Y' & 0 \\ 0 & Y'' \end{pmatrix}.$$ 

The first summand is positive semidefinite. By inductive hypothesis, $Y'$ and $Y''$ are positive definite, so the second summand is positive definite. Hence $Z$ is positive definite.

Also by inductive hypothesis, $Y'$ and $Y''$ have positive weightings $v'$ and $v''$ respectively. Let $v$ be the concatenation of $v'$ and $v''$. It is straightforward to verify that

$$\frac{v}{z(|Y'| + |Y''|) + 1}$$
The Magnitude of Metric Spaces

is a weighting on \( Z \), and it is positive since \( v' \) and \( v'' \) are positive and \(|z|, |Y'|, |Y''| \geq 0\). □

**Corollary 2.4.19** If \( A \) is a finite ultrametric space then \(|A| \leq e^{\text{diam } A}\).

**Proof** Let \( \Delta \) be the metric space with the same point-set as \( A \) and \( d(a, b) = \text{diam } A \) for all distinct points \( a, b \). By Proposition 2.1.5, \(|\Delta| \leq e^{\text{diam } A} \) and \( \Delta \) has a positive weighting. But \( \Delta \) is an expansion of \( A \), so \(|A| \leq |\Delta| \) by Lemma 2.2.5. □

A homogeneous space always has a positive weighting, by Proposition 2.1.5. However, Example 2.1.7 and Corollary 2.4.4 together show that a homogeneous space need not be positive definite. A homogeneous space need not even have Möbius inversion: \((\log 2)K_{3,3}\) is an example. In particular, a finite metric space may have magnitude but not Möbius inversion.

Magnitude can be understood in terms of entropy or diversity. For every finite metric space \( A \) and \( q \in [0, \infty) \), there is a function \( qD^A \) assigning to each probability distribution \( p \) on \( A \) a real number \( qD^A(p) \), the diversity of order \( q \) of the distribution [26]. An ecological community can be modelled as a finite metric space (as explained in Section 2.1) together with a probability distribution \( p \) on \( A \) (representing the relative abundances of the species). Then \( qD^A(p) \) is a measure of the biodiversity of the community. In the special case that \( A \) is discrete, the diversities are the exponentials of the Rényi entropies [39], and in particular, the diversity of order 1 is the exponential of the Shannon entropy.

It is a theorem [23] that for each finite metric space \( A \), there is some probability distribution \( p \) maximizing \( qD^A(p) \) for all \( q \in [0, \infty] \) simultaneously. Moreover, the maximal value of \( qD^A(p) \) is independent of \( q \); call it \( D_{\text{max}}(A) \). If \( A \) is positive definite with nonnegative weighting then, in fact, \(|A| = D_{\text{max}}(A)\); magnitude is maximum diversity.

2.5 Subsets of Euclidean space

Here we show that every finite subspace of Euclidean space \( \ell_2^N \) is positive definite. In particular, every such space has well-defined magnitude.

Write \( L_1(\mathbb{R}^N) \) for the space of Lebesgue-integrable complex-valued functions on \( \mathbb{R}^N \). Define the Fourier transform \( \hat{f} \) of \( f \in L_1(\mathbb{R}^N) \) by

\[ \hat{f}(\xi) = \int_{\mathbb{R}^N} e^{-2\pi i \langle \xi, x \rangle} f(x) \, dx \]

(\( \xi \in \mathbb{R}^N \)). Define functions \( g \) and \( \psi \) on \( \mathbb{R}^N \) by

\[ g(x) = e^{-\|x\|^2}, \quad \psi(\xi) = C_N / (1 + 4\pi^2\|\xi\|^2)^{(N+1)/2} \]

where \( C_N \) is the constant \( 2^N \pi^{(N-1)/2} \Gamma((N + 1)/2) > 0 \).

**Lemma 2.5.1** \( \psi \in L_1(\mathbb{R}^N) \) and \( \hat{\psi} = g \).
The first statement is straightforward. Theorem 1.14 of [47] states that \( \hat{g} = \psi \); but \( g \) is continuous and even, so the second statement follows by Fourier inversion.

The next lemma is elementary and standard (e.g. [15], Theorem VI.2.8).

**Lemma 2.5.2** Let \( \phi \in L_1(\mathbb{R}^N) \), let \( A \) be a finite subset of \( \mathbb{R}^N \), and let \( v \in \mathbb{R}^A \). Then

\[
\sum_{a, b \in A} v(a)\hat{\phi}(a - b)v(b) = \int_{\mathbb{R}^N} \left| \sum_{a \in A} v(a)e^{-2\pi i \langle \xi, a \rangle} \right|^2 \phi(\xi) d\xi.
\]

In analytic language, our task is to show that the function \( g \) is strictly positive definite. This would follow from the easy half of Bochner’s theorem [15], except that Bochner’s theorem concerns non-strict positive definiteness. We therefore need to refine the argument slightly.

**Theorem 2.5.3** Every finite subspace of Euclidean space is positive definite.

**Proof** Let \( A \) be a finite subspace of \( \ell_1^N \). Let \( v \in \mathbb{R}^A \). Then

\[
v^*\zeta_A v = \sum_{a, b \in A} v(a)g(a - b)v(b) = \int_{\mathbb{R}^N} \left| \sum_{a \in A} v(a)e^{-2\pi i \langle \xi, a \rangle} \right|^2 \psi(\xi) d\xi \geq 0
\]

by Lemmas 2.5.1 and 2.5.2. Suppose that \( v \neq 0 \). The characters \( e^{-2\pi i \langle \cdot, a \rangle} \) \( (a \in A) \) are linearly independent, so the squared term is positive (that is, strictly positive) for some \( \xi \in \mathbb{R}^N \). By continuity, the squared term is positive for all \( \xi \) in some nonempty open subset of \( \mathbb{R}^N \). Moreover, \( \psi \) is continuous and everywhere positive. So the integral is positive, as required.

On the other hand, some of the weights on a finite subspace of Euclidean space can be negative; see Willerton [50] for examples.

**Corollary 2.5.4** Every finite subspace of Euclidean space has magnitude.

A similar argument gives an alternative proof of Theorem 2.4.14, that finite subspaces of \( \ell_1^N \) are positive definite. For this we use the explicit formula for the Fourier transform of \( x \mapsto e^{-\|x\|_1} \). For \( p \neq 1,2 \) there is no known formula for the Fourier transform of \( e^{-\|x\|^p_p} \), so matters become more difficult. Nevertheless, Meckes [31, Section 3] has shown that every finite subspace of \( \ell_p^N \) is positive definite whenever \( 0 < p \leq 2 \), and that this is false for \( p > 2 \).
3 Compact metric spaces

To extend the notion of magnitude from finite to infinite spaces, there are broadly speaking two strategies. In the first, we approximate an infinite space by finite spaces. As an initial attempt, given a compact metric space $A$, we might take a sequence $(A_k)$ of finite metric spaces converging to $A$ in the Gromov–Hausdorff metric, and try to define $|A|$ as the limit of the sequence $(|A_k|)$. However, this ‘definition’ is inconsistent; recall Example 2.2.8. We might respond by constraining the sequence $(A_k)$: for example, by taking $(A_k)$ to be a sequence of subsets of $A$ converging to $A$ in the Hausdorff metric.

The second strategy is to work directly with the infinite space, replacing finite sums by integrals. Weightings are now measures, or perhaps distributions. For example, a weight measure on a metric space $A$ is a finite signed Borel measure $w$ such that $\int_A e^{-d(a,b)} \, dw(b) = 1$ for all $a \in A$. If $A$ admits a weight measure $w$ then an argument similar to Lemma 1.1.2 shows that $w(A)$ is independent of the choice of $w$, and we may define the magnitude of $A$ to be $w(A)$. This was the definition used by Willerton in [51].

Meckes [31] has shown that to a large extent, these different approaches produce the same result. Here we implement the first strategy, defining the magnitude of a space to be the supremum of the magnitudes of its finite subspaces. This works well when the space is compact and its finite subspaces are positive definite.

3.1 The magnitude of a positive definite compact metric space

Definition 3.1.1 A metric space is positive definite if every finite subspace is positive definite. The magnitude of a compact positive definite space $A$ is

$$|A| = \sup \{|B| : B \text{ is a finite subspace of } A\} \in [0, \infty].$$

These definitions are consistent with the definitions for finite metric spaces, by Lemma 2.4.2(iii) and Corollary 2.4.4.

There may even be non-compact spaces for which this definition of magnitude is sensible. For example, let $t > 0$, and let $A$ be a space with infinitely many points and $d(a,b) = t$ for all $a \neq b$; then every finite subspace of $A$ is positive definite, and the supremum of their magnitudes is $e^t < \infty$. In any case, we confine ourselves to compact spaces.

A metric space $A$ is stably positive definite if $tA$ is positive definite for all $t > 0$, or equivalently if every finite subspace of $A$ is stably positive definite. (A further equivalent condition, due to Meckes, is that $A$ is of negative type [31, Theorem 3.3].) We already know that $\ell^1_N$ and $\ell^2_N$ are stably positive definite; much of the rest of this paper concerns the magnitudes of their compact subspaces. Ultrametric spaces are also stably positive definite (Proposition 2.4.18), and, if compact, have finite magnitude (Corollary 2.4.19). Many other com-
monly occurring spaces, such as hyperbolic space, are stably positive definite too; see [31, Theorem 3.6].

**Definition 3.1.2** Let $A$ be a stably positive definite compact metric space. The **magnitude function** of $A$ is the function

$$(0,\infty) \rightarrow [0,\infty)$$

$t \rightarrow |tA|$.

**Lemma 3.1.3** Let $A$ be a positive definite compact metric space. Then:

i. Every closed subspace $B$ of $A$ is positive definite, and $|B| \leq |A|$.

ii. If $A$ is nonempty then $|A| \geq 1$. \hfill \Box

**Proposition 3.1.4** Let $A$ and $B$ be positive definite compact spaces. Then

$A \otimes B$ is positive definite and compact, and

$|A \otimes B| = |A||B|$.

In the case $A = \emptyset$ and $|B| = \infty$, we interpret $0 \cdot \infty$ as $0$.

**Proof** Let $C$ be a finite subspace of $A \otimes B$. Then $C \subseteq A' \otimes B'$ for some finite subspaces $A' \subseteq A$ and $B' \subseteq B$. Since $A$ and $B$ are positive definite, so are $A'$ and $B'$. By Lemma 2.4.2, $A' \otimes B'$ is positive definite, so $C$ is positive definite. Hence $A \otimes B$ is positive definite. A similar argument shows that $|A \otimes B| = |A||B|$, using Proposition 2.3.6 and Corollary 2.4.4. \hfill \Box

Similarly, Proposition 2.3.2 on unions extends to the compact setting.

**Proposition 3.1.5** Let $X$ be a metric space and $A,B \subseteq X$, with $A$ and $B$ compact and $A \cup B$ positive definite. Suppose that $A$ projects to $B$ and $B$ projects to $A$. Then

$|A \cup B| + |A \cap B| = |A| + |B|$.

**Proof** Let $\varepsilon > 0$. Choose finite sets $E \subseteq A \cup B$ and $H \subseteq A \cap B$ such that $|A \cup B| \leq |E| + \varepsilon$ and $|A \cap B| \leq |H| + \varepsilon$. For each $a \in E \cap A$, choose $\pi_A(a) \in A \cap B$ satisfying the condition of Definition 2.3.1, and similarly $\pi_B(b)$ for $b \in E \cap B$. Put

$H' = H \cup \pi_A(E \cap A) \cup \pi_B(E \cap B), \quad F = (E \cap A) \cup H', \quad G = (E \cap B) \cup H'$.

Then $F$ and $G$ are finite subsets of $X$, each projecting to the other. Also $E \subseteq F \cup G$ and $H \subseteq F \cap G$. Applying Proposition 2.3.2 to $F$ and $G$ gives $|A \cup B| + |A \cap B| \leq |A| + |B| + 2\varepsilon$. Since $\varepsilon$ was arbitrary, $|A \cup B| + |A \cap B| \leq |A| + |B|$. For the opposite inequality, again let $\varepsilon > 0$, and choose finite sets $F \subseteq A$ and $G \subseteq B$ such that $|A| \leq |F| + \varepsilon$ and $|B| \leq |G| + \varepsilon$. For each $a \in F$, choose...
π_A(a) ∈ A ∩ B satisfying the condition of Definition 2.3.1, and similarly π_B(b) for b ∈ G. Put

F' = F ∪ π_AF ∪ π_BG,  \quad G' = G ∪ π_AF ∪ π_BG.

Then F" and G' are finite subsets of X, each projecting to the other; also F ⊆ F' ⊆ A and G ⊆ G' ⊆ B. A similar argument proves that |A| + |B| ≤ |A ∪ B| + |A ∩ B| + 2ε. □

3.2 Subsets of the real line

As soon as we ask about the magnitude of real intervals, connections with geometric measure begin to appear.

**Proposition 3.2.1** Let t ≥ 0 and let (A_k) be a sequence of finite subsets of \( \mathbb{R} \) converging to [0, t] in the Hausdorff metric. Then (|A_k|) converges to \( 1 + \frac{t}{2} \).

This result was announced in [22], and also appears, with a different proof, as Proposition 6 of [27].

**Proof** Given \( A = \{a_0 < \cdots < a_n\} \subseteq \mathbb{R} \), we have

\[
(1 + t/2) - |A| = \sum_{i=1}^{n} \left\{ \frac{a_i - a_{i-1}}{2} - \tanh\left(\frac{a_i - a_{i-1}}{2}\right) \right\} + \frac{t - (a_n - a_0)}{2}
\]

by Corollary 2.3.4. The result will follow from the facts that tanh(0) = 0 and tanh'(0) = 1. Indeed, write \( f(x) = (x - \tanh(x))/x \), so that \( f(x) \to 0 \) as \( x \to 0 \). Then

\[
|(1 + t/2) - |A|| \leq \max_{1 \leq i \leq n} \left| f\left(\frac{a_i - a_{i-1}}{2}\right) \right| + \left| \frac{t - (a_n - a_0)}{2} \right|.
\]

But max\(_i\)\((a_i - a_{i-1})\) → 0 and \( a_n - a_0 \to t \) as \( A \to [0, t] \), proving the proposition. □

**Theorem 3.2.2** The magnitude of a closed interval [0, t] is \( 1 + t/2 \).

**Proof** Proposition 3.2.1 immediately implies that \( ||[0, t]|| \geq 1 + t/2 \). Now let A be a finite subset of [0, t]. We may choose a sequence (A_k) of finite subsets of \( \mathbb{R} \) such that \( \lim_{k \to \infty} A_k = [0, t] \) and \( A \subseteq A_k \) for all k. Then \( |A| \leq |A_k| \to 1 + t/2 \) as \( k \to \infty \), so \( |A| \leq 1 + t/2 \) as \( k \to \infty \), so \( |A| \leq 1 + t/2 \). □

Schanuel [41] argued from basic geometric intuition that the ‘size’ of a closed interval of length \( t \) inches ought to be \( (t \text{ inches} + 1) \). Ignoring the factor of \( 1/2 \) (which is purely a product of convention), Theorem 3.2.2 makes his idea rigorous.

As noted by Willerton [51], there is a weight measure on [0, t]. It is \( w = (\delta_0 + \lambda + \delta_t)/2 \), where \( \delta_x \) is the Dirac measure at \( x \) and \( \lambda \) is Lebesgue measure.
on $[0, t]$. Then $w([0, t]) = 1 + t/2$. Hence $w([0, t]) = [[0, t]]$, as guaranteed by Theorems 2.3 and 2.4 of Meckes [31].

The magnitude of subsets of $\mathbb{R}$ is also described by the following formula, which has no known analogue in higher dimensions.

**Proposition 3.2.3** Let $A$ be a compact subspace of $\mathbb{R}$. Then

$$|A| = \frac{1}{2} \int_{\mathbb{R}} \text{sech}^2 d(x, A) \, dx$$

where $d(x, A) = \inf_{a \in A} d(x, a)$.

**Proof** First we prove the identity for finite spaces $A \subseteq \mathbb{R}$, by induction on $n = \# A$. It is elementary when $n \leq 2$. Now suppose that $n \geq 3$, writing the points of $A$ as $a_1 < \cdots < a_n$. Put $B = \{a_1, \ldots, a_{n-1}\}$ and $C = \{a_{n-1}, a_n\}$. Then

$$\frac{1}{2} \int_{\mathbb{R}} \text{sech}^2 d(x, A) \, dx = \frac{1}{2} \int_{-\infty}^{a_{n-1}} \text{sech}^2 d(x, B) \, dx + \frac{1}{2} \int_{a_{n-1}}^{\infty} \text{sech}^2 d(x, C) \, dx.$$

Since $\int_0^\infty \text{sech}^2 u \, du = 1$, this in turn is equal to

$$\frac{1}{2} \left( \int_{\mathbb{R}} \text{sech}^2 d(x, B) \, dx - 1 \right) + \frac{1}{2} \left( \int_{\mathbb{R}} \text{sech}^2 d(x, C) \, dx - 1 \right)$$

which by inductive hypothesis is $|B| + |C| - 1$. On the other hand, $|A| = |B| + |C| - 1$ by Corollary 2.3.3. This completes the induction.

Now take a compact space $A \subseteq \mathbb{R}$. We know that

$$|A| = \sup \left\{ \frac{1}{2} \int_{\mathbb{R}} \text{sech}^2 d(x, B) \, dx : B \text{ is a finite subset of } A \right\}.$$

Since $\text{sech}^2$ is decreasing on $[0, \infty)$, this implies that

$$|A| \leq \frac{1}{2} \int_{\mathbb{R}} \text{sech}^2 d(x, A) \, dx.$$

To prove the opposite inequality, choose a sequence $(B_k)$ of finite subsets of $A$ converging to $A$ in the Hausdorff metric. We have $0 \leq \text{sech}^2 d(x, B_k) \leq \text{sech}^2 d(x, A)$ for all $x$ and $k$, so

$$\lim_{k \to \infty} \int_{\mathbb{R}} \text{sech}^2 d(x, B_k) \, dx = \int_{\mathbb{R}} \text{sech}^2 d(x, A) \, dx$$

by the dominated convergence theorem. The result follows. \qed

\textbf{Documenta Mathematica 18 (2013) 857–905}
3.3 Background on integral geometry

To go further, we need some concepts and results from integral geometry. Those concerning $\ell^N_2$ can be found in standard texts such as [17]. Those concerning $\ell^N_1$ can be found in [24].

Write $K^N$ for the set of compact convex subsets of $\mathbb{R}^N$. A valuation on $K^N$ is a function $\phi: K^N \to \mathbb{R}$ such that

$$\phi(\emptyset) = 0, \quad \phi(A \cup B) = \phi(A) + \phi(B) - \phi(A \cap B)$$

whenever $A, B, A \cup B \in K^N$. It is continuous if continuous with respect to the Hausdorff metric on $K^N$, and invariant if $\phi(gA) = \phi(A)$ for all $A \in K^N$ and isometries $g: \ell^N_2 \to \ell^N_2$ (not necessarily fixing the origin).

**Examples 3.3.1**

i. $N$-dimensional Lebesgue measure is a continuous invariant valuation on $K^N$, denoted by Vol.

ii. Euler characteristic $\chi$ is a continuous invariant valuation on $K^N$. Since the sets are convex, $\chi(A)$ is 0 or 1 according as $A$ is empty or not.

The continuous invariant valuations on $K^N$ form a real vector space, $\text{Val}_N$. When $A \subseteq \ell^N_p$ and $t > 0$, the abstract metric space $tA$ can be interpreted as the subspace $\{ta : a \in A\}$ of $\ell^N_p$. A valuation $\phi$ is homogeneous of degree $i$ if $\phi(tA) = t^i \phi(A)$ for all $A \in K^N$ and $t > 0$.

**Theorem 3.3.2 (Hadamard [12])** The vector space $\text{Val}_N$ has dimension $N + 1$ and a basis $V_0, \ldots, V_N$ where $V_i$ is homogeneous of degree $i$. □

This description determines the valuations $V_i$ uniquely up to scale factor. They can be uniquely normalized to satisfy two conditions. First, $V_N(A) = \text{Vol}(A)$ for $A \in K^N$. Second, whenever $\ell^N_2$ is embedded isometrically into $\ell^{N+1}_2$ and $0 \leq i \leq N$, the value $V_i(A)$ is the same whether $A$ is regarded as a subset of $\ell^N_2$ or of $\ell^{N+1}_2$. With this normalization, $V_i$ is called the $i$th intrinsic volume. For example, $V_0 = \chi$. When $A \in K_2$, $V_1(A)$ is half of the perimeter of $A$; when $A \in K_3$, $V_2(A)$ is half of the surface area.

Here is a general formula for the intrinsic volumes. For each $0 \leq i \leq N$, there is an $O(N)$-invariant measure $\nu_{N,i}$ on the Grassmannian $\text{Gr}_{N,i}$, unique up to scale factor. Given $P \in \text{Gr}_{N,i}$, write $\pi_P: \mathbb{R}^N \to P$ for orthogonal projection. Then for $A \in K_N$,

$$V_i(A) = c_{N,i} \int_{\text{Gr}_{N,i}} \text{Vol}(\pi_P A) d\nu_{N,i}(P)$$

where $c_{N,i}$ is a positive constant chosen so that the normalizing conditions are satisfied.

Hadamard’s theorem solves the classification problem for valuations on $\ell^N_2$. More generally, we can try to classify the valuations on any metric space, in the following sense.
A metric space \( A \) is geodesic \([37]\) if for all \( a, b \in A \) there exists an isometry \( \gamma: [0, d(a, b)] \to A \) with \( \gamma(0) = a \) and \( \gamma(d(a, b)) = b \). Given a metric space \( X \), write \( \mathcal{K}(X) \) for the set of compact subsets of \( X \) that are geodesic with respect to the subspace metric. For example, \( \mathcal{K}(\ell^N_2) = \mathcal{K}_N \).

A valuation on \( \mathcal{K}(X) \) is a function \( \phi: \mathcal{K}(X) \to \mathbb{R} \) satisfying equations (4) whenever \( A, B, A \cup B, A \cap B \in \mathcal{K}(X) \). It is continuous if continuous with respect to the Hausdorff metric, and invariant if \( \phi(gA) = \phi(A) \) for all isometries \( g \) of \( X \). Write \( \text{Val}(X) \) for the vector space of continuous invariant valuations on \( \mathcal{K}(X) \). For example, \( \text{Val}(\ell^N_2) = \text{Val}_N \).

Given any metric space \( X \), one can attempt to describe the vector space \( \text{Val}(X) \). Here we will need to know the answer for \( \ell^N_1 \), as well as \( \ell^N_2 \). To state it, we write \( \mathcal{K}'_N = \mathcal{K}(\ell^N_1) \) and call its elements compact \( \ell_1 \)-convex sets; similarly, we write \( \text{Val}'_N = \text{Val}(\ell^N_1) \).

There are far more \( \ell_1 \)-convex sets than convex sets. On the other hand, there are far fewer isometries of \( \ell^N_1 \) than of \( \ell^N_2 \); they are generated by translations, coordinate permutations, and reflections in coordinate hyperplanes. The following Hadwiger-type theorem is proved in \([24]\).

**Theorem 3.3.3** The vector space \( \text{Val}'_N \) has dimension \( N + 1 \) and a basis \( V'_0, \ldots, V'_N \) where \( V'_i \) is homogeneous of degree \( i \). \( \square \)

Again, this determines the valuations \( V'_i \) uniquely up to scaling. They can be described as follows. For \( 0 \leq i \leq N \), let \( \text{Gr}'_N,i \) be the set of \( i \)-dimensional vector subspaces of \( \mathbb{R}^N \) spanned by some subset of the standard basis. For \( A \in \mathcal{K}'_N \), put

\[
V'_i(A) = \sum_{P \in \text{Gr}'_N,i} \text{Vol}(\pi_P A).
\]

These valuations \( V'_0, \ldots, V'_N \) are called the \( \ell_1 \)-intrinsic volumes, and satisfy two normalization conditions analogous to those in the Euclidean case.

The intrinsic volumes of a product space are given by the following formula, proved in \([24, \text{Proposition } 8.1]\) and precisely analogous to the classical Euclidean formula \([17, \text{Theorem } 9.7.1]\).

**Proposition 3.3.4** Let \( A \in \mathcal{K}'_M \) and \( B \in \mathcal{K}'_N \). Then \( A \times B \in \mathcal{K}'_{M+N} \), and

\[
V'_k(A \times B) = \sum_{i+j=k} V'_i(A) V'_j(B)
\]

whenever \( 0 \leq k \leq M + N \). \( \square \)

### 3.4 Subsets of \( \ell^N_1 \)

Our investigation of the magnitude of subsets of \( \ell^N_1 \) begins with sets of a particularly amenable type.
Definition 3.4.1 A cuboid in $\ell^1_N$ is a subspace of the form $[x_1, y_1] \times \cdots \times [x_N, y_N]$, where $x_r, y_r \in \mathbb{R}$ with $x_r \leq y_r$.

As an abstract metric space, a cuboid is a tensor product $[x_1, y_1] \otimes \cdots \otimes [x_N, y_N]$.

Theorem 3.4.2 For cuboids $A \subseteq \ell^1_N$,

$$|A| = \sum_{i=0}^{N} 2^{-i} V'_i(A).$$

(5)

Proof First let $I = [x, y] \subseteq \mathbb{R}$ be a nonempty interval. By Theorem 3.2.2,

$$|I| = 1 + (y - x)/2 = \chi(I) + \text{Vol}(I)/2 = V'_0(I) + 2^{-1} V'_1(I).$$

This proves the theorem for $N = 1$. The theorem also holds for $N = 0$. It now suffices to show that if $A \in \mathcal{K}'_M$ and $B \in \mathcal{K}'_N$ satisfy (5) then so does $A \times B \in \mathcal{K}'_{M+N}$. Indeed, as a metric space, $A \times B \subseteq \ell^1_{M+N}$ is $A \otimes B$, and the result follows from Propositions 3.1.4 and 3.3.4. □

In fact, $V'_i(\prod [x_r, y_r])$ is the $i$th elementary symmetric polynomial in $(y_r - x_r)_{r=1}^N$, again by Proposition 3.3.4. It is also equal to $V_i(\prod [x_r, y_r])$, the Euclidean intrinsic volume. But in general, the Euclidean and $\ell_1$-intrinsic volumes of a convex set are not equal.

Corollary 3.4.3 The magnitude function of a cuboid $A \subseteq \ell^1_N$ is given by

$$|tA| = \sum_{i=0}^{N} 2^{-i} V'_i(A)t^i.$$ 

In particular, the magnitude function of a cuboid $A$ is a polynomial whose degree is the dimension of $A$, and whose coefficients are proportional to the $\ell_1$-intrinsic volumes of $A$. □

The moral is that for spaces belonging to this small class, the dimension and all of the $\ell_1$-intrinsic volumes can be recovered from the magnitude function. In this sense, magnitude encodes those invariants. For the rest of this work we advance the conjectural principle—first set out in [27]—that the same is true for a much larger class of spaces, in both $\ell^1_N$ and $\ell^2_N$.

We begin by showing that the principle holds for subspaces of $\ell^1_N$ when the invariant concerned is dimension.

Definition 3.4.4 The growth of a function $f: (0, \infty) \to \mathbb{R}$ is

$$\inf\{\nu \in \mathbb{R} : f(t)/t^\nu \text{ is bounded for } t \gg 0\} \in [-\infty, \infty].$$

For example, the growth of a polynomial is its degree.
Definition 3.4.5 The \textit{(magnitude) dimension} \( \dim A \) of a stably positive definite compact metric space \( A \) is the growth of its magnitude function.

Examples 3.4.6 i. The magnitude dimension of a cuboid in \( \ell^N_1 \) is its dimension in the usual sense, by Corollary 3.4.3.

ii. The magnitude dimension of a nonempty finite space is 0, by Proposition 2.2.6(v).

Lemma 3.4.7 Let \( A \) be a stably positive definite compact space. Then:

1. Every closed subspace \( B \subseteq A \) satisfies \( \dim B \leq \dim A \).

2. If \( A \neq \emptyset \) then \( \dim A \geq 0 \).

Proof For (i), we have \( 0 \leq |tB| \leq |tA| \) for all \( t > 0 \), so \( \dim B \leq \dim A \).

For (ii), take \( B \) to be a one-point subspace of \( A \). \( \square \)

Recall that the magnitude of a positive definite compact space can in principle be infinite (although there are no known examples).

Theorem 3.4.8 Let \( A \) be a compact subset of \( \ell^N_1 \). Then:

1. \( |A| < \infty \).

2. \( \dim A \leq N \), with equality if \( A \) has nonempty interior.

We will show in Theorem 3.5.8 that the hypothesis 'nonempty interior' can be relaxed to 'nonzero measure'.

Proof \( A \) is a subset of some cuboid \( B \subseteq \ell^N_1 \), which has finite magnitude by Theorem 3.4.2, so \( |A| \leq |B| < \infty \). Also \( \dim A \leq \dim B \leq N \) by Lemma 3.4.7 and Example 3.4.6(i). If \( A \) has nonempty interior then it contains an \( N \)-dimensional cuboid, giving \( \dim A \geq N \). \( \square \)

We now ask whether the \( \ell_1 \)-intrinsic volumes of an \( \ell_1 \)-convex set can be extracted from its magnitude function.

Let \( \mathcal{C}_N \) be the smallest class of compact subsets of \( \ell^N_1 \) containing all cuboids and closed under unions of the type in Proposition 3.1.5. By that proposition and Theorem 3.4.2, equation (5) holds for all \( A \in \mathcal{C}_N \).

Example 3.4.9 Let \( T \) be a solid, compact triangle in \( \ell^2_1 \) with two edges parallel to the coordinate axes. We compute \( |T| \) by exhaustion. For each \( k \geq 1 \), let \( I_k \) be the union of \( k \) rectangles approximating \( T \) from the inside as in Fig. 5; similarly, let \( E_k \) be the exterior approximation by \( k \) rectangles. Then \( T, I_k \).
and $E_k$ are all $\ell_1$-convex with $I_k, E_k \in \mathcal{C}_2$, and $\lim_{k \to \infty} I_k = T = \lim_{k \to \infty} E_k$, so

$$\lim_{k \to \infty} |I_k| = \lim_{k \to \infty} \sum_{i=0}^{2} 2^{-i}V'_i(I_k) = \sum_{i=0}^{2} 2^{-i}V'_i(T)$$

$$= \lim_{k \to \infty} \sum_{i=0}^{2} 2^{-i}V'_i(E_k) = \lim_{k \to \infty} |E_k|$$

(using continuity of $V'_i$ in the second and third equalities). But $|I_k| \leq |T| \leq |E_k|$ for all $k$, so $|T| = \sum_{i=0}^{2} 2^{-i}V'_i(T)$. Similar arguments prove this identity for all compact convex polygons in $\ell_1^2$.

These and other examples suggest the following conjecture.

**Conjecture 3.4.10** Let $A$ be a compact $\ell_1$-convex subspace of $\ell_1^N$. Then

$$|A| = \sum_{i=0}^{N} 2^{-i}V'_i(A).$$

If the conjecture holds then $|tA| = \sum_{i=0}^{N} 2^{-i}V'_i(A)t^i$ for all $t > 0$ and $A \in \mathcal{K}'_N$. Hence we can recover all of the $\ell_1$-intrinsic volumes of an $\ell_1$-convex set from its magnitude function.

### 3.5 Subsets of Euclidean space

We now prove results for $\ell_2^N$ similar to some of those for $\ell_1^N$. Our first task is to prove that the magnitude of a compact subset of Euclidean space is finite. Given $A \subseteq \mathbb{R}^N$, write

$$\mathcal{S}(A) = \{\text{Schwartz functions } \phi: \mathbb{R}^N \to \mathbb{R} \text{ such that } \hat{\phi}(a-b) = 1 \text{ for all } a, b \in A\}.$$ 

**Lemma 3.5.1** Let $A \subseteq \mathbb{R}^N$ be a bounded set. Then $\mathcal{S}(A) \neq \emptyset$. 

**References**

Documenta Mathematica 18 (2013) 857–905
Proof. Since $A$ is bounded, there is a real even Schwartz function $f$ such that $f(a-b) = 1$ for all $a, b \in A$; then there is a unique real Schwartz function $\phi$ such that $\hat{\phi} = f$. □

The rest of the proof uses the function $\psi$ from Section 2.5. For a Schwartz function $\phi$ on $\mathbb{R}^N$, write

$$c(\phi) = \sup_{\xi \in \mathbb{R}^N} \left| \frac{\phi(\xi)}{\psi(\xi)} \right| < \infty.$$ 

Lemma 3.5.2 Let $A$ be a compact subspace of $\ell_2^N$ and $\phi \in \mathcal{S}(A)$. Then $|A| \leq c(\phi)$.

Proof. Let $B$ be a finite subset of $A$; thus, $B$ is positive definite by Theorem 2.5.3. Then for all $v \in \mathbb{R}^B$, using Lemma 2.5.2 twice,

$$c(\phi) \cdot v^* \zeta_B v = c(\phi) \int_{\mathbb{R}^N} \left| \sum_{a \in B} v(a) e^{-2\pi i \langle \xi, a \rangle} \right|^2 \psi(\xi) d\xi$$

$$\geq \int_{\mathbb{R}^N} \left| \sum_{a \in B} v(a) e^{-2\pi i \langle \xi, a \rangle} \right|^2 \phi(\xi) d\xi$$

$$= \sum_{a,b \in B} v(a) \hat{\phi}(a-b)v(b) = \left( \sum_{a \in B} v(a) \right)^2.$$

Taking $v$ to be the unique weighting on $B$ gives $c(\phi) \geq |B|$. □

Proposition 3.5.3 The magnitude of a compact subspace of $\ell_2^N$ is finite. □

We can extract more from the argument. For a compact set $A \subseteq \mathbb{R}^N$, write

$$\langle A \rangle = \inf \{ c(\phi) : \phi \in \mathcal{S}(A) \} < \infty.$$

Lemma 3.5.2 states that $|A| \leq \langle A \rangle$.

Lemma 3.5.4 Let $A$ be a compact subset of $\mathbb{R}^N$ and $t \geq 1$. Then $\langle tA \rangle \leq t^N \langle A \rangle$.

Proof. Let $\phi \in \mathcal{S}(A)$. Define $\theta : \mathbb{R}^N \to \mathbb{R}$ by $\theta(\xi) = t^N \phi(t\xi)$. Then $\theta$ is Schwartz, and if $a, b \in tA$ then $\hat{\theta}(a-b) = \hat{\phi}((a-b)/t) = 1$. Hence $\theta \in \mathcal{S}(tA)$.

I now claim that $c(\theta) \leq t^N c(\phi)$. Indeed, using the fact that $\psi(t\xi) \geq \psi(\xi)$ for all $\xi \in \mathbb{R}^N$,

$$c(\theta) = t^N \sup_{\xi \in \mathbb{R}^N} \left| \frac{\phi(t\xi)}{\psi(t\xi)} \right| \leq t^N \sup_{\xi \in \mathbb{R}^N} \left| \frac{\phi(\xi)}{\psi(\xi)} \right| = t^N c(\phi).$$

This proves the claim, and the result follows. □

Theorem 3.5.5 A compact subspace of $\ell_2^N$ has dimension at most $N$. 

*Documenta Mathematica* 18 (2013) 857–905
The Magnitude of Metric Spaces

Theorem 3.5.6 Let \( \| \cdot \| \) be a norm on \( \mathbb{R}^N \) whose induced metric is positive definite. Write \( B = \{ x \in \mathbb{R}^N : \|x\| \leq 1 \} \). For a compact set \( A \subseteq \mathbb{R}^N \), equipped with the subspace metric,
\[
|A| \geq \frac{\text{Vol}(A)}{N! \text{Vol}(B)}.
\]

Before proving this, we state some consequences. Write \( \omega_N \) for the volume of the unit Euclidean \( N \)-ball.

Corollary 3.5.7 Let \( A \) be a compact subset of \( \mathbb{R}^N \).

i. If \( A \) is given the subspace metric from \( \ell_2^N \) then \( |A| \geq \text{Vol}(A)/N! \omega_N \).

ii. If \( A \) is given the subspace metric from \( \ell_1^N \) then \( |A| \geq \text{Vol}(A)/2^N \).

Proof Part (i) is immediate. Part (ii) follows from the fact that the unit ball in \( \ell_1^N \) has volume \( 2^N/N! \), or can be derived from Lemma 3.5.9 below.

Theorem 3.5.8 Let \( p \in \{1,2\} \) and let \( A \) be a compact subspace of \( \ell_p^N \). Then
\[
\dim A \leq N, \quad \text{with equality if } A \text{ has nonzero Lebesgue measure.}
\]

Proof The inequality follows from Theorems 3.4.8 and 3.5.5. Now suppose that \( \text{Vol}(A) > 0 \). By Corollary 3.5.7, there is a constant \( K_A > 0 \) such that \( |tA| \geq K_A t^N \) for all \( t > 0 \), giving \( \dim A \geq N \).

Generalizations of these theorems have been proved by Meckes, using more sophisticated methods ([31], Theorems 4.4 and 4.5). In particular, Theorem 3.5.8 is extended to \( \ell_p^N \) for all \( p \in [1,2] \).

To prove Theorem 3.5.6, we first need a standard calculation.

Lemma 3.5.9 Let \( \| \cdot \| \) be a norm on \( \mathbb{R}^N \). Write \( B \) for the unit ball. Then
\[
\int_{\mathbb{R}^N} e^{-\|x\|} \, dx = N! \text{Vol}(B).
\]

Proof
\[
\int_{\mathbb{R}^N} e^{-\|x\|} \, dx = \int_0^\infty e^{-r} \, d(\text{Vol}(rB)) = \int_0^\infty e^{-r} N r^{N-1} \text{Vol}(B) \, dr = N! \text{Vol}(B).
\]

\( \square \)
Proof of Theorem 3.5.6 We use the result of Meckes [31, Theorem 2.4] that for a compact positive definite space $A$ and a finite Borel measure $v$ on $A$,

$$|A| \geq v(A)^2 \int_A \int_A e^{-d(a,b)} \, dv(a) \, dv(b).$$

Let $A \subseteq \mathbb{R}^N$ be a compact set and take $v$ to be Lebesgue measure: then

$$|A| \geq \text{Vol}(A)^2 \int_A \int_A e^{-\|a-b\|} \, da \, db \geq \frac{\text{Vol}(A)^2}{\int_A \int_{\mathbb{R}^N} e^{-\|a-b\|} \, da \, db} \int_A \int_{\mathbb{R}^N} e^{-\|a\|} \, da \, db.$$

The theorem follows from Lemma 3.5.9.

This proof is a rigorous rendition of part of Willerton’s bulk approximation argument [50]. There is an alternative proof in the same spirit, not depending on the results of Meckes but instead working with finite approximations. We sketch it now.

Alternative proof of Theorem 3.5.6 For $\delta > 0$, write

$$S_\delta = \left\{ x \in \delta \mathbb{Z}^N : A \cap \prod_{r=1}^N [x_r, x_r + \delta) \neq \emptyset \right\}.$$ 

Define $\alpha : \delta \mathbb{Z}^N \to \mathbb{R}^N$ by choosing for each $x \in S_\delta$ an element $\alpha(x) \in A \cap \prod_{r=1}^N [x_r, x_r + \delta)$, and putting $\alpha(x) = x$ for $x \in \delta \mathbb{Z}^N \setminus S_\delta$. A calculation similar to that in the first proof of Theorem 3.5.6 shows that for all $\delta > 0$,

$$|A| \geq \frac{\#S_\delta}{\sum_{x \in \delta \mathbb{Z}^N} E_\delta(x)}$$

where

$$E_\delta(x) = \frac{1}{\#S_\delta} \sum_{y \in S_\delta} e^{-\|\alpha(x+y) - \alpha(y)\|} \approx e^{-\|x\|}.$$

(Apply Proposition 2.4.3 to the finite space $\alpha S_\delta$.) Since Lebesgue measure is outer regular, $\lim_{\delta \to 0} (\delta^N (\#S_\delta)) = \text{Vol}(A)$. From the fact that $\|\alpha(x) - x\| \leq \text{diam}(0, \delta)^N$ for all $\delta > 0$ and $x \in \delta \mathbb{Z}^N$, it also follows that

$$\lim_{\delta \to 0} \left( \delta^N \sum_{x \in \delta \mathbb{Z}^N} E_\delta(x) \right) = \int_{\mathbb{R}^N} e^{-\|x\|} \, dx.$$

The theorem now follows from Lemma 3.5.9.

These results suggest the following conjecture, first stated in [27]:

Documenta Mathematica 18 (2013) 857–905
Conjecture 3.5.10  Let $A$ be a compact convex subspace of $\ell_2^N$. Then

$$|A| = \sum_{i=0}^{N} \frac{1}{i!\omega_i} V_i(A).$$

Assuming the conjecture, the magnitude function of a compact convex set $A \subseteq \ell_2^N$ is a polynomial:

$$|tA| = \sum_{i=0}^{N} \frac{1}{i!\omega_i} V_i(A)t^i. \quad (6)$$

All of the intrinsic volumes, as well as the dimension, can therefore be recovered from the magnitude function.

The evidence for Conjecture 3.5.10 is as follows.

- The two sides of equation (6) have the same growth (by Theorem 3.5.8).
- The left-hand side of (6) is greater than or equal to the leading term of the right-hand side (by Corollary 3.5.7).
- Both sides of (6) are monotone increasing in $A$ (by Lemma 3.1.3).
- The conjecture holds for $N = 1$ (by Theorem 3.2.2).
- It is closely analogous to Conjecture 3.4.10, which, while itself a conjecture, is known to hold for a nontrivial class of examples. (To see the analogy, note that in both cases the $i$th coefficient is $1/i!\text{Vol}(B_i)$, where $B_i$ is the $i$-dimensional unit ball.)
- There is good numerical evidence, due to Willerton [50], when $A$ is a disk, square or cube.

One strategy for proving Conjecture 3.5.10 would be to apply Hadwiger’s theorem (3.3.2). There are currently two obstacles. First, it is not known that magnitude is a valuation on compact convex sets. Certainly it is not a valuation on all compact subsets of $\ell_2^N$: consider the union of two points. Second, even if we knew that magnitude was a valuation on convex sets, the conjecture would still not be proved. We would know that magnitude was an invariant valuation, monotone and therefore continuous by Theorem 8 of McMullen [30]. By Hadwiger’s theorem, there would be constants $c_i$ such that $|A| = \sum c_i V_i(A)$ for all convex sets $A$. However, current techniques provide no way of computing those constants. Knowing the magnitude of balls or cubes would be enough; but apart from subsets of the line, there is not a single convex subset of Euclidean space whose magnitude is known.

References


*Documenta Mathematica* 18 (2013) 857–905


Convergence of Voevodsky’s Slice Tower

Marc Levine

Received: May 7, 2012
Revised: March 3, 2013
Communicated by Alexander Merkurjev

Abstract. We consider Voevodsky’s slice tower for a finite spectrum $\mathcal{E}$ in the motivic stable homotopy category over a perfect field $k$. In case $k$ has finite cohomological dimension, we show that the slice tower converges, in that the induced filtration on the bi-graded homotopy sheaves $\Pi_{a,b}f_n\mathcal{E}$ is finite, exhaustive and separated at each stalk (after inverting the exponential characteristic of $k$). This partially verifies a conjecture of Voevodsky.

2010 Mathematics Subject Classification: Primary 14F42; Secondary 55P42
Keywords and Phrases: Morel-Voevodsky stable homotopy category, slice filtration, motivic homotopy theory

Contents

Introduction 908
1. Background and notation 911
2. Voevodsky’s slice tower 914
3. The homotopy coniveau tower 917
4. The simplicial filtration 919
5. The bottom of the filtration 923
6. Finite spectra and cohomologically finite spectra 925
7. The proof of the convergence theorem 930
Appendix A. Norm maps 933
Appendix B. Inverting integers in a triangulated category 936
References 939

1Research supported by the Alexander von Humboldt Foundation
We continue our investigation, begun in [22], of the slice filtration on the bigraded homotopy sheaves $\Pi_{a,b}(E)$ for objects $E$ in the motivic stable homotopy category $\mathcal{SH}(k)$. We refer the reader to [23] for the notation to be used in this introduction.

Let $k$ be a perfect field, let $\mathcal{SH}(k)$ denote Voevodsky’s motivic stable homotopy category of $T$-spectra over $k$, $\mathcal{SH}$ the classical stable homotopy category of spectra. For a spectrum $E \in \mathcal{SH}$, the Postnikov tower of $E$,

$$\ldots \rightarrow E^{(n+1)} \rightarrow E^{(n)} \rightarrow \ldots \rightarrow E$$

consists of the $n-1$-connected covers $E^{(n)} \rightarrow E$ of $E$, that is, $\pi_mE^{(n)} \rightarrow \pi_mE$ is an isomorphism for $m \geq n$ and $\pi_mE^{(n)} = 0$ for $m < n$. Sending $E$ to $E^{(n)}$ defines a functor from $\mathcal{SH}$ to the full subcategory $\Sigma^n\mathcal{SH}^{eff}$ of $n-1$-connected spectra that is right adjoint to the inclusion $\Sigma^n\mathcal{SH}^{eff} \rightarrow \mathcal{SH}$.

Replacing $\Sigma^n\mathcal{SH}^{eff}$ with a certain triangulated subcategory $\Sigma^n_1\mathcal{SH}^{eff}(k)$ of $\mathcal{SH}(k)$ that measures a kind of “$\mathbb{P}^1$-connectedness” (in a suitable sense, see [30, 31, 23, 24] or §2 of this paper), Voevodsky has defined a motivic analog $\Sigma^n_1\mathcal{SH}^{eff}$ of the Postnikov tower; for an object $E$ of $\mathcal{SH}(k)$ this yields the Tate-Postnikov tower (or slice tower)

$$\ldots \rightarrow f_{n+1}E \rightarrow f_nE \rightarrow \ldots \rightarrow E$$

for $E$. For integers $a, b$, we have the stable homotopy sheaf $\Pi_{a,b}(E)$, defined as the Nisnevich sheaf associated to the presheaf

$$U \in \text{Sm}/k \mapsto [\Sigma^\infty_{T_1} \Sigma^b_{\mathbb{G}_m} \Sigma^\infty_{T_1} U_+, E]|_{\mathcal{SH}(k)}$$

and the Tate-Postnikov tower for $E$ gives rise to the filtration

$$\text{Fil}_T^n \Pi_{a,b}(E) := \text{im}(\Pi_{a,b} f_nE \rightarrow \Pi_{a,b}E).$$

Let $\mathcal{SH}_{\text{fin}}(k) \subset \mathcal{SH}(k)$ be the thick subcategory of $\mathcal{SH}(k)$ generated by the objects $\Sigma^\infty_{T_1}X_+$, with $X$ smooth and projective over $k$, $n \in \mathbb{Z}$. For example, the motivic sphere spectrum $\mathbb{S}_k := \Sigma^\infty_\mathbb{P}\text{Spec } k$ is in $\mathcal{SH}_{\text{fin}}(k)$.

Voevodsky has stated the following conjecture:

**Conjecture 1** ([29, conjecture 13]). Let $k$ be a perfect field. Then for $E \in \mathcal{SH}_{\text{fin}}(k)$, the Tate-Postnikov tower of $E$ is convergent in the following sense: for all $a, b, m \in \mathbb{Z}$, one has

$$\cap_n F^n_T \Pi_{a,b} f_mE = 0.$$  

The cases $E = \Sigma^4_{\mathbb{G}_m} \mathbb{S}_k$, $a = m = 0$ gives some evidence for this conjecture, as we shall now explain.

For $k$ a perfect field, Morel has given a natural isomorphism of $\Pi_{0,-p}(\mathbb{S}_k)$ with the Milnor-Witt sheaf $K^M_{p}k$; this is a certain sheaf on $\text{Sm}/k$ with value on each field $F$ over $k$ given by the Milnor-Witt group $K^M_{p}F$. For $F$ a field, $K^M_pF$ is canonically isomorphic to the Grothendieck-Witt group $GW(F)$

---

1A presentation of the graded ring $K^M_pF$ may be found in [14, definition 3.1].

---

2Marc Levine
of non-degenerate symmetric bilinear forms over $F$ \cite[lemma 3.10]{[14]}. More generally, Morel has constructed a natural isomorphism\footnote{This follows from \cite[theorem 6.13, theorem 6.40]{[14]}, using the argument of \cite[theorem 6.43]{[14]}.} for $p, q \in \mathbb{Z}$

$$\Pi_{0,p}(\Sigma^q_{\text{gm}} S_k) \cong k_{q-p}.$$ 

The isomorphism $K^{\text{MW}}_M(F) \cong GW(F)$ makes $K^{\text{MW}}_M(F)$ a $GW(F)$-module; let $I(F) \subset GW(F)$ denote the augmentation ideal. Our main result of loc. cit. is

**Theorem 2** (\cite[theorem 1]{[12]}). Let $F$ be a perfect field extension of $k$ of characteristic $\neq 2$. Then

$$\text{Fil}^p_{\text{Tate}} \Pi_{0,p}(\Sigma^q_{\text{gm}} S_k)(F) = I(F)^m K^{\text{MW}}_{q-p}(F) \subset K^{\text{MW}}_{q-p}(F) = \Pi_{0,p}(\Sigma^q_{\text{gm}} S_k)(F)$$

where $M = 0$ if $n \leq p$ or $n \leq q$, and $M = \min(n-p, n-q)$ if $n \geq p$ and $n \geq q$.

The following consequence of theorem \cite{[2]} gives some evidence for Voevodsky’s convergence conjecture:

**Proposition 3.** Let $k$ be a perfect field with $\text{char.} k \neq 2$. For all $p, q \geq 0$, and all perfect field extensions $F$ of $k$, we have

$$\cap_n \text{Fil}^p_{\text{Tate}} \Pi_{0,p}(\Sigma^q_{\text{gm}} S_k)(F) = 0.$$ 

**Proof.** In light of theorem \cite{[2]} this is asserting that the $I(F)$-adic filtration on $K^{\text{MW}}_{q-p}(F)$ is separated. By \cite[théorème 5.3]{[16]}, for $m \geq 0$, $K^{m}_{m}(F)$ fits into a cartesian square of $GW(F)$-modules

$$
\begin{array}{ccc}
K^{\text{MW}}_m(F) & \rightarrow & K^M_m(F) \\
\downarrow & & \downarrow \text{Pf} \\
I(F)^m & \rightarrow & I(F)^m/I(F)^{m+1},
\end{array}
$$

where $K^M_m(F)$ is the Milnor $K$-group, $q$ is the quotient map and $\text{Pf}$ is the map sending a symbol $\{u_1, \ldots, u_m\}$ to the class of the Pfister form $\langle <u_1, \ldots, u_m> \rangle \mod I(F)^m$. For $m < 0$, $K^{\text{MW}}_m(F)$ is isomorphic to the Witt group $W(F)$ of $F$, that is, the quotient of $GW(F)$ by the ideal generated by the hyperbolic form $x^2 - y^2$. Also, the map $GW(F) \rightarrow W(F)$ gives an isomorphism of $I(F)^r$ with its image in $W(F)$ for all $r \geq 1$. Thus, for $n \geq 1$,

$$I(F)^n K^{\text{MW}}_m(F) = \begin{cases} I(F)^n \subset W(F) & \text{for } m < 0 \\ I(F)^{n+m} \subset W(F) & \text{for } m \geq 0. \end{cases}$$

The fact that $\cap_n I(F)^n = 0$ in $W(F)$ is a theorem of Arason and Pfister \cite{[1]} Korollar 1.

**Remarks.** 1. The proof in \cite{[16]} that $K^{\text{MW}}_m(F)$ fits into a cartesian square as above relies on the Milnor conjecture. 2. As pointed out to me by Igor Kriz, Voevodsky’s convergence conjecture in the generality as stated above is false. In fact, take $\mathcal{E}$ to be the Moore spectrum.
$S_k/\ell$ for some prime $\ell \neq 2$. Since $\Pi_{a,q}S_k = 0$ for $a < 0$, proposition 6.9 below shows that $\Pi_{a,q}f_nS_k = 0$ for $a < 0$, and thus we have the right exact sequence for all $n \geq 0$

\[
\Pi_{0,0}f_nS_k \times_{\ell} \Pi_{0,0}f_nS_k \to \Pi_{0,0}f_nE \to 0.
\]

In particular, we have

\[
F^n_{\text{Tate}}\Pi_{0,0}E(k) = \text{im} \left( F^n_{\text{Tate}}\Pi_{0,0}S_k \to \Pi_{0,0}S_k(k)/\ell \right) =
\]

\[
= \text{im} \left( I(k)^n \to GW(k)/\ell \right).
\]

Take $k = \mathbb{R}$. Then $I(\mathbb{R}) \subset GW(\mathbb{R})$ is isomorphic to $\mathbb{Z}$ via the virtual negative index, and $I(\mathbb{R})^n = (2^{n-1}) \subset \mathbb{Z} = I(\mathbb{R})$. Thus $\Pi_{0,0}E = \mathbb{Z}/\ell \oplus \mathbb{Z}/\ell$ and the filtration $F^n_{\text{Tate}}\Pi_{0,0}E$ is constant, equal to $\mathbb{Z}/\ell = I(\mathbb{R})/\ell$, and is therefore not separated.

The convergence property is thus not a "triangulated" one in general, and therefore seems to be a subtle one. However, if the $I$-adic filtration on $GW(F)$ is finite for all finitely generated $F$ over $k$ (possibly of varying length depending on $F$), then the augmentation ideal in $GW(F)$ is two-primary torsion. Our computations (at least in characteristic $\neq 2$) show that the filtration $F^n_{\text{Tate}}\Pi_{0,p}\Sigma_{r,q}^{\infty}G_m^{a,q}$ is in this case at least locally finite, and thus has better triangulated properties. In particular, for $\ell \neq 2$,

\[
\Pi_{0,0}(S_k/\ell) = \mathbb{Z}/\ell, \quad F^n_{\text{Tate}}\Pi_{0,0}(S_k/\ell) = 0 \text{ for } n > 0.
\]

One can therefore ask if Voevodsky’s convergence conjecture is true if one assumes the finiteness of the $I(F)$-adic filtration on $GW(F)$ for all finitely generated fields $F$ over $k$. The main theorem of this paper is a partial answer to the convergence question along these lines.

**Theorem 4.** Let $k$ be a perfect field of finite cohomological dimension and let $p$ denote the exponential characteristic.\(^4\) Take $E$ in $\mathcal{SH}_{\text{fin}}(k)$ and take $x \in X \in \mathcal{S}m/k$ with $X$ irreducible. Let $d = \dim_k X$. Then for every $r, q, m \in \mathbb{Z}$, there is an integer $N = N(E, r, d, q)$ such that

\[
(\text{Fil}_{\text{Tate}}^n\Pi_{r,q}f_mE)_{x[1/p]} = 0
\]

for all $n \geq N$. In particular, if $F$ is a field extension of $k$ of finite transcendence dimension $d$ over $k$, then $\text{Fil}_{\text{Tate}}^n\Pi_{r,q}f_mE(F)[1/p] = 0$ for all $n \geq N$.

For a more detailed and perhaps more general statement, we refer the reader to theorem 7.3.

**Remarks.** 1. The proof of theorem 4 relies on the Bloch-Kato conjecture.

2. As we have seen, Voevodsky’s convergence conjecture is not true for all base fields $k$. An interesting class of fields strictly larger than the class of fields of finite cohomological dimension is those of finite virtual cohomological dimension (e.g., $\mathbb{R}$). We suggest the following formulation:

\(^4\)That is, $p = \text{char.} \; k$ if $\text{char.} \; k > 0$, $p = 1$ if $\text{char.} \; k = 0$. 
Conjecture 5. Let \( k \) be a field of finite virtual 2-cohomological dimension. Then the \( I(k) \)-completed slice tower is weakly convergent: after \( I(k) \)-completion, the filtration \( \text{Fil}^r_{\text{Tate}} \Pi_{r,q,f_m}E \) is stalkwise separated for each \( E \) in \( \mathcal{SH}_{\text{fin}}(k) \) and each \( r,q,m \).

This modified conjecture is equivalent to Voevodsky’s convergence conjecture in case \( k \) has finite 2-cohomological dimension, as in this case \( I(k) \) is nilpotent. One could also ask for a weaker version, in which one assumes that \( k \) has finite \( p \)-cohomological dimension for all odd primes \( p \).

3. It would be interesting to be able to say something about the \( p \)-torsion in \( (\text{Fil}^r_{\text{Tate}} \Pi_{r,q}E)_x \).

The paper is organized as follows: We set the notation in §1. In §2 we recall some basic facts about the slice tower, the truncation functors \( f_n \) in \( \mathcal{SH}(k) \) and \( \mathcal{SH}_{S^1}(k) \), and the associated filtration \( \text{Fil}^r_{\text{Tate}} \Pi_{a,b} \). We recall the construction and basic properties of the homotopy coniveau tower, a simplicial model for the slice tower in \( \mathcal{SH}_{S^1}(k) \), in §3. In §4 we use the simplicial nature of the homotopy coniveau tower to analyze the terms in the slice tower. This leads to the main inductive step in our argument (Lemma 4.5), and isolates the particular piece that we need to study. This is analyzed further in §5, where we more precisely identify this piece in terms of a \( K_{\text{MW}} \)-module structure on the bi-graded homotopy sheaves (see Theorem 5.3). In §6 we use a decomposition theorem of Morel and results of Cisinski-Déglise to prove some boundedness properties of the homotopy sheaves \( \Pi_{p,q}E \) and their \( \mathbb{Q} \)-localizations \( \Pi_{p,q}E_\mathbb{Q} \) for \( E \) in \( \mathcal{SH}_{\text{fin}}(k) \), under the assumption that the base-field \( k \) has finite 2-cohomological dimension. In the final section §7 we assemble all the pieces and prove our main result. We conclude with two appendices; the first collects some results on norm maps for finite field extensions that are used throughout the paper and the second assembles some basic facts on the localization of compactly generated triangulated categories with respect to a collection of non-zero integers.

I am grateful to the referee for a number of comments and suggestions for improving an earlier version of this paper and to S. Kelly for discussions on aspects of localization.

1. Background and notation

Unless we specify otherwise, \( k \) will be a fixed perfect base field, without restriction on the characteristic. For details on the following constructions, we refer the reader to [5, 7, 8, 14, 15, 17, 18].

We write \([n]\) for the set \([0, \ldots, n]\) with the standard order (including \([-1] = \emptyset\) and let \( \Delta \) be the category with objects \([n] \), \( n \in \mathbb{N} \), and morphisms \([n] \to [m] \) the order-preserving maps of sets. Given a category \( \mathcal{C} \), the category of simplicial objects in \( \mathcal{C} \) is as usual the category of functors \( \Delta^{\text{op}} \to \mathcal{C} \), and the category of cosimplicial objects the functor category \( \mathcal{C}_{\Delta} \).

\( \mathbf{Spc} \) will denote the category of simplicial sets, \( \mathbf{Spc}_* \) the category of pointed simplicial sets, \( \mathcal{H} := \mathbf{Spc}[W^{-1}] \) the classical unstable homotopy category and
\( \mathcal{H}_\bullet := \text{Spc}_\bullet [W^{-1}] \) the pointed version; here \( WE \) is the usual class of weak equivalences, that is, morphisms \( A \to B \) that induce an isomorphism on all \( \pi_n \) for all choice of base-point. \( \text{Spt} \) is the category of spectra, that is, spectrum objects in \( \text{Spc}_\bullet \) with respect to the left suspension functor \( \Sigma_{S^1} := S^1 \wedge (\_). \)

With \( sWE \) denoting the class of stable weak equivalences, that is, morphisms \( f : E \to F \) in \( \text{Spt} \) that induce an isomorphism on all stable homotopy groups, \( \mathcal{SH} := \text{Spt}[sWE^{-1}] \) is the classical stable homotopy category.

For a simplicial object in \( \text{Spc} \), resp. \( \text{Spc}_\bullet \), resp. \( \text{Spt} \), \( S : \Delta^{op} \to \text{Spc}, \text{Spc}_\bullet, \text{Spt} \), we let \( \lbrack S \rbrack \in \text{Spc}, \text{Spc}_\bullet, \text{Spt} \) denote respective homotopy colimit \( \text{hocolim}_{\Delta^{op}} S \).

The motivic versions are as follows: \( \text{Sm}/k \) is the category of smooth finite type \( k \)-schemes. \( \text{Spc}(k) \) is the category of \( \text{Spc} \)-valued presheaves on \( \text{Sm}/k \). \( \text{Spc}_\bullet(k) \) the \( \text{Spc}_\bullet \)-valued presheaves, and \( \text{Spt}_{\Sigma_1}(k) \) the \( \text{Spt} \)-valued presheaves. These all come with “motivic” model structures as simplicial model categories (see for example [8]); we denote the corresponding homotopy categories by \( \mathcal{H}(k), \mathcal{H}_\bullet(k) \) and \( \mathcal{SH}_{S^1}(k) \), respectively. Sending \( X \in \text{Sm}/k \) to the sheaf of sets on \( \text{Sm}/k \) represented by \( X \) (which we also denote by \( X \)) gives an embedding of \( \text{Sm}/k \) to \( \text{Spc}(k) \); we have the similarly defined embedding of the category of smooth pointed schemes over \( k \) into \( \text{Spc}_\bullet(k) \). Sending a (pointed) simplicial set \( A \) to the constant presheaf with value \( A \) (also denoted by \( A \)) defines an embedding of \( \text{Spc} \) in \( \text{Spc}(k) \) and of \( \text{Spc}_\bullet \) in \( \text{Spc}_\bullet(k) \).

Let \( \mathbb{G}_m \) be the pointed \( k \)-scheme \( (k^1 \setminus 0, 1) \). We let \( T := A^1/(A^1 \setminus \{0\}) \) and let \( \text{Spt}_T(k) \) denote the category of \( T \)-spectra, i.e., spectra in \( \text{Spc}_\bullet(k) \) with respect to the left \( T \)-suspension functor \( \Sigma^T_T := T \wedge (\_). \). \( \text{Spt}_T(k) \) has a motivic model structure (see [8]) and \( \mathcal{SH}(k) \) is the homotopy category. We can also form the category of spectra in \( \text{Spt}_{\Sigma_1}(k) \) with respect to \( \Sigma^T_{\Sigma_1} \); with an appropriate model structure the resulting homotopy category is equivalent to \( \mathcal{SH}(k) \). We will identify these two homotopy categories without further mention.

For each \( A \in \text{Spc}_\bullet(k) \), the suspension functor \( \Sigma_A : \text{Spc}_\bullet(k) \to \text{Spc}_\bullet(k) \), \( \Sigma_A(B) := B \wedge A \), extends to the suspension functor \( \Sigma_A : \text{Spt}_{S^1}(k) \to \text{Spt}_{S^1}(k) \) or \( \Sigma_A : \text{Spt}_T(k) \to \text{Spt}_T(k) \). For \( A \) cofibrant, this gives the suspension functors \( \Sigma_A : \mathcal{H}_\bullet(k) \to \mathcal{H}_\bullet(k), \mathcal{S} \mathcal{H}_{S^1}(k) \to \mathcal{S} \mathcal{H}_{S^1}(k) \) and \( \Sigma_A : \mathcal{S} \mathcal{H}(k) \to \mathcal{S} \mathcal{H}(k) \) by applying \( \Sigma_A \) to a cofibrant replacement.

Both \( \mathcal{S} \mathcal{H}_{S^1}(k) \) and \( \mathcal{S} \mathcal{H}(k) \) are triangulated categories with translation functor \( \Sigma_{S^1} \). On \( \mathcal{H}_\bullet(k), \mathcal{S} \mathcal{H}_{S^1}(k) \) and \( \mathcal{S} \mathcal{H}(k) \), we have \( \Sigma_T \cong \Sigma_{S^1} \circ \Sigma_{\mathbb{G}_m} \); the suspension functors \( \Sigma_T \) and \( \Sigma_{\mathbb{G}_m} \) on \( \mathcal{S} \mathcal{H}(k) \) are invertible. For \( A \in \text{Spc}_\bullet(k) \), we have an enriched Hom on \( \text{Spt}_{S^1}(k) \) and \( \text{Spt}_T(k) \) with values in spectra; we denote the enriched Hom functor by \( \text{Hom}(A, \_). \) This passes to the homotopy categories \( \mathcal{H}_\bullet(k), \mathcal{S} \mathcal{H}_{S^1}(k) \) and \( \mathcal{S} \mathcal{H}(k) \) to give for \( A \in \mathcal{H}_\bullet(k) \) an enriched Hom \( \text{Hom}(A, \_). \) with values in \( \mathcal{S} \mathcal{H} \). For \( X \in \text{Sm}/k, E \in \text{Spt}_{S^1}(k) \), \( \text{Hom}(X_+, E) = E(X) \).
We have the triangle of infinite suspension functors $\Sigma^\infty$ and their right adjoints $\Omega^\infty$:

$$
\begin{array}{ccc}
\mathcal{H}(k) & \xrightarrow{\Sigma^\infty} & \mathcal{H}_{S^1}(k) \\
\downarrow & & \downarrow \\
\mathcal{H}(k) & \xrightarrow{\Omega^\infty} & \mathcal{H}_{S^1}(k)
\end{array}
$$

both commutative up to natural isomorphism. These are all left, resp. right derived versions of Quillen adjoint pairs of functors on the underlying model categories.

For $X \in \mathcal{H}(k)$, we have the bi-graded homotopy sheaf $\Pi_{a,b}X$, defined for $a, b \geq 0$, as the Nisnevich sheaf associated to the presheaf on $\text{Sm}/k$

$$
U \mapsto \text{Hom}_{\mathcal{H}(k)}(\Sigma_{a}^{b} \Sigma_{G_m}^{\infty} U_+, X);
$$

note the perhaps non-standard indexing. We have the bi-graded homotopy sheaves $\Pi_{a,b}E$ for $E \in \mathcal{S}H_{S^1}(k)$, $b \geq 0$, $a \in \mathbb{Z}$, and $\Pi_{a,b}E$ for $E \in \mathcal{H}(k)$, $a, b \in \mathbb{Z}$, by taking the Nisnevich sheaf associated to

$$
U \mapsto \text{Hom}_{\mathcal{S}H_{S^1}(k)}(\Sigma_{a}^{b} \Sigma_{G_m}^{\infty} U_+, E) \text{ or } U \mapsto \text{Hom}_{\mathcal{H}(k)}(\Sigma_{a}^{b} \Sigma_{G_m}^{\infty} U_+, E),
$$

as the case may be. We write $\pi_n$ for $\Pi_{n,0}$; for $E \in \text{Spt}_{S^1}(k)$ fibrant, $\pi_n E$ is the Nisnevich sheaf associated to the presheaf $U \mapsto \pi_n(E(U))$.

$\mathcal{S}H(k)$ has the set of compact generators

$$
\{\Sigma_{a}^{n} \Sigma_{G_m}^{\infty} X_+, n, m \in \mathbb{Z}, X \in \text{Sm}/k\}
$$

and $\mathcal{S}H_{S^1}(k)$ has the set of compact generators

$$
\{\Sigma_{a}^{n} \Sigma_{G_m}^{\infty} X_+, n \in \mathbb{Z}, m \geq 0, X \in \text{Sm}/k\}.
$$

For $\mathcal{S}H(k)$, this is \cite[Theorem 9.2]{[4]}; the proof of this result goes through without change to yield the statement for $\mathcal{S}H_{S^1}(k)$. As these triangulated categories are both homotopy categories of stable model categories, both admit arbitrary small coproducts.

For $F$ a finitely generated field extension of $k$, we may view $\text{Spec} F$ as the generic point of some $X \in \text{Sm}/k$ (since $k$ is perfect). Thus, for a Nisnevich sheaf $\mathcal{S}$ on $\text{Sm}/k$, we may define $\mathcal{S}(F)$ as the stalk of $\mathcal{S}$ at $\text{Spec} F \in X$. For an arbitrary field extension $F$ of $k$ (not necessarily finitely generated over $k$), we define $\mathcal{S}(F)$ as the colimit over $\mathcal{S}(F_n)$, as $F_n$ runs over subfields of $F$ containing $k$ and finitely generated over $k$. For a finitely generated field $F$ over $k$, we consider objects such as $\text{Spec} F$, or $\mathcal{A}_F^+$ as pro-objects in $\text{Spec}(k)$ by the usual system of finite-type models; the same holds for related objects such as $\text{Spec} F_+$ in $\mathcal{H}(k)$ or $\Sigma_{G_m}^{\infty} \text{Spec} F_+$ in $\mathcal{S}H_{S^1}(k)$, etc. We extend this to arbitrary field extensions of $k$ by taking the system of finitely generated subfields. We will usually not explicitly insert the "pro-" in the text, but all such objects, as well as morphisms and isomorphisms between them, should be so understood.
2. VOEVODSKY’S SLICE TOWER

We begin by recalling definition and basic properties of the Tate-Postnikov tower in $\mathcal{S}H_S^1(k)$ and in $\mathcal{S}H(k)$. We then define the main object of our study: the filtration on the bi-graded homotopy sheaves of a $T$-spectrum or an $S^1$-spectrum induced by the respective Tate-Postnikov towers.

For $n \geq 0$, we let $\Sigma^n_T \mathcal{S}H^1_S(k)$ be the localizing subcategory of $\mathcal{S}H^1_S(k)$ generated by the (compact) objects $\Sigma^n_T \Sigma^m_S X_+$, with $X \in \text{Sm}/k$ and $m \geq n$. We note that $\Sigma^n_T \mathcal{S}H^1_S(k) = \mathcal{S}H_S^1(k)$. The inclusion functor $i_n : \Sigma^n_T \mathcal{S}H^1_S(k) \rightarrow \mathcal{S}H^1_S(k)$ admits, by results of Neeman [21, theorem 4.1], a right adjoint $r_n$; define the functor $f_n : \mathcal{S}H^1_S(k) \rightarrow \mathcal{S}H^1_S(k)$ by $f_n := i_n \circ r_n$. The co-unit for the adjunction gives us the natural morphism

$$\rho_n : f_n E \rightarrow E$$

for $E \in \mathcal{S}H^1_S(k)$; similarly, the inclusion $\Sigma^n_T \mathcal{S}H^1_S(k) \subset \mathcal{S}H_S^1(k)$ for $n < m$ gives the natural transformation $f_m E \rightarrow f_n E$, forming the Tate-Postnikov tower

$$\ldots \rightarrow f_{n+1} E \rightarrow f_n E \rightarrow \ldots \rightarrow f_0 E = E;$$

we define $f_n := \text{id}$ for $n < 0$. We complete $f_{n+1} E \rightarrow f_n E$ to a distinguished triangle

$$f_{n+1} E \rightarrow f_n E \rightarrow s_n E \rightarrow f_{n+1} E[1];$$

this distinguished triangle actually characterizes $s_n E$ up to unique isomorphism, hence this defines a distinguished triangle that is functorial in $E$. The object $s_n E$ is the $n$th slice of $E$.

There is an analogous construction in $\mathcal{S}H(k)$: For $n \in \mathbb{Z}$, let

$$\Sigma^n_T \mathcal{S}H^{e/f}_1(k) \subset \mathcal{S}H(k)$$

be the localizing category generated by the $T$-suspension spectra $\Sigma^n_T \Sigma^m_S X_+$, for $X \in \text{Sm}/k$ and $m \geq n$; write $\mathcal{S}H^{e/f}_1(k)$ for $\Sigma^n_T \mathcal{S}H^{e/f}_1(k)$. As above, the inclusion $i_n : \Sigma^n_T \mathcal{S}H^{e/f}_1(k) \rightarrow \mathcal{S}H(k)$ admits a right adjoint $r_n$, giving us the truncation functor $f_n$, $n \in \mathbb{Z}$, and the Tate-Postnikov tower

$$\ldots \rightarrow f_{n+1} \mathcal{E} \rightarrow f_n \mathcal{E} \rightarrow \ldots \rightarrow \mathcal{E}.$$ 

We define the layer $s_n \mathcal{E}$ by a distinguished triangle as above. For integers $N \geq n$, we let $\rho_{n,N} : f_N \rightarrow f_n$ and $\rho_n : f_n \rightarrow \text{id}$ denote the canonical natural transformations. We mention the following elementary but useful result.

**Lemma 2.1.** For integers $N, n$, the diagram of natural endomorphisms of $\mathcal{S}H(k)$

$$\begin{array}{ccc}
\Delta & \rightarrow & \Delta \\
\rho_n & \downarrow & \downarrow \rho_N \\
n \circ f_N & \rightarrow & f_N \\
\end{array}$$

$$\begin{array}{ccc}
\Delta & \rightarrow & \Delta \\
r_n & \downarrow & \downarrow \text{id} \\
n \circ f_N & \rightarrow & f_N \\
\end{array}$$
commutes. Moreover, for $N \geq n$, the map $\rho_n(f_N)$ is a natural isomorphism, and for $N \leq n$, the map $f_n(\rho_N)$ is a natural isomorphism. The same holds with $\mathcal{SH}_{S^1}(k)$ replacing $\mathcal{SH}(k)$.

Proof. The first assertion is just the naturality of $\rho_n$ with respect to the morphism $\rho_N : f_N \to \id$.

Suppose $N \geq n$. Then $\Sigma^N \mathcal{SH}^{eff}(k) \subset \Sigma^n \mathcal{SH}^{eff}(k)$ and thus for all $E \in \mathcal{SH}(k)$, $\id : f_N E \to f_N E$ satisfies the universal property of $\rho_n(f_N E) : f_n(f_N E) \to f_N E$, namely, $f_N E$ is in $\Sigma^N \mathcal{SH}^{eff}(k)$ and $\id : f_N E \to f_N E$ is universal for maps $T \to f_N E$ with $T \in \Sigma^N \mathcal{SH}^{eff}(k)$. Thus, $\rho_n(f_N E)$ is an isomorphism.

If $N \leq n$, then for $E \in \mathcal{SH}(k)$, $f_n(f_N E)$ is in $\Sigma^n \mathcal{SH}^{eff}(k)$ and $\rho_n(f_N E) : f_n(f_N E) \to f_N E$ is universal for maps $T \to f_N E$ with $T \in \Sigma^n \mathcal{SH}^{eff}(k)$. Since $\Sigma^n \mathcal{SH}^{eff}(k) \subset \Sigma^N \mathcal{SH}^{eff}(k)$, the universal property of $\rho_N(E) : f_N E \to E$ shows that $\rho_N(E) \circ \rho_n(f_N E) : f_n(f_N E) \to E$ is universal for maps $T \to E$ with $T \in \Sigma^n \mathcal{SH}^{eff}(k)$, and thus $f_n(\rho_N(E))$ is an isomorphism. The proof for $\mathcal{SH}_{S^1}(k)$ is the same.

\begin{lemma}
For $n \in \mathbb{Z}$, there is a natural isomorphism
\begin{equation}
\label{eq:iso}
f_n \Omega^\infty_E \cong \Omega^\infty_{f_n E}.
\end{equation}
\end{lemma}

Proof. First suppose that $n \geq 0$. It follows from [10] theorem 7.4.1 that $\Omega^\infty_{f_n E}$ is in $\Sigma^n \mathcal{SH}_{S^1}(k)$ and thus we need only show that $\Omega^\infty_{\rho_n} : \Omega^\infty_{f_n E} \to \Omega^\infty_{f_n E}$ satisfies the universal property of $f_n \Omega^\infty E \to \Omega^\infty_{f_n E}$ in $\Sigma^n \mathcal{SH}_{S^1}(k)$ generated as a localizing subcategory of $\mathcal{SH}_{S^1}(k)$ by objects $\Sigma^N \mathcal{SH}_{S^1}(k)$, $G \in \mathcal{SH}_{S^1}(k)$, so it suffices to check for objects of this form. We have

$$
\text{Hom}_{\mathcal{SH}_{S^1}(k)}(\Sigma^n \mathcal{SH}_{S^1}(k), \Omega^\infty_{f_n E}) \cong \text{Hom}_{\mathcal{SH}(k)}(\Sigma^n \mathcal{SH}(k), f_n E)
$$

$$
\cong \text{Hom}_{\mathcal{SH}(k)}(\Sigma^n \mathcal{SH}(k), f_n E) \xrightarrow{\sim} \text{Hom}_{\mathcal{SH}(k)}(\Sigma^n \mathcal{SH}(k), E)
$$

$$
\cong \text{Hom}_{\mathcal{SH}(k)}(\Sigma^n \mathcal{SH}(k), E) \cong \text{Hom}_{\mathcal{SH}_{S^1}(k)}(\Sigma^n \mathcal{SH}_{S^1}(k), E).
$$

It is easy to check that this sequence of isomorphisms is induced by $(\Omega^\infty_{\rho_n}).$

Now suppose that $n < 0$. Then $f_n \Omega^\infty E \cong f_0 \Omega^\infty E \cong \Omega^\infty f_0 E$, so it suffices to show that the map $f_0 E \to f_n E$ induces an isomorphism $\Omega^\infty_{f_0 E} \to \Omega^\infty f_n E$. But for $F \in \mathcal{SH}_{S^1}(k)$, $\Omega^\infty_{S^1} F$ is in $\mathcal{SH}^{eff}(k)$ and

$$
\text{Hom}_{\mathcal{SH}_{S^1}(k)}(F, \Omega^\infty_{f_0 E}) \cong \text{Hom}_{\mathcal{SH}(k)}(F, f_0 E)
$$

$$
\xrightarrow{\rho_n \circ \sim} \text{Hom}_{\mathcal{SH}(k)}(F, f_n E) \cong \text{Hom}_{\mathcal{SH}_{S^1}(k)}(F, \Omega^\infty_{f_n E}).
$$

For $E \in \mathcal{SH}_{S^1}(k)$, we have (by [10] theorem 7.4.2] the canonical isomorphism

\begin{equation}
\label{eq:can_iso}
\Omega^\infty_{\mathcal{SH}_{S^1}(k)} f_n E \cong f_{n-1} \Omega^\infty_{\mathcal{SH}_{S^1}(k)} E
\end{equation}
for $r \geq 0$. As $\Omega_{S_m} : \SH(k) \to \SH(k)$ is an auto-equivalence, and restricts to an equivalence
\[ \Omega_{S_m} : \Sigma^n S \SH^{eff}(k) \to \Sigma^{n-1} S \SH^{eff}(k), \]
the analogous identity in $\SH(k)$ holds as well, for all $r \in \mathbb{Z}$.

**Definition 2.3.** For $a \in \mathbb{Z}$, $b \geq 0$, $E \in \SH_{S^1}(k)$, define the filtration $F^n_{\text{Tate}} \Pi_{a,b} E$ of $\Pi_{a,b} E$ by
\[ F^n_{\text{Tate}} \Pi_{a,b} E := \text{im}(\Pi_{a,b} f_n E \to \Pi_{a,b} E); \quad n \in \mathbb{Z}. \]
Similarly, for $E \in \SH(k)$, $a, b, n \in \mathbb{Z}$, define
\[ F^n_{\text{Tate}} \Pi_{a,b} E := \text{im}(\Pi_{a,b} f_n E \to \Pi_{a,b} E). \]
The main object of this paper is to understand $F^n_{\text{Tate}} \Pi_{a,b} E$ for suitable $E$. For later use, we note the following:

**Lemma 2.4.** 1. For $E \in \SH_{S^1}(k)$, $n, p, a, b \in \mathbb{Z}$ with $p, b - p \geq 0$, the adjunction isomorphism $\Pi_{a,b} E \cong \Pi_{a,b-p} \Omega_{G_m}^p E$ induces an isomorphism
\[ F^n_{\text{Tate}} \Pi_{a,b} E \cong F^n_{\text{Tate}} \Pi_{a,b-p} \Omega_{G_m}^p E. \]
Similarly, for $E \in \SH(k)$, $n, p, a, b \in \mathbb{Z}$, the adjunction isomorphism $\Pi_{a,b} E \cong \Pi_{a,b-p} \Omega_{G_m}^p E$ induces an isomorphism
\[ F^n_{\text{Tate}} \Pi_{a,b} E \cong F^n_{\text{Tate}} \Pi_{a,b-p} \Omega_{G_m}^p E. \]
2. For $E \in \SH(k)$, $a, b, n \in \mathbb{Z}$, with $b \geq 0$, we have a canonical isomorphism
\[ \varphi_{E,a,b,n} : \Pi_{a,b} f_n E \to \Pi_{a,b} \Omega_{G_m}^\infty E, \]
inducing an isomorphism $F^n_{\text{Tate}} \Pi_{a,b} E \cong F^n_{\text{Tate}} \Pi_{a,b} \Omega_{G_m}^\infty E$.

**Proof.** (1) By (2.2), adjunction induces isomorphisms
\[ F^n_{\text{Tate}} \Pi_{a,b} E := \text{im}(\Pi_{a,b} f_n E \to \Pi_{a,b} E) \cong \text{im}(\Pi_{a,b-p} \Omega_{G_m}^p f_n E \to \Pi_{a,b} E) \]
\[ = \text{im}(\Pi_{a,b-p} \Omega_{G_m}^p E \to \Pi_{a,b} \Omega_{G_m}^\infty E) = F^n_{\text{Tate}} \Pi_{a,b} \Omega_{G_m}^p E. \]
The proof for $E \in \SH(k)$ is the same.

For (2), the isomorphism $\varphi_{E,a,b,n}$ arises from (2.1) and the adjunction isomorphism
\[ \text{Hom}_{\SH_{S^1}(k)}(\Sigma^a \Sigma_{G_m} \Sigma_{S^1} \Sigma_{G_m} \Sigma_{S^1} \Sigma_{G_m} \Sigma_{S^1} U_+, f_n \Omega_{G_m}^\infty E) \cong \text{Hom}_{\SH_{S^1}(k)}(\Sigma^a \Sigma_{G_m} \Sigma_{S^1} \Sigma_{G_m} \Sigma_{S^1} U_+, f_n \Omega_{G_m}^\infty E) \cong \text{Hom}_{\SH(k)}(\Sigma^a \Sigma_{G_m} \Sigma_{S^1} \Sigma_{G_m} \Sigma_{S^1} U_+, f_n \Omega_{G_m}^\infty E). \]
\[ \square \]
3. The homotopy coniveau tower

Our computations rely heavily on our model for the Tate-Postnikov tower in \( S\text{H}_{\text{St}}(k) \), which we briefly recall (for details, we refer the reader to [10]).

We start with the cosimplicial scheme \( n \mapsto \Delta^n \), with \( \Delta^n \) the algebraic \( n \)-simplex \( \text{Spec} k[t_0, \ldots, t_n]/\sum t_i - 1 \). The cosimplicial structure is given by sending a map \( g : [n] \to [m] \) to the map \( \Delta(g) : \Delta^n \to \Delta^m \) determined by

\[
\Delta(g)^*(t_i) = \begin{cases} 
\sum_{j, g(j) = i} t_j & \text{if } g^{-1}(i) \neq \emptyset \\
0 & \text{else}.
\end{cases}
\]

A face of \( \Delta^m \) is a closed subscheme \( F \) defined by equations \( t_1 = \ldots = t_{i_*} = 0 \); we let \( \partial \Delta^n \subset \Delta^n \) be the closed subscheme defined by \( \prod_{i=0}^n t_i = 0 \), i.e., \( \partial \Delta^n \) is the union of all the proper faces.

Take \( X \in \text{Sm}/k \). We let \( S_X^{(q)}(m) \) denote the set of closed subsets \( W \subset X \times \Delta^m \) such that

\[
\text{codim}_{X \times F} W \cap X \times F \geq q
\]

for all faces \( F \subset \Delta^m \) (including \( F = \Delta^m \)). We make \( S_X^{(q)}(m) \) into a partially ordered set via inclusions of closed subsets. Sending \( m \) to \( S_X^{(q)}(m) \) and \( g : [n] \to [m] \) to \( (\Delta(g))^{-1} : S_X^{(q)}(m) \to S_X^{(q)}(n) \) gives us the simplicial poset \( S_X^{(q)} \).

Now take \( E \in \text{Spt}_{\text{St}}(k) \). For \( X \in \text{Sm}/k \) and closed subset \( W \subset X \), we have the spectrum with supports \( E^W(X) \) defined as the homotopy fiber of the restriction map \( E(X) \to E(X \setminus W) \). This construction is functorial in the pair \( (X, W) \), where we define a map \( f : (Y, T) \to (X, W) \) as a morphism \( f : Y \to X \) in \( \text{Sm}/k \) with \( f^{-1}(W) \subset T \). We usually denote the map induced by \( f = id_X : (X,T) \to (X,W) \) by \( f^* : E^W(X) \to E^T(Y) \), but for \( f = id_X : (X,T) \to (X,W), i : W \to T \) the resulting inclusion, we write \( i_* : E^W(X) \to E^T(X) \) for \( id_X \).

Define

\[
E^{(q)}(X, m) := \text{hocolim} \ E^W(X \times \Delta^m).
\]

The fact that \( m \mapsto S_X^{(q)}(m) \) is a simplicial poset, and \( (Y, T) \mapsto E^T(Y) \) is a functor from the category of pairs to spectra shows that \( m \mapsto E^{(q)}(X, m) \) defines a simplicial spectrum. We define the spectrum \( E^{(q)}(X) \) by

\[
E^{(q)}(X) := \{ m \mapsto E^{(q)}(X, m) \} := \text{hocolim}_{\Delta^0} E^{(q)}(X, -).
\]

For \( q \geq q' \), the inclusions \( S_X^{(q)}(m) \subset S_X^{(q')}(m) \) induce a map of simplicial posets \( S_X^{(q)} \subset S_X^{(q')} \) and thus a morphism of spectra \( i_{q', q} : E^{(q)}(X) \to E^{(q')}(X) \). Since \( E^{(0)}(X, 0) = E(X) \), we have the canonical map

\[
\epsilon_X : E(X) \to E^{(0)}(X),
\]

Documenta Mathematica 18 (2013) 907–941
which is a weak equivalence if $E$ is homotopy invariant. Together, this forms the augmented homotopy coniveau tower

$$E^{(q)}(X) := \ldots \rightarrow E^{(q+1)}(X) \xrightarrow{i_q} E^{(q)}(X) \xrightarrow{i_{q-1}} \ldots \rightarrow E^{(1)}(X) \xrightarrow{i_0} E^{(0)}(X) \xrightarrow{\epsilon} E(X)$$

with $i_q := i_{q,q+1}$. For homotopy invariant $E$, this gives us the homotopy coniveau tower in $\mathcal{SH}$

$$E^{(q)}(X) := \ldots \rightarrow E^{(q+1)}(X) \xrightarrow{i_q} E^{(q)}(X) \xrightarrow{i_{q-1}} \ldots \rightarrow E^{(1)}(X) \xrightarrow{i_0} E^{(0)}(X) \cong E(X).$$

Letting $\mathbf{Sm}/k$ denote the subcategory of $\mathbf{Sm}/k$ with the same objects and with morphisms the smooth morphisms, it is not hard to see that sending $X$ to $E^{(q)}(X)$ defines a functor from $\mathbf{Sm}/k^{op}$ to augmented towers of spectra. On the other hand, for $E \in \mathbf{Spt}_{S^1}(k)$, we have the (augmented) Tate-Postnikov tower

$$f_*E := \ldots \rightarrow f_{q+1}E \rightarrow f_qE \rightarrow \ldots \rightarrow f_0E \cong E$$

in $\mathcal{SH}_{S^1}(k)$, which we may evaluate at $X \in \mathbf{Sm}/k$, giving the tower $f_*E(X)$ in $\mathcal{SH}$, augmented over $E(X)$.

Call $E \in \mathbf{Spt}_{S^1}(k)$ quasi-fibrant if, for $E \rightarrow E^{fib}$ a fibrant replacement in the motivic model structure, the map $E(X) \rightarrow E^{fib}(X)$ is a stable weak equivalence in $\mathbf{Spt}$ for all $X \in \mathbf{Sm}/k$. As a general rule, we will represent an $E \in \mathcal{SH}_{S^1}(k)$ by a fibrant object in $\mathbf{Spt}_{S^1}(k)$, also denoted $E$, without making explicit mention of this choice.

As a direct consequence of [10] theorem 7.1.1] we have

**Theorem 3.1.** Let $E$ be a quasi-fibrant object in $\mathbf{Spt}_{S^1}(k)$, and take $X \in \mathbf{Sm}/k$. Then there is an isomorphism of augmented towers in $\mathcal{SH}$

$$(f_*E)(X) \cong E^{(q)}(X)$$

over the identity on $E(X)$, which is natural with respect to smooth morphisms in $\mathbf{Sm}/k$.

In particular, we may use the model $E^{(q)}(X)$ to understand $(f_*E)(X)$.

**Remark 3.2.** For $X, Y \in \mathbf{Sm}/k$ with given $k$-points $x \in X(k), y \in Y(k)$, we have a natural isomorphism in $\mathcal{SH}_{S^1}(k)$

$$\Sigma_{S^1}^\infty(X \wedge Y) \oplus \Sigma_{S^1}^\infty(X \vee Y) \cong \Sigma_{S^1}^\infty(X \times Y),$$

using the additivity of the category $\mathcal{SH}_{S^1}(k)$. Thus, $\Sigma_{S^1}^\infty(X \wedge Y)$ is a canonically defined summand of $\Sigma_{S^1}^\infty(X \times Y)$. In particular for $E$ a quasi-fibrant object of $\mathbf{Spt}_{S^1}(k)$, we have a natural isomorphism in $\mathcal{SH}$

$$\mathbf{Hom}(X \wedge Y, E) \cong \mathbf{h}(\mathbf{f}(E(X) \times Y) \rightarrow \mathbf{f}(E(X) \oplus E(Y) \rightarrow E(k)))$$

where the maps are induced by the evident restriction maps. In particular, we may define $E(X \wedge Y)$ via the above isomorphism, and our comparison results
for Tate-Postnikov tower and homotopy coniveau tower extend to values at smash products of smooth pointed schemes over \( k \).

4. The simplicial filtration

In this section, we study the filtration on \( \pi_r E(X) \) induced by the simplicial structure of the model \( E^{(n)}(X) \).

**Lemma 4.1.** Let \( S \) be a smooth \( k \)-scheme, \( W \subset S \times \mathbb{A}^1 \) a closed subset such that \( p : W \to S \) is finite. Let \( E \in \text{Spf}_{S^1}(k) \) be quasi-fibrant. Then the map induced by the inclusion \( i : W \to p^{-1}(p(W)) \) induces the zero map

\[
i_* : \pi_r(E^W(S \times \mathbb{A}^1)) \to \pi_r(E^{p^{-1}(p(W))}(S \times \mathbb{A}^1)).
\]

**Proof.** We steal a proof of Morel’s: Let \( Z = p(W) \), and let \( j_0 : S \times \mathbb{A}^1 \to S \times \mathbb{P}^1 \) be the standard open neighborhood of \( S \times 0 \) in \( S \times \mathbb{P}^1 \). Since \( W \) is finite over \( S \), \( W \) is closed in \( S \times \mathbb{P}^1 \), so we have the following commutative diagram

\[
\begin{array}{c}
\pi_r(E^W(S \times \mathbb{P}^1)) \\
\downarrow j_0^* \\
\pi_r(E^W(S \times \mathbb{A}^1)) \\
\uparrow j_* \\
\pi_r(E^Z(S \times \mathbb{A}^1))
\end{array}
\]

where \( j : W \to Z \times \mathbb{P}^1 \) is the inclusion. Let \( i : S \to S \times \mathbb{P}^1 \) be the infinity section. Since \( W \cap S \times \infty = \emptyset \), the composition

\[
\pi_r(E^W(S \times \mathbb{P}^1)) \xrightarrow{i_*} \pi_r(E^Z(S \times \mathbb{P}^1)) \xrightarrow{i_*} \pi_r(E^Z(S \times \infty))
\]

is the zero map. Letting \( j_\infty : S \times \mathbb{A}^1 \to S \times \mathbb{P}^1 \) be the standard open neighborhood of \( S \times \infty \) in \( S \times \mathbb{P}^1 \), the restriction map

\[
i_\infty^* : \pi_r(E^Z(S \times \mathbb{A}^1)) \to \pi_r(E^Z(S \times \infty))
\]

is an isomorphism, hence

\[
j_\infty^* \circ i_* : \pi_r(E^W(S \times \mathbb{P}^1)) \to \pi_r(E^Z(S \times \mathbb{A}^1))
\]

is the zero map. Write \( j_\infty \) for the inclusions of \( S \times \mathbb{P}^1 \setminus \{0, \infty\} \) into \( j_0(S \times \mathbb{A}^1) \) and \( j_0 \) for the inclusions of \( S \times \mathbb{P}^1 \setminus \{0, \infty\} \) into \( j_\infty(S \times \mathbb{A}^1) \). Combining (4.1) with the commutativity of the diagram

\[
\begin{array}{c}
\pi_r(E^W(S \times \mathbb{P}^1)) \\
\downarrow j_0^* \\
\pi_r(E^W(S \times \mathbb{A}^1))
\end{array}
\]

\[
\begin{array}{c}
\pi_r(E^Z(S \times \mathbb{A}^1)) \\
\downarrow j_\infty^* \\
\pi_r(E^Z(S \times \mathbb{P}^1 \setminus \{0, \infty\}))
\end{array}
\]

\[
\begin{array}{c}
\pi_r(E^Z(S \times \infty)) \\
\downarrow j_* \\
\pi_r(E^Z(S \times \infty))
\end{array}
\]
we see that $j^*_\infty \circ i_* = 0$. From the long exact localization sequence
\[ \ldots \to \pi_r(E^{Z \times 0}(S \times A^1)) \xrightarrow{i_*} \pi_r(E^{Z \times A^1}(S \times A^1)) \]
we see that
\[ i_*(\pi_r(E^W(S \times A^1))) \subset i_0*(\pi_r(E^{Z \times 0}(S \times A^1))) \subset \pi_r(E^{Z \times A^1}(S \times A^1)). \]

**Lemma 4.2.** Suppose $F$ is infinite. Take $W \in S_F^{(n)}(p)$ and suppose $\text{codim}_{\Delta_F^p}(W) > n$. Then the canonical map $E^W(\Delta_F^p) \to E^{(n)}(\text{Spec} F, p)$ induces the zero map on $\pi_*$. 

**Proof.** We identify $\Delta^p$ with $A^p$ via the barycentric coordinates $t_1, \ldots, t_p$. Suppose $W$ has dimension $d < p - n$. Then $d \leq p - 1$ and, as $F$ is infinite, a general linear projection $L : A^p \to A^{p-1}$ restricts to $W$ to a finite morphism $W \to A^{p-1}$. In addition, $W' := L^{-1}(L(W))$ is in $S_F^{(n)}(p)$ for $L$ suitably general. Letting $i : W \to W'$ be the inclusion, it suffices to show that the map
\[ i_* : \pi_*(E^W(\Delta_F^p)) \to \pi_*E^{W'}(\Delta_F^p) \]
is the zero map. Via an affine linear change of coordinates on $\Delta^p$, we may identify $\Delta^p$ with $A^{p-1} \times A^1$ and $L : A^p \to A^{p-1}$ with the projection $A^{p-1} \times A^1 \to A^{p-1}$. The result thus follows from Lemma 4.1. \qed

Let $(\Delta_F^p, \partial \Delta^p)^{(n)}$ be the set of codimension $n$ points $w$ of $\Delta_F^p$ such that $\overline{\{w\}}$ is in $S_F^{(n)}(p)$. 

**Lemma 4.3.** Let $F$ be an infinite field. Then the restriction maps
\[ E^W(\Delta_F^p) \to \oplus_{w \in (\Delta_F^p, \partial \Delta^p)^{(n)} \cap W} E^w(\text{Spec} O_{\Delta_F^p, w}) \]
for $W \in S_F^{(n)}(p)$ defines an injection
\[ \pi_r(E^{(n)}(F, p)) \to \oplus_{w \in (\Delta_F^p, \partial \Delta^p)^{(n)}} \pi_r E^w(\text{Spec} O_{\Delta_F^p, w}) \]
for each $r \in \mathbb{Z}$. 

**Proof.** Take $W \in S_F^{(n)}(p)$. Since $\Delta_F^p$ is affine, we can find a $W' \in S_F^{(n)}(p)$ of pure codimension $n$ with $W' \supset W$: just take a sufficiently general collection of $n$ functions $f_1, \ldots, f_n$ vanishing on $W$ and let $W'$ be the common zero locus of the $f_i$. Thus the set of pure codimension $n$ subsets $W'$ of $\Delta_F^p$ with $W' \in S_F^{(n)}(p)$ is cofinal in $S_F^{(n)}(p)$.

*Documenta Mathematica* 18 (2013) 907–941
Convergence of Voevodsky’s Slice Tower

Let $W \in \mathcal{S}_{F}^{(n)}(p)$ have pure codimension $n$ on $\Delta_{F}^{p}$ and let $W_{0} \subseteq W$ be any closed subset. Then $W_{0}$ is also in $\mathcal{S}_{F}^{(n)}(p)$ and we have the long exact localization sequence

$$\cdots \to \pi_{r}E^{W_{0}}(\Delta_{F}^{p}) \xrightarrow{i_{\ast}} \pi_{r}E^{W}(\Delta_{F}^{p}) \to \pi_{r}E^{W\setminus W_{0}}(\Delta_{F}^{p} \setminus W_{0}) \to \cdots$$

Let $\mathcal{S}_{F}^{(n)}(p)_{0} \subseteq \mathcal{S}_{F}^{(n)}(p)$ be the set of all $W_{0} \in \mathcal{S}_{F}^{(n)}(p)$ with $\text{codim}_{\Delta_{F}^{p}}W_{0} > n$. Let

$$E^{(n)}(F, p)_{0} = \text{hocolim}_{W_{0} \in \mathcal{S}_{F}^{(n)}(p)_{0}} E^{W_{0}}(\Delta_{F}^{p})\text{.}$$

Passing to the limit over the above localization sequences gives us the long exact sequence

$$\cdots \to \pi_{r}E^{(n)}(F, p)_{0} \xrightarrow{i_{\ast}} \pi_{r}E^{(n)}(F, p) \xrightarrow{-} \bigoplus_{w \in (\Delta_{F}^{p}, \partial \Delta_{F}^{p})^{(n)}} \pi_{r} E^{w}((\text{Spec} \mathcal{O}_{\Delta_{F}^{p}, w})) \to \cdots$$

By lemma 4.2 the map $i_{\ast}$ is the zero map, which proves the lemma. $\square$

Let $S : \Delta^{op} \to \text{Spt}$ be a simplicial spectrum, $|S| = \text{hocolim}_{\Delta^{op}} S \in \text{Spt}$ the associated spectrum, giving us the spectral sequence

$$E^{1}_{p, q} = \pi_{q}S(p) \Longrightarrow \pi_{p+q}|S|\text{.}$$

This spectral sequence induces an increasing filtration $\text{Fil}^{\text{simp}}_{\ast} \pi_{r}|S|$ on $\pi_{r}|S|$. We have the $q$-truncated simplicial spectrum $S_{\leq q}$ and $\text{Fil}^{\text{simp}}_{\ast} \pi_{r}|S|$ is just the image of $\pi_{r}|S_{\leq q}|$ in $\pi_{r}|S|$. In particular $\text{Fil}^{\text{simp}}_{\ast} \pi_{r}|S| = 0$ and $\cup_{q=0}^{\infty} \text{Fil}^{\text{simp}}_{q} \pi_{r}|S| = \pi_{r}|S|$, so the spectral sequence is weakly convergent, and is strongly convergent if for instance there is an integer $q_{0}$ such that $S(p)$ is $q_{0}$-connected for all $p$.

The isomorphism of theorem 3.7 thus gives us the weakly convergent spectral sequence

$$E^{1}_{p, q}(X, E, n) = \pi_{q}E^{(n)}(X, p) \Longrightarrow \pi_{p+q}f_{n}E(X)$$

which is strongly convergent if $\pi_{q}E^{(n)}(X, p) = 0$ for $q \leq q_{0}$, independent of $p$. This defines the increasing filtration $\text{Fil}^{\text{simp}}_{\ast}E(\pi_{r}f_{n}E(X))$ of $\pi_{r}f_{n}E(X)$ with associated graded $gr^{\text{simp}}_{p}(E)\pi_{r}f_{n}E(X) = E^{\text{simp}}_{p-r, p}$.

**Lemma 4.4.** Suppose that $k$ is infinite and that $\Pi_{a, \ast}E(K) = 0$ for $a < 0$ and all fields $K$ over $k$. Let $F \supset k$ be a field extension of $k$. Then

1. $E^{1}_{p-r, p}(F, E, n) = 0$ for $p > r + n$ and $\text{Fil}^{\text{simp}}_{r+n}E(\pi_{r}f_{n}E(F)) = \pi_{r}f_{n}E(F)$.
2. $E^{1}_{p, q}(F, E, n)$ is isomorphic to a subgroup of $\bigoplus_{w \in (\Delta_{F}^{p}, \partial \Delta_{F}^{p})^{(n)}} \Pi_{q+n, \ast}E(F(w))$.
3. The spectral sequence 4.2 is strongly convergent.

**Proof.** Since the spectral sequence is weakly convergent, to prove (3) it suffices to show that $E^{1}_{p, q}(F, E, n) = 0$ for $q < -n$ and for (1) it suffices to show that
$E_{p,r-p}^1 = 0$ for $p > r + n$. These both follow from (2) as our hypothesis implies that $\Pi_{a,*} E(F(w)) = 0$ for $a < 0$, $w \in \Delta_r^p$.

For (2), lemma 4.3 gives us an inclusion

$$E_{p,q}^1 = \pi_q E^{(n)}(F, p) \subset \oplus_{w \in (\Delta_r^p, \partial \Delta_r^p)} \pi_q E^w(\Spec \mathcal{O}_{\Delta_r^p, w}).$$

Take $w \in (\Delta_r^p, \partial \Delta_r^p)(n)$. By the Morel-Voevodsky purity isomorphism [18, loc.cit.], we have $E^w(\Spec \mathcal{O}_{\Delta_r^p, w}) \cong \Hom(\Sigma^n_{\mathcal{O}_{\Delta_r^p, w}}, E)$, hence

$$\pi_q E^w(\Spec \mathcal{O}_{\Delta_r^p, w}) \cong \Pi_{q+n,n} E(F(w)),$$

which proves (2). \hfill \square

For a field extension $K$ of $k$, we write $\text{tr. dim}_k K$ for the transcendence dimension of $K$ over $k$.

LEMMA 4.5. Let $E$ be in $\mathcal{SH}_{S^1}(k)$ and suppose $\Pi_{a,*} E(K) = 0$ for $a < 0$ and all fields $K$ over $k$. Let $p$ be the exponential characteristic of $k$. Let $r$ be an integer. Suppose we have functions

$$(d, q) \mapsto N_j(d, q; E) \geq 0; \ d, q \geq 0, j = 0, \ldots, r - 1,$$

such that, for each field extension $K$ of $k$ with $\text{tr. dim}_k K \leq d$, each $j = 0, \ldots, r - 1$, and all integers $q, M \geq 0$, $m \geq 0$, we have

$$F^m_{\text{tr. dim}_k F} \Pi_j q f_m E(K)/[1/p] = 0.$$

Let $F$ be a field extension of $k$ with $\text{tr. dim}_k F \leq d$ and fix an integer $n \geq 0$. For $r > 0$, let $N = \max_{j=0}^{r-1} N_j(r - j + d, n; E)$; for $r \leq 0$, let $N = 0$. Then for all integers $m \geq N$, $n \geq 0$, we have

$$\text{tr. dim}_k E^r(f_n(f_m E))(F)/[1/p] = \text{Fil}^\text{simp}_n(f_m E) \text{tr. dim}_k E^r(f_n(f_m E))(F)/[1/p].$$

Proof. We delete the $[1/p]$ from the notation in the proof, using the convention that we have inverted the exponential characteristic $p$ throughout.

If $k$ is a finite field, fix a prime $\ell$ and let $k_\ell$ be the union of all $\ell$-power extensions of $k$. If we know the result for $k_\ell$, then using proposition [3, 2] for each $k \subset k_\ell$ with $k_\ell'$ finite over $k$ proves the result for $k$, after inverting $\ell$. Doing the same for some $\ell \neq \ell$, we reduce to the case of an infinite field $k$.

By [12, proposition 3.2], the hypothesis $\Pi_{a,*} E(K) = 0$ for $a < 0$ and all $K$ implies $\Pi_{a,*} f_{m n} E(K) = 0$ for $a < 0$, all $K$, and all $m \geq 0$. In particular, $\pi_r(f_n(f_m E))(F) = 0$ for $r < 0$ and all $n, m \geq 0$, so for $r < 0$, the lemma is trivially true. We therefore assume $r \geq 0$. In addition, it follows from lemma 4.4(1) that the spectral sequence $E^r_{*,*}(F, f_m E, n)$, is strongly convergent for all $m, n \geq 0$ and $E_{p,r-p}^1(F, f_m E, n) = 0$ for $p > r + n$. Since $\text{Fil}^\text{simp}_n(f_m E) \pi_r(f_n(f_m E))(F)$ is by definition the filtration on $\pi_r(f_n(f_m E))(F)$ induced by this spectral sequence, we need only show that

$$E_{p,r-p}^1(F, f_m E, n) = 0 \text{ for } n < p \leq r + n \text{ and } m \geq N.$$

In particular, the result is proved for $r = 0$; we now assume $r > 0$.
Let \( p \) be an integer, \( n < p \leq r + n \). By lemma [14](2),
\[
E^p_{r+p}(F, f_m E, n) \subset \bigoplus_{w \in (\Delta^p_F, \partial \Delta^p_F)^n} \Pi_{r+p+n, n} f_m E(F(w)).
\]
For \( w \in (\Delta^p_F, \partial \Delta^p_F)^n \), \( w \) has codimension \( n \) on \( \Delta^p_F \), hence \( \text{tr. dim}_F F(w) = p - n \) and thus \( \text{tr. dim}_k F(w) \leq p - n + d \). We have \( m \geq N_{r-p+n}(p - n + d, n; E) \) since \( 0 \leq r - p + n < r \), and so our hypothesis (1.3) implies
\[
F^m_{\text{Tate}} \Pi_{r-p+n, n} f_m E(F(w)) = 0.
\]
But
\[
E^m_{\text{Tate}} \Pi_{a,b} f_m E(F(w)) = \text{im}(\Pi_{a,b} f_m E(F(w)) \rightarrow \Pi_{a,b} f_m E(F(w)))
\]
by definition, and \( \rho_m(f_m E) \) is an isomorphism (lemma 2.1), hence
\[
E^m_{\text{Tate}} \Pi_{a,b} f_m E(F(w)) = \Pi_{a,b} f_m E(F(w)).
\]
Thus \( \Pi_{r-p+n, n} f_m E(F(w)) = 0 \) and hence \( E^1_{p,r-p} = 0 \) for \( n < p \leq r + n \), as desired.

5. The bottom of the filtration

In this section, \( k \) will be a fixed perfect base field. We study the subgroup \( \Fil^n_{q}(E) \pi_r f_n E(F) \) isolated in lemma [13].

**Lemma 5.1.** Let \( E \) be in \( \SH_{S^1}(k) \). Then \( \Fil^n_{q-1}(E) \pi_r f_n E(F) = 0 \) for all fields \( F \) over \( k \).

**Proof.** For any \( X \in \Sm/k \), \( \Fil^n_{q-1}(X) \) is defined as the image of
\[
\pi_r[E^n(X, - \leq q)] \rightarrow \pi_r E^n(X),
\]
where \( E^n(X, - \leq q) \) is the \( q \)-truncation (or \( q \)-skeleton) of \( E^n(X, -) \). For \( X = \Spec F \), we clearly have \( S^{(1)}_p(\psi) = 0 \) for \( p < n \), as \( \Delta^p_F \) has no closed subsets of codimension \( > p \). Thus \( |E^n(X, - \leq q)| \) is the 0-spectrum for \( q < n \) and hence \( \Fil^n_{q-1}(X) \) is 0.

To study the first non-zero layer \( \Fil_{q-1}(E) \pi_r f_n E(F) \) in \( \Fil_{q-1}(E) \pi_r f_n E(F) \), we apply the results of [13] §5. For this, we recall some of these results and constructions.

We let \( V_n = (\Delta^p_F \setminus \partial \Delta^p_F)^n \). The function \(-t_1/t_0\) on \( \Delta^2 \) gives an open immersion \( \rho_n : V_n \rightarrow \mathbb{A}^p_k \), identifying \( V_n \) with \( (\mathbb{A}^p_k \setminus \{0, 1\})^n \).

Suppose that \( E \) is an \( n \)-fold \( T \)-loop spectrum, that is, there is an object \( \omega_T^{-n} E \in \Spt(k) \) and an isomorphism \( E \cong \Omega_T^n \omega_T^{-n} E \) in \( \SH_{S^1}(k) \). Given an \( n \)-fold delooping \( \omega_T^{-n} E \) of \( E \), we have explained in [13] §5 how to construct a “transfer map”
\[
\text{Tr}_{F(w)/F} : \pi_\ast E(w) \rightarrow \pi_\ast E(F),
\]
for each closed point \( w \in \mathbb{A}^p_k \), separable over \( F \).

If now \( E = \Omega_T^n E \) for some \( T \)-spectrum \( E \in \SH(k) \), then the bi-graded homotopy sheaves \( \Pi_{p,q} E \) admit a canonical right action by the bi-graded homotopy sheaves of the sphere spectrum \( S_k \in \SH(k) \):
\[
\Pi_{a,b} E \otimes \Pi_{p,q}(S_k) \rightarrow \Pi_{a+p,b+q} E.
\]
Hopkins and Morel [17, theorem 6.3.3] have defined a graded ring homomorphism

\[ \theta_*(F) : \oplus_{n \in \mathbb{Z}} K_n^{MW}(F) \rightarrow \oplus_{n \in \mathbb{Z}} \Pi_{0,-n} \mathcal{S}_k(F), \]

natural in the field extension \( F / \mathbb{F} \). Via \( \theta_*(F) \), the right \( \Pi_{p,q}(\mathcal{S}_k) \)-action gives a right action, natural in \( F \),

\[ \Pi_{a,b} E(F) \otimes K_n^{MW}(F) \rightarrow \Pi_{a,b-n} E(F). \]

This gives us the filtration \( F^n_{MW, a,b} E(F) \) of \( \Pi_{a,b} E \), defined by

\[ F^n_{MW, a,b} E(F) := \text{im}[\Pi_{a,b+n} E(F) \otimes K_n^{MW}(F) \rightarrow \Pi_{a,b} E(F)]; \quad n \geq 0. \]

Completing with respect to the transfer maps yields the filtration \( F^n_{MW-T, a,b} E(F) \):

**Definition 5.2** (see [12, definition 7.9]). Let \( E = \Omega_{\mathbb{F}}^E \) for some \( E \in \mathcal{SH}(k) \), \( F \) a field extension of \( k \). Take integers \( a, b, n \) with \( n, b \geq 0 \). Let \( F^n_{MW-T, a,b} E(F) \) denote the subgroup of \( \Pi_{a,b} E(F) \) generated by elements of the form

\[ T_{\tau, E}(w)^*(x); \quad x \in F^n_{MW} \Pi_{a,b} E(F(w)) \]

as \( w \) runs over closed points of \( \mathbb{A}^n_F \), separable over \( F \). We write \( F^n_{MW-T, \tau, E} E(F) \) for the filtration \( F^n_{MW-T, \tau, 0} E(F) \) on \( E \).

Theorem 7.11 of [12] expresses the “Tate-Postnikov” filtration \( F^n_{Tate} \pi_0 E(F) \) on \( \pi_0 E(F) \) in terms of the “Milnor-Witt” filtration \( F^n_{MW} \pi_0 E(F) \), under the connectivity assumption \( \Pi_{a,*} E = 0 \) for \( a < 0 \). With some minor changes, the proof of this result goes through to show:

**Theorem 5.3.** Let \( E \) be in \( \mathcal{SH}(k) \) and let \( E = \Omega_{\mathbb{F}}^E \). Suppose that \( \Pi_{a,*} E = 0 \) for \( a < 0 \). Let \( F \) be a perfect field extension of \( k \). Then for all \( r \geq 0, n \geq 0 \), we have an equality of subgroups of \( \Pi_{r,0} E(F) \):

\[ \rho_n(E)(\text{Fil}^{simp}_n(E) \pi_0 f_0(E)) = F^n_{MW-T, \pi_0 E(F)} \]

**Sketch of proof.** We briefly indicate the changes that need to be made in the arguments for theorem 7.11 loc. cit.: In [12] prop. 4.3 and thm. 7.6], replace \( \pi_0 \) with \( \pi_r \) and \( F^n_{Tate} \pi_0 E(F) \) with \( \rho_n(E)(\text{Fil}^{simp}_n(E) \pi_r f_0(E)) \). Also, instead of using the fact that \( \theta_*(F) \) is an isomorphism, we need only use that \( \theta_*(F) \) is a ring homomorphism, and that for \( u \in F^x, \theta_1([u]) \) is the element of \( \Pi_{0,-n} \mathcal{S}_k(F) \) coming from the map \( u : \text{Spec} F \rightarrow \mathbb{G}_m \) corresponding to \( u \) (see [17, theorem 6.3.3]). \( \square \)

A result of Morel [13, corollary 6.41] implies that \( \theta_*(F) \) is an isomorphism for all fields \( F \), but we will not need this.
6. Finite spectra and cohomologically finite spectra

In this section we will have occasion to use some localizations of \(SH(k)\) with respect to multiplicatively closed subsets of \(\mathbb{Z} \setminus \{0\}\). To make our discussion precise, we have collected notations and some elementary facts concerning such localizations in Appendix B; we will use these without further mention in this section.

For a field \(F\), let \(F^{sep}\) denote the separable closure of \(F\) and \(G_F\) the absolute Galois group \(\text{Gal}(F^{sep}/F)\). For \(p\) a prime, let \(cd_p(G_F)\) denote the \(p\)-cohomological dimension of the profinite group \(G_F\), as defined in [26, I, §3.1]. We write \(cd_p(F)\) for \(cd_p(G_F)\), and call \(cd_p(F)\) the \(p\)-cohomological dimension of \(F\). The cohomological dimension of \(F\), \(cd(F)\), is the supremum of the \(cd_p(F)\) over all primes \(p\).

We turn to our main topic in this section: the study of finite and cohomologically finite objects in \(SH(k)\).

**Definition 6.1.** 1. Let \(SH(k)_{\text{fin}}(k)\) be the thick subcategory of \(SH(k)\) generated by objects \(\Sigma^n_T \Sigma^\infty_X X_+\) for \(X\) a smooth projective \(k\)-scheme and \(n \in \mathbb{Z}\).

2. Let \(SH(k)_{\text{coh.fin}}(k)\) be the full subcategory of \(SH(k)\) with objects those \(E\) such that
   - (i) there is an integer \(d\) such that, for \(n > d\), \(\Pi_{r,n}(E)_{\mathbb{Q}} = 0\) for all \(r\),
   - (ii) there is an integer \(c\) such that \(\Pi_{r,q}(E) = 0\) for \(r \leq c\), \(q \in \mathbb{Z}\).

3. For \(E \in SH(k)_{\text{coh.fin}}(k)\), define \(d(E)\) to be the infimum among integers \(d\) such that \(\Pi_{r,n}(E)_{\mathbb{Q}} = 0\) for all \(r\) and for all \(n > d\).

4. We say that \(E \in SH(k)\) is *topologically \(c\)-connected* if \(\Pi_{r,E} = 0\) for \(r \leq c\).

5. For \(E \in SH_S^1(k)\), we say that \(E\) is topologically \(c\)-connected if \(\Pi_{r,E} = 0\) for \(r \leq c\) and \(n \geq 0\); we say that \(E\) is \(c\)-connected if \(\pi_r E = 0\) for \(r \leq c\).

**Remarks 6.2.** 1. \(SH(k)_{\text{coh.fin}}(k)\) is a thick subcategory of \(SH(k)\).

2. As \(\Sigma^n_T \Sigma^\infty_X X_+\) is compact for all \(n \in \mathbb{Z}\), \(X \in \text{Sm}/k\), \(SH(k)_{\text{fin}}(k)\) is contained in the category \(SH(k)^c\) of compact objects in \(SH(k)\). If \(k\) admits resolution of singularities, it is not hard to show that \(\Sigma^n_T \Sigma^\infty_X X_+\) is in \(SH(k)_{\text{fin}}(k)\) for all \(X \in \text{Sm}/k\), \(n \in \mathbb{Z}\); as these objects (and their translates) form a set of compact generators for \(SH(k)\), it follows that \(SH(k)_{\text{fin}}(k) = SH(k)^c\) if \(k\) admits resolution of singularities, that is, if \(k\) has characteristic zero.

**Remark 6.3.** For a presheaf \(A\) of abelian groups on \(\text{Sm}/k\), we have the presheaf \(A_{-1}\) defined by

\[
A_{-1}(U) := \ker(i^*_r : A(U \times \mathcal{A}^1 \setminus \{0\}) \to A(U)).
\]

Let \(E\) be in \(SH_S^1(k)\). By [17, lemma 4.3.11], there is a natural isomorphism \(\pi_n(\Omega_{G_n}^m E) \cong \pi_n(E)_{-1}\). In particular, if \(E\) is \(c\)-connected, then so is \(\Omega_{G_n}^m E\). As \(\Pi_{a,b} E = \Pi_{a,b-1} \Omega_{G_n} E\) and \(\Pi_{r,0} = \pi_r\), we see that \(E\) is \(c\)-connected if and only if \(E\) is topologically \(c\)-connected.
Lemma 6.4. Take $E \in \mathcal{SH}_{\text{fin}}(k)$. Then there is an integer $n(E)$ such that $\rho_n(E) : f_n(E) \to E$ is an isomorphism for all $n \leq n(E)$.

Proof. As $f_n$ is exact, it suffices to prove the result for $E = \Sigma_T^{n} \Sigma_{\infty}^T X_+$ with $X$ smooth and projective over $k$ and $m \in \mathbb{Z}$. As $f_{n} \circ \Sigma_T^m = \Sigma_T^m \circ f_{n-m}$, we need only prove the result for $E = \Sigma_T^{n} X_+$. But $\Sigma_T^{\infty} X_+$ is in $\mathcal{SH}^{e/f}(k)$, so $\rho_n(\Sigma_T^{\infty} X_+)$ is an isomorphism for all $n \leq 0$. □

Lemma 6.5. Let $E$ be in $\mathcal{SH}(k)_{\text{coh.fin}}(k)$.
1. For $U \in \text{Sm}/k$, we have $\text{Hom}_{\mathcal{SH}(k)_{\text{q}}}((\Sigma_S^{p}, \Sigma_{\infty}^S \Sigma_T^S U_+)_{\text{q}}, E_{\text{q}}) = 0$ for all $p \in \mathbb{Z}$, $q > d(E)$.
2. For all $n > d(E)$, $(f_n E)_{\text{q}} \cong 0$ in $\mathcal{SH}(k)_{\text{q}}$.

Proof. Let $i_n : \Sigma_T^n \mathcal{SH}^{e/f}(k) \to \mathcal{SH}(k)$ be the inclusion, $r_n$ the right adjoint to $i_n$; recall that $f_n := i_n \circ r_n$. By Example 8.2, the $\mathbb{Q}$-localization $\Sigma_T^p \mathcal{SH}^{e/f}(k)_{\mathbb{Q}}$ of $\Sigma_T^p \mathcal{SH}^{e/f}(k)$ is the localizing subcategory of $\mathcal{SH}(k)_{\mathbb{Q}}$ generated by the objects $(\Sigma_T^q \Sigma_{\infty}^T U_+)_{\mathbb{Q}}$, $q \geq n$, $U \in \text{Sm}/k$, the inclusion $\Sigma_T^n \mathcal{SH}^{e/f}(k)_{\mathbb{Q}} \to \mathcal{SH}(k)_{\mathbb{Q}}$ is given by $i_n$, the right adjoint to $i_n$ is $r_n$ and $f_n(E)_{\mathbb{Q}} = (f_n E)_{\mathbb{Q}}$ for all $E \in \mathcal{SH}(k)$.

Using this, we see that (1) implies (2), as (1) implies that for $F \in \Sigma_T^n \mathcal{SH}^{e/f}(k)_{\mathbb{Q}}$, $n > d(E)$,

$$\text{Hom}_{\mathcal{SH}(k)_{\text{q}}}(F, E_{\text{q}}) = 0.$$ 

Since $(f_n E)_{\mathbb{Q}} \to E_{\mathbb{Q}}$ is universal for maps $F \to E_{\mathbb{Q}}$ with $F \in \Sigma_T^n \mathcal{SH}^{e/f}(k)_{\mathbb{Q}}$, it follows that $(f_n E)_{\mathbb{Q}} = 0$ for all $n > d(E)$.

For (1), by Lemma 5.2, we have

$$\text{Hom}_{\mathcal{SH}(k)_{\text{q}}}((\Sigma_S^{p}, \Sigma_{\infty}^S \Sigma_T^S U_+)_{\text{q}}, E_{\text{q}}) = \text{Hom}_{\mathcal{SH}(k)}((\Sigma_S^{p}, \Sigma_{\infty}^S \Sigma_T^S U_+)_{\text{q}}, E_{\text{q}}).$$

Thus, tensoring the usual Gersten-Quillen spectral sequence with $\mathbb{Q}$ gives the strongly convergent spectral sequence

$$E_1^{a,b} = \bigoplus_{u \in U} \Pi_{-b+q} \mathcal{E}(k(u))_{\text{q}} \Rightarrow \text{Hom}_{\mathcal{SH}(k)_{\text{q}}}((\Sigma_S^{a-b} \Sigma_{\infty}^S \Sigma_T^S U_+)_{\text{q}}, E_{\text{q}}),$$

concentrated in the range $0 \leq a \leq \dim U$. The assumption $q > d(E)$ implies that $E_1^{a,b} = 0$ for all $a, b$, proving (1). □

Lemma 6.6. Let $E$ be in $\mathcal{SH}(k)_{\text{coh.fin}}(k)$. Then $f_n E$ is in $\mathcal{SH}(k)_{\text{coh.fin}}(k)$, $d(f_n E) \leq d(E)$ and $c(f_n E) \geq c(E)$, for all $n \in \mathbb{Z}$.

Proof. We have shown in 12 that, for $E \in \mathcal{SH}_{S^1}(k)$, if $E$ is topologically $-1$-connected, then $f_n E$ is also topologically $-1$-connected for all $p$. For $E := \Omega_T^\infty \Sigma_{-1}^{q-1} \Sigma_{\infty}^T X_+$, we have

$$\Pi_{r-c-1,m+q} f_n E \cong \Pi_{r,m+q} \Sigma_{\infty}^T \Sigma_{\infty}^T f_n E \cong \Pi_{r,m} f_n E$$

for $m \geq -q$. Similarly, $\Pi_{r-c-1,m+q} E \cong \Pi_{r,m} E$ for $m \geq -q$. Thus, if $\Pi_{r,m} E = 0$ for $r \leq c$ and $m \geq -q$, the same holds for $f_n E$. As $q$ was arbitrary, we see that if $E$ is topologically $c$-connected, so is $f_n E$ for all $n$, that is, $c(f_n E) \geq c(E)$.

We have already seen in Lemma 5.3 that $f_n E_{\text{q}} = 0$ for $n > d(E)$, hence $f_n E$ is in $\mathcal{SH}(k)_{\text{coh.fin}}(k)$ and $d(f_n E) = -\infty$ for $n > d(E)$. For $n \leq d(E)$, take
We also write \( \tau \) for the induced automorphism of the sphere spectrum \( S_k \). Morel [17, section 6.1] has considered the action of \( \mathbb{Z}/2 \) on the sphere spectrum \( S_k \) arising from the exchange of factors

\[ \tau : T \land T \to T \land T. \]

We also write \( \tau \) for the induced automorphism of the sphere spectrum \( S_k \). Morel [17, remark 6.3.5] identifies the corresponding element of \( \text{End}(S_k) \) (denoted \(-e\) in loc. cit.) as

\[ \tau := \theta_k(1 + \eta \cdot [-1]). \]

After inverting 2, the action of \( \tau \) decomposes \( S_k[\frac{1}{2}] \) into its +1 and -1 eigenspaces

\[ S_k[\frac{1}{2}] = S_k^+ \oplus S_k^-; \]

as \( S_k[\frac{1}{2}] \) is the unit in the tensor category \( S(h)[\frac{1}{2}] \), this induces a decomposition of \( S(h)[\frac{1}{2}] \) as

\[ S(h)[\frac{1}{2}] = S(h)^+ \times S(h)^-. \]

This extends to a decomposition of \( S(h)_\mathbb{Q} \) as \( S(h)_\mathbb{Q}^+ \times S(h)_\mathbb{Q}^- \). For an object \( E \) of \( S(h)(k) \), we write the corresponding factors of \( E_\mathbb{Q} \) as \( E_\mathbb{Q}^+, E_\mathbb{Q}^- \).

**Lemma 6.7.** Suppose that either

1. char. \( k = 0 \) and \( cd_2(k) < \infty \), or
2. char. \( k > 0 \).

Then \( 2^{n+1} \cdot \eta = 0 \), where in case (1), \( N = cd_2(k) \), and in case (2), \( N = 0 \) if char. \( k = 2 \) or if char. \( k = p \) is odd and \( p \equiv 1 \mod 4 \), \( N = 1 \) if \( p \equiv 3 \mod 4 \). Moreover, letting \( I \subset GW(k) \) be the augmentation ideal, we have \( I^n = 0 \) for \( n > cd_2(k) \), assuming \( k \) has characteristic \( \neq 2 \).

**Proof.** Suppose that char. \( k \neq 2 \). Sending \( (a_1, \ldots, a_n) \in (k^\times)^n \) to the \( n \)-fold Pfister form \( \langle a_1, \ldots, a_n \rangle \) descends to a well-defined surjective homomorphism \( p_n : K^M_n(k)/2 \to I^n/I^{n+1} \) [13, §4.1], in particular \( 2I^n \subset I^{n+1} \) for all \( n \geq 0 \). By the Milnor conjecture [29, theorem 6.6 and theorem 7.4] and [19, theorem 4.1], \( p_n \) is an isomorphism, and induces an isomorphism

\[ I^n/I^{n+1} \cong H_{et}^{2n}(k, \mu_{2^n}). \]
Thus $I^n = I^{n+1}$ for $n \geq N + 1$, where $N := \text{cd}_2 k$. By the theorem of Arason-Pfister [I] Korollar 1, $\cap_n I^n = \{0\}$, hence $I^{N+1} = 0$, and thus $2^{N+1}$ kills the Witt group $W(k)$.

As a $K_0^{MW}(k)$-module, $K_{-1}^{MW}(k)$ is cyclic with generator $\eta$. However, $K_{-1}^{MW}(k) \cong W(k)$ by [14] lemma 3.10, and thus $2^{N+1}\eta = 0$. This handles the case (1).

In case $k$ has characteristic $p > 0$, then as $\eta$ comes from base extension from the prime field $\mathbb{F}_p$, it suffices to show that $2^{N+1}\eta = 0$ in $\mathcal{SH}(\mathbb{F}_p)$, with $N$ as in the statement of the lemma.

For $p$ odd, we have $GW(\mathbb{F}_p) \cong K_0^{MW}(\mathbb{F}_p)$, with $1 + \eta[-1]$ corresponding to the form $-x^2$; the hyperbolic form $x^2 - y^2$ corresponds to $2 + \eta[-1]$. If $p \equiv 1 \mod 4$, then $-1$ is a square, hence $-x^2$ and $x^2$ are isometric forms and $1 + \eta[-1] = 1$ in $K_0^{MW}(\mathbb{F}_p)$. The relation $\eta(2 + \eta[-1]) = 0$ in $K_{-1}^{MW}(\mathbb{F}_p)$ thus simplifies to $2\eta = 0$.

If $p \equiv 3 \mod 4$, then $-1$ is a sum of two squares, and hence the quadratic form $x^2 + y^2 + z^2$ is isotropic. Thus the Pfister form $x^2 + y^2 + z^2 + w^2$ is also isotropic and hence hyperbolic, and hence the form $-x^2 - y^2 - z^2 - w^2$ is hyperbolic as well. Translating this back to $K_0^{MW}(\mathbb{F}_p)$ gives the relation $4(1 + \eta[-1]) = 2(2 + \eta[-1])$ or $2\eta[-1] = 0$. Combining this with relation $\eta(2 + \eta[-1]) = 0$ yields $4\eta = 0$.

In case $k$ has characteristic 2, $-1 = +1$. The relation $\eta(2 + \eta[-1]) = 0$ simplifies to $2\eta = 0$ (as $[1] = 0$ in $K_1^{MW}(F)$ for all fields $F$ [14] lemma 3.5).

□

Lemma 6.7 and (6.1) imply:

**Lemma 6.8.** Suppose that $k$ has finite 2-cohomological dimension or that char. $k > 0$. Then $\mathcal{SH}(k)[1/2] = \mathcal{SH}(k)^{+}$.

**Proposition 6.9.** Take $X \in \text{Sm}/k$. Then

1. $\Pi_{r,n}(\Sigma^r_\mathbb{G}_m X_+) = 0$ for $r < 0$ and $n \in \mathbb{Z}$

2. Suppose $X$ is smooth and projective over $k$ of dimension $d$ and that $k$ has finite 2-cohomological dimension or char. $k > 0$. Then $\Pi_{r,n}(\Sigma^r \mathbb{G}_m X_+)_Q = 0$ for $n > d$, $r \in \mathbb{Z}$.

3. Suppose $k$ has finite 2-cohomological dimension or char. $k > 0$. Then each $E$ in $\mathcal{SH}(k)_{\text{fin}}(k)$ is also in $\mathcal{SH}(k)_{\text{coh.fin}}(k)$.

**Proof.** Noting that $\mathcal{SH}(k)_{\text{coh.fin}}(k)$ is a thick subcategory of $\mathcal{SH}(k)$, and that

$$\Pi_{r,n}(\Sigma^r_\mathbb{G}_m X_+) \cong \Pi_{r-m,n-m}(\Sigma^r_\mathbb{G}_m X_+),$$

we see that (3) follows from (1) and (2).

We first prove (1). We have the $\mathcal{S}^1$-spectrum $\Sigma^\infty_\mathbb{S} X_+ \in \mathcal{SH}_{\text{S}}(k)$. By [8] theorem 10 we have

$$\Pi_{r,n}(\Sigma^\infty_\mathbb{S} X_+) \cong \lim_{\rightarrow b} \Pi_{r,n+b}(\Sigma^b X_+)$$

**Documenta Mathematica** 18 (2013) 907–941
so it suffices to see that \( \Sigma_{G_m}^b X_+ \) is topologically \(-1\) connected for all \( b \geq 0 \). By remark 6.3, we need only see that \( \Sigma_{G_m}^b X_+ \) is \(-1\) connected. This follows from Morel’s \( S^1 \)-stable \( \mathbb{A}^1 \)-connectedness theorem [17, theorem 4.2.10].

For (2), lemma 6.8 and our assumption on \( k \) imply that \( SH(k)_Q = SH(k)^+_Q \). By a result of Morel (proved in detail by Cisinski-Déglise [3, theorem 15.2.13]), there is an equivalence of triangulated categories \( SH(k)_Q^+ \cong DM(k)_Q \), and hence

\[
\text{Hom}_{SH(k)}(\Sigma_{G_m}^b \Sigma_{T}^\infty U_+, \Sigma_{T}^\infty X_+)_{Q} \\
\cong \text{Hom}_{SH(k)}(\Sigma_{G_m}^b \Sigma_{T}^\infty U_+, (\Sigma_{T}^\infty X_+)_{Q}) \\
\cong \text{Hom}_{DM(k)_Q}(M(U)(b)_{Q}[a+b], M(X)_{Q}),
\]

where \( M : Sm/k \to DM(k) \) is the canonical functor. Since \( M(U)(b) \) is compact in \( DM(k) \), we have

\[
\text{Hom}_{DM(k)_Q}(M(U)(b)_{Q}[a+b], M(X)_{Q}) \cong \text{Hom}_{DM(k)}(M(U)(b)[a+b], M(X)_{Q}).
\]

As \( X \) is smooth and projective, we may use duality (via, e.g., duality in Chow motives over \( k \) and [33, V, prop. 2.1.4])

\[
\text{Hom}_{DM(k)}(M(U)(b)[c], M(X)) \cong \text{Hom}_{DM(k)}(M(U \times X), Z(d-b)[2d-c])
\]

\[= H^{2d-c}(U \times X, Z(d-b)).\]

But for all \( Y \in Sm/k \), \( H^p(Y, Z(q)) = 0 \) for \( q < 0, p \in \mathbb{Z} \), by [28, corollary 2]. Thus the presheaf

\[
U \mapsto \text{Hom}_{SH(k)}(\Sigma_{G_m}^b \Sigma_{T}^\infty U_+, \Sigma_{T}^\infty X_+)_Q
\]

is zero for \( b > d \), and hence the associated sheaf \( \Pi_{a,b}(\Sigma_{T}^\infty X_+)_Q \) is zero as well.

\[
\square
\]

Remark 6.10. The statement of the result of Cisinski-Déglise [3, theorem 16.2.13] cited above is not the same as given here; Cisinski-Déglise show that the \( \mathbb{A}^1 \)-derived category over a base-scheme \( S \), with \( \mathbb{Q} \)-coefficients, \( D_{\mathbb{A}^1,Q}(S) \), has plus part \( D_{\mathbb{A}^1,Q}(S)^+ \) equivalent to the category of “Beilinson motives” \( DM_{Q}(S) \). For \( S \) geometrically unibranched, they show [3, theorem 16.1.4] that \( DM_{Q}(S) \cong DM(S)_{Q} \).

Furthermore, the well-known adjunction between simplicial sets and chain complexes extends to give an equivalence of \( SH(S)_Q \) with \( D_{\mathbb{A}^1,Q}(S) \) (see e.g. [3, §5.3.35]). Putting these equivalences all together gives us the equivalence \( SH(k)_{Q}^+ \cong DM(k)_{Q} \) we use in the proof of proposition 6.5.

Alternatively, one can also repeat the argument used by Cisinski-Déglise, replacing \( D_{\mathbb{A}^1,Q}(S) \) with \( SH(k)_Q \) and \( DM_{Z}(S) \) with \( Ho MZ-Mod_{Q} \), where \( MZ \) is the commutative monoid in symmetric motivic spectra over \( k \) defined in [25], and representing motivic cohomology in \( SH(k) \). This shows that the forgetful functor \( Ho MZ-Mod_{Q} \to SH(k)_{Q} \) induces an equivalence \( Ho MZ-Mod_{Q} \to SH(k)_{Q}^+ \). One can then use the theorem of Röndigs-Ostvær [25 theorem 1.1, discussion preceding theorem 1.2], which gives an equivalence of \( Ho MZ-Mod_{Q} \) with \( DM(k)_{Q} \) for \( k \) a perfect field.
7. The proof of the convergence theorem

The following result combines with lemma 2.3 to form the heart of the proof of our main theorem.

**Lemma 7.1.** Let \( k \) be a perfect field. Suppose that \( k \) has finite cohomological dimension \( D_k \). Let \( F \) be a perfect field extension of \( k \) with tr. dim \( F \leq d \). Let \( E = \Omega^2 \otimes k \) for some \( E \in SH_{coh, fin}(k) \) and suppose that \( \Pi_{a.} E = 0 \) for \( a < 0 \). Consider the map

\[
\rho_n(f_M E) : \pi_r f_n(f_M E)(F) \to \pi_r f_M E(F).
\]

Then, for all \( r, M \geq 0 \), \( n > \max(D_k + d, d(E)) \), we have

\[
\rho_n(f_M E)(\text{Fil}^{simp}_n(f_M E)\pi_r f_n(f_M E)(F)) = 0.
\]

**Proof.** Let \( A = \max(D_k + d, d(E)) \geq 0 \). By theorem 5.3 and the definition of the filtration \( F^n_{MW}, \pi_r f_M E(F) \), it suffices to show that, for every finite extension \( F' \) of \( F \), the product

\[
\bigcup: \Pi_{r,n} f_M E(F') \otimes K^{MW}_n(F') \to \Pi_{r,0} f_M E(F')
\]

is zero if \( n > A \).

We note that \( f_M E \cong \Omega^2 \otimes k \) for \( M \geq 0 \) (lemma 2.2). Thus, by lemma 6.6, \( \langle \Pi_{r,n} f_M E \rangle_0 = 0 \) for \( n > \max(d(E), -1) \) and the image of \( \bigcup \) is the same as the subgroup generated by the images of the maps

\[
\bigcup_N: \Pi_{r,n} f_M E(F') \otimes K^{MW}_n(F')/N \to \pi_r f_M E(F'); \quad N \in N.
\]

First suppose \( \text{char.} k \neq 2 \). By Morel’s theorem [16, theorem 5.3], we have a cartesian diagram

\[
\begin{array}{ccc}
K^{MW}_n(F') & \longrightarrow & K^{MW}_n(F') \\
\downarrow & & \downarrow \\
I(F')^n & \longrightarrow & I(F')^n/I(F')^{n+1}
\end{array}
\]

But as tr. dim \( F' \leq d \), we have \( cd_2(F') \leq cd(F) \leq D_k + d \) [26, II, §4.2, proposition 11]. Thus, for \( n > D_k + d \), lemma 6.7 tells us that \( K^{MW}_n(F') = K^{MW}_n(F') \). Furthermore, for \( N \) prime to the characteristic, the Bloch-Kato conjecture\(^6\) gives the isomorphism

\[
K^{MW}_n(F')/N \cong H^1_n(F', \mathbb{Z}_N),
\]

and hence \( K^{MW}_n(F')/N = 0 \) for \( n > D_k + d \). For \( N = p = \text{char.} k \), it follows from [2, theorem 2.1] that \( K^{MW}_n(F')/p = 0 \), since \( F' \) is perfect. Thus, for \( n > D_k + d \), \( K^{MW}_n(F')/N = 0 \), and hence the image of \( \bigcup \) is zero if \( n > A \).

If \( \text{char.} k = 2 \), then as \( F' \) is perfect, each \( u \in F' \) is a square. Thus, by [14, proposition 3.13], the quotient map \( K^{MW}_n(F') \to K^{MW}_n(F') \) is an isomorphism for \( n \geq 0 \). The remainder of the discussion is the same. \( \square \)

---

\(^6\)Proved by Rost and Voevodsky [32]. For a presentation of some of Rost’s results that go into the proof of Bloch-Kato, see, e.g., [3] [27].
Relying on lemma 7.4, here is the main step in the proof of theorem 4.

**Proposition 7.2.** Let $k$ be a perfect field of finite cohomological dimension and let $p$ be the exponential characteristic of $k$. Then there is a function $N_k : \mathbb{Z}^3 \to \mathbb{Z}$, with $N_k(e, r, d) \geq 0$ for all $e, d, r$, such that, given an integer $e$, and an $E$ in $\mathcal{SH}_{coh,\text{fin}}(k)$ with $c(E) \geq -1$ and $d(E) \leq e$, we have

$$\text{Fil}_n^N \Pi_{r,q} f_M \Omega^n_\pi \varepsilon(E)[1/p] = 0$$

for all $r, M \in \mathbb{Z}$, all $N \geq N_k(e - q, r, d) + q$, all $q \geq 0$, and all field extension $F$ of $k$ with $\text{tr. dim} F \leq d$.

**Proof.** For a field $F$, let $F^{per}$ be the perfect closure of $F$. By proposition A.1, $\text{Fil}_n^N \Pi_{r,q} G(F)[1/p] \cong \text{Fil}_n^N \Pi_{r,q} G(F^{per})[1/p]$ for all $G \in \mathcal{SH}(k)$, so we may replace $F$ with $F^{per}$ whenever necessary.

From our hypothesis $c(E) \geq -1$ and lemma 6.6 follows $c(f_M E) \geq -1$, so $\Pi_{r,q} \Omega^n_\pi f_M E = \Pi_{r,q} f_M \Omega^n_\pi E = 0$ for $r < 0, q \geq 0$. Thus, taking $N_k(e, r, d) = 0$ for $r < 0$, and all $d$ and $e$ settles the cases $r < 0$, so we may proceed by induction on $r$, defining $N_k(e, r, d)$ recursively. We omit the $[1/p]$ in the notation, using the convention that we invert $p$ throughout the proof of this proposition.

Assume $r \geq 0$ and that we have defined $N_k(e, j, d)$ for $j \leq r - 1$ (and all $e, d$) so that the proposition holds for $\text{Fil}_n^N \Pi_{r,q} f_M \Omega^n_\pi E(F)$, $r < j$; we assume in addition that $N_k(e, j, d) = 0$ for $j < 0$. Define $N_j(d, q; E) := N_k(e - q, j, d) + q$ for $j < r$. Then (1.3) is satisfied for $q, M \geq 0, m \geq N_j(d, q; E)$, $j = 0, \ldots, r - 1$. By lemma 1.5, it follows that for all $n \geq 0$, all fields $F$ with $\text{tr. dim} F \leq d$ and all $m \geq \max_{j=0}^{r-1} N_j(r - j + d, n; E)$ (or $m \geq 0$ in case $r = 0$), we have

$$\pi_r(f_n f_m E(F)) = \text{Fil}_n^{simp} (f_n E) \pi_r(f_n f_m E(F)).$$

As the filtration $\text{Fil}_n^{simp}(G) \pi_r(f_n G)(F)$ is functorial in $G$, this implies that

$$f_n(\rho_m(E))(\pi_r(f_n f_m E(F)) \subset \text{Fil}_n^{simp}(E) \pi_r(f_n E(F))$$

for $n, m, F$ as above.

If we take $n \geq \max(D_k + d, d(E)) + 1$, then by lemma 7.1,

$$\text{im}(\text{Fil}_n^{simp}(E) \pi_r(f_n E)(F^{per}) \xrightarrow{\rho_m(E)} \pi_r(E)(F^{per})) = 0.$$

For $m \geq n$, we have $f_m f_n = f_m$ and $\rho_n(E) \circ f_n(\rho_m(E)) = \rho_m(E)$ (lemma 2.1). Thus, for

$$n = \max(D_k + d, e) + 1, \ m \geq \max_{j=0}^{r-1} N_j(r - j + d, n; E),$$

we have

$$0 = \rho_n(E) \circ f_n(\rho_m(E))(\pi_r(f_m f_n E)(F^{per})) = \rho_m(\pi_r(f_m E)(F^{per})) \subset \pi_r E(F^{per}).$$

Let $n = \max(D_k + d, e) + 1$ and define

$$N_k(e, r, d) := n + \max_{j=0}^{r-1} N_k(e - n, j, r - j + d).$$

**Documenta Mathematica** 18 (2013) 907–941
Using the isomorphism $F^N_{\text{Tate}} \pi_r E(F) \cong F^N_{\text{Tate}} \pi_r E(F^{pr})$, we thus have

$$F^N_{\text{Tate}} \pi_r E(F) = 0$$

for $N \geq N_k(e, r, d)$ and $F$ a field extension of $k$ with tr. dim$_k F \leq d$.

Fix $q \geq 0$ and let $E' = \Omega^q_F E$, $E' = \Omega^q_F E'$. Then $\Pi_{a,b+q} E' \cong \Pi_{a,b} E$ for all $b \in \mathbb{Z}$, hence $c(E') \geq -1$. Also, $d(E') = d(E) - q \leq e - q$, and we may apply our result with $E'$ replacing $E$, and $e - q$ replacing $e$. Since $\pi_r E' = \Pi_{r,q} E$ and $F^N \pi_r E' = F^N \Pi_{r,q} E$ (lemma 2.4), we thus have

$$F^N_{\text{Tate}} \Pi_{r,q} E(F) = 0$$

for $N \geq N_k(e - q, r, d) + q$, $q \geq 0$, and $F$ as above.

Now take an integer $M$. Then $c(f_M E) \geq -1$ and $d(f_M E) \leq d(E)$ (by lemma 6.0), so the same result holds for $E_M := \Omega^q_{F_M} E$. By lemma 2.2, $E_M \cong f_M E$, so

$$F^N_{\text{Tate}} \Pi_{r,q} f_M E(F)[1/p] = 0$$

for $q \geq 0$, $N \geq N_k(e - q, r, d) + q$, and all field extensions $F$ of $k$ with tr. dim$_k F \leq d$. As $M$ was arbitrary, the induction thus goes through, completing the proof. □

Now for the proof of the main theorem. For $x \in X \in \text{Sm}/k$, and $\mathcal{F}$ a sheaf on $\text{Sm}/k_{\text{Nis}}$, we let $\mathcal{F}_x$ denote the Nisnevich stalk of $\mathcal{F}$ at $x$.

**Theorem 7.3.** Let $k$ be a perfect field of finite cohomological dimension. Let $p$ be the exponential characteristic of $k$, let $x \in X \in \text{Sm}/k$ and let $d = \text{dim}_k X$.

Let $\mathcal{E}$ be in $\mathcal{SH}_{\text{coh}, \text{fin}}(k)$. Then for all integers $r, q$, and $M$, we have

$$(\text{Fil}^n_{\text{Tate}} \Pi_{r,q} f_M \mathcal{E})_x[1/p] = 0$$

for $n \geq N_k(d(\mathcal{E}) - q, r - c(\mathcal{E}) - 1, d) + q$, where $N_k$ is the integer-valued function given by proposition 7.2.

As $\mathcal{SH}(k)_{\text{fin}}$ is a subcategory of $\mathcal{SH}(k)_{\text{coh}, \text{fin}}$ in case $k$ has finite cohomological dimension (proposition 6.9(3)), our main theorem 4 follows immediately from theorem 7.3 with $N(\mathcal{E}, r, d, q) := N_k(d(\mathcal{E}) - q, r - c(\mathcal{E}) - 1, d) + q$.

**Proof of theorem 7.3.** As before, we invert $p$ throughout the proof and omit the "$[1/p]$" from the notation.

The proof of [15] lemma 6.1.4 shows that, for $x \in X \in \text{Sm}/k$, $X_x := \text{Spec} \mathcal{O}_{U,x}$ and $U \subset X_x$ open, the map $X_x \to X_x/U$ in $\mathcal{H}(k)$ is equal to the map sending $X_x$ to the base-point of $X_x/U$. This implies that for any $\mathcal{F} \in \mathcal{SH}(k)$, $a, b \in \mathbb{Z}$, the restriction map

$$[\Sigma^a_{\mathcal{S}_k}, \Sigma^b_{\mathcal{G}_m} \Sigma^\infty X_{x+}, \mathcal{F}]_{\mathcal{SH}(k)} \to [\Sigma^a_{\mathcal{S}_k}, \Sigma^b_{\mathcal{G}_m} \Sigma^\infty U_+, \mathcal{F}]_{\mathcal{SH}(k)}$$

is injective. Passing to the limit over $U$, this shows that the restriction map

$$\Pi_{a,b}(\mathcal{F})_x \to \Pi_{a,b}(\mathcal{F})(k(X))$$

is injective. From this it easily follows that the restriction map

$$(\text{Fil}^n_{\text{Tate}} \Pi_{r,q} f_M \mathcal{E})_x \to (\text{Fil}^n_{\text{Tate}} \Pi_{r,q} f_M \mathcal{E})(k(X))$$

is injective.
Thus, it suffices to show that $\text{Fil}^n_{\text{Tate}} \Pi_{r,q} f_M E(F) = 0$ for $M \in \mathbb{Z}$, for $n \geq N(d(E) - q, r - c(E) - 1, d) + q$, and for all finitely generated fields $F$ over $k$ with $\text{tr. dim}_k F \leq d$.

As

$$\text{Fil}^n_{\text{Tate}} \Pi_{r+p,q} \Sigma_{S^1} E = \text{Fil}^n_{\text{Tate}} \Pi_{r,q} E$$

we may replace $E$ with $\Sigma_{S^1}^{-c(E) - 1} E$ and assume that $c(E) \geq -1$. Similarly, if $q < 0$ we may replace $E$ with $\Sigma_{G_m}^{-q} E$, since

$$\text{Fil}^n_{\text{Tate}} \Pi_{r+q} f_M E = \text{Fil}^n_{\text{Tate}} \Pi_{r,q} f_M E$$

and $d(\Sigma_{G_m}^{-q} E) = d(E) - q$. This reduces us to the case $q \geq 0$ and $c(E) \geq -1$.

Letting $E = \Omega^2_{\mathbb{F}_p} E$, we have

$$\text{Fil}^n_{\text{Tate}} \Pi_{r,q} f_M E = \text{Fil}^n_{\text{Tate}} \Pi_{r,q} \Omega^\infty_{\mathbb{F}_p} f_M E = \text{Fil}^n_{\text{Tate}} \Pi_{r,q} f_M E$$

for all $M \in \mathbb{Z}$, $q \geq 0$, $n \geq 0$, so the result follows from proposition 7.2.

**Corollary 7.4.** Take $k$, $x$, $d$ and $E$ as in theorem 7.3. Then for all $q \in \mathbb{Z}$, $(\Pi_{r,q} f^*_M E)_{x[1/p]} = 0$ for all $M \geq N_k(d(E) - q, r - c(E) - 1, d) + q$.

**Proof.** Indeed, for $M \geq n$, $f_n f_M \cong f_M$, hence $\text{Fil}^n_{\text{Tate}} \Pi_{r,q} f_M E = \Pi_{r,q} f_M E$. The result thus follows from theorem 7.3.

**Remark 7.5.** Although $N_k(e, r, d)$ is defined recursively, it has a simple expression: For $D_k + d \geq 0$,

$$N_k(e, r, d) = \begin{cases} (r + 1)(D_k + d) + \frac{1}{2}(r + 1)(r + 2) & \text{if } e \leq D_k + d, r \geq 0, \\ e + r(D_k + d) + \frac{1}{2}(r + 1)(r + 2) & \text{if } e > D_k + d, r \geq 0, \\ 0 & \text{if } r < 0. \end{cases}$$

In particular, $N_k(e, r, d)$ is an increasing function in each variable (assuming $D_k + d \geq 0$). Thus, one can apply theorem 7.3 or corollary 7.4 even if one only has an upper bound for $d(E)$ and a lower bound for $c(E)$.

For instance, for $Y \in \text{Sm}/k$, we have $d(\Sigma_{G_m}^{-q} Y) \leq \dim_k Y$ and $c(\Sigma_{G_m}^{-q} Y) \geq -1$. Thus, for $x \in X \in \text{Sm}/k$, $r \geq 0$, $M \in \mathbb{Z}$, we have $(F^N_{\text{Tate}} \Pi_{r,q} f_M \Sigma_{G_m}^{-q} Y)_x = 0$ for

$$N \geq \max(\dim_k Y - q, D_k + \dim_k X) + r(D_k + \dim_k X) + \frac{1}{2}(r + 1)(r + 2) + q.$$ 

We do not know if any of these bounds are sharp.

**Appendix A. Norm maps**

Suppose our perfect base-field $k$ has characteristic $p > 0$. For an abelian group $A$, we write $A'$ for $A \otimes_{\mathbb{Z}} \mathbb{Z}[1/p]$. Our task in this section is to prove

**Proposition A.1.** 1. Suppose that $k$ is a perfect field of characteristic $p > 0$. Take $E \in \mathcal{SH}(k)$ and let $\alpha : F \to L$ be a purely inseparable extension of finitely generated fields over $k$. Then the map

$$\alpha^* : \Pi_{a,b} E(F)' \to \Pi_{a,b} E(L)'$$
is injective.
2. Suppose that \( k \) is a finite field. Take \( E \in \mathcal{SH}(k) \) and let \( \alpha : k \to k' \) be a finite extension of degree \( n \), let \( F' \) be a field extension of \( k \), and let \( F'' = F \otimes_k k' \). Then the kernel of
\[
\alpha^* : \Pi_{a,b}E(F) \to \Pi_{a,b}E(F''
\]
is \( n \)-torsion.

As one would expect, we construct a quasi-inverse to \( \alpha^* \) by constructing transfers. We first discuss (1). Take \( a \in F^\times \setminus (F^\times)^p \) and consider the extension \( F_a := F(a^{1/p}) \). We have the corresponding closed point \( p(a) \) of \( \mathbb{A}_k^1 \) defined by the homogeneous ideal \( (X^p - a) \subset F[X] \). This gives us the closed point \( p(a) \) of \( \mathbb{P}_F^1 \) via the standard open immersion \( x \mapsto (1 : x) \) of \( \mathbb{A}_k^1 \) into \( \mathbb{P}_k^1 \).

Suppose \( F = k(U) \) for some finite type \( k \)-scheme \( U \). Since \( k \) is perfect, \( U \) has a dense open subscheme smooth over \( k \); shrinking \( U \) and changing notation, we may assume that \( U \) is affine and smooth over \( k \), and also that \( a \) is a global unit on \( U \). Let \( V \subset \mathbb{A}_k^1 \times U \) be the closed subscheme defined by \( X^p - a \). Then \( V \) is reduced and irreducible (since \( a \notin (F^\times)^p \) and is finite over \( U \). Shrinking \( U \) again, we may assume that \( V \) is also smooth and affine over \( k \). Let \( I_V \subset k[U][t] \) be the ideal defining \( V \) in \( \mathbb{A}_k^1 \times U \).

Using \( X^p - a \) as generator for \( I_V/I_V^2 \), we have the Morel-Voevodsky purity isomorphism [18 theorem 3.2.23]
\[
\mathbb{A}_k^1 \times U/(\mathbb{A}_k^1 \times U \setminus V) \cong \mathbb{P}_V^1.
\]

Combining with the excision isomorphism \( \mathbb{A}_k^1 \times U/(\mathbb{A}_k^1 \times U \setminus V) \cong \mathbb{P}_U^1 \times U/(\mathbb{P}_U^1 \times U \setminus V) \) and passing to the limit over open subschemes of \( U \) gives us the sequence of maps of pro-objects in \( \mathcal{H}_\ast(k) \):
\[
(\mathbb{P}_U^1, \infty) \to (\mathbb{P}_F^1/\mathbb{P}_F^1 \setminus \{p(a)\}) \cong (\mathbb{P}_{F_a}^1, \infty);
\]
we denote the composition by \( \text{Tr}_{p(a)/F} \). Passing to \( \mathcal{SH}(k) \) gives us the morphism
\[
\text{Tr}_{p(a)/F} : S_k \otimes \text{Spec} F_+ \to S_k \otimes p(a)_+ = S_k \otimes \text{Spec} F_{a+},
\]
where we consider these objects as pro-objects in \( \mathcal{SH}(k) \). Composing with the map induced by the structure morphism \( \pi_a : p(a) \to \text{Spec} F \) gives us the endomorphism \( \pi_a \circ \text{Tr}_{p(a)/F} \) of \( S_k \otimes \text{Spec} F_+ \).

Proposition [A.1.1] is an immediate consequence of

**Lemma A.2.** After inverting \( p \), the endomorphism
\[
\pi_a \circ \text{Tr}_{p(a)/F} : S_k \otimes \text{Spec} F_+ \to S_k \otimes \text{Spec} F_+
\]
is an isomorphism.

**Proof.** We consider a deformation of \( \text{Tr}_{F_a/F} \). Let \( P(a) \subset \text{Spec} F[t][X] \) be the closed subscheme defined by the ideal \((X^p + t(X + 1) + (t - 1)a) \). We have
\[
d(X^p + t(X + 1) + (t - 1)a) = (X + a + 1)dt + tdX \in \Omega_{F[t][X]/F}.
\]
Thus the singular locus of \( P(a) \) (over \( F \)) is given by
\[
(X + a + 1 = t = 0) \cap (X^p + t(X + 1) + (t - 1)a = 0);
\]
Convergence of Voevodsky’s Slice Tower

935

since \( a \notin (F^{+})^p \), the singular locus is empty, i.e., \( P(a) \) is smooth over \( F \). Using \( X^p + t(X + 1) + (t - 1)a \) as the generator for \( I_{P(a)}/I_{P(a)}^2 \), we have as above the map

\[
\text{Tr}_{P(a)/F[t]} : \mathbb{P}^1_k \wedge \text{Spec } F[t]_+ \to \mathbb{P}^1_k \wedge P(a)_+
\]
of pro-objects in \( \mathcal{H}_\bullet(k) \), and the endomorphism

\[
\pi_{P(a)/F[t]} \circ \text{Tr}_{P(a)/F[t]} : S_k \wedge \text{Spec } F[t]_+ \to S_k \wedge \text{Spec } F[t]_+
\]
of pro-objects in \( \mathcal{SH}(k) \).

Setting \( t = 1 \) gives us the closed subscheme \( p_0 \) of \( \text{Spec } F[X] \) defined by the ideal \( (X^p + X + 1) \). Clearly \( \pi_0 : p_0 \to \text{Spec } F \) is finite and étale of degree \( p \).

Using \( X^p + X + 1 \) as a generator of \( I_{p_0}/I_{p_0}^2 \), gives us the map

\[
\text{Tr}_{p_0/F} : (\mathbb{P}^1_k, \infty) \wedge \text{Spec } F_+ \to (\mathbb{P}^1_k, \infty) \wedge p_0+
\]
of pro-objects in \( \mathcal{H}_\bullet(k) \), and the endomorphism

\[
\pi_0 \circ \text{Tr}_{p_0/F} : S_k \wedge \text{Spec } F_+ \to S_k \wedge \text{Spec } F_+
\]
of pro-objects in \( \mathcal{SH}(k) \).

The map \( \pi_{P(a)/F[t]} \circ \text{Tr}_{P(a)/F[t]} \) thus gives us an \( A^1 \)-homotopy between the maps \( \pi_0 \circ \text{Tr}_{P(a)/F} \) and \( \pi_0 \circ \text{Tr}_{p_0/F} \). Thus, these maps are equal in \( \mathcal{SH}(k) \), and it suffices to show that \( \pi_0 \circ \text{Tr}_{p_0/F} \) is an isomorphism after inverting \( p \).

As \( X^p + X + 1 \) has coefficients in \( F_p \), \( \pi_0 \circ \text{Tr}_{p_0/F} \) arises as base-extension from the similarly defined map

\[
\pi_0 \circ \text{Tr}_{p_0/F_p} : S_{F_p} \to S_{F_p}
\]
in \( \mathcal{SH}(F_p) \), that is, from the corresponding element \( [\pi_0 \circ \text{Tr}_{p_0/F_p}] \in [S_{F_p}, S_{F_p}]_{\mathcal{SH}(F_p)} \).

By Morel’s theorem \([S_{F_p}, S_{F_p}]_{\mathcal{SH}(F_p)} \cong \text{GW}(F_p)\) (see [14] lemma 3.10, corollary 6.41). Since \( [p_0(p_0)] : F_p = p \), it follows that the image of \( \pi_0 \circ \text{Tr}_{p_0/F_p} \in [S_{F_p}, S_{F_p}]_{\mathcal{SH}(F_p)} \) under the rank homomorphism \( \text{GW}(F_p) \to \mathbb{Z} \) is \( p \). Since the augmentation ideal \( I \subset \text{GW}(F_p) \) is nilpotent (in fact \( I^2 = 0 \)), it follows that \( \pi_0 \circ \text{Tr}_{p_0/F_p} \) is a unit in \( \text{GW}(F_p)[1/p] \), completing the proof of (1).

The proof of proposition [A.1(2)] is similar. Write \( k' = k(a) \), let \( f(x) \in k[x] \) be the minimal polynomial of \( a \) and let \( p \in \mathbb{A}^1_k \) be the closed point defined by the ideal \( (f(x)) \). Using \( f(x) \) as the generator for \( I_p/I_p^2 \), the Morel-Voevodsky purity isomorphism [18 loc. cit.] defines the map

\[
\text{Tr}_{p/k} : (\mathbb{P}^1_k, \infty) \to (\mathbb{P}^1_k, \infty) \wedge p_+
\]
in \( \mathcal{H}_\bullet(k) \), and the endomorphism

\[
\pi_p \circ \text{Tr}_{p/k} : S_k \to S_k
\]
in \( \mathcal{SH}(k) \). We can use Morel’s theorem again and identify \( \text{End}(S_k) \) with \( \text{GW}(k) \).

Under this identification, \( \pi_p \circ \text{Tr}_{p/k} \) corresponds to an element of \( \gamma \in \text{GW}(k) \), mapping to the degree \( n \) under the rank homomorphism \( \text{GW}(k) \to \mathbb{Z} \). Since the augmentation ideal \( I \subset \text{GW}(k) \) is nilpotent, this implies that there is an element \( \beta \in \text{GW}(k) \) with \( \beta \cdot \gamma = n \). Viewing \( \beta \) as an endomorphism of \( S_k \), this
categories, there are a number of definitions possible, as well as some questions what such a localization is and how to define it, however for “large” triangulated If every object of the triangulated category is compact, there is no issue about

Although this material is quite elementary, these facts on localization do not

In what follows, \( \mathcal{T} \) will be a triangulated category admitting arbitrary small coproducts.

**Remark** B.1. Suppose \( \mathcal{R} \) is a localizing subcategory of \( \mathcal{T} \) with a set of compact generators. Then (see e.g. [22, construction 1.6]), the Verdier localization \( g : \mathcal{T} \to \mathcal{T}/\mathcal{R} \) exists and admits a right adjoint \( r : \mathcal{T}/\mathcal{R} \to \mathcal{T} \). \( r \) induces an equivalence of \( \mathcal{T}/\mathcal{R} \) with \( \mathcal{R}^\perp \), this being the full subcategory of objects \( X \in \mathcal{T} \) such that \( \text{Hom}_\mathcal{T}(A,X) = 0 \) for all \( A \in \mathcal{R} \). If \( R \) is a set of compact generators for \( \mathcal{R} \), then \( \mathcal{R}^\perp = R^\perp \).

Let \( R \) be a set of compact objects in \( \mathcal{T} \) and \( \mathcal{R} \) the localizing subcategory of \( \mathcal{T} \) generated by \( R \), that is, the smallest localizing subcategory of \( \mathcal{T} \) containing \( R \). We recall from [21, theorem 2.1] that \( R \) is a set of generators for \( \mathcal{R} \), that is, if \( X \) is an object of \( \mathcal{R} \) such that \( \text{Hom}_\mathcal{R}(A,X) = 0 \) for all \( A \in \mathcal{R} \), then \( X \cong 0 \). Furthermore, the subcategory \( \mathcal{R}^c \) of compact objects in \( \mathcal{R} \) is the thick subcategory of \( \mathcal{R} \) generated by \( R \).

Let \( S \) be a multiplicatively closed subset of \( \mathbb{Z}\setminus\{0\} \) containing 1. Call an object \( X \) in \( \mathcal{T} \) \( S \)-torsion if \( n\cdot\text{id}_X = 0 \) for some \( n \in S \). We let \( \mathcal{T}_{S\text{-tor}} \) be the localizing subcategory of \( \mathcal{T} \) generated by the compact \( S \)-torsion objects of \( \mathcal{T} \) and let \( \mathcal{T}_{S^{-1}2} \) denote the Verdier localization \( \mathcal{T}/\mathcal{T}_{S\text{-tor}} \). If \( S \) is the set of powers of some integer \( n \), we write \( \mathbb{Z}[\frac{1}{n}] \) for \( S^{-1}\mathbb{Z} \), \( \mathcal{T}_{S\text{-tor}} \) for \( \mathcal{T}_{S\text{-tor}} \) and \( \mathcal{T}[\frac{1}{n}] \) for \( \mathcal{T}_{S^{-1}2} \). For an object \( A \) of \( \mathcal{T} \) write \( A_{S^{-1}2} \) or \( A[\frac{1}{n}] \) for the image of \( A \) in \( \mathcal{T}_{S^{-1}2} \) or \( \mathcal{T}[\frac{1}{n}] \); for an abelian group or (pre)sheaf of abelian groups \( M \), we write \( M_{S^{-1}2} \) or \( M[\frac{1}{n}] \) for \( M \otimes_{\mathbb{Z}} S^{-1}\mathbb{Z} \). For \( S = \mathbb{Z}\setminus\{0\} \), we will of course write \( \mathbb{Q} \) for \( S^{-1}\mathbb{Z} \).

If \( \mathcal{T} \) is a compactly generated tensor triangulated category such that the tensor product of compact objects is compact, then \( \mathcal{T}_{S\text{-tor}} \) is a tensor ideal (since \( n\cdot\text{id}_{A\otimes B} = (n\cdot\text{id}_A) \otimes \text{id}_B \) so \( \mathcal{T}_{S^{-1}2} \) inherits a tensor structure from \( \mathcal{T} \), and the localization functor \( \mathcal{T} \to \mathcal{T}_{S^{-1}2} \) is an exact tensor functor of tensor triangulated categories.
For $A$ an object in $\mathcal{T}$, let $A \otimes^L \mathbb{Z}/n$ denote an object fitting into a distinguished triangle

$$A \xrightarrow{n \cdot \text{id}} A \to A \otimes^L \mathbb{Z}/n \to A[1];$$

this defines $A \otimes^L \mathbb{Z}/n$ up to non-unique isomorphism. In addition, $A \otimes^L \mathbb{Z}/n$ is an $n^e$-torsion object.

**Lemma B.2.** Let $S$ be a subset of $\mathbb{Z} \setminus \{0\}$ containing 1, and suppose $\mathcal{T}$ has a set $C$ of compact generators.

1. The set $C_S := \{A \otimes^L \mathbb{Z}/n \mid A \in C, n \in S\}$ is a set of compact generators for $\mathcal{T}_{S\text{-tor}}$.

2. For $A$, $X$ objects in $\mathcal{T}$ with $A$ compact, there is a canonical isomorphism

$$\text{Hom}_\mathcal{T}(A, X)_{S^{-1}\mathbb{Z}} \cong \text{Hom}_{\mathcal{T}_{S^{-1}\mathbb{Z}}}(A_{S^{-1}\mathbb{Z}}, X_{S^{-1}\mathbb{Z}}).$$

In addition, for $A$ compact in $\mathcal{T}$, $A_{S^{-1}\mathbb{Z}}$ is compact in $\mathcal{T}_{S^{-1}\mathbb{Z}}$, and $\mathcal{T}_{S^{-1}\mathbb{Z}}$ is compactly generated, with $C_S := \{A_{S^{-1}\mathbb{Z}}, A \in C\}$ a set of compact generators.

3. $\mathcal{T}_{S\text{-tor}}$ is equal to the full subcategory $\mathcal{T}(S)$ of $\mathcal{T}$ with objects $X$ such that $\text{Hom}_\mathcal{T}(A, X)_{S^{-1}\mathbb{Z}} = 0$ for all compact $A$ in $\mathcal{T}$.

4. $\mathcal{SH}(k)_{S\text{-tor}}$ is the full subcategory of $\mathcal{SH}(k)$ with objects $\mathcal{K}$ such that $\Pi_{a,b}(\mathcal{K})_{S^{-1}\mathbb{Z}} = 0$ for all $a, b \in \mathbb{Z}$.

5. A map $f : \mathcal{E} \to \mathcal{F}$ in $\mathcal{SH}(k)$ becomes an isomorphism in $\mathcal{SH}(k)_{S^{-1}\mathbb{Z}}$ if and only if the induced map on the localized homotopy sheaves

$$f_* \otimes \text{id} : \Pi_{a,b}(\mathcal{E})_{S^{-1}\mathbb{Z}} \to \Pi_{a,b}(\mathcal{F})_{S^{-1}\mathbb{Z}}$$

is an isomorphism for all $a, b \in \mathbb{Z}$.

6. $\mathcal{SH}(k)_{S\text{-tor}}$ is a tensor ideal, hence $\mathcal{SH}(k)_{S^{-1}\mathbb{Z}}$ is a tensor triangulated category and the localization functor $\mathcal{SH}(k) \to \mathcal{SH}(k)_{S^{-1}\mathbb{Z}}$ is an exact tensor functor.

**Proof.** For (1), clearly $C_S$ consists of compact objects of $\mathcal{T}_{S\text{-tor}}$. If $X$ is $n$-torsion in $\mathcal{T}$, then $X$ is a summand of $X \otimes^L \mathbb{Z}/n$, hence the localizing subcategory of $\mathcal{T}$ generated by $C_S$ contains all the compact objects in $\mathcal{T}_{S\text{-tor}}$, hence is equal to $\mathcal{T}_{S\text{-tor}}$, proving (1).

To prove (2), let $Y = A \xrightarrow{w} Z \to Y[1]$ be a distinguished triangle in $\mathcal{T}$ with $A$ compact and $Z$ in $\mathcal{T}_{S\text{-tor}}$. Then $n \cdot u = 0$ for some $n \in S$, so there is a map $w : A \to Y$ with $v \circ w = n \cdot \text{id}_A$. Thus the category of maps $\{n \cdot \text{id} : A \to A\}$ is cofinal in the category of maps $Y \to A$ with cone in $\mathcal{T}_{S\text{-tor}}$, from which the isomorphism in (2) follows directly.

As the localization functor $q : \mathcal{T} \to \mathcal{T}_{S^{-1}\mathbb{Z}}$ admits a right adjoint, $q$ preserves all small coproducts; the isomorphism we have just proved shows that $A_{S^{-1}\mathbb{Z}}$ is compact if $A$ is compact. If $\text{Hom}_\mathcal{T}(A, X)_{S^{-1}\mathbb{Z}} = 0$ for all $A \in C$, then clearly $X$ is in $\mathcal{T}_{S\text{-tor}}$ hence $X_{S^{-1}\mathbb{Z}} = 0$, completing the proof of (2). (3) follows immediately from (2).

Clearly (5) follows from (4). To prove (4), let $\mathcal{SH}(k)_{S\text{-tor}} \subset \mathcal{SH}(k)$ be the full subcategory of objects $\mathcal{K}$ as in (4). We recall that $\mathcal{SH}(k)$ has the set of compact generators $\{\Sigma^n_{a,b} \Sigma^k_{i,j} X+ | a, b \in \mathbb{Z}, X \in \text{Sm}/k\}$.
Take $\mathcal{K} \in \mathcal{SH}(k)_{S,\text{tor}}$, $X \in \text{Sm}/k$, and let $f : \Sigma_{2}^{a} \Sigma_{T}^{b} \Sigma_{T}^{\infty} X_{+} \to \mathcal{K}$ be a morphism. Using the Gersten spectral sequence on $X$ and the assumption that $\mathcal{K}$ is in $\mathcal{SH}(k)_{S,\text{tor}}$, we see there is an $n \in S$ such that $n \cdot f = 0$. Thus $\mathcal{SH}(k)_{S,\text{tor}} \subset \mathcal{SH}(k)(S)$; the reverse inclusion follows from the definition of the homotopy sheaf $\Pi_{n,0}(\mathcal{K})$ as the Nisnevich sheaf associated to the presheaf

$$X \mapsto [\Sigma_{S}^{a} \Sigma_{T}^{b} \Sigma_{T}^{\infty} X_{+}, \mathcal{K}]_{\mathcal{SH}(k)}.$$ 

By (3), this shows that $\mathcal{SH}(k)_{S,\text{tor}} = \mathcal{SH}_{S,\text{tor}}$.

For (6), we note that $\Sigma_{T}^{\infty} X_{+} \wedge \Sigma_{T}^{\infty} Y_{+} \cong \Sigma_{T}^{\infty} X \times Y_{+}$. As the subcategory of compact objects of $\mathcal{SH}(k)$ is the thick subcategory generated by $C$ (see remark B.1), it follows that the $\wedge$-product of compact objects is compact, which suffices to prove (6).

**Remark B.3.** Lemma B.2(4)-(6) hold with $\mathcal{SH}(k)$ replaced by $\mathcal{SH}_{S,\text{tor}}(k)$, with the obvious modification in the statements.

**Lemma B.4.** Let $L : \mathcal{T}_{1} \to \mathcal{T}_{2}$ be an exact functor of compactly generated triangulated categories, with right adjoint $R : \mathcal{T}_{2} \to \mathcal{T}_{1}$; we assume that $\mathcal{T}_{1}$ and $\mathcal{T}_{2}$ admit arbitrary small coproducts. Suppose that $L(A)$ is compact for $A$ in $\mathcal{T}_{1}$.

1. $R$ is compatible with small coproducts.
2. Let $S \subset \mathbb{Z}\setminus\{0\}$ be a multiplicatively closed subset. Then $(L, R)$ descends to an adjoint pair $L_{S^{-1}2} : \mathcal{T}_{1 S^{-1}2} \rightleftarrows \mathcal{T}_{2 S^{-1}2} : R_{S^{-1}2}$.
3. Suppose that $L$ is the inclusion functor for a full triangulated subcategory $\mathcal{T}_{1}$ of $\mathcal{T}_{2}$. Then $L_{S^{-1}2} : \mathcal{T}_{1 S^{-1}2} \to \mathcal{T}_{2 S^{-1}2}$ is an isomorphism of $\mathcal{T}_{1 S^{-1}2}$ with its image in $\mathcal{T}_{2 S^{-1}2}$.

**Proof.** For (1), it is easy to show that $L$ sends compact objects of $\mathcal{T}_{1}$ to compact objects of $\mathcal{T}_{2}$ if and only if $R$ preserves small coproducts; we leave the details of the proof to the reader.

For (2), if $C_{2}$ is a set of compact generators for $\mathcal{T}_{2}$, and $A$ is in $C_{2}$, then for $n \in S$, $R(A) \otimes^{L} \mathbb{Z}/n \cong R(A \otimes^{L} \mathbb{Z}/n)$, hence $R$ maps $C_{2S}$ to $\mathcal{T}_{1 S^{-1}2}$. As $R$ is compatible with small coproducts, it follows that $R(\mathcal{T}_{2 S^{-1}2}) \subset \mathcal{T}_{1 S^{-1}2}$. As $L$ is a left adjoint, $L$ is compatible with small coproducts, so the same argument shows that $L(\mathcal{T}_{1 S^{-1}2}) \subset \mathcal{T}_{2 S^{-1}2}$, giving the induced functors on the localizations

$$L_{S^{-1}2} : \mathcal{T}_{1 S^{-1}2} \rightleftarrows \mathcal{T}_{2 S^{-1}2} : R_{S^{-1}2}.$$ 

To show that $(L_{S^{-1}2}, R_{S^{-1}2})$ is an adjoint pair, we may replace $\mathcal{T}_{1 S^{-1}2}$ with the equivalent full subcategory $C_{2S}^{i}$ of $\mathcal{T}_{1}$, $i = 1, 2$. $R$ maps $C_{2S}^{i}$ to $\mathcal{T}_{1 S^{-1}2}^{i}$, and $R_{S^{-1}2}$ is identified with the functor $R_{S} : C_{2S}^{i} \to C_{1S}^{i}$ induced from $R$. Letting $f : \mathcal{T}_{2} \to C_{2S}^{i}$ be the left adjoint to the inclusion $i_{2} : C_{2S}^{i} \to \mathcal{T}_{2}$, $L_{S^{-1}2}$ is identified with the restriction $L_{S}$ of $f \circ L$ to $C_{1S}^{i}$. Letting $i_{1} : C_{1S}^{i} \to \mathcal{T}_{1}$ be the inclusion, we have

$$\text{Hom}_{C_{2S}^{i}}(L_{S}(X), Y) \cong \text{Hom}_{\mathcal{T}_{2}}(L(i_{1}X), i_{2}Y) \cong \text{Hom}_{\mathcal{T}_{2}}(i_{1}X, R(i_{2}Y)) = \text{Hom}_{C_{1S}^{i}}(X, R_{S}(Y))$$

**Documenta Mathematica** 18 (2013) 907–941
for $X \in C^+_S$, $Y \in C^2_S$.

For (3), as $L$ and $L_{S^{-1}Z}$ are the same on objects, we need only show that $L_{S^{-1}Z}$ is fully faithful. Let $X, Y$ be objects of $T_1$. If $X$ is compact, then $L(X)$ is compact by assumption, and thus

\[
\text{Hom}_{T_{S^{-1}Z}}(L_{S^{-1}Z}(X_{S^{-1}Z}), L_{S^{-1}Z}(Y_{S^{-1}Z})) = \text{Hom}_{T_{S^{-1}Z}}(L(X)_{S^{-1}Z}, L(Y)_{S^{-1}Z}) \\
\cong \text{Hom}_{T_1}(L(X), L(Y))_{S^{-1}Z} \cong \text{Hom}_{T_1}(X, Y)_{S^{-1}Z} \\
\cong \text{Hom}_{T_{S^{-1}Z}}(X_{S^{-1}Z}, Y_{S^{-1}Z}).
\]

Furthermore, as $L_{S^{-1}Z}$ is a left adjoint, it preserves small coproducts, hence for fixed $Y$, the full subcategory $T^Y_{S^{-1}Z}$ of $X_{S^{-1}Z}$ in $T_{S^{-1}Z}$ such that

\[
\text{Hom}_{T_{S^{-1}Z}}(L_{S^{-1}Z}(X_{S^{-1}Z}), L_{S^{-1}Z}(Y_{S^{-1}Z})) \cong \text{Hom}_{T_{S^{-1}Z}}(X_{S^{-1}Z}, Y_{S^{-1}Z})
\]

is a localizing subcategory of $T_{S^{-1}Z}$ containing the objects $A_{S^{-1}Z}$ for $A$ a compact object of $T_1$. As these form a set of compact generators for $T_{S^{-1}Z}$, we see that $T^Y_{S^{-1}Z} = T_{S^{-1}Z}$, hence $L_{S^{-1}Z}$ is fully faithful. □

**Example B.5.** We consider the example of $i_n : \Sigma^n_S \mathcal{SH}^{eff} (k) \to \mathcal{SH} (k) : r_n$.

Lemma B.3 gives us the full subcategory $\Sigma^n_S \mathcal{SH}^{eff} (k)_{S^{-1}Z}$ of $\mathcal{SH} (k)_{S^{-1}Z}$ with inclusion functor $i_{nS^{-1}Z}$, the adjoint pair of functors

\[
i_{nS^{-1}Z} : \Sigma^n_S \mathcal{SH}^{eff} (k)_{S^{-1}Z} \to \mathcal{SH} (k)_{S^{-1}Z} : r_{nS^{-1}Z},
\]

the truncation functor $f_{nS^{-1}Z} := i_{nS^{-1}Z} \circ r_{nS^{-1}Z}$, and for $\mathcal{E} \in \mathcal{SH} (k)$, the canonical isomorphism

\[
(f_n \mathcal{E})_{S^{-1}Z} \cong f_{nS^{-1}Z} (\mathcal{E}_{S^{-1}Z}).
\]

In addition, by lemma B.2 $\Sigma^n_S \mathcal{SH}^{eff} (k)_{S^{-1}Z}$ is equal to the localizing subcategory of $\mathcal{SH}^{eff} (k)_{S^{-1}Z}$ generated by the set of compact objects $(\Sigma^q_S X_+)_{S^{-1}Z}$, for $q \geq n$ and $X \in \text{Sm}/k$. The analogous results hold for $\mathcal{SH}^S (k)$.

**References**


[12] Levine, M., The slice filtration and Grothendieck-Witt groups, PAMQ 7 no. 4.


Convergence of Voevodsky’s Slice Tower


Marc Levine
Universität Duisburg-Essen
Fakultät Mathematik
Campus Essen
45117 Essen
Germany
marc.levine@uni-due.de


**Abstract.** $p$-jets of finite flat maps of schemes are generally neither finite nor flat. However, for $p$-isogenies, and in particular for $p$-divisible groups, this pathology tends to disappear “in the limit”. We illustrate this in the case of $\mathbb{G}_m$, elliptic curves, and formal groups of finite height.

2010 Mathematics Subject Classification: 14 L 05, 13 K 05
Keywords and Phrases: Formal groups, $p$-divisible groups, Witt vectors

1. Introduction

This paper is the first in a series of papers where we investigate $p$-jet spaces (in the sense of [3]) of finite flat schemes/algebras. The understanding of such $p$-jet spaces seems to hold the key to a number of central questions about *arithmetic differential equations* [4]. The present paper deals with $p$-isogenies, and in particular with $p$-divisible groups; a sequel to this paper [8] will deal with algebras of Witt vectors of finite length.

Let $p$ be an odd prime and let $R = \hat\mathbb{Z}_p$ be the $p$-adic completion of the maximum unramified extension of the ring $\mathbb{Z}_p$ of $p$-adic numbers. (Throughout the paper the symbol $\hat{}$ means $p$-adic completion.) Let $k = R/pR$ be the residue field of $R$. Then for each integer $n \geq 0$ a functor $J^n$ was introduced in [3] that attaches to any scheme of finite type $X/R$ a (Noetherian) $p$-adic formal scheme $J^n(X)$ over $R$ called the $p$-jet space of $X$ of order $n$. For each $X$ there are morphisms $J^n(X) \to J^{n-1}(X)$, $n \geq 1$, functorial in $X$, and $J^0(X) = \hat{X}$. We refer to [3, 4] for an exposition of the theory and for some of the applications of these spaces; see also [6, 2] for a several prime version of the theory.

The functors $J^n$ behave nicely on smooth schemes and étale morphisms: in particular if $X$ is smooth over $R$ then $J^n(X)$ are $p$-adic completions of smooth schemes over $R$; and if $X \to Y$ is an étale morphism then $J^n(X) \simeq J^n(Y) \times_Y X$. So, in particular, if $X \to Y$ is finite and étale then the map $J^n(X) \to J^n(Y)$ is
again finite and étale. However if $X \to Y$ is, say, finite and flat then $J^n(X) \to J^n(Y)$ is generally neither finite nor flat. This basic pathology can be seen, in its simplest form, for $X$ a finite flat group scheme of degree a $p$-power over $Y = \text{Spec } R$; or for $p$-isogenies $X \to Y$ (i.e. isogenies of degree a $p$-power) between smooth group schemes over $R$. The present paper offers an analysis of $p$-jets of $p$-isogenies $[p^\nu] : X \to X$ and of $p$-divisible groups $(X_\nu; \nu \geq 1)$, $X_\nu := \text{Ker } [p^\nu]$, where $X$ is either the multiplicative group, or an elliptic curve, or a (one dimensional) formal group of finite height. (For the latter case one still has at one’s disposal a $p$-jet space theory.) The main moral of the story will be that although the $p$-jets of order $n$ of the individual $X_\nu$’s are generally highly pathological order tends to be restored “in the limit”, when either $n \to \infty$ or $\nu \to \infty$. One of our main motivations for trying to understand the $p$-jets of $p$-isogenies is their apparent link with the problem of understanding the $U$-operator (and the Hecke operator $T(p)$) on differential modular forms. Discussing this link here would lead us too far afield; the interested reader can see a hint of this in [5, 7].

In order to state some of our main results let us recall/introduce some basic notation. For any scheme of finite type $X/R$ the rings of global functions $\mathcal{O}^n(X) := \mathcal{O}(J^n(X))$ form an inductive system; its direct limit is denoted by $\mathcal{O}^\infty(X)$. There are canonical (non-linear) operators $\delta : \mathcal{O}^n(X) \to \mathcal{O}^{n+1}(X)$ that can be viewed as arithmetic analogues of the total derivative operator in differential geometry/mechanics.

Here is a basic example that we are going to be interested in. Let $x, x', x'', ...$ be variables and consider the rings

$$A^n := R[x, x', \ldots, x^{(n)}] \subset A := R[x] := R[x, x', x'', ...]$$

whose elements are referred to as $\delta$-polynomials. Let $\phi : R \to R$ be the unique ring automorphism that lifts the $p$-power Frobenius on $k$ and let $\phi : A \to A$ be the unique ring homomorphism which is the $\phi$ above on $R$ and sends $x, x', x'', ...$ into $x^p + px', (x')^p + px'', (x'')^p + px''', ...$ respectively. Then one defines the following map of sets (the Fermat quotient operator):

$$\delta : A \to A, \quad \delta F = \frac{\phi(F) - F^p}{p}. $$

This map induces maps $\delta : A^n \to A^{n+1}$, and by continuity, maps $\delta : (A^n)^* \to (A^{n+1})^*$ (where $^*$ always denotes in this paper the $p$-adic completion). Note that if $\mathbb{A}^1 = \text{Spec } R[x] = \text{Spec } A^0$ is the affine line over $R$ then $J^n(\mathbb{A}^1) = \text{Spf } (A^n)$ and the arithmetic analogues of the total derivatives $\delta : \mathcal{O}^n(\mathbb{A}^1) \to \mathcal{O}^{n+1}(\mathbb{A}^1)$ identify with the Fermat quotient operators $\delta : (A^n)^* \to (A^{n+1})^*$ we just introduced.

A related example is $\mathbb{G}_m = \text{Spec } R[x, x^{-1}]$ in which case $J^n(\mathbb{G}_m) = \text{Spf } A^n[x^{-1}]^*$ and $\delta : \mathcal{O}^n(\mathbb{G}_m) \to \mathcal{O}^{n+1}(\mathbb{G}_m)$ is induced by the $\delta$ above. If $[p^\nu] : \mathbb{G}_m \to \mathbb{G}_m$ is the $p$-isogeny defined by $x \mapsto x^{p^\nu}$ then the induced morphism $[p^\nu]^* : J^n(\mathbb{G}_m) \to J^n(\mathbb{G}_m)$ is given by the map

$$[p^\nu]^* : A^n[x^{-1}]^* \to A^n[x^{-1}]^*, \quad x^{(i)} \mapsto \delta^i(x^{p^\nu}).$$
Moreover if \( \mu^{p^s} = G_m[p^s] \) is the kernel of \([p^s]: G_m \to G_m\) then
\[
\mathfrak{O}^n(\mu^{p^s}) = \frac{A^n[x^{-1}]}{(xp^c - 1, \delta(xp^c), \ldots, \delta^n(xp^c))}
\]
So it becomes crucial to compute the \( \delta^n(xp^c) \).
Consider the filtration of \( A \) by the subrings:
\[
A^{(n)} = A^n + pA^{n+1} + p^2 A^{n+2} + \ldots \subset A
\]
and consider the ideal \( I = (x', x'', \ldots) \subset A \). Also consider the ideals \( I[p^s] \) of \( A \) generated by all \( \delta \)-polynomials of the form \( p^s(x^{(s)})x^{(s)} \), with \( s \geq 1, i, j \geq 0, i + j = \nu \). By abuse of notation, we will often denote by \( [S] \) an element of a set \( S \). The starting point of this paper will be the following “leading term computation”. Let \( n, \nu \geq 1 \).

**Theorem 1.1.**
\[
\delta^n(xp^c) = \begin{cases} 
\frac{p^{\nu-n+1}xp^n(p^{\nu-1})\phi^{n-1}(x') + \left([p^{\nu-n+2}A^{(1)}] \cap I[p^s]\right)}{x^n(p^{\nu-1})\phi^n(x^{(n-\nu)}) + \left[A^n \cap I[p^s]\right]} & \text{if } n \leq \nu + 1 \\
\frac{x^n(p^{\nu-1})\phi^n(x^{(n-\nu)}) + \left[A^{(n-\nu-1)} \cap I[p^s]\right]}{x^{n+p^c} - 1} & \text{if } n \geq \nu + 2.
\end{cases}
\]
This computation will have a number of consequences (both in characteristic zero and in characteristic \( p \)). Here is a consequence in characteristic zero. Let \( J^n(\mu^{p^s}) \) be the kernel of the projection \( J^n(\mu^{p^s}) \to J^0(\mu^{p^s}) \) and write \( \mathfrak{O}^n(\mu^{p^s})_1 := O(\mu^{p^s})_1 \). Then for \( n \geq 1 \) we have:

**Corollary 1.2.**
\[
\lim_{n \to \infty} \mathfrak{O}^n(\mu^{p^s})_1 = R[x', \ldots, x^{(n)}].
\]
Let us mention some consequences in characteristic \( p \). Before doing so we introduce some notation. For any ring \( B \) we denote \( \overline{B} = B/(p) \); and for any \( b \in B \) we let \( \overline{b} \in \overline{B} \) be the image of \( B \). In particular for any scheme of finite type \( X/R \) we set
\[
(1.1) \quad \overline{\mathfrak{O}^n(X)} := \mathfrak{O}^n(X)/(p), \quad \overline{\mathfrak{O}^\infty(X)} := \mathfrak{O}^\infty(X)/(p).
\]
Note that the morphisms \( \overline{\mathfrak{O}^n(X)} \to \overline{\mathfrak{O}^\infty(X)} \) are generally not injective! (They are injective, however, if \( X/R \) is smooth \([3]\).) It turns out that for non-smooth \( X/R \) a special role is then played by the rings:
\[
(1.2) \quad \overline{\mathfrak{O}^n(X)} := Im(\overline{\mathfrak{O}^n(X)} \to \overline{\mathfrak{O}^\infty(X)}).
\]
According to our notation above we may consider the ring \( \overline{A} = k[x, x', x'', \ldots] \) and its filtration with subrings \( \overline{A}^{(n)} := \overline{A}^{(n)} = k[x, x', \ldots, x^{(n)}] \). Also we may consider the reduction mod \( p \), \( \overline{T} = (x', x'', \ldots) \subset \overline{A} \), of the ideal \( I = (x', x'', \ldots) \subset A \). Then \( \overline{\mathfrak{O}^n(T)} \) coincides with the ideal \( \overline{\mathfrak{O}^n(T)} \) in \( \overline{A} \) generated by \( (x')p^c, (x'')p^c, \ldots \). Moreover clearly \( \overline{A}^{(n)} \cap \overline{\mathfrak{O}^{n[p^s]}} \) is generated in \( \overline{A}^{(n)} \) by \( (x')p^c, (x'')p^c, \ldots, (x^{(n)})p^c \). So the reduction mod \( p \) of the morphism \( J^n([p^s]): J^n(G_m) \to J^n(G_m) \) is given by the homomorphism
\[
[p^s]^* : \overline{\mathfrak{O}^n(G_m)} = k[x, x^{-1}, x', \ldots, x^{(n)}] \to k[x, x^{-1}, x', \ldots, x^{(n)}], \quad x^{(i)} \to \delta^{(i)}(xp^c)
\]
where, by Theorem 1.1:

**Corollary 1.3.** The element \( \delta^n(x^p) \in \overline{A} \) satisfies

\[
\delta^n(x^p) = \begin{cases} 
0 & \text{if } 1 \leq n \leq \nu \\
x^{p^{\nu+1}(p-1)}(x')^p & \text{if } n = \nu + 1 \\
x^{p^{\nu}(p-1)}(x'^{p-\nu})^p + \prod_{i=0}^{\nu-1} T_i & \text{if } n \geq \nu + 2
\end{cases}
\]

The “smallest” interesting case is \( \nu = 1, n = 3 \),

\[\delta^3(x^p) = x^{p^3(p-1)}(x')^p - \frac{1}{2}x^{p^2(p-2)}(x')^2p.\]

**Remark 1.4.** By Corollary 1.3, for \( n \geq \nu + 1 \), the map \( \overline{[p]}^* : \overline{O}^n(G_m) \to \overline{O}^n(G_m) \) induces injective finite flat maps

\[\overline{[p]}^* : \overline{O}^n(G_m)/(x', ... , x^{(\nu)}) \to \overline{O}^{n-\nu}(G_m).\]

Indeed finiteness is clear; injectivity follows by looking at dimensions; and flatness follows from the general fact that finite surjective maps of non-singular (irreducible) varieties are automatically flat (cf., say, [9], Theorem 18.16.).

Corollary 1.3 trivially implies the following determination of \( \overline{O}^n(\mu_{p^\nu}) \).

Let \( n, \nu \geq 1 \).

**Corollary 1.5.**

\[\overline{O}^n(\mu_{p^\nu}) \simeq \frac{k[x, x', x'', ..., x^{(n)}]}{(x - 1)^{p^\nu}, (x')^{p^\nu}, ..., (x^{(n)})^{p^\nu}}.\]

The statement of the corollary above should be contrasted with the fact that, as we shall see, for all \( n, \nu \geq 1 \) the the rings \( \overline{O}^n(\mu_{p^\nu}) \) have positive Krull dimension (actually they are polynomial rings in \( \min\{n, \nu\} \) variables over some explicit local Artin rings).

**Remark 1.6.** As we saw the mod \( p \) Corollary 1.5 follows trivially from our characteristic zero Theorem 1.1. We will also present an alternative proof of this mod \( p \) result using a Witt vector computation; we are indebted to the referee for this alternative proof. We included both approaches because each has its own advantage: the Witt vector computation yields a shorter argument (but apparently working only mod \( p \)) whereas the computation in Theorem 1.1 is valid in characteristic zero and has other consequences as well.

The above theory has an analogue for formal groups of height \( \geq 2 \) which we now explain. We consider the rings

\[A^n = R[[x]][x', ..., x^{(n)}], \quad A := \bigcup_{n \geq 0} A^n.\]

Consider the filtration of \( A \) by the subrings:

\[A^{(n)} = A^n + pA^{n+1} + p^2A^{n+2} + ... \subset A\]
and consider the ideals $\mathcal{I}^{[\nu]}$ of $A$ generated by all $\delta$-polynomials of the form $p^i(x^{(s)})p^j$, with $s \geq 0$, $i, j \geq 0$, $i + j = \nu$. (Note that, unlike in the case of the ideals $I^{[\nu]}$, the superscript $s$ here is allowed to be 0! So, for instance $x^{\nu} \in \mathcal{I}^{[\nu]}$ but $x^{2\nu} \notin I^{[\nu]}$. The lift of Frobenius $\phi: A \to A$ on $A = R[x]$ and the Fermat quotient operator $\delta: A \to A$ induce obvious maps $\phi: A \to A$ and $\delta: A \to A$.

Now let $\mathcal{F} \in R[[x]]$ be a formal group law (in one variable) of finite height and let $\mathcal{F}[p^\nu]$ be the kernel of the multiplication by $p^\nu$ viewed as a finite flat group scheme over $R$. As we shall see it turns out that

$$\mathcal{O}^n(\mathcal{F}[p^\nu]) = \frac{A^n}{\langle F^{\alpha}, \delta(F^{\alpha}), \ldots, \delta^n(F^{\alpha}) \rangle},$$

where $F(x) = [p]x(x) \in R[[x^p]] + pR[[x]]$ is the series giving the multiplication by $p$ in $\mathcal{F}$ and $F^{\alpha}$ is its $\nu$-th iterate. So we shall be interested in computing $\delta^n(F^{\alpha})$.

More generally start with any series $F \in R[[x^p]] + pR[[x]], F(0) = 0$; any $F$ of the form $[p]x(x)$ ($\mathcal{F}$ a formal group) has this shape. Then one easily sees that

$$(1.3) \quad F^{\alpha}(x) = \sum_{j=0}^\nu p^{\nu-j}G_j(x^{p^j}),$$

where $G_j \in xR[[x]], \ j \geq 0$. So the computation of $\delta^n(F^{\alpha})$ boils down to computing the quantities $\delta^m(p^iG_j(x^{p^j}))$ for $i + j = \nu$ and $m \leq n$. Here is our main characteristic zero “leading term computation” of such quantities.

Assume $G(x) \in xR[[x]], m \geq 1, i + j = \nu \geq 1, i \geq 0, j \geq 0$. Then:

**Theorem 1.7.**

$$\delta^m(p^iG(x^{p^j})) = \begin{cases} p^{i-m}G(x^{p^{j-m}}) & \text{if } m \leq i \\ \phi^m \{ \frac{dG^{(i)}}{dx^{(i)}}(x^{p^j}) \} \bigg|_{x=0} - \delta^{m-i}(x^{p^j}) & \text{if } m > i \end{cases}$$

Here for $a$ an integer we set $a^+ = \max\{a, 0\}$.

The above (and indeed a much weaker statement) implies in particular that the natural homomorphism

$$(1.4) \quad A^n \to \lim_{\nu} \mathcal{O}^n(\mathcal{F}[p^\nu])$$

is injective; this is in the spirit of Corollary 1.2. A more precise consequence of the above can be obtained by combining Theorems 1.1 and 1.7 to give a “leading term computation” for $\delta^m(F^{\alpha})$ in characteristic zero: all one has to do is to replace the expression $\delta^{m-i}(x^{p^j})$ in the formula of Theorem 1.7 by its value given in Theorem 1.1. Rather than stating this consequence in characteristic zero we look at some of its effects in characteristic $p$.

According to our conventions we may consider the ring $\mathcal{F} = k[[x]][x', x'', \ldots]$ and its filtration with subrings $\mathcal{F}^n := \mathcal{F}^n = k[[x]][x', \ldots, x^{(n)}]$. Also we may consider the reduction mod $p$ of the ideal $\mathcal{I}, \mathcal{J} = (x, x', x'', \ldots) \subset \mathcal{F}$. Then $\mathcal{F}[p^\nu]$ coincides
with the ideal \( \mathfrak{J}^{[p^{\nu}]} \) generated by \( x^{p^\nu}, (x')^{p^\nu}, \ldots \). Moreover clearly \( \mathfrak{A}^n \cap \mathfrak{J}^{[p^{\nu}]} \) is generated by \( x^{p^\nu}, \ldots, (x^{(n)})^{p^\nu} \) in \( \mathfrak{A}^n \).

For the next Corollaries we continue to denote by \( F \) any series in \( R[[x^p]] + pR[[x]] \) with \( F(0) = 0 \) and to write its \( \nu \)-th iterate as in Equation 1.3. Note that \( \frac{dF^\nu}{dx} \in p^\nu R[[x]] \) so we may consider the series

\[
\nu^{-\nu}\frac{dF^\nu}{dx} \in \mathfrak{A}^1 = k[[x]].
\]

Then our Theorem 1.7 will imply the following. Let \( \nu \geq 1 \) and \( n \geq 0 \).

**Corollary 1.8.** The element \( \delta^n(F^{\nu}) \in \mathfrak{A}^1 \) is given by

\[
\delta^n(F^{\nu}) = \begin{cases} 
(1_{n-n}(x^{p^{n-n}}))^{p^n} + [x^{2p^n} \mathfrak{J}] & \text{if } 0 \leq n \leq \nu \\
(\nu^{-\nu}\frac{dF^\nu}{dx})^{p^n} (x^{(n-n)})^{p^n} + [\mathfrak{A}^{n-1} \cap \mathfrak{J}^{[p^{\nu}]}] & \text{if } n \geq \nu + 1.
\end{cases}
\]

**Remark 1.9.** The case \( \nu = 1 \) of Corollary 1.8 above can be interpreted as follows. Let us consider \( \mathbb{A}^1 = \text{Spec } R[x] \), the affine line over \( R \), and its reduction mod \( p \), \( \overline{\mathbb{A}^1} = \text{Spec } k[x] \). Then the \( R \)-morphism \( \Phi : \overline{\mathbb{A}^1} \to \overline{\mathbb{A}^1} \) defined by \( x \mapsto \Phi(x) = F(x) : = x^p + pxf(x), f(x) \in xR[x] \), is the most general \( R \)-morphism lifting the relative \((k\text{-linear})\) Frobenius \( \mathbb{A}^1 \to \mathbb{A}^1 \) and sending \( 0 \) into \( 0 \); Corollary 1.8 provides then, in particular, a description of the reduction mod \( p \) of the induced map

\[
J^n(\Phi) : J^n(\mathbb{A}^1) \to J^n(\mathbb{A}^1) = \text{Spf } R[x, x', \ldots, x^{(n)}],
\]

(which sends \( x, x', \ldots, x^{(n)} \) into \( F, \delta F, \ldots, \delta^n F \)). Note that the map \( J^n(\Phi) \) is generally neither finite nor flat and its behavior depends in an essential way on the series \( f(x) \).

Another immediate consequence of Corollary 1.8 is the following structure theorem for \( \mathfrak{O}^n(\mathfrak{F}^{[p^n]}) \). We actually prove a slightly more general result covering cases that do not come from formal groups. Let \( n, \nu \geq 1 \).

**Corollary 1.10.** Assume \( F(x) \equiv px \mod x^2 \). For all \( \nu \geq 1 \) consider the scheme \( X_n := \text{Spec } \frac{R[[x]]}{(px^n)} \). Then for \( n \geq 1 \) we have:

\[
\mathfrak{O}^n(X_n) = \frac{k[x, x', \ldots, x^{(n)}]}{(x^{p^n}, (x')^{p^n}, \ldots, (x^{(n)})^{p^n})}.
\]

Recall that if \( F = [p]_{(x)} \) for some formal group \( \mathfrak{F} \) then the condition \( F(x) \equiv px \mod x^2 \) is automatic and \( X_n = \mathfrak{F}^{[p^n]} \).

Again Corollary 1.10 should be contrasted to the fact that \( \mathfrak{O}^n(X_n) \) have positive Krull dimension (they are, again, polynomial rings in \( \min\{n, \nu\} \) variables over some explicit local Artin rings).

**Remark 1.11.** We already mentioned that Corollary 1.5 can be proved independently of Theorem 1.1 via a Witt vector computation argument. It would
be very interesting to find a similar Witt vector argument for Corollary 1.10 which is independent of Theorem 1.7, at least in the case when $F(x)$ is of the form $[p]^\mathcal{J}(x)$ for some formal group $\mathcal{J}$.

**Remark 1.12.** Results similar to Corollaries 1.5 and 1.10 above are obtained in the body of the paper for the $p$-divisible groups of elliptic curves. The case of ordinary elliptic curves is deduced from (a twisted version of) the results for $\mathbb{G}_m$ while the case of supersingular elliptic curves is deduced from the results on formal groups. In the ordinary case the shape of the results depends on the value of the Serre-Tate parameter.

**Remark 1.13.** It would be interesting to have a generalization of our computations (in characteristic zero or at least in characteristic $p$) to the case of arbitrary $p$-divisible groups.

**Remark 1.14.** It is interesting to note the following phenomenon. Let $X_0, \ldots, X_\nu$ be closed subschemes of the affine line $\mathbb{A}^1$ over $R$ which are, say, finite and flat over $R$, and let

$$X = \bigcup_{i=0}^{\nu} X_i$$

(scheme theoretic union inside $\mathbb{A}^1$, defined by the intersection of the defining ideals). Then, in general,

$$J^n(X) \neq \bigcup_{i=0}^{\nu} J^n(X_i)$$

as closed subschemes of $J^n(\mathbb{A}^1)$. An example is provided by the case when $X_i = \text{Spec } R[\zeta_{p^i}]$ where $\zeta_{p^i}$ is a $p^i$-th root of unity. In this case $J^n(X_0) = \text{Spec } R$ and as we shall see later in the paper, $J^n(X_i) = \emptyset$ for $i \geq 1$ and $n \geq 1$; on the other hand $X = \mu_{p^n}$ (kernel of multiplication by $p^n$ on $\mathbb{G}_m$ over $R$) and hence $J^n(\mu_{p^n})$ has a non-reduced reduction mod $p$ by Theorem 1.10. It would be interesting to understand this phenomenon more generally when, for instance, $X_i = \text{Spec } R[\alpha_i]$ with $\alpha_i$ integers in a finite ramified extension of the fraction field of $R$.

**Remark 1.15.** In a sequel to this paper [8] we shall investigate the $p$-jet spaces of another remarkable example of finite flat schemes over $R$ namely schemes of the form $\text{Spec } W_m(R)$ where $W_m(R)$ are the rings of Witt vectors of finite length on $R$.

A few words about the structure of the paper. We begin by recalling from [3, 4] some of the basic concepts we shall be dealing with. Then we will study the filtrations $A^1$ and $I^1$ in a general setting and we will prove Theorem 1.1. Then, in subsequent sections, we will investigate the $p$-jets of the divisible groups of $\mathbb{G}_m$, ordinary elliptic curves, formal groups of finite height, and supersingular elliptic curves respectively. The $\mathbb{G}_m$ case will be used as a step in the analysis of all the other cases.
In particular if \( Y \) is a closed immersion so is \( J \) and then \( J \) is necessarily reflect the views of the National Science Foundation, IHES, or MPI. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation, IHES, or MPI.

2. Review of some basic concepts \([3, 4]\)

Rings in this paper will always be assumed commutative with unity. A \( p \)-derivation \( \delta : A \to A \) on a ring \( A \) is a set theoretic map satisfying
\[
\delta(x + y) = \delta x + \delta y + C_p(x, y) \\
\delta(xy) = x^p\delta y + y^p\delta x + p\delta x\delta y, 
\]
where \( C_p \) is the polynomial:
\[
C_p(X, Y) = p^{-1}(X^p + Y^p - (X + Y)^p) \in \mathbb{Z}[X, Y].
\]
If \( \delta \) is as above then \( \phi : A \to A, \phi(x) = x^p + p\delta x \), is a ring homomorphism. Note that \( \delta(xy) = x^p\delta y + \phi(y)\delta x = y^p\delta x + \phi(x)\delta y \). Also \( \delta \) and \( \phi \) commute. If \( A \) is \( p \)-torsion free then \( \delta \) is, of course, uniquely determined by \( \phi \); also
\[
\delta(x_1 + \ldots + x_m) = \delta x_1 + \ldots + \delta x_m + C_p(x_1, \ldots, x_m),
\]
where
\[
C_p(X_1, \ldots, X_m) := p^{-1}\left(\sum_{i=1}^{m}X_i^p - \left(\sum_{i=1}^{m}X_i\right)^p\right) \in \mathbb{Z}[X_1, \ldots, X_m].
\]
Now the ring \( R = \widehat{\mathbb{Z}_p^n} \) has a unique \( p \)-derivation defined by \( \delta x = (\phi(x) - x^p)/p \) where \( \phi : R \to R \) is the unique ring automorphism lifting the \( p \)-power Frobenius on \( R/pR \). Let \( x \) be a variable (or more generally an \( N \)-tuple of variables \( x_1, \ldots, x_N \)) We consider the \( \delta \)-polynomial ring \( R[x] = R[x', x'', \ldots] \); this is the polynomial ring in variables \( x, x', x'', \ldots, x^{(n)}, \ldots \), where \( x', x'', \ldots \) are variables (or \( N \)-tuples of variables), equipped with the unique \( p \)-derivation \( \delta : R\{x\} \to R\{x\} \) such that \( \delta x = x', \delta x' = x'', \text{etc.} \). For \( X \) a scheme of finite type over \( R \) one defines the \( p \)-jet spaces \( J^n(X) \), \( n \geq 0 \) \([3]\). The latter are \( p \)-adic formal schemes over \( R \) fitting into a projective system
\[
\ldots \to J^n(X) \to J^{n-1}(X) \to \ldots \to J^0(X) = \widehat{X}.
\]
Note that \( X \mapsto J^n(X) \) are functors commuting with open immersions and more generally with étale maps in the sense that if \( X \to Y \) is étale then \( J^n(X) \simeq J^n(Y) \times_Y X \) in the category of \( p \)-adic formal schemes. If \( X = \text{Spec } R[x]/(f) \) for a tuple of variables \( x \) and a tuple of polynomials \( f \) and then
\[
J^n(X) = \text{Spf } R[x, x', \ldots, x^{(n)}]/(f, \delta f, \ldots, \delta^n f).
\]
In particular if \( Y \to X \) is a closed immersion so is \( J^n(Y) \to J^n(X) \) for all \( n \). Moreover \( J^n \) commutes with fiber products: \( J^n(Y \times_X Z) \simeq J^n(Y) \times_{J^n(X)} \)

Acknowledgment. The author would like to thank the referee for communicating to him the Witt vector computation argument yielding an alternative proof of Corollary 1.5. The material in this paper is based upon work supported by the National Science Foundation under Grant No. 0852591 and by the IHES, France and MPI, Bonn. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation, IHES, or MPI.
$J^n(Z)$. The rings $\mathcal{O}^n(X) := \mathcal{O}(J^n(X))$ form an inductive system, the $p$-derivation $\delta$ on $R[x]$ induces operators $\delta : \mathcal{O}^n(X) \to \mathcal{O}^{n+1}(X)$, and the direct limit of these $\delta$’s induces a $p$-derivation $\delta$ on the direct limit $\mathcal{O}^\infty(X)$ of the rings $\mathcal{O}^n(X)$. The following universality property holds. Assume for simplicity $X$ is affine. Then any $R$-algebra homomorphism of $\mathcal{O}(X)$ into a $p$-adically complete ring $B$ equipped with a $p$-derivation $\delta$ is induced by a unique $R$-algebra homomorphism $\mathcal{O}^\infty(X) \to B$ that commutes with $\delta$.

3. Filtrations

In this section we will introduce and study some basic filtrations, especially on rings equipped with $p$-derivations. In the next section we will specialize to the case when the ring in question is the ring $R(x)$ of $\delta$-polynomials.

Let $A$ be a ring which for simplicity we assume $p$-torsion free and let $I$ be an ideal in $A$. For any integer $\nu \geq 0$ we denote by $I^{[\nu]}$ the ideal generated by all the elements of the form $p^i f^{p^\nu}$ with $f \in I$, $i, j \geq 0$, $i + j = \nu$. (N.B. Sometimes the superscript $[\cdot]$ is used to denote divided powers of ideals; our use of this superscript here has nothing to do with divided powers but rather generalizes the notation used for Frobenius powers of ideals in characteristic $p$.) In particular $I^{[1]} = I$. Also note that $I^{[\nu]} \subset I(pA + I)^\nu$, where $J^\nu$ is, as usual, the $\nu$-th power of an ideal $J$.

**Lemma 3.1.**

1) $I^{[\nu+1]} \subset I^{[\nu]}$.

2) If $f \in I^{[\nu]}$ then $pf \in I^{[\nu+1]}$.

3) If $f \in I^{[\nu+1]}$ then $pf \in I^{[\nu+1]}$.

4) If $I$ is generated by a family $\{f_s; s \in S\}$ then $I^{[\nu]}$ is generated by the family $\{f_s^{p^\nu}; s \in S, i, j \geq 0, i + j = \nu\}$.

**Proof.** Assertions 1 and 2 are clear. For assertion 3 if $f = \sum_{s=1}^{N} p^{i_s} f_s^{p^{\nu_s}} g_s$ with $f_s \in I$, $g_s \in A$, and $i_s + j_s = \nu$ for all $s$ then

$$f^p \in \sum_{s=1}^{N} (p^{i_s} f_s^{p^{\nu_s}} g_s)^p + pC_p(p^{i_1} f_1^{p^{\nu_1}} g_1, \ldots, p^{i_N} f_N^{p^{\nu_N}} g_N) \subset I^{[\nu+1]}.$$  

To check assertion 4 it is sufficient to prove that if $g \in I$ then $g^{p^t}$ is in the ideal generated by the family $\{p^t f_s^p; s \in S, i, j \geq 0, i + j = t\}$. One proves this by induction on $t \geq 0$. The case $t = 0$ is clear. Now if the statement is true for $t = \nu$ and we set $f = g^{p^t}$ then we are done by equation (3.1). \qed

In what follows we assume we are given a $p$-derivation $\delta : A \to A$.

**Lemma 3.2.** Assume $\delta(I) \subset I$. Then:

1) $\delta(I^{[\nu]}) \subset I^{[\nu]}$;

2) $\delta(I^{[\nu]}) \subset I^{[\nu+1]}$.
Proof. First, for any \( f \in A \), we have the following computation:

\[
\begin{align*}
\delta(f^{p^\nu}) &= \frac{1}{p} (\phi(f^{p^\nu}) - f^{p^{\nu+1}}) \\
&= \frac{1}{p} ((f^p + p\delta f)^{p^\nu} - f^{p^{\nu+1}}) \\
&= \ p^\nu f^{p^{\nu-1}} \delta f + p^{\nu+1} (\delta f)^2 P(f, \delta f)
\end{align*}
\]

where \( P \) is a polynomial with \( \mathbb{Z} \)-coefficients; indeed this is because for \( 2 \leq m \leq p^\nu \) (and since \( p \) is odd) we have \( p^{\nu+1} | (g^m f^{p^\nu}) \).

In particular if \( f \in I \) then \( \delta(f^{p^\nu}) \in p^\nu I \subset I[p^\nu] \).

Let’s prove assertion 1. In view of the equation (2.1) it is enough to note that for \( i + j = \nu \), \( f \in I \), \( g \in A \) we have \( \delta(p^i f^{p^j} g) \in I[p^\nu] \). Now

\[
\begin{align*}
\delta(p^i f^{p^j} g) &= \delta(p^i) f^{p^{i+1}} g^p + p^i \delta(f^{p^j} g) \\
&= \delta(p^i) f^{p^{i+1}} g^p + p^i (\delta(f^{p^j}) g^p + f^{p^{i+1}} \delta g + p(\delta(f^{p^j}))(\delta g)) \\
&\in I[p^\nu],
\end{align*}
\]

by Lemma 3.1 and because \( \delta(p^i) \) is either 0 or in \( p^{i-1} A \) according as \( i = 0 \) or \( i \geq 1 \).

To prove assertion 2 note that if \( f \in I[p^\nu] \) then \( \phi(f) = f^p + p\delta f \in I[p^{\nu+1}] \) by Lemma 3.1.

In what follows we assume we are given, in addition, a filtration on \( A \)

\[
A^0 \subset A^1 \subset A^2 \subset \ldots \subset A^n \subset \ldots \subset A,
\]

by subrings \( A^n \) such that \( \delta A^n \subset A^{n+1} \) for all \( n \geq 0 \). Then we define a new filtration by subrings

\[
A^{[n]} := \sum_{s=0}^{\infty} p^s A^{n+s} = A^n + pA^{n+1} + p^2 A^{n+2} + \ldots
\]

**Lemma 3.3.**

1) \( pA^{[n+1]} \subset A^{[n]} \);
2) \( \delta(A^{[n]}) \subset A^{[n+1]} \);
3) \( \phi(A^{[n]}) \subset A^{[n]} \).

**Proof.** A trivial exercise. \( \square \)

We will also need the following general:

**Lemma 3.4.** For any \( f, g \in A^0 \) we have the following equality of ideals in \( A^n \):

\[
(f - g, \delta(f - g), \ldots, \delta^n(f - g)) = (f - g, \delta f - \delta g, \ldots, \delta^n f - \delta^n g).
\]
Proof. Induction on $n$. The induction step follows from the congruence
\[
\delta(\delta^{n-1} f - \delta^{n-1} g) = (\delta^n f - \delta^n g) = C_{\nu}(\delta^{n-1} f, -\delta^{n-1} g) \\
\equiv C_{\nu}(\delta^{n-1} g, -\delta^{n-1} g) \mod \delta^{n-1} f - \delta^{n-1} g \\
= 0.
\]
\[\square\]

4. $p$-jets of $p$-isogenies of $\mathbb{G}_m$

In this section we specialize the discussion of the previous section to the case when
\[(4.1) \quad A = R[x], \quad A^n = R[x, x', \ldots, x^{(n)}], \quad I = (x', x'', \ldots).
\]

So in this case, explicitly,
\[A^{(n)} = R[x, x', \ldots, x^{(n)}] + pR[x, x', \ldots, x^{(n+1)}] + p^2 R[x, x', \ldots, x^{(n+2)}] + \ldots,
\]

while $I[p^n]$ is the ideal of $R[x]$ generated by all $\delta$-polynomials of the form $p^i(x^{(s)})p^j$, with $s \geq 1$, $i, j \geq 0$, $i + j = \nu$; cf. assertion 4 in Lemma 3.1.

We start by proving Theorem 1.1 in the Introduction.

Proof of Theorem 1.1. First note that for $n \leq \nu + 1$ we have $\phi^{\nu-1}(x') \in I[p^{n-1}]$ by Lemma 3.2 and hence
\[(4.2) \quad p^{\nu-n+1}x^\nu(x'(p^{n-1})\phi^{\nu-1}(x') \in I[p^n],
\]

by Lemma 3.1. Similarly, for $n \geq \nu + 2$, we have
\[(4.3) \quad x^\nu(x'(p^{n-1})\phi^{\nu}(x^{(n-\nu)}) \in I[p^n],
\]

by Lemma 3.2. We also claim that
\[(4.4) \quad \delta^n(x^p) \in I[p^n].
\]

To check (4.4) it is enough, by Lemma 3.2, to check that $\delta(x^p) \in I[p^n]$; this however follows from equation (3.2). In view of (4.2), (4.3), (4.4), in order to prove our theorem it is enough to prove that
\[
\delta^n(x^p) \in \begin{cases} 
p^{\nu-n+1}x^\nu(x'(p^{n-1})\phi^{\nu-1}(x') + p^{\nu-n+2}A^{(1)}, & \text{if } n \leq \nu + 1 \\
x^\nu(x'(p^{n-1})\phi^{\nu}(x^{(n-\nu)}) + A^{(n-\nu-1)}, & \text{if } n \geq \nu + 2.
\end{cases}
\]

We fix $\nu$ and proceed by induction on $n \geq 1$. For $n = 1$ we are done by (3.2). Next assume the theorem is true for $n$ and we prove it for $n + 1$.

Assume first $n \leq \nu + 1$. By Lemma 3.3, $\phi^{\nu-1}(x') \in A^{(1)}$. So we have
\[
\delta^{n+1}(x^p) \in \begin{cases} 
\delta(p^{\nu-n+1}x^\nu(p^{n-1})\phi^{\nu-1}(x')) + \delta(p^{\nu-n+2}A^{(1)}) \\
+ C_{\nu}(p^{\nu-n+1}A^{(1)}, p^{\nu-n+2}A^{(1)}).
\end{cases}
\]
Clearly the last term in the last equation is in $p^{\nu-n+1}A^{(1)}$. Also, by Lemma 3.3:
\[
\delta(p^{\nu-n+2}A^{(1)}) \subset \delta(p^{\nu-n+2}A^{(1)} + p^{\nu-n+2}A^{(1)}) \\
\subset \delta(p^{\nu-n+1}A^{(1)} + p^{\nu-n+2}A^{(2)}) \\
\subset p^{\nu-n+1}A^{(1)}.
\]

Now for $n \leq \nu$ we have
\[
\delta(p^{\nu-n+1}) = p^{\nu-n} - p^{\nu-n+1-1} \in p^{\nu-n} + p^{\nu-n+1}\mathbb{Z}
\]
hence:
\[
\delta(p^{\nu-n+1}x^{\nu-1}(x')) = \delta(p^{\nu-n+1}x^{\nu-1}(x'))p \\
+ p^{\nu-n+1}\delta(x^{\nu-1}(x')) \\
\in \delta(x^{\nu-1}(x')) + p^{\nu-n+1}A^{(1)} \\
+ p^{\nu-n+1}\delta(x^{\nu-1}(x')) \\
+ p^{\nu-n+1}\delta(x^{\nu-1}(x')) \\
\subset p^{\nu-n+1}A^{(1)} \\
+ p^{\nu-n}(p^{\nu-1} + pA^{(1)}(\nu^{n-1}(px'''))) \\
\subset p^{\nu-n}x^{\nu-1}(x') + p^{\nu-n+1}A^{(1)}
\]
because $\delta \circ \phi^{n-1} = \phi^{n-1} \circ \delta$, $(x')p + px'' = \phi(x')$, and
\[
p^{\nu-n} \cdot pA^{(1)}(\nu^{n-1}(px''')) \subset p^{\nu-n+1}A^{(1)}(A^{(1)} \cdot pA^{(1)}) \subset p^{\nu-n+1}A^{(1)}.
\]
So for $n \leq \nu$ we get
\[
\delta^{n+1}(x^{p^n}) = p^{\nu-n}x^{p^n-1}(x') + p^{\nu-n+1}A^{(1)}
\]
which ends the induction step in case $n \leq \nu$.

For $n = \nu + 1$ we get
\[
\delta(p^{\nu-n+1}x^{\nu-1}(x')) = \delta(x^{\nu-1}(x'))p \\
+ \phi(x^{\nu-1}(x')) \\
\in A^{(1)} + (x^{\nu+1}(p^{\nu-1} + pA^{(1)}(\nu^{n-1}(px'''))) \\
= x^{\nu+1}(p^{\nu-1} + pA^{(1)}(\nu^{n-1}(px'''))) \\
\subset p^{\nu-n+1}A^{(1)}
\]
by Lemma 3.3. Hence
\[
\delta^{n+1}(x^{p^n}) = x^{p^n-1}(x^{n+1-n'}) + A^{(n-n')}
\]
which ends the induction step in case $n = \nu + 1$.

Assume now $n \geq \nu + 2$; then, by Lemma 3.3,
\[
\delta^{n+1}(x^{p^n}) = \delta(x^{p^n}(p^{\nu-1})) + \delta(A^{(n-n-1)}) \\
+ C_p(A^{(n-n-1)}) \\
\in \delta(x^{p^n}(p^{\nu-1})) + A^{(n-n-1)} \\
= x^{p^n-1}(p^{\nu-1}) + A^{(n-n-1)} \\
+ \phi^{n+1}(x^{n-n')(p^{\nu-1}) + A^{(n-n-1)} \\
= x^{p^n-1}(p^{\nu-1}) + A^{(n-n-1)}.
\]
This ends the induction step in case $n \geq \nu + 2$. □
Corollary 4.1.

\[ \delta^n(x^{p^n}) \in \begin{cases} A^{(0)} \cap J^{[p^n]} & \text{if } n \leq \nu, \\ A^{(n-\nu)} \cap J^{[p^n]} & \text{if } n \geq \nu + 1. \end{cases} \]

The following will also be useful later.

Lemma 4.2.

1) \( \delta^n(x^n) \in R[x, x', ..., x^{(n)}] \) for \( n \geq 0 \).

2) \( \delta^n(x_1x_2) \in R[x_1^p, x_2^p, x_1', x_2', x_1^{(n)}, x_2^{(n)}] \).

Proof. Trivial induction on \( n \). \( \square \)

5. \( p \)-Jets of \( \mu_{p^n} \)

Start again with the multiplicative group \( \mathbb{G}_m \) and the isogeny \([p^n] : \mathbb{G}_m \to \mathbb{G}_m\). Let

\[ \mu_{p^n} := \mathbb{G}_m[p^n] := \text{Ker}([p^n]_{\mathbb{G}_m}) = \text{Spec } R[x, x^{-1}]/(x^{p^n}-1) \]

be the kernel of \([p^n]_{\mathbb{G}_m}\). More generally, (for the purpose of looking later at \( p \)-divisible groups of elliptic curves) we consider, for any \( a \in R^\times \), the finite flat scheme

\[ \mu_{p^n}^a := \text{Spec } R[x]/(x^{p^n}-a). \]

Its functor of points is given by \( \mu_{p^n}^a(S) = \{ s \in S; s^{p^n} = a \} \) for any \( R \)-algebra \( S \). Then \( \mu_{p^n}^a \) has a natural structure of \( \mu_{p^n}^a \)-torsor. More generally, for any \( a, b \in R^\times \) we have a natural morphism

\[ \mu_{p^n}^a \times \mu_{p^n}^b \xrightarrow{\text{can}} \mu_{p^n}^{ab} \]

given on \( S \)-points by \( (s, t) \mapsto st \). We also have a natural isomorphism

\[ \mu_{p^n}^{b\nu} \xrightarrow{\text{can}} \mu_{p^n}^a \]

given on points by \( s \mapsto sb^\nu \).

Consider the group \( U_m = 1 + p^m R = U_1^m \).

So if \( a \in U_{\nu+1} \) then \( a = b^\nu \), \( b \in U_1 \), so division by \( b \) gives an isomorphism

\[ \mu_{p^n}^{a\nu} \cong \mu_{p^n}^a. \]

Finally the system \((\mu_{p^n}; \nu \geq 1)\) is a \( p \)-divisible group with embeddings \( \mu_{p^n} \subset \mu_{p^{n+1}} \) given by the inclusions on points. More generally for any \( a \in R^\times \) and any \( \nu_0 \geq 1 \), the schemes \((\mu_{p^n}^{a\nu_0}; \nu \geq \nu_0)\) form an inductive system with embeddings

\[ \mu_{p^n}^{a\nu_0} \subset \mu_{p^n}^{a\nu_0+1} \subset \ldots \subset \mu_{p^n}^{a\nu_0-\nu} \subset \mu_{p^n}^{a\nu_0+1} \subset \ldots \]

given by the inclusions on points. Note that if \( a \in U_{\nu_0+1} \) then we can write \( a = b^{\nu_0} \) and hence division by \( b \) gives an isomorphism between the inductive system (5.1) and the inductive system

\[ \mu_{p^{\nu_0}} \subset \mu_{p^{\nu_0}+1} \subset \ldots \subset \mu_{p^n} \subset \mu_{p^{n+1}} \subset \ldots \]
Recall that \( J^n(\mathbb{G}_m) = \text{Spf} \ R[x,x^{-1},x',...,x^{(n)}] \) and that \([p^n]\) \( J^n(\mathbb{G}_m) : J^n(\mathbb{G}_m)\) is given at the level of rings by \( x \mapsto x^{p^n}, \quad x' \mapsto \delta(x^{p^n}), \) etc. By the commutation of \( J^n \) with fiber products it follows that

\[
J^n([p^n]) = \text{Ker}(J^n([p^n]_{\mathbb{G}_m}) = \text{Ker}([p^n]_{J^n(\mathbb{G}_m)}) =: J^n(\mathbb{G}_m)[p^n].
\]

More generally, if \( a \in R^\times \), and if we still denote by \( a : \text{Spec} \ R \rightarrow \mathbb{G}_m \) the point defined by \( x \mapsto a \) then

\[
J^n(a) = J^n([p^n_{\mathbb{G}_m}]) = (J^n([p^n]_{\mathbb{G}_m})^{-1}(J^n(a)) = ([p^n]_{J^n(\mathbb{G}_m)})^{-1}(J^n(a))
\]

where \( J^n(a) : \text{Spec} \ R \rightarrow J^n(\mathbb{G}_m) \) is given, at the level of rings, by

\[
x \mapsto a, \quad x' \mapsto \delta a, \quad ..., x^{(n)} \mapsto \delta^n a.
\]

It follows that:

**Proposition 5.1.**

\[
\mathcal{O}^n(a) = \frac{R[x,x',...,x^{(n)}]}{(x^{p^n} - a, \delta(x^{p^n}) - \delta a, ..., \delta^n(x^{p^n}) - \delta^n a)}
\]

In particular

\[
\mathcal{O}^n(a) = \frac{R[x,x',...,x^{(n)}]}{(x^{p^n} - 1, \delta(x^{p^n}), ..., \delta^n(x^{p^n}))}
\]

Alternatively Proposition 5.1 follows from Lemma 3.4.

**Remark 5.2.** Let \( J^n(a)_{1} \) be the kernel of the projection \( J^n(a) \rightarrow J^n(a) = \mu_{p^n} \) and write \( \mathcal{O}^n(a)_{1} := (J^n(a)(a)). \) Since \( \mu_{p^n} \) has a lift of Frobenius \( x \mapsto x^{p^n} \) in the category of group schemes we get an isomorphism

\[
\mathcal{O}^n(a) \approx \mathcal{O}(a) \otimes \mathcal{O}^n(a)_{1}
\]

compatible with the group laws. Equivalently we have

\[
J^n(a) = \mu_{p^n} \times J^n(a)_{1}
\]

as groups in the category of formal \( p \)-adic schemes over \( R \).

**Proposition 5.3.** For all \( n \geq 1 \) we have

\[
\lim_{\nu} \mathcal{O}^n(a)_{1} = R[x',...,x^{(n)}].
\]

**Proof.** By Proposition 5.1

\[
\mathcal{O}^n(a)_{1} = \frac{R[x',...,x^{(n)}]}{(\delta(x^{p^n})_{x=1}, ..., \delta^n(x^{p^n})_{x=1})}.
\]

Since, by Theorem 1.1 the denominator in the last equation is in \((p^n+1)\) we are done by the following well known fact (whose proof we recall).

**Lemma 5.4.** If \( A \) is a Noetherian ring, \( I \) is an ideal, \( A \) is \( I \)-adically complete, and \((L_n)\) is a descending sequence of ideals such that \( L_n \subset I^n \) then

\[
A = \lim_{n} A/L_n.
\]
Proof. The map from $A$ to the projective limit is clearly injective. It is surjective because if $f_n \in A$, $f_{n+1} - f_n \in L_n$ then $f_{n+1} - f_n \in I^n$ hence there exists $f \in A$ such that $f - f_n \in I^n$. Now fix $m$; since for $n \geq m$, $f - f_m = (f - f_n) + (f_n - f_m) \in I^n + L_m$ and (by [13], Theorems 8.2 and 8.14) $\cap_{n \geq 1}(I^n + L_m) = L_m$ we get $f - f_m \in L_m$. □.

According to our general notation (Equations 1.1 and 1.2) we next recall the rings
\[
\mathcal{O}^n(\mu_{p^n}), \quad \mathcal{O}^n(\mu_{p^n}).
\]
Also recall we set $U_m = 1 + p^m R = U_1^{p^{m-1}}, m \geq 1$.

**Proposition 5.5.** Let $n, \nu \geq 1$ and $a \in U_1$.

1) If $a \not\in U_{\nu+1}$ then $\mathcal{O}^n(\mu_{p^n}) = 0$.

2) If $a \in U_{\nu+1}$ then
\[
\mathcal{O}^n(\mu_{p^n}) \simeq \mathcal{O}^n(\mu_{p^n}) \simeq \frac{k[x, x', x'', \ldots, x^{(n)}]}{(x - 1)^{p^n}, (x')^{p^n}, \ldots, (x^{(n)})^{p^n}}.
\]
\[
\mathcal{O}^n(\mu_{p^n}) \simeq \mathcal{O}^n(\mu_{p^n}) \simeq \frac{k[x, x', x'', \ldots, x^{(n)}]}{(x^{p^n} - 1)} \quad \text{if } n \leq \nu,
\]
\[
\mathcal{O}^n(\mu_{p^n}) \simeq \mathcal{O}^n(\mu_{p^n}) \simeq \frac{k[x, x', x'', \ldots, x^{(n)}]}{(x^{p^n} - 1, (x')^{p^n}, \ldots, (x^{(n)})^{p^n})} \quad \text{if } n \geq \nu + 1.
\]

**Proof.** To prove assertion 1 let $a \in U_m \setminus U_{m+1}, 1 \leq m \leq \nu$. Then a simple induction shows that $\delta^m a \in R^\nu$ hence, by Corollary 1.3, the reduction mod $p$ of $\delta^m(x^{p^m}) - \delta^m a$ is in $k^\times$. By Proposition 5.1 $\mathcal{O}^N(\mu_{p^n}) = 0$ for all $N \geq m$ and hence $\mathcal{O}^n(\mu_{p^n}) = 0$ for all $n \geq 1$.

To prove assertion 2 note that the equalities (5.4) and (5.5) follow from Corollary 1.3 and Proposition 5.1; in particular we have
\[
\lim_{m} \mathcal{O}^n(\mu_{p^n}) = \frac{k[x, x', x'', \ldots]}{(x^{p^n} - 1, (x')^{p^n}, (x')^{p^n}, \ldots)}
\]
Now (5.3) follows from the fact that the intersection
\[
k[x, x', \ldots, x^{(n)}] \cap (x^{p^n} - 1, (x')^{p^n}, (x')^{p^n}, \ldots)
\]
in the ring $k[x, x', x'', \ldots]$ equals the ideal
\[
(x^{p^n} - 1, (x')^{p^n}, \ldots, (x^{(n)})^{p^n}).
\]
By the above Proposition we get:

**Corollary 5.6.** Let $n, \nu_0 \geq 1, a \in U_1$. Then
\[
\lim_{\nu} \mathcal{O}^n(\mu_{p^{\nu-\nu_0}}) = \begin{cases} 0 & \text{ if } a \not\in U_{\nu_0+1} \\ k[[x - 1, x', \ldots, x^{(n)}]] & \text{ if } a \in U_{\nu_0+1} \end{cases}
\]
We also remark the following:
Proposition 5.7. Assume $n, \nu \geq 1$. Then

1) $O^n(\mu_{\nu^p})$ is not a finite $R$-algebra.

2) $O^n(\mu_{\nu^p})$ is not a flat $R$-algebra.

Proof. If $O^n(\mu_{\nu^p})$ is a finite $R$-algebra then $O^n(\mu_{\nu^p})$ is a finite $k$-algebra which is not the case, cf. equations (5.4) and (5.5). If $O^n(\mu_{\nu^p})$ is a flat $R$-algebra then it is torsion free. Since, by equation (3.2), $\delta(x^{\nu^p}) \in p^2 x^{\nu(p^2-1)} x' + p^{\nu+1} A_1$ it follows that an element in $x' + pA_1$ is zero in $O^n(\mu_{\nu^p})$ hence $x'$ is in the denominator of the ring in equation (5.3), a contradiction. □

Remark 5.8. The author is indebted to A. Saha for pointing out assertion 2 in Proposition 5.7 above.

Remark 5.9. We end this section by providing an alternative argument for Corollary 1.5, hence of Proposition 5.5 in case $a = 1$; the author is indebted to the referee for this argument.

We start by recalling that for finitely generated $R$-algebras $B$ and $k$-algebras $C$ there are isomorphisms

$$\text{Hom}_k(\mathcal{J}^n(B), C) \simeq \text{Hom}_R(B, W_n(C))$$

functorial in both $B$ and $C$ and compatible with varying $n$. (Here $W_n(C) = (R^{n+1}, +, \times)$ is the ring of $p$-typical Witt vectors of length $n+1$.) This follows from the theory in [1]; cf. Section 3.4 of that paper. (The indexing of rings of Witt vectors is also taken from [1] and is not the classical one: the above $W_n$ are usually denoted by $W_{n+1}$.) Using the isomorphism above Corollary 1.5 is easily seen to be equivalent to the following statement.

Proposition 5.10. For any $k$-algebra $C$ and any Witt vector $x = (x_0, \ldots, x_n) \in W_n(C)$ we have

$$x^{\nu^p} = 1 \iff (x_0 - 1)^{\nu^{p^i}} = x_1^{\nu^{p^i}} = \ldots = x_{n-\nu}^{p^{\nu}} = 0$$

Here ... has the obvious meaning if $\nu \geq n$.

Proof. By multiplying $x$ with the inverse of the Teichmüller lift $(x_0, 0, \ldots, 0)$ of $x_0$ one may reduce to the case when $x_0 = 1$. Assume this from now on. Then by induction it is enough to show that for all $0 \leq i \leq n-\nu-1$ and under the assumption $x_1^{\nu^p} = \ldots = x_i^{\nu^p} = 0$ we have

$$x^{\nu^p} = 1 \iff x_{i+1}^{\nu^p} = 0$$

To show the latter we may further assume that $i = n - \nu - 1$. Now write $x = 1 + V(z)$ where $V$ is the Verschiebung and $z = (x_1, \ldots, x_n)$. Then we have

$$x^{\nu^p} = (1 + V(z))^{\nu^p} = 1 + p^i V(z) + \sum_{j=2}^{p^i} \left( \begin{array}{c} p^i \\ j \end{array} \right) (V(z))^j.$$
Since $C$ has characteristic $p$, if $F$ is the Frobenius, we have $FV = VF = p$ hence
\begin{align}
p^n V(z) &= V^{n+1} F^n V(z) \\
&= V^{n+1} (x_1^{p^n}, \ldots, x_{n-\nu}^{p^n}) \\
&= (0, \ldots, 0, x_1^{p^n}, \ldots, x_{n-\nu}^{p^n}) \\
&= (0, \ldots, 0, x_{n-\nu}^{p^n}).
\end{align}
(5.8)
By Equations 5.7 and 5.8 we are left to prove that for all $j \geq 2$,
\begin{align}
(p_j)(V(z))^j = 0 \in W_n(C).
\end{align}
(5.9)
Let $r = \text{ord}_p \left( \begin{bmatrix} p \\ j \end{bmatrix} \right)$. Then Equation 5.9 is equivalent (due to the general identity $V(a)V(b) = pV(ab)$) to
\begin{align}
0 = p'(V(z))^j = p^{r+j-1}V(z^j) = V^{r+j} F^{r+j-1}(z^j) = V^{r+j}(F^{r+j-1}(z))^j.
\end{align}
(5.10)
It is therefore enough to show that
\begin{align}
0 = F^{r+j-1}(z) = (x_1^{p^{r+j-1}}, \ldots, x_{n-r-j+1}^{p^{r+j-1}}).
\end{align}
Since we know that $x_1^{p^{\nu+1}} = \ldots = x_{n-\nu-1}^{p^{\nu+1}} = 0$, it is enough to show that $\nu + 2 \leq r + j$ or, equivalently,
\begin{align}
\text{ord}_p \left( \begin{bmatrix} p^n \\ j \end{bmatrix} \right) \geq \nu - j + 2
\end{align}
for $j \geq 2$, which is true (for $p \geq 3$) for elementary reasons. □

6. $p$-jets of the irreducible components of $\mu_{p^n}$

Next note that $\mu_{p^n}$ is connected and has $\nu + 1$ irreducible components:
\begin{align}
\mu_{p^n} = \bigcup_{i=0}^\nu \mu_{p^n,i}, \quad \mu_{p^n,i} := \text{Spec } R[\zeta_{p^n}],
\end{align}
where $\zeta_{p^n}$ is a primitive $p$-root of unity. So $\zeta_1 = 1$, $R[\zeta_1] = R = R[x]/(x - 1)$, and
\begin{align}
R[\zeta_{p^n}] = R[x]/(\Phi_{p^n}(x)), \quad \Phi_{p^n}(x) := \frac{x^{p^n} - 1}{x^{p^n-1} - 1}, \quad i \geq 1.
\end{align}
The scheme theoretic intersection of these components is
\begin{align}
\text{Spec } R[x]/(x - 1, p) = \text{Spec } k.
\end{align}
In deep contrast with Proposition 5.5 the $p$-jets of these components are completely uninteresting:

Proposition 6.1.
1) $\Theta^n(\mu_{p^n,0}) = R$ for $n \geq 1$;
2) $\Theta^n(\mu_{p^n,i}) = 0$ for $i \geq 1$ and $n \geq 1$.

Proof. The first equality is clear. The second follows from the fact that $\Phi_{p^n}(x + 1)$ is an Eisenstein polynomial plus the following general fact:
□
Proposition 6.2. Let \( f(x) = x^e + a_1x^{e-1} + \ldots + a_{e-1}x + a_e \in R[x] \) be an Eisenstein polynomial (i.e. \( e \geq 2, a_1, \ldots, a_e \in pR, a_e \notin p^2R \)) and let \( X = \text{Spec } R[x]/(f(x)) \). Then \( \mathcal{O}^n(X) = 0 \) for \( n \geq 1 \).

Proof. It is enough to show that \( \delta f(x) \) is invertible in \( \mathcal{O}^1(X) \otimes k \). Now we have:

\[
\delta f(x) = \delta(x^e) + \delta(a_1x^{e-1}) + \ldots + \delta(a_{e-1}x) + \delta a_e + C_p(x^e, a_1x^{e-1}, \ldots, a_e)
\]

\[
\equiv e^{p^{e-1}x} + x^{p^{e-1}}(\delta a_1) + \ldots + x^p(\delta a_{e-1}) + \delta a_e \mod p.
\]

Since the image of \( x \) in \( \mathcal{O}^1(X) \otimes k \) is nilpotent and the image of \( \delta a_e \) in the same ring is invertible it follows that the image of \( \delta f(x) \) in this ring is invertible which ends the proof. \( \square \)

7. \( p \)-jets of \( E[p^r] \) for ordinary elliptic curves

We start with a review of extensions of \( p^{-\nu}\mathbb{Z}/\mathbb{Z} \) by \( \mu_{p^r} \). For any group (respectively group scheme) \( G \) we denote by \( G[N] \) the kernel of the multiplication by \( N \) map. For a finite group \( \Gamma \) we continue to denote by \( \Gamma \) the étale group scheme over \( R \) attached to \( \Gamma \); so for any connected \( R \)-algebra \( S \), \( \Gamma(S) = \Gamma \). In particular we have the connected \( R \)-group scheme \( \mu_{p^r} = \mathbb{G}_m[p^r] \). Also one can consider the étale \( R \)-group scheme \( p^{-\nu}\mathbb{Z}/\mathbb{Z} \). Let \( R_m = R/p^mR, m \geq 1 \). We also view, when appropriate, \( \mu_{p^r} \) and \( p^{-\nu}\mathbb{Z}/\mathbb{Z} \) as \( R_m \)-group schemes via base change. Then, by Kummer theory,

\[
\text{Ext}_R^1(p^{-\nu}\mathbb{Z}/\mathbb{Z}, \mu_{p^r}) \simeq R_m^\times/(R_m^\times p^\nu)
\]

\[
\simeq (1 + pR_m)/(1 + pR_m p^\nu)
\]

\[
\simeq (1 + pR_m)/(1 + p^{\nu+1}R_m).
\]

We will need to recall the following explicit description of the above isomorphism. Let \( q \in 1 + pR \). Consider the finite flat \( R \)-scheme

\[
\Gamma_{p^r}^q = \prod_{i=0}^{p^r-1} \mu_{p^r}^q.
\]

This is a group scheme with multiplication given by

\[
\mu_{p^r}^q \times \mu_{p^r}^q \xrightarrow{can} \mu_{p^r}^{q+j} \xrightarrow{can} \mu_{p^r}^q,
\]

where \( 0 \leq l < p^r, i + j = l \mod p^r \). The functor of points of \( \Gamma_{p^r}^q \) is given by

\[
\Gamma_{p^r}^q(S) = \{(s, i); s \in S^\times, 0 \leq i < p^r, s^{p^r} = q^i\}
\]

for any \( R \)-algebra \( S \) with connected spectrum; the multiplication on points is given by \( (s, i) \cdot (t, j) = (st, i + j) \) if \( i + j < p^r \) and \( (s, i) \cdot (t, j) = (st/q, i + j - p^r) \) if \( i + j \geq p^r \). We have an extension

\[
0 \to \mu_{p^r} \to \Gamma_{p^r}^q \to p^{-\nu}\mathbb{Z}/\mathbb{Z} \to 0
\]

(7.2)

(the second map being given on points by \( (s, i) \mapsto \frac{i}{p^r} + \mathbb{Z} \)). Then Kummer theory gives:
Lemma 7.1. The isomorphism (7.1) is given by attaching to the class of $q \in 1 + pR$ in $(1 + pR)/Q(1 + p^{q+1}R)$ the class of the extension (7.2) in $\text{Ext}_{R_m}^1(p^{-\nu}Z/\mathbb{Z}, \mu_{p^n})$.

Note that the system $(\Gamma_{p^r}: \nu \geq 1)$ is a $p$-divisible group via the morphisms $\Gamma_{p^r} \to \Gamma_{p^{r+1}}$ given on points by $(s, i) \mapsto (s, pi)$ and given on schemes by the inclusions $\mu_{p^r}^\nu \subset \mu_{p^{r+1}}^\nu$. The $p$-divisible group $(\Gamma_{p^r}: \nu \geq 1)$ is an extension of the $p$-divisible group $(p^{-\nu}Z/\mathbb{Z}: \nu \geq 1)$ by the $p$-divisible group $(\mu_{p^n}: \nu \geq 1)$, where the latter are viewed as $p$-divisible groups with respect to the natural inclusions.

Next we consider an elliptic curve $E/R$. References for this are [14, 10]. Let $\bar{E}/k$ be its reduction mod $p$ and let $\bar{E}^{\text{for}}$ be the formal group attached to $E$. (We use the superscript for rather than $\tilde{\text{ }}$ because the latter is used in the present paper to denote $p$-adic completion). Let $E^{\text{for}}[p^n]$ be the kernel of the multiplication by $p^n$ on $E^{\text{for}}$, viewed as a finite flat group scheme over $R$. Assume in what follows that $E$ is ordinary. Then

$$E^{\text{for}} \simeq \bar{G}^{\text{for}}_m;$$

we fix such an isomorphism. So we have induced isomorphisms $E^{\text{for}}[p^n] \simeq \mu_{p^n}$. Moreover we fix isomorphisms

$$\bar{E}(k)[p^n] \simeq \mathbb{Z}/p^n\mathbb{Z} \simeq p^{-\nu}Z/\mathbb{Z}.$$ 

With the isomorphisms (7.3) and (7.4) fixed one defines the Serre-Tate parameter $q = q(E) \in 1 + pR$ of $E$ as follows. The isomorphisms (7.4) define a basis $(\alpha_{\nu})$ of the Tate module $T_{\nu, \mathbb{E}} = \lim_{\nu} \mathbb{E}(k)[p^n]$, $\alpha_{\nu} \in \mathbb{E}(k)[p^n]$ a generator, $p\alpha_{\nu} = \alpha_{\nu+1}$. If $A_{\nu} \in E(R)$ lifts $\alpha_{\nu}$ then one defines the Serre-Tate parameter $q(E) \in 1 + pR$ as the image of $\lim p^n A_{\nu} \in E^{\text{for}}(R)$ via the isomorphism $E^{\text{for}}(R) \simeq 1 + pR$ induced by (7.3); cf. [10], section 2. On the other hand with the isomorphisms (7.3) and (7.4) fixed there are induced exact sequences of finite flat group schemes over $R$:

$$0 \to \mu_{p^{n}} \to E[p^n] \to p^{-\nu}Z/\mathbb{Z} \to 0.$$ 

 Cf., say, [10]. Also by loc.cit. we have

Lemma 7.2. The class of the extension (7.5) is the image of the Serre-Tate parameter $q(E) \in 1 + pR$ under the isomorphism (7.1).

We conclude by Lemmas 7.1 and 7.2 that if $q = q(E)$ then $E[p^n]$ and $\Gamma_{p^r}$ are isomorphic as extensions over $R_m$ for any $m$; the isomorphisms are compatible as $m$ varies so we get the following:

Corollary 7.3. $E[p^n]$ and $\Gamma_{p^r}$ are isomorphic as extensions over $R$ for $q = q(E)$. In particular if $0 \leq i < p^n$ and $\theta = \frac{1}{p^n} \mathbb{Z} \subset p^{-\nu}Z/\mathbb{Z}$ then the connected component $E[p^n]_{\theta}$ of $E[p^n]$ lying above $\theta$ is isomorphic to $\mu_{p^n}$. Consequently if
we fix $\nu_0$ and an index $0 \leq i_0 < p^{\nu_0}$ and if $\theta = \frac{i_0}{p^{\nu_0}} + \mathbb{Z} \in p^{-\nu_0}\mathbb{Z}/\mathbb{Z}$ then the inductive system

$$E[p^{\nu_0}]_0 \subset E[p^{\nu_0+1}]_0 \subset E[p^{\nu_0+2}]_0 \subset ... \subset E[p^\nu]_0 \subset ...$$

identifies with the inductive system

$$\mu_{p^{\nu_0}} \subset \mu_{p^{\nu_0+1}} \subset \mu_{p^{\nu_0+2}} \subset ... \subset \mu_{p^\nu} \subset ...$$

Putting together Proposition 5.5 and Corollary 7.3 (and making the change of variables $x \mapsto x + 1$) we get:

**Proposition 7.4.** Let $E/R$ be an elliptic curve with ordinary reduction and Serre-Tate parameter $q = q(E) \in U_1$. Let $n \geq 1$, let $\theta \in \bigcup_{m \geq 1} p^{-m}\mathbb{Z}/\mathbb{Z}$, and let $\nu_0 \geq 1$ be minimal with the property that $\theta \in p^{-\nu_0}\mathbb{Z}/\mathbb{Z}$. Let $\nu \geq \nu_0$ and let $E[p^\nu]_0$ be the connected component of $E[p^\nu]$ lying over $\theta$. Then:
1) If $q \not\in U_{\nu_0+1}$ and $\theta \neq 0$ then $\mathcal{O}^n(E[p^\nu]_0) = 0$.
2) If $q \in U_{\nu_0+1}$ or $\theta = 0$ then

$$\mathcal{O}^n(E[p^\nu]_0) \simeq \frac{k[x, x', x'', ... , x^{(n)}]}{(x^{(p^n)}, (x')^{p^n}, ... , (x^{(n)})^{p^n})}.$$  

$$\mathcal{O}^n(E[p^\nu]_0) \simeq \frac{k[x, x', x'', ... , x^{(n)}]}{(x^{(p^n)})} \quad \text{if } n \leq \nu,$$

$$\mathcal{O}^n(E[p^\nu]_0) \simeq \frac{k[x, x', x'', ... , x^{(n)}]}{(x^{(p^n)}, (x')^{p^n}, ... , (x^{(n-v)})^{p^n})} \quad \text{if } n \geq \nu + 1.$$  

**Corollary 7.5.** Let $n, \nu_0 \geq 1$, $q \in U_1$. Then

$$\lim_{\nu} \mathcal{O}^n(E[p^\nu]_0) = \begin{cases} 0 & \text{if } q \not\in U_{\nu_0+1} \text{ and } \theta \neq 0, \\ k[[x, x', ... , x^{(n-1)}]] & \text{if } q \in U_{\nu_0+1} \text{ or } \theta = 0. \end{cases}$$

Also Proposition 5.7 and Corollary 7.3 imply

**Corollary 7.6.** Let $n, \nu \geq 1$. Then the map

$$J^n([p^\nu]_E) = [p^\nu]_J^n(E) : J^n(E) \to J^n(E)$$

is neither finite not flat.

An analogue of Remark 1.4 should hold nevertheless in the elliptic case as well.

8. $p$-jets of $J[p^n]$

Recall that a series $F(x) \in xR[[x]]$ without constant term is said to have finite height if $F \not\equiv 0 \mod p$; if this is the case the height of $F$ is defined as the largest integer $h \geq 0$ such that $F \in R[[x^p]] + pR[[x]]$.  

*Documenta Mathematica 18 (2013) 943–969*
Remark 8.1. The main example we have in mind here arises as follows. Consider a formal group law $F \in R[[x_1, x_2]]$. By [14] the multiplication by $p$ in $F$ is given by a series $F(x) := [p]_F(x)$ satisfying $F(x) \equiv px \mod x^2$. The height of $F$ is defined to be the height of $F(x)$; if the height is finite then it is $\geq 1$. Not every series of height $\geq 1$ which is $\equiv px \mod x^2$ is the multiplication by $p$ of a formal group law $F$; indeed if $F(x) = [p]_F$ has height $h$ then one knows that the $x$-adic valuation of the reduction mod $p$ of $F$ in $k[[x]]$ is exactly $p^h$; cf. [14], p.127.

Let $F(x) \in xR[[x]]$ be a series of finite height $h \geq 1$ and let $F^{\circ \nu} = F \circ \cdots \circ F$ be the $\nu$ fold composition of $F$ with itself for $\nu \geq 1$. Let $e p^h$ be the $x$-adic valuation of the reduction mod $p$ of $F$. (So if $F(x) = [p]_F(x)$ for some formal group law $F$ then $e = 1$.) By “Weierstrass preparation” (cf. [11], p. 130) $F^{\circ \nu} = U_\nu \cdot P_\nu$ where $U_\nu \in R[[x]]^\times$ and $P_\nu \in R[x]$ is monic of degree $e^\nu p^{h\nu}$, $P_\nu \equiv x^{e^\nu p^{h\nu}} \mod p$. Consider the scheme:

$$X_\nu := \text{Spec } R[[x]] / (F^{\circ \nu}) = \text{Spec } R[[x]] / (P_\nu) \cong \text{Spec } R[x] / (P_\nu).$$

the latter isomorphism follows from “Euclid division” by $P_\nu$ in $R[[x]]$; cf. [11], p. 129. So $X_\nu$ is a finite flat scheme over $R$ of degree $e^\nu p^{h\nu}$ and we have a natural sequence of closed immersions

$$X_1 \subset X_2 \subset \cdots \subset X_\nu \subset \cdots$$

Our aim in this section is to understand the rings $\mathcal{O}^n(X_\nu)$. Note that if $F(x) = [p]_F(x)$ is the multiplication by $p$ on some formal group law $F$ then $X_\nu = \mathcal{F}[p^\nu]$, where the latter is the kernel of $[p^\nu]_F$ on $F$ and indeed the inductive system (8.1) coincides with the $p$-divisible group

$$\mathcal{F}[p] \subset \mathcal{F}[p^2] \subset \cdots \subset \mathcal{F}[p^\nu] \subset \cdots$$

of $F$; cf. [15].

Remark 8.2. Recall (cf. [12], p. 480) that any formal group law $F$ over $R$ of height $h = 1$ is isomorphic to the multiplicative formal group law hence in particular $\mathcal{F}[p^\nu] \simeq \mu_{p^\nu}$. Hence our analysis in Section 5 applies to $\mathcal{O}^n(\mathcal{F}[p^\nu])$ in the height one case. We will consider in what follows the case of formal groups of arbitrary height $\geq 1$. More generally we will treat the case of iterates of series of height $\geq 1$ which are not necessarily coming from formal groups; so even for height 1 our analysis below will not be covered by Section 5.

To understand the rings $\mathcal{O}^n(X_\nu)$ we will first perform some computations in characteristic zero culminating with a proof of Theorem 1.7. Then we will reduce mod $p$ the outcome of these computations.
We begin by noting that:
\[ \mathcal{O}^n(X_v) = R[x, x', ..., x^{(n)}]/(P_v, \delta P_v, ..., \delta^n P_v) \]
\[ (8.2) \]
\[ = R[[x]][x', ..., x^{(n)}]/(P_v, \delta P_v, ..., \delta^n P_v) \]
\[ = R[[x]][x', ..., x^{(n)}]/(F_v^\circ, \delta(F_v^\circ), ..., \delta^n(F_v^\circ)) \]

So we will be concerned from now on with understanding the structure of the expressions \( \delta^i(F_v^\circ) \). To do this we need to develop some filtration machinery on power series.

We start by considering the decreasing filtration of \( A^0 := R[[x]] \) by the subrings \( A^0_v, \nu \geq 1 \), defined by
\[ A^0_v = R[[x^{(n)}]] + pR[[x^{(n-1)}]] + p^2 R[[x^{(n-2)}]] + ... + p^\nu R[[x]] \subset R[[x]]. \]

Let \( v_p \) be the \( p \)-adic valuation on \( R \).

**Lemma 8.3.**
1. \( A^0_v = \{ \sum_{n \geq 0} a_n x^n \in R[[x]] : v_p(a_n) \geq \nu - v_p(n) \} \).
2. If \( G_1, G_2, G_3, ... \in A^0_v, G_m \in x^m R[[x]] \). Then \( \sum_{m \geq 1} G_m \in A^0_v \).
3. If \( H \in A^0_v, H(0) = 0, \) and \( G \in R[[x]] \) then \( G(H(x)) \in A^0_v \).
4. \( pA^0_v \subset A^0_{\nu^2} \).
5. If \( G \in A^0_v, \) then \( G^n \in A^0_{\nu^2} \).
6. If \( F \in A^0_v \) and \( F(0) = 0 \) then \( F^\circ \in A^0_v \).

**Proof.** Assertion 1 is easy. Assertion 2 clearly follows from assertion 1. Assertion 3 clearly follows from assertion 2. Assertions 4 and 5 are clear. Assertion 6 follows from assertions 3, 4, 5. \( \square \)

We continue by considering the filtration
\[ A^n = R[[x]][x', ..., x^{(n)}]/, \quad n \geq 0 \]
on
\[ \mathcal{A} := \bigcup_{n \geq 0} A^n. \]
(Here \( A^0 = R[[x]] \).) There is a natural \( p \)-derivation \( \delta \) on \( \mathcal{A} \) sending \( \delta x = x', \delta x' = x'' \), etc. Note that \( \delta A^n \subset A^{n+1} \) for all \( n \). So according to equation (3.3) we may then consider the filtration
\[ A^{[n]} = A^n + pA^{n+1} + p^2 A^{n+2} + ... \]
on \( \mathcal{A} \). Finally let
\[ \mathcal{I} = (x, x', x'', ...) \subset \mathcal{A}. \]

So we may consider the descending filtration of \( \mathcal{I} \) by ideals \( \mathcal{I}^{[\nu]} \), \( \nu \geq 0 \). Note that with \( A^n, \mathcal{A}, I \) as in 4.1 we have \( A^n \subset A^n, \mathcal{A} \subset A, I \subset \mathcal{I} \), and hence \( I^{[\nu]} \subset \mathcal{I}^{[\nu]} \). Let \( n, i \geq 1 \). Note also that Lemma 3.2 immediately implies the injectivity of the map in Equation 1.4 because \( \mathcal{I}^{[\nu]} \subset (pA^n + \mathcal{I})^\nu \) and because \( A^n \) is separated in the topology given by the maximal ideal \( pA^n + \mathcal{I} \).

In what follows we prove a series of lemmas that will lead to Theorem 1.7.
Lemma 8.4.
\[
\delta^n(p^i x) = \begin{cases} 
    p^{i-n} \phi^n(x) + [(p^{i-n+1}A^{(0)}) \cap J[p^{i+1}]] & \text{if } n \leq i \\
    \phi^n(x^{(n-i)}) + [A^{(n-i-1)} \cap J[p^{i+1}]] & \text{if } n \geq i + 1
\end{cases}
\]

Proof. Induction on \( n \). The case \( n = 1 \) is clear. Now assume the above is true for some \( n \geq 1 \). If \( n \leq i - 1 \) we have
\[
\delta^{n+1}(p^i x) \subseteq \delta(p^{i-n} \phi^n(x)) + \delta((p^{i-n+1}A^{(0)}) \cap J[p^{i+1}]) + C_p(p^{i-n}A^{(0)}, (p^{i-n+1}A^{(0)}) \cap J[p^{i+1}]) \\
\subseteq p^{i-n-1} \phi^{n+1}(x) - p^{i-(n-1)} \phi^n(x) + (p^{i-n}A^{(0)}) \cap J[p^{i+1}]
\]
If \( n \geq i + 1 \) we have
\[
\delta^{n+1}(p^i x) \subseteq \delta(\phi^{(n-i)}(x)) + \delta(A^{(n-i-1)} \cap J[p^{i+1}]) + C_p(A^{(n-i)}, A^{(n-i-1)} \cap J[p^{i+1}]) \\
\subseteq \phi^{(n-i)}(x^{(n-1-i)}) + A^{(n-i-1)} \cap J[p^{i+1}].
\]
The case \( n = i \) is similar. □

Lemma 8.5. Let \( G \in xR[[x]] \). Then for all \( n \geq 1 \):
\[
\delta^n(G(x)) = \left( \frac{dG}{dx} \right) x^{(n)} + [A^{(n-1)} \cap J[p]].
\]

Proof. We proceed by induction on \( n \). To check the statement for \( n = 1 \) write \( G(x) = \sum_{m \geq 1} a_m x^m \); then
\[
\delta(G(x)) = \frac{1}{p} \left[ (\sum_{m \geq 1} a_m)(x^p + px')^m - (\sum_{m \geq 1} a_m x^m)^p \right] \\
= \frac{1}{p} \left[ (\sum_{m \geq 1} a_m + p \sum_{m \geq 1} a_m)(x^p + px^{p(m-1)}x' + ...) - (\sum_{m \geq 1} a_m x^m)^p \right] \\
\subseteq (\frac{dG}{dx}) x' + A^{(0)} \cap J[p],
\]
which settles the case \( n = 1 \). For the induction step, assuming the statement true for some \( n \geq 1 \), we have
\[
\delta^{n+1}(G(x)) \subseteq \delta(\left( \frac{dG}{dx} \right) x^{(n)}) + \delta(A^{(n-1)} \cap J[p]) + C_p(A^{(n)}, A^{(n-1)} \cap J[p]) \\
\subseteq \left( \frac{dG}{dx} \right) x^{(n+1)} + \phi(x^{(n)}) \cdot \delta(\left( \frac{dG}{dx} \right) x^{(n)}) + A^{(n)} \cap J[p].
\]
Now, using Theorem 1.1, we have
\[
\phi(x^{(n)}) \cdot \delta(\left( \frac{dG}{dx} \right) x^{(n)}) \subseteq A^{(n)} \cap J[p]
\]
and we are done. □

Documenta Mathematica 18 (2013) 943–969
Next consider any series \( \Sigma = \Sigma(x) \in xR[[x]] \) and consider the unique ring endomorphism \( \Sigma^* : A \to A \) such that \( \Sigma^* x = \Sigma(x) \), \( \Sigma^* x' = \delta(\Sigma(x)) \), \( \Sigma^* x'' = \delta^2(\Sigma(x)) \), etc. Clearly \( \Sigma^*(A^n) \subset A^n \) for \( n \geq 0 \) and hence \( \Sigma^*(A^{(n)}) \subset A^{(n)} \) for \( n \geq 0 \). It is trivial to see that \( \Sigma^* \) and \( \delta \) commute on \( A \); similarly \( \Sigma^* \) and \( \phi \) commute on \( A \). Moreover for any two series \( \Sigma_1, \Sigma_2 \) we have the following compatibility of upper * with composition: \((\Sigma_1 \circ \Sigma_2)^* = \Sigma_2^* \circ \Sigma_1^* \).

Recall that for any integer \( a \in \mathbb{Z} \) we write \( a^+ = \max\{a, 0\} \).

**Lemma 8.6.** Assume \( \Sigma(x) = x^p^m \), \( m \geq 1 \), \( \nu \geq 0 \), \( n \geq 0 \); then:

1) \( \Sigma^*[j^p^\nu] \subset j^[[p^\nu+m]] \); 
2) \( \Sigma^*[A^n] \subset A^{(n-m)+} \).

**Proof.** To check assertion 1 it is enough to check it for \( m = 1 \). Now assertion 1 follows from the following computations in which \( i + j = \nu \):

\[
\Sigma^*(p^i(x^p)^j) = p^i(\delta^j(x^p)) = p^i(A^{(n-1)}) + j^[[p^\nu+1]] = j^[[p^\nu+1]]; 
\]

in the above we used the fact that since \( x^p \in j^[[p]] \) we have \( \delta^j(x^p) \in j^[[p]] \) hence \( (\delta^j(x^p)) \in j^[[p^\nu+1]]; \) cf. Lemma 3.1. To prove assertion 2 it is enough, by the compatibility with composition, to prove these two statements for \( m = 1 \) and \( n \geq 1 \) which we now assume. Now by Theorem 1.1 we have \( \delta^0(x^p) \in A^{(n-1)} \) for \( n \geq 1 \). Consequently, for \( n \geq 1 \) and \( F \in A^{n+i} \) we have

\[
\Sigma^*(p^i F(x, \ldots, x^{n+1})) = p^i F(x^p, \ldots, \delta^0(x^p)) \in p^i A^{(n-i)} \subset A^{(n-1)} 
\]

which proves assertion 2. \( \Box \)

We are ready to prove Theorem 1.7:

**Proof of Theorem 1.7.** Set \( \Sigma(x) = x^p^m \). Using Lemmas 8.4, 8.5, 8.6 we have the following computation for \( 1 \leq m \leq i \):

\[
\delta^m(p^i G(x^p)) = \delta^m(\Sigma^* G(p^i x)) \\
= \Sigma^* G^* (\delta^m (p^i x)) \\
= \Sigma^* G^* (\delta^m (p^i x)) \\
\in \Sigma^* G^* (\{p^i-m \delta^m (x) + (p_i-m+1) A^{(0)} \cap j^[[p^\nu+1]]\}) \\
\subset \Sigma^* G^* (\{p^i-m \delta^m (x) + (p_i-m+1) A^{(0)} \cap j^[[p^\nu+1]]\}) \\
\subset \{p^{i-m} \delta^m (x^p) + (p_i-m+1) A^{(0)} \cap j^[[p^\nu+1]]\}.
\]

For \( m \geq i+1 \) we have:

\[
\delta^m(p^i G(x^p)) = \Sigma^* G^* (\delta^m (p^i x)) \\
\in \Sigma^* G^* (\delta^m (p^i x)) \\
\subset \Sigma^* G^* (\phi(x^{(m-i)}) \cap j^[[p^\nu+1]]) \\
\subset \phi \Sigma^* (\delta^m (p^i G) + A^{(m-\nu-1)+}) \cap j^[[p^\nu+1]] \\
\subset \phi \Sigma^* (\delta^m (p^i G) + A^{(m-\nu-1)+}) \cap j^[[p^\nu+1]] \\
\subset \phi \Sigma^* (\{\frac{d}{dx} x^p \}^{(m-i)} + A^{(m-\nu-1)+}) \cap j^[[p^\nu+1]] \\
\subset \phi \Sigma^* (\{\frac{d}{dx} x^p \}^{(m-i)} + A^{(m-\nu-1)+}) \cap j^[[p^\nu+1]].
\]

\( \Box \)
Now Theorem 1.7 and Corollary 1.3 trivially imply:

**Corollary 8.7.** Let \( m \geq 0, \nu \geq 1, i + j = \nu, i, j \geq 0, G \in xR[[x]]. \) Then the element \( \delta^m(p^\nu G(x^\nu)) \in \mathcal{A}^n \) is given by

\[
\delta^m(p^\nu G(x^\nu)) = \begin{cases} 
0 & \text{if } m < i \\
G(x^\nu)^{p^i} & \text{if } m = i \\
\left( x^{p^i-1} \frac{d}{dx}(x^\nu) \right)^{m_{\nu}} (x^{(n-\nu)}p^\nu + [\mathcal{A}^{m-\nu-1} \cap \mathcal{I}^{p^\nu}]) & \text{if } m > \nu 
\end{cases}
\]

In particular \( \delta^m(p^\nu G(x^\nu)) \in \mathcal{A}^{(m-\nu)^+} \cap \mathcal{I}^{p^\nu} \).

**Lemma 8.8.** Let \( y \) be an \( N \)-tuple \( y_1, \ldots, y_N \) of variables. Then, for \( n \geq 1 \)

\[
\delta^n \left( \sum_{i=1}^{N} \lambda_i \right) = \sum_{i=1}^{N} \lambda_i^{(n)} + P_{N,n}(y, y', \ldots, y^{(n-1)})
\]

in \( R[y] \) where \( P_{N,n} \) is a polynomial with \( \mathbb{Z} \)-coefficients without constant term or linear terms.

**Proof.** Induction on \( n \). \( \square \)

Note that Corollary 8.7 and Lemma 8.8 immediately imply Corollary 1.8. Also Corollary 1.8 immediately implies Corollary 1.10; one can prove a slightly more precise result:

**Proposition 8.9.** Let \( F(x) \in xR[[x]] \) be a series of finite height \( h \geq 1 \) satisfying \( F(x) \equiv px \mod x^2 \). For all \( \nu \geq 1 \) consider the scheme \( X_{\nu} := \text{Spec} \left( \frac{R[[x]]}{(x^\nu)} \right) \). Then we have:

\[
\mathcal{O}^n(X_{\nu}) = k[x, x', \ldots, x^{(n)}] \quad \text{if } n \geq 1,
\]

\[
\mathcal{O}^n(X_{\nu}) = \frac{k[x, x', \ldots, x^{(n)}]}{(x^{p^\nu}, (x')^{p^\nu}, \ldots, (x^{(n-\mu)})^{p^\nu})} \quad \text{if } n \geq \nu,
\]

\[
\mathcal{O}^n(X_{\nu}) = \frac{k[x, x', \ldots, x^{(n)}]}{(x^\mu)} \quad \text{if } 1 \leq n \leq \nu - 1,
\]

where \( \mu \geq p^\nu \).

**Proof.** By assertion 6 in Lemma 8.3 we may write \( E^{\nu \mu} = \sum_{j=0}^\nu p^{\nu-j} G_j(x^{p^j}), \) \( G_j \in R[[x]], j \geq 0. \) We may choose the \( G_j \)'s in \( xR[[x]] \) and then \( G_0(x) \equiv x \mod x^2. \) Also \( p^{\nu-\mu} \frac{d^2}{dx^2} \equiv 1 \mod x. \) We conclude by Corollary 1.8 and equation (8.2). \( \square \)
9. \( p \)-jets of \( E[p^n] \) for supersingular elliptic curves

**Proposition 9.1.** Let \( E/R \) be an elliptic curve with supersingular reduction and \( E[p^n] \) the kernel of the multiplication by \( p \). Then for any \( n \geq 1 \) we have:

\[
\mathcal{O}^n(E[p^n]) = \begin{cases} 
    k[x, x', \ldots, x^{(n)}] & \text{if } n \geq 1, \\
    k[x, x', \ldots, x^{(n)}] & \text{if } n \geq \nu, \\
    k[x, x', \ldots, x^{(n)}] & \text{if } 1 \leq n \leq \nu - 1,
\end{cases}
\]

where \( \mu \geq p^\nu \).

**Proof.** Since \( E \) has supersingular reduction \( E[p^n] \) is connected so it is equal to \( \mathcal{F}[p^n] \) where \( \mathcal{F} \) is the formal group law of \( E \) and we conclude by Proposition 8.9. \( \square \)

**References**

$p$-Jets of Finite Algebras, I

Alexandru Buium
Department of Mathematics and Statistics
University of New Mexico
Albuquerque NM 87131
USA
buium@math.unm.edu
\textbf{Abstract.} We study the structure of the \( p \)-jet spaces of the \( p \)-typical Witt rings of the \( p \)-adics. We also study the \( p \)-jets of the comonad map. These data can be viewed as an arithmetic analogue, for the ring \( \mathbb{Z} \), of the Lie groupoid of the line and hence as an infinitesimal version of the Galois group of \( \mathbb{Q} \) over \( \mathbb{F}_1 \).

2010 Mathematics Subject Classification: 13K05
Keywords and Phrases: Witt vectors

\section{Introduction}

1.1. Motivation. This paper is the second in a series of papers where we investigate \( p \)-jet spaces (in the sense of \([6]\)) of finite flat schemes/algebras. The understanding of such \( p \)-jet spaces seems to hold the key to a number of central questions about \textit{arithmetic differential equations} \([7]\). This paper is logically independent of its predecessor \([8]\). In \([8]\) we dealt with the case of \( p \)-divisible groups; in the present paper we investigate the case of algebras of Witt vectors of finite length. Another example of a class of finite algebras whose \( p \)-jet spaces are arithmetically significant is that of Hecke algebras; we hope to undertake the study of this example in a subsequent work.

The present paper is partly motivated by the quest for “absolute geometries” (the so-called “geometries over the field with one element, \( \mathbb{F}_1 \)”; cf. \([12]\) for an overview of various approaches and some history. In particular, according to Borger’s approach \([3]\), the geometry over \( \mathbb{F}_1 \) should correspond to \( \lambda \)-\textit{geometry} (i.e. algebraic geometry in which all rings appear equipped with a structure of \( \lambda \)-ring in the sense of Grothendieck). For the case of one prime \( p \) the “\( p \)-adic completion” of \( \lambda \)-geometry is the \( \delta \)-\textit{geometry} developed by the author \([6, 7]\), where \( \delta \) is a \( p \)-\textit{derivation} (morally a “Fermat quotient operator”). Now Borger established in \([3]\) an elegant categorical framework which predicts what actual objects should correspond to the basic hypothetical constructions over \( \mathbb{F}_1 \). According to his framework the hypothetical tensor product \( \mathbb{Z} \otimes_{\mathbb{F}_1} \mathbb{Z} \)
(which was one of the first objects sought in the quest for $\mathbb{F}_1$) should correspond to the big Witt ring $\mathcal{W}(\mathbb{Z})$ of the integers. Then the hypothetical groupoid structure on $\mathbb{Z} \otimes \mathbb{F}_1 \mathbb{Z}$ should correspond to the comonad structure $\Delta : \mathcal{W}(\mathbb{Z}) \rightarrow \mathcal{W}(\mathcal{W}(\mathbb{Z}))$. The main interest in $\mathbb{Z} \otimes \mathbb{F}_1 \mathbb{Z}$ comes from the fact that this tensor product should be viewed as an arithmetic analog of a surface $X \times X$ where $X$ is a curve (algebraic, analytic, $C^\infty$). With this analogy in mind one is immediately tempted to ask for an arithmetic analogue of the Lie groupoid of $X$, in the sense of Lie and Cartan, and more recently Malgrange [11]. (Recall that, roughly speaking, a point of the Lie groupoid of $X$ is by definition a pair of points of $X$ together with a formal isomorphism between the germs of $X$ at these two points.) Since the Lie groupoid of $X$ is the infinitesimal version of an automorphism group we should view any arithmetic analogue of the Lie groupoid of $X$ as an infinitesimal version of the “Galois group of $\mathbb{Q}/\mathbb{F}_1$”. Now the Lie groupoid of $X$ is an open set in the inverse limit, as $n \rightarrow \infty$, of the manifolds $J^n(X \times X/X)$ of “$n$-jets of formal sections, at various points, of the second projection $X \times X \rightarrow X$”. Since the arithmetic analogue of the second projection $X \times X \rightarrow X$ is the structure morphism $\text{Spec} \mathcal{W}(\mathbb{Z}) \rightarrow \text{Spec} \mathbb{Z}$, one candidate for an arithmetic analogue of the Lie groupoid of $X$ could be the jet spaces (in the sense of [4]) of the Witt ring $\mathcal{W}(\mathbb{Z})$ over $\mathbb{Z}$. We will not recall the definition of these jet spaces here (because we don’t need it) but let us note that they are constructed using derivations and knowing them essentially boils down (in this easy case) to knowing the Kähler differentials $\Omega_{\mathcal{W}(\mathbb{Z})/\mathbb{Z}}$. By the way, the module of Kähler differentials $\Omega_{\mathcal{W}(\mathbb{Z})/\mathbb{Z}}$ is also the starting point for the construction of the deRham-Witt complex of $\mathbb{Z}$ [9]. However using Kähler differentials (equivalently usual derivations) arguably looks like “going arithmetic only half way”. What we propose in this paper is to “go arithmetic all the way” and consider $p$-jet spaces (in the sense of [6]) of Witt rings rather than usual jet spaces (in the sense of [4]) of the same Witt rings. The former are an arithmetic analogue of the latter in which usual derivations are replaced by $p$-derivations.

A few adjustments are in order. First since $p$-jet spaces are “local at $p$” we replace the big Witt ring functor $\mathcal{W}(\ )$ by the $p$-typical Witt functor $\mathcal{W}_p(\ )$. Also we replace $\mathbb{Z}$ by $\mathbb{Z}_p$ or, more generally, by the Witt ring $R = W(k)$ on a perfect field $k$ of characteristic $p$. Finally since $W(R)$ is not of finite type over $R$ we replace $W(R)$ by its truncations $W_m(R)$ (where we use the labeling in [1], so $W_m(R) = R_{m+1}$ as sets.) So after all what we are going to study are the $p$-jet algebras $J^n(W_m(R))$ and the $p$-jet maps

$$J^n(\Delta) : J^n(W_m+m')(R) \rightarrow J^n(W_m(W_m'(R)))$$

induced by the comonad maps $\Delta : W_{m+m'}(R) \rightarrow W_m(W_m'(R))$; cf. the review of $J^n$ and $W_m$ in the next subsection. Since $W_m(R)$ and $W_m(W_m'(R))$ are finite flat $R$-algebras our investigation here is part of the more general effort to study $p$-jets $J^n(C)$ of finite flat $R$-algebras $C$; the case when $\text{Spec} C$ is a finite flat $p$-group scheme was addressed in [8].

Documenta Mathematica 18 (2013) 971–996
1.2. Main concepts and results. For $R = W(k)$ the Witt ring on a perfect field $k$ of characteristic $p \neq 2, 3$ we let $\phi = W(\text{Frob})$ be the automorphism of $R$ defined by the $p$-power Frobenius $\text{Frob}$ of $k$. (The main examples we have in mind are the ring of $p$-adic integers $\mathbb{Z}_p = W(F_p)$ and the completion of the maximum unramified extension of $\mathbb{Z}_p$, $\mathbb{Z}_p^\ur = W(F_p^\ur)$; here $F_p^\ur$ is the algebraic closure of $F_p$.)

Let $x, x', x'', \ldots, x^{(n)}, \ldots$ be families of variables $x = (x_α)_{α \in Ω}$, $x' = (x'_α)_{α \in Ω}$, etc., indexed by the same set $Ω$, and let $\phi : R[x, x', x'', \ldots] \to R[x, x', x'', \ldots]$ be the unique endomorphism extending $\phi$ on $R$ and satisfying $\phi(x_α^{(r)}) = (x_α^{(r)})^p + px_α^{(r+1)}$ for $r \geq 0$. Following [6] we define the map of sets (referred to as a $p$-derivation) $δ : R[x, x', x'', \ldots] \to R[x, x', x'', \ldots]$ by the formula

$$δF = \frac{\phi(F) - F^p}{p}.$$ 

Then for any $R$-algebra $C = R[x]/(f)$, where $f$ is a family of polynomials, we define the $p$-jet algebras of $C$:

$$J^n(C) = \frac{R[x, x', x'', \ldots, x^{(n)}]}{(f, δf, \ldots, δ^n f)}, \quad J^\infty(C) = \frac{R[x, x', x'', \ldots]}{(f, δf, δ^2 f, \ldots)}.$$ 

Note that each $J^{n+1}(C)$ has a natural structure of $J^n(C)$-algebra and we have naturally induced set theoretic maps $δ : J^n(C) \to J^{n+1}(C)$ and $δ : J^\infty(C) \to J^\infty(C)$. Note also that $ϕ$ on $R[x, x', x'', \ldots]$ induces ring homomorphisms $ϕ : J^n(C) \to J^n(C)$ and $ϕ : J^\infty(C) \to J^\infty(C)$. (For $C$ of finite type over $R$ we also defined in [6] the $p$-jet spaces of $\text{Spec} C$ as the formal schemes $J^n(\text{Spec} C) := \text{Spf } (J^n(C)^\wedge)$ where $\wedge$ means $p$-adic completion; these spaces are very useful when one further looks at non-affine schemes but here we will not need to take this step.)

We need one more piece of terminology. First, for any ring $B$ and element $b \in B$ we let $\overline{B} = B/pB$ and we let $\overline{b} \in \overline{B}$ be the image of $b$. Assume now the finitely generated $R$-algebra $C$ comes equipped with an $R$-algebra homomorphism $C \to R$ which we refer to as an augmentation. Then there is a unique lift of the augmentation to an $R$-algebra homomorphism $J^\infty(C) \to R$ that commutes with $δ$. Composing the latter with the natural homomorphism $J^n(C) \to J^\infty(C)$ and reducing mod $p$ we get an induced homomorphism $J^n(C) \to k$. Let $P_\prime$ be the kernel of the latter. Consider the ring $J^n(C)$ defined (up to isomorphism) by asking that $\text{Spec } J^n(C)$ be the connected component of $\text{Spec } J^n(C)$ that contains the prime ideal $P_\prime$; we refer to $J^n(C)$ as the identity component of $J^n(C)$. If $J^n(C)'$ is “the” ring such that

$$\text{Spec } J^n(C)' \simeq (\text{Spec } J^n(C)) \setminus (\text{Spec } J^n(C))'$$ 

then we call $J^n(C)'$ the complement of the identity component of $J^n(C)$. Clearly

$$J^n(C) \simeq J^n(C) \times J^n(C)'.$$
Let now $C$ be the Witt ring $W_m(R)$, $m \geq 1$. Recall that $W_m(R)$ is the set $R^{m+1}$ equipped with the unique ring structure which makes the ghost map $w : R^{m+1} \to R^{m+1}$, $w_i(a_0, \ldots, a_m) = a'_0 + pa'_1 + \ldots + p^ia_i$, a ring homomorphism. Let $v_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in W_m(R)$, (1 preceded by $i$ zeroes, $i = 1, \ldots, m$), set $\pi = 1 - \delta v_1 \in J^1(W_m(R))$, and let $\Omega_m = \{1, \ldots, m\}$. As usual we denote by $(J^n(W_m(R)))_{\pi}$ the ring of fractions of $J^n(W_m(R))$ with denominators powers of $\pi$. The ring $W_m(R)$ comes with a natural augmentation $W_m(R) \to R$ given by the first projection. So we may consider the identity component of $J^n(W_m(R))$. The following gives a compete description of this component and also shows this component is $(J^n(W_m(R)))_{\pi}$. 

**Theorem 1.1.** For $n \geq 2$ the image of $\pi^p$ in $(J^n(W_m(R)))_{\pi}$ is idempotent and we have an isomorphism

$$
(J^n(W_m(R)))_{\pi} \simeq \frac{k[x_i^{(r)}; i \in \Omega_m; 0 \leq r \leq n]}{(x_i^{(r)}, (x_i^{(r)})^p; i \in \Omega_m, 1 \leq r \leq n - 1)}
$$

sending each $\delta^r v_i$ into the class of the variable $x_i^{(r)}$.

Indeed by the theorem Spec $(J^n(W_m(R)))_{\pi}$ is connected (indeed irreducible) and contains $P_n$ hence $(J^n(W_m(R)))_{\pi}$ is isomorphic to the identity component of $J^n(W_m(R))$. By the way, since $(1 - \pi)^p = 1 - \pi^p$ is also idempotent in $J^n(W_m(R))$ it follows that $(J^n(W_m(R)))_{1-\pi}$ is isomorphic to the complement of the identity component of $J^n(W_m(R))$.

Next let $C$ be one of the iterated Witt rings $W_m(W_{m'}(R))$, $m, m' \geq 1$, (cf. the next section for more details). Set

$$
v_{i,i'} = (0, \ldots, 0, v_{i'}, 0, \ldots, 0) \in W_m(W_{m'}(R)),
$$

with $v_{i'} \in W_{m'}(R)$ preceded by $i$ zeroes in $W_{m'}(R)$ and set

$$
\Pi = (1 - \delta v_1)(1 - \delta v_{0,1}) \in J^1(W_m(W_{m'}(R))),
$$

$$
\Omega_{m,m'} = \{(0, \ldots, m) \times \{0, \ldots, m'\}\} \setminus \{(0,0)\}.
$$

There is a natural augmentation of $W_m(W_{m'}(R))$ given by composing the two obvious first projections. Then we have the following complete description for the identity component of $J^n(W_m(W_{m'}(R)))$.

**Theorem 1.2.** For $n \geq 2$ the image of $\Pi^p$ in $(J^n(W_m(W_{m'}(R))))_{\pi}$ is idempotent and we have an isomorphism

$$
(J^n(W_m(W_{m'}(R))))_{\pi} \simeq \frac{k[x_{i,i'}^{(r)}; (i, i') \in \Omega_{m,m'}; 0 \leq r \leq n]}{(x_{i,i'}^{(r)}, (x_{i,i'}^{(r)})^p; (i, i'), (j, j') \in \Omega_{m,m'}, 1 \leq r \leq n - 1)}
$$

sending each $\delta^r v_{i,i'}$ into the class of the variable $x_{i,i'}^{(r)}$. 

**DOCUMENTA MATHEMATICA 18 (2013) 971–996**
Again the theorem shows that
\[ \left( J^n(W_m(W_{m'}(R))) \right)_\Pi \text{ and } \left( J^n(W_m(W_{m'}(R))) \right)_{1-\pi} \]
are isomorphic to the identity component, respectively to the complement of the identity component, of \( J^n(W_m(W_{m'}(R))) \).
Finally we have the following complete description of the reduction mod \( p \) of the map induced by the comonad map:

**Theorem 1.3.** The map
\[ J^\infty(\Delta) : J^\infty(W_{m+m'}(R))_\pi \to J^\infty(W_m(W_{m'}(R)))_\Pi \]
sends \( x_{i,i'}^{(r)} \) into the class of
\[ \sum_{i+i'=i'} x_{i,i'}, \quad \text{if } r = 0 \]
and into the class of
\[ \delta^{r-1} \left( \sum_{i+i'=i'} x_{i,i'}^{i'} \right), \quad \text{if } r \geq 1. \]

**Remark 1.4.** The above results give a complete description of the identity components of our objects. On the other hand one can ask for a description of the complements of the identity components. Take for instance \( J^n(W_m(R)) \). This is not a group object so the components different from the identity component cannot be expected to necessarily “look like” the identity component. And this is indeed what happens (in spite of the comonad structure floating around): the complement \( \left( J^n(W_m(R)) \right)_{1-\pi} \) of the identity component looks quite differently (more degenerate) than the identity component \( \left( J^n(W_m(R)) \right)_\pi \). Indeed the identity component is a polynomial ring in \( m \) variables over a local Artin ring (cf. Theorem 1.1) and hence has Krull dimension \( m \); by contrast, for the complement of the identity component, we have:

**Theorem 1.5.** For \( n \geq 3 \) and \( m \geq 2 \) the ring \( \left( J^n(W_m(R)) \right)_{1-\pi} \) has Krull dimension \( \geq 2m - 1 \).

**Remark 1.6.** The simplicity modulo \( p \) of all these \( p \)-jet rings and maps may be deceptive. The structure of these objects in characteristic zero is actually extremely complicated and, as in [8], the whole point of this paper is to manage the complexity of the situation in such a way that, eventually, mod \( p \), the situation becomes transparent. On a conceptual level the results of this paper are best understood as an attempt to unravel the “differential geometry” of the “automorphisms of \( Z \) over \( F_1 \)”; cf. the beginning of our Introduction. The objects introduced and studied in the present paper could then be viewed as an “infinitesimal” replacement (at \( p \)) for the elusive absolute Galois group of \( Q \) over \( F_1 \).
1.3. Plan of the paper. In Section 2 we review (and give some complements to) the basic theory of Witt vectors. Section 3 is devoted to computing \( J^n(C) \) for certain finite flat \( R \)-algebras \( C \) whose structure constants satisfy some simple divisibility/vanishing axioms. These axioms are in particular satisfied in the cases \( C = W_m(R) \) and \( C = W_m(W_m'(R)) \). Using this we derive, in Section 4, the main results of this note, stated above.

1.4. Acknowledgment. This material is based upon work supported by the National Science Foundation under Grant No. 0852591, and by IHES, Bures sur Yvette, and MPI, Bonn. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation, IHES, or MPI. Also the author is indebted to James Borger for explaining some of his insights (and also some technical points) about Witt vectors.

2. Witt rings

In this section we review some basic facts about the rings of Witt vectors which we are going to need in the sequel. For most proofs we refer to [9] and [1]. However, for the convenience of the reader, we will provide proofs for the facts for which we could not find an explicit reference. Note that the labeling of Witt rings in [9] and [1] are different (\( W_m \) in [9] is \( W_{m-1} \) in [1]); we follow here the labeling in [1].

Fix a prime \( p \) and \( m \) a non-negative integer or \( \infty \). For any ring \( A \) we may consider the ghost maps \( w_i : A^{m+1} \rightarrow A^m, 0 \leq i \leq m, i < \infty \),

\[
w_i(a_0, ... , a_m) = a_0^{p^i} + pa_1^{p^{i-1}} + ... + p^i a_i.
\]

Then there is a unique functor \( W_m \) from the category of rings to itself such that, for any ring \( A \), we have that \( W_m(A) = A^{m+1} \) as sets and the ghost map \( w : W_m(A) \rightarrow A^{m+1} \) is a ring homomorphism where the target \( A^{m+1} \) is given the product ring structure. We use the convention \( \infty + 1 = \infty - 1 = \infty \) and we write \( W(A) = W_\infty(A) \). The ghost map \( w : W_m(A) \rightarrow A^{m+1} \) is an integral ring homomorphism and has a nilpotent kernel if \( 0 \leq m < \infty \); it is injective if \( A \) is \( p \)-torsion free and \( 0 \leq m \leq \infty \). The rings \( W_m(A) \) are called the \((p \text{-typical}) \) rings of Witt vectors of length \( m+1 \). There are natural additive maps, functorial in \( A \), called Verschiebung maps \( V : W_{m-1}(A) \rightarrow W_m(A) \) defined by

\[
V(a_0, a_1, a_2, ...) = (0, a_0, a_1, a_2, ...).
\]

Also there are unique ring homomorphisms, functorial in \( A \), called Frobenius maps, \( F : W_m(A) \rightarrow W_{m-1}(A) \), such that \( w \circ F = F^w \circ w \) where \( F^w : A^{m+1} \rightarrow A^m \) is the shift

\[
F^w(b_0, b_1, b_2, ...) = (b_1, b_2, ...).
\]

(If \( pA = 0 \) and \( m = \infty \) we have \( F = W(\text{Frob}) \) where \( \text{Frob} : A \rightarrow A \) is the \( p \)-power Frobenius.) For \( m < \infty \) one has also ring homomorphisms \( \rho : W_m(A) \rightarrow \)
Finally one has the multiplicative map, called the Teichmüller map, \([ \ ] = [ \ ]_m : A \to W_m(A)\), 

\[[a] = (a, 0, 0, \ldots).\]

These maps are related by the following identities:

1) \(F(V(u)) = pu\),

2) \(uV(u') = V(F(u)u')\),

3) \(F([a]) = [a^p]\),

4) If \(pA = 0\) then \(V(F(u)) = pu\).

It is also convenient to introduce the maps \(V^i_m = V^i \circ [ \ ]_{m-i} : A \to W_m(A)\), 

\(0 \leq i \leq m < \infty\). Then \(V^i_m(a) = (0, \ldots, 0, a, 0, \ldots, 0)\) where \(a\) is preceded by \(i\) zeroes. We have the identities:

\[
\begin{align*}
V^i_m(a) \cdot V^j_m(b) &= p^i \cdot V^j_m(a^{p^{i-1}} b), \quad 0 \leq i \leq j \leq m, \\
W(V^i_m(a)) &= (0, \ldots, 0, p^i a, p^i a^p, p^i a^p, \ldots).
\end{align*}
\]

Also, for any \(N \in \mathbb{Z}\) we have the following formula for the Teichmüller map \([9]\):

\[
[N]_m = \sum_{t=0}^{m} c_t(N)V^t_m(1)
\]

where \(c_0(N) = N\) and

\[
c_t(N) = \frac{N^{p^t} - N^{p^{t-1}}}{p^t}, \quad t \geq 1.
\]

If \(A\) is \(p\)-torsion free so is \(W_m(A)\). Now if \(A\) is \(p\)-torsion free and \(\phi : A \to A\) is a ring homomorphism lifting the \(p\)-power Frobenius on \(A/pA\) then there is a unique ring homomorphism \(\lambda_\phi : A \to W_m(A)\) such that \(v_i(\lambda_\phi(a)) = \phi^i(a)\) for all \(i\); if in addition \(A/pA\) is perfect then \(\lambda_\phi\) induces an isomorphism \(A/p^{m+1}A \simeq W_m(A/pA)\).

**Lemma 2.1.** Let \(A\) be a \(p\)-torsion free ring equipped with a ring automorphism \(\phi : A \to A\) lifting the \(p\)-power Frobenius on \(A/pA\). Let \(0 \leq m < \infty\) and view \(W_m(A)\) as an \(A\)-algebra via the homomorphism \(\lambda_\phi : A \to W_m(A)\). Set \(v_i = V^i_m(1), 0 \leq i \leq m\). Then \(\{v_i; 0 \leq i \leq m\}\) is an \(A\)-basis for \(W_m(A)\) and \(v_i v_j = p^i v_j\) for \(i \leq j\).

**Proof.** The case \(A = \mathbb{Z}_p\) is in [1]. The general case is similar but for convenience we recall the argument. If \(w : W_m(A) \to A^{m+1}\), by (2.2) and by the injectivity of \(\phi\), we have that \(w(v_i)\) are \(A\)-linearly independent in \(A^{m+1}\) (the latter viewed as an \(A\)-algebra via \((1, \phi, \ldots, \phi^m) : A \to A^{m+1}\)). Hence \(v_i\) are \(A\)-linearly independent. To check that \(v_i\) span \(W_m(A)\) we proceed by induction on \(m\). For \(m = 0\) this is clear. Assume spanning holds for \(m - 1\). The kernel of the map \(W_m(A) \to W_{m-1}(A)\) is \(V^m_{m}(A) = \{0, \ldots, 0, a; a \in A\}\). By induction the
images of \(v_0, ..., v_{m-1}\) in \(W_{m-1}(A)\) generate \(W_{m-1}(A)\) so it is enough to show that \(v_m\) generates \(V^m_{m}(A)\) as an \(A\)-module. This follows from the equality
\[
\lambda_{\phi}(a) \cdot v_m = (0, ..., 0, \phi^m(a)),
\]
plus the fact that \(\phi\) is surjective. The last assertion of the Lemma follows from (2.1).

\section{Lemma 2.2}

With notation as in Lemma 2.1 the Frobenius map \(F : W_m(A) \to W_{m-1}(A)\) is the unique \(\phi\)-linear map with \(F(v_i) = p \cdot \rho(v_{i-1}), 1 \leq i \leq m, v_0 = 1\).

**Proof.** The equalities \(F(v_i) = p \cdot \rho(v_{i-1})\) follow from the identity \(FV = p \cdot \text{id}\). We are left to prove that \(F \circ \lambda_{\phi} = \lambda_{\phi} \circ \phi : A \to W_{m-1}(A)\). It is enough to show that \(w \circ F \circ \lambda_{\phi} = \lambda_{\phi} \circ \phi\). This follows from the following computation:
\[
\begin{align*}
w \circ F \circ \lambda_{\phi} &= F^w \circ w \circ \lambda_{\phi} \\
&= F^w \circ (1, \phi, ..., \phi^m) \\
&= (\phi, ..., \phi^m) \\
&= (1, ..., \phi^{m-1}) \circ \phi \\
&= w \circ \lambda_{\phi} \circ \phi.
\end{align*}
\]

**Lemma 2.3.** Let \(A\) be Noetherian a \(p\)-torsion free ring equipped with a ring automorphism \(\phi : A \to A\) lifting the \(p\)-power Frobenius on \(A/pA\). Let \(w : A \to B\) be a \(p\)-torsion free \(A\)-algebra, let \(0 \leq m < \infty\), and view \(W_m(B)\) as an \(A\)-algebra via the homomorphism \(A \xrightarrow{\lambda_{\phi}} W_m(A) \xrightarrow{\lambda_{\phi}} W_m(B)\). If \(B\) is a finitely generated \(A\)-algebra (respectively a finite \(A\)-module) then \(W_m(B)\) is also a finitely generated \(A\)-algebra (respectively a finite \(A\)-module).

**Proof.** The ghost map \(w : W_m(B) \to B^{m+1}\) is injective and integral. Now if \(B\) is a finitely generated \(A\)-algebra (respectively a finite \(A\)-module) then so is \(B^{m+1}\) (with the \(A\)-algebra structure given by \((1, \phi, ..., \phi^m) : A \to A^{m+1} \to B^{m+1}\)), because \(\phi\) is bijective. In the finite case, by Noetherianity \(W_m(B)\) is a finite \(A\)-algebra. In the finitely generated case it follows that \(B^{m+1}\) is finite over \(W_m(B)\) and hence, by the Artin-Tate lemma \(W_m(B)\) is a finitely generated \(A\)-algebra.

Next we discuss iterated Witt vectors. One proves (cf. e.g. [9]) that \(F : W(A) \to W(A)\) lifts the \(p\)-power Frobenius on \(W(A)/pW(A)\). So for \(A\) \(p\)-torsion free, since \(W(A)\) is also \(p\)-torsion free, we have at our disposal a ring homomorphism \(\Delta = \lambda_F : W(A) \to W(W(A))\) which composed with any ghost map \(w_i : W(W(A)) \to W(A)\) equals the \(i\)-th iterate \(F^i\). Then one trivially checks that the composition
\[
W(A) \xrightarrow{\Delta} W(W(A)) \xrightarrow{w} W(A)^\infty \xrightarrow{w^\infty} (A^\infty)^\infty
\]
equals the composition
\[
W(A) \xrightarrow{w} A^\infty \xrightarrow{\Delta^w} (A^\infty)^\infty,
\]
where, if we write the elements of \((A^\infty)^\infty\) as
\[
((a_{00}, a_{01}, a_{02}, ...), (a_{10}, a_{11}, a_{12}, ...), (a_{20}, a_{21}, a_{22}, ...), ...)
\]
then
\[
\Delta^w(a_0, a_1, a_2, ...) = \begin{pmatrix}
a_0 & a_1 & a_2 & ... \\
a_1 & a_2 & a_3 & ... \\
a_2 & a_3 & a_4 & ... \\
... & ... & ... & ...
\end{pmatrix}.
\]
Using this plus the injectivity of the map \(w\) one immediately checks that if \(a_i = 0\) for \(i \leq m + m'\) then \(\Delta(a_0, a_1, a_2, ...)\) is in the kernel of \(W(W(A)) \to W_m(W_{m'}(A))\). So we have induced ring homomorphisms
\[
\Delta_{(m,m')} : W_{m+m'}(A) \to W_m(W_{m'}(A)).
\]
These homomorphisms (and \(\Delta\)) were constructed for \(A\) \(p\)-torsion free but, as usual, one extends this construction uniquely to all rings in a functorial manner.

Also one immediately checks (composing with ghost maps) that the following diagram is commutative:

\[
\begin{array}{ccc}
W(A) & \xrightarrow{\Delta} & W(W(A)) \\
F \downarrow & & \downarrow F \\
W(A) & \xrightarrow{\Delta} & W(W(A))
\end{array}
\]

**Lemma 2.4.** Let \(R = W(k)\) be the Witt ring on a perfect field of characteristic \(p\) and \(\phi : R \to R\) the Frobenius. Let \(0 \leq m, m' < \infty\). Then:

1) \(W_m(W_{m'}(R))\) is a finite \(R\)-algebra, where the structure morphism is given by \(R \xrightarrow{\lambda_{\phi}} W_m(R) \xrightarrow{W_m(\lambda_{\phi})} W_m(W_{m'}(R))\).

2) If \(W_{m+m'}(R)\) is viewed as an \(R\)-algebra via \(\lambda_{\phi} : R \to W_{m+m'}(R)\) then the morphism \(\Delta_{(m,m')} : W_{m+m'}(R) \to W_m(W_{m'}(R))\) is an \(R\)-algebra homomorphism.

**Proof.** The first assertion follows from Lemma 2.3. The second assertion follows from the “coassociativity” property in [9], p 15. \(\square\)

**Lemma 2.5.** For any \(a \in A, s \in \mathbb{Z}_+,\) and \(0 \leq i \leq m < \infty\) we have the following formula in \(W_m(A)\):
\[
V_i^s(p^s a) = \sum_{t=0}^{m-i} c_t(p^s) V_{m+i+t}^{i+t}(a^{p^s}).
\]
Proof.

\[ V_m^i (p^s a) = V_i([p^s a]_{m-i}) = V_i([p^s]_{m-i}[a]_{m-i}) = V_i(\sum_{t=0}^{m-i} c_i(p^s)V_m^t(1)V_{m-t}^0(a)) = \sum_{t=0}^{m-i} c_i(p^s)V_i(V_m^t(1)V_{m-t}^0(a))) = \sum_{t=0}^{m-i} c_i(p^s)V_i(V_m^t(a')) = \sum_{t=0}^{m-i} c_i(p^s)V_i^{i+t}([a^s]_{m-i-t}) = \sum_{t=0}^{m-i} c_i(p^s)V_i^{i+t}(a') \].

\[ \Box \]

Lemma 2.6. Let \( R = W(k) \), \( k \) a perfect field of characteristic \( p \geq 5 \), \( \phi : R \to R \) the lift of Frobenius on \( R \), \( \sigma : R \to A \) a \( p \)-torsion free finite \( R \)-algebra (e.g. \( A = W_{m'}(R) \) for some \( 0 \leq m' < \infty \), let \( 0 \leq m < \infty \), and \( W_m(A) \) be viewed as an \( R \)-algebra via \( R \to W_m(R) \to W_m(A) \). Moreover let \( a \in A \), \( a^2 = p^s a \) (e.g. \( a = V_m^i(1) \), in which case \( \nu = i' \)). Then for any \( s \geq 0 \) and \( 0 \leq i \leq m \) we have

\[ V_m^i (p^s a) \in p^s V_m^i (a) + p^{s+1} \sum_{t=0}^{m-i-1} R \cdot V_m^{i+t+1}(a). \]

(For \( i = m \) the sum in the right hand side is, by definition, zero.)

Proof. For \( 0 \leq i \leq m \) consider the \( R \)-modules

\[ M_m^i = \sum_{t=0}^{m-i} \sum_{r \geq 0} R \cdot V_i^{r+t}(p^s a) \subset W_m(A) \]

\[ N_m^i = \sum_{t=0}^{m-i} R \cdot V_i^{r+t}(a) \subset M_m^i. \]

Also set \( M_m^i = N_m^i = 0 \) for \( i > m \). Since \( W_m(A) \) is a finite \( R \)-algebra the modules \( M_m^i \) and \( N_m^i \) are finitely generated. By Lemma 2.5, for \( s \geq 1 \) we have

\[ V_m^i (p^s a) = \sum_{t=0}^{m-i} c_i(p^s)V_m^{i+t}(a^s) = p^s V_m^i (a) + \sum_{t=1}^{m-i} c_i(p^s)V_m^{i+t}(p^{s+1} a) = p^s V_m^i (a) + \sum_{t=1}^{m-i} c_i(p^s) c_t(p^{s+1} a) V_m^{i+t+r}(a^s) = p^s V_m^i (a) + \sum_{t=1}^{m-i} c_i(p^s) c_t(p^{s+1} a) V_m^{i+t+r}(p^{s+1} a) \in p^s V_m^i (a) + p^{s+1} M_m^{i+1}. \]
because for \( p \geq 5 \), \( s \geq 1 \), \( t \geq 1 \), \( \nu \geq 1 \), \( r \geq 0 \) we have \( p^{s-1}|c_t(p^s) \) and \( p^2|c_t(p^{s-1}) \). In particular
\[
M^i_m \subset N^i_m + pM^i_m
\]
\[
\subset N^i_m + p(N^i_m + pM^i_m) = N^i_m + p^2M^i_m
\]
\[
\subset N^i_m + p^2(N^i_m + pM^i_m) = N^i_m + p^3M^i_m, \quad \text{etc.}
\]
Hence
\[
N^i_m \subset M^i_m \subset \bigcap_{r=1}^{\infty} (N^i_m + p^rM^i_m) = N^i_m,
\]
because \( M^i_m \) is a finitely generated \( R \)-module and hence \( N^i_m \) is \( p \)-adically separated. So \( M^i_m = N^i_m \). So for all \( s \geq 0 \) we have
\[
V^i_m(p^s a) \in p^sV^i_m(a) + p^{s+1}N^i_m,
\]
which is what we had to prove. \( \square \)

**Lemma 2.7.** Let \( R = W(k) \), \( k \) a perfect field of characteristic \( p \). For \( 0 \leq i \leq m < \infty \) and \( 0 \leq i' \leq m' < \infty \) set
\[
(2.5) \quad v_{i,i'} = V^i_m(V^{i'}_{m'}(1)) \in W_m(W_{m'}(R)).
\]
Then the family \( \{ v_{i,i'} \} \) is \( R \)-linearly independent in \( W_m(W_{m'}(R)) \) where the latter is viewed as an \( R \)-algebra via the map
\[
R \overset{\lambda}{\longrightarrow} W_m(R) \overset{W_m(\lambda)}{\longrightarrow} W_m(W_{m'}(R))
\]
*Proof.* First it is trivial to check that the composition
\[
(2.6) \quad R \overset{\lambda}{\longrightarrow} W(R) \overset{W(\lambda)}{\longrightarrow} W(W(R)) \overset{w}{\longrightarrow} W(R) \overset{w}{\longrightarrow} (R^\infty)^\infty
\]
is given by
\[
\alpha \mapsto \begin{pmatrix}
a & \phi(a) & \phi^2(a) & \ldots \\
\phi(a) & \phi^2(a) & \phi^3(a) & \ldots \\
\phi^2(a) & \phi^3(a) & \phi^4(a) & \ldots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
\]
Next note that the images of \( w^\infty(w(v_{i,i'})) \) in \( (R^{m+1})^{m+1} \) are \( R \)-linearly independent (where \( (R^{m+1})^{m+1} \) is an \( R \)-algebra via the map (2.6)); indeed the matrix
\[
(2.7) \quad w^\infty(w(v_{i,i'})) = \begin{pmatrix}
0 & \ldots & 0 & 0 & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 0 & p^{i+i'} & \ldots \\
0 & \ldots & 0 & p^{i+i'p} & \ldots \\
0 & \ldots & 0 & p^{i+i'p^2} & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots
\end{pmatrix}
\]
with the first $i$ rows and $i'$ columns zero. The assertion of the Lemma now follows.

**Lemma 2.8.** Let $R = W(k)$, $k$ a perfect field of characteristic $p \geq 5$. For $0 \leq i \leq m < \infty$ and $0 \leq i' \leq m' < \infty$ let $v_{i,i'}$ be as in (2.5). Then for $0 \leq i \leq j \leq m$, $i', j' \in \{0, \ldots, m'\}$, and $1 \leq t \leq m - j$ there exist unique elements $c_{i,i',j,j'} \in R$ such that the following equalities hold in $W_m(W_{m'}(R))$:

$$v_{i,i'} \cdot v_{j,j'} = \begin{cases} p^{i + i'} v_{j,j'} + p^{i + i' + 1} \sum_{t=1}^{m-j} c_{i,i',j,j'} v_{j+t,j'}, & i' \leq j' \\
p^{i + i' - 1} v_{j,j'} + p^{i + i' - 1 + j} \sum_{t=1}^{m-j} c_{i,i',j,j'} v_{j+t,j'}, & i' \geq j' \end{cases}$$

**Proof.** Uniqueness of the $c$'s follows from Lemma 2.7. Let us prove the existence of the $c$'s. We have

$$v_{i,i'} \cdot v_{j,j'} = p^i V_m^j((V_m^i(1))^{p^{i'-1}} V_m^j(1))$$

The latter equals $p^i V_m^j(p^{i'-1} V_m^j(1))$ if $i' \leq j'$ and $p^i V_m^j(p^{i'-1} + V_m^j(1))$ if $i' \geq j'$. We conclude by Lemma 2.6.

**Lemma 2.9.** Let $R = W(k)$, $k$ a perfect field of characteristic $p \geq 5$, let $m, m' < \infty$, and view $W_m(W_{m'}(R))$ as an $R$-algebra via the homomorphism $W_m(\lambda_0) \circ \lambda_0 : R \to W_m(R) \to W_m(W_{m'}(R))$. Then $\{v_{i,i'}; 0 \leq i \leq m, 0 \leq i' \leq m'\}$ is an $R$-basis for $W_m(W_{m'}(R))$.

**Proof.** Linear independence was proved in Lemma 2.7. To prove generation we fix $m'$, set $A = W_m(R)$, and proceed by induction on $m$. The case $m = 0$ is Lemma 2.1. For the induction step we need to show that the kernel of the map $W_m(A) \to W_{m-1}(A)$ (which equals $V^m(A) = \{(0, \ldots, 0, a); a \in A\}$) is generated as an $R$-module by $v_{m,j'}$. By Lemma 2.1 the $R$-module $A$ is generated by the $v_{j'}$'s and note that $v_{m,j'} = V_m(v_{j'})$. So to conclude it is enough to show that the map $V^m : A_{\phi^m} \to V^m(A)$ is an isomorphism of $R$-modules where $A_{\phi^m}$ is $A$ viewed as an $R$-module via the map $R \xrightarrow{\phi^m} R \xrightarrow{\lambda_0} A$. The map $V^m : A_{\phi^m} \to V^m(A)$ is clearly a bijection. So we are reduced to check that $V^m : A_{\phi^m} \to W_m(A)$ is an $R$-module homomorphism. It is enough to check that the composition $w \circ V^m : A_{\phi^m} \to W_m(A) \to A^{m+1}$ (which by the way is given by $a \mapsto (0, \ldots, 0, p^m a)$), is an $R$-module homomorphism where $A^{m+1}$ is an $R$-algebra via the map $(\lambda_0)^{m+1} \circ (1, \phi, \ldots, \phi^m) : R \to R^{m+1} \to A^{m+1}$. This is however trivial to check.

**Lemma 2.10.** With notation as in Lemmas 2.1 and 2.9 the comultiplication $\Delta = \Delta_{(m,m')} : W_{m+m'}(R) \to W_m(W_{m'}(R))$ is given by

$$\Delta v_{i,i'} = \sum_{i,i'} a_{i,i',i'} v_{i,i'}, \quad 0 \leq i', \leq m + m',$$

where $a_{i,i',i'} \in \mathbb{Z}$. Moreover we have the following relations:

1) $a_{i,i',i''} = 0$ for $i + i' \leq i''$.
2) \( a_{i,i',i''} = 1 \) for \( i + i' = i'' \),
3) \( a_{i,0,i''} = 0 \) for \( i > i'' \),
4) \( a_{i,i',i''} = 0 \) for \( i' > i'' \),
5) \( a_{i,i',i''} \in p\mathbb{Z} \) for \( i + i' > i'' \),
6) For \( i + i' \geq i'' + 1 \), and \( i, i' \geq 1 \),

\[
(2.9) \quad a_{i,i',i''} = - \sum_{j=1}^{i-1} a_{j,i',i''} p^{j+i'-i-i'}.
\]

Note that the relations above allow one to recurrently determine all the coefficients \( a_{i,i',i''} \).

**Proof of Lemma 2.10.** Let \( K = \text{Frac}(R) \) and let \( M(i'') \) be the linear subspace of the space of all \((m+1)\times(m'+1)\)-matrices \((K^{m+1})^{m+1}\) consisting of all matrices \((r_{i,j})\) with \( r_{i,i'} = 0 \) for \( i + i' < i'' \). Since the elements \( w^\infty(w(v_{i,i''})) \in M(i'') \) for \( i + i' \geq i'' \) and since these elements are \( K \)-linearly independent it follows that these elements form a basis of \( M(i'') \). By (2.2) and (2.3) we have that

\[
w^\infty(w(\Delta(v_{i,i''}))) = \\
\begin{pmatrix}
0 & \ldots & 0 & p^{i''} & p^{i''} & \ldots \\
0 & \ldots & 0 & p^{i''} & p^{i''} & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots
\end{pmatrix}
\]

with \( i'' \) zeros on the first line. So \( w^\infty(w(\Delta(v_{i,i''}))) \) belongs to \( M(i'') \), hence we get an equality as in (2.8) with \( a_{i,i',i''} \in K \) and relation 1) holding. Since \( v_{i,i'} \) form a basis of \( W_m(W_m(R)) \) we get that \( a_{i,i',i''} \in R \). Picking out the \((i,i'')\)-entry in (2.8) and using (2.7) we get the relation

\[
(2.10) \quad p^{i''} = \sum_{j+j' \
\geq i'' \} j \leq i', j' \leq i'} a_{j,j',i''} p^{j+j'-i'}. 
\]

Relations 2) follows immediately. Relation 3) follows by induction. To prove relation 6) subtract the equality (2.10) with \( i' \) replaced by \( i' - 1 \) from the equality (2.10) and divide by \( p^{i'+1} \). Relation 4) for \( i = 0 \) follows by induction from (2.10). Relation 4) for arbitrary \( i \) follows by induction from 6). Relations 1), 2), and 6) imply relation 5) by induction. By 1), 2), and 5) we have \( a_{i,i',i''} \in \mathbb{Z} \) for all \( i, i', i'' \). \( \square \)

**Remark 2.11.** It is easy to see directly from the definitions that for any \( \mathbb{Z}_p \)-algebra \( C \) we have an isomorphism of \( \mathbb{Q}_p \)-algebras

\[
J^n(C) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \simeq (C \otimes_{\mathbb{Z}_p} \mathbb{Q}_p) \otimes (n+1)
\]

which for any \( c \in C \) sends \( c^s \) to \( 1 \otimes \ldots \otimes 1 \otimes c \). Hence we have \( \mathbb{Q}_p \)-algebra
isomorphisms

\[ J^n(W_m(\mathbb{Z}_p)) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \cong (W_m(\mathbb{Z}_p) \otimes \mathbb{Q}_p)^{(n+1)} \]

(2.11)

\[
\cong (\prod_{i=1}^{m+1} \mathbb{Q}_p)^{(n+1)} \\
\cong \prod_{i=1}^{(m+1)^{n+1}} \mathbb{Q}_p
\]

where \( \prod \) means \( N \)-fold product in the category of rings. If we set \( v_0 = 1 \) and \( v_{m+1} = 0 \) then the isomorphism \( W_m(\mathbb{Z}_p) \otimes \mathbb{Q}_p \cong \prod_{i=1}^{m+1} \mathbb{Q}_p \) is defined by the family of orthogonal idempotents

\[ \frac{v_j}{p^j} - \frac{v_{j+1}}{p^{j+1}} \in W_m(\mathbb{Z}_p) \otimes \mathbb{Q}_p, \quad 0 \leq j \leq m. \]

Hence the isomorphism (2.11) is defined by the family of orthogonal idempotents

\[ \prod_{j=0}^{n} \left( \frac{v_{j_0}}{p^{j_0}} - \frac{v_{j_0+1}}{p^{j_0+1}} \right) \in J^n(W_m(\mathbb{Z}_p)) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p, \quad j_0, \ldots, j_n \in \{0, \ldots, m\}. \]

(2.12)

In particular the whole complexity of \( J^n(W_m(\mathbb{Z}_p)) \) disappears after tensorization with \( \mathbb{Q}_p \) and hence it is an “integral” phenomenon. On the other hand, by the above, the \( \mathbb{Z}_p \)-algebra

\[ J^n(W_m(\mathbb{Z}_p))/\text{torsion} \]

is a free \( \mathbb{Z}_p \)-module of rank \((m+1)^{n+1}\) that retains most (but not all) of the complexity of \( J^n(W_m(\mathbb{Z}_p)) \). For instance if one considers the surjection

\[ J^n(W_m(\mathbb{Z}_p)) \rightarrow J^n(W_m(\mathbb{Z}_p))/\text{torsion} \]

(2.13)

then the target of this surjection is an \( \mathbb{F}_p \)-vector space of dimension \((m+1)^{n+1}\) whereas the source of this surjection is, by Theorem 1.1, an infinite dimensional \( \mathbb{F}_p \)-vector space; in fact this source, \( J^n(W_m(\mathbb{Z}_p)) \), is a product of two algebras: the identity component of \( J^n(W_m(\mathbb{Z}_p)) \) and the complement of the identity component. By Theorem 1.1, the identity component is a polynomial algebra in

\[ \delta^n v_{1}, \ldots, \delta^n v_{m} \in J^n(W_m(\mathbb{Z}_p)) \]

over an Artin local subring of \( J^n(W_m(\mathbb{Z}_p)) \) whose dimension as an \( \mathbb{F}_p \)-vector space is \( 2p^{n-1}m^n \). Indeed one can take as an \( \mathbb{F}_p \)-vector space basis for this Artin ring the elements

\[ v_{j_0}^e_0 (\delta v_{j_1})^e_1 \cdots (\delta^{n-1} v_{j_{n-1}})^e_{n-1} \in J^n(W_m(\mathbb{Z}_p)), \]

\[ e_0 \in \{0,1\}, \quad e_1, \ldots, e_{n-1} \in \{0, \ldots, p-1\}. \]

(2.14)

It is interesting to compare the two families (2.12) and (2.14).
Remark 2.12. We end by discussing the link between $p$-jets and Witt vectors. The discussion that follows will be helpful to set up notation for later and to put things into the right perspective. However, the adjunction properties that will be explained below have, by themselves, very little impact on the unraveling of the structure of $p$-jet spaces.

The following concept was introduced independently by Joyal [10] and the author [6]. A $p$-derivation from a ring $A$ into an $A$-algebra $B$ is a map of sets $\delta : A \to B$ such that the map $A \to W_1(B)$, $a \mapsto (a, \delta a)$ is a ring homomorphism. (Here we identify $a$ with $a \cdot 1_B$.) A $\delta$-ring is a ring $A$ equipped with a $p$-derivation $A \to A$. The category of $\delta$-rings is the category whose objects are the $\delta$-rings and whose morphisms are the ring homomorphisms that commute with $\delta$. By definition a $p$-derivation $\delta : A \to B$ satisfies

$$\begin{align*}
\delta(x + y) &= \delta x + \delta y + C_p(x, y) \\
\delta(xy) &= x^p\delta y + y^p\delta x + p\delta x\delta y,
\end{align*}$$

where $C_p$ is the polynomial:

$$C_p(X, Y) = p^{-1}(X^p + Y^p - (X + Y)^p) \in \mathbb{Z}[X, Y].$$

If $\delta$ is as above then $\phi : A \to B$, $\phi(x) = x^p + p\delta x$, is a ring homomorphism. Note that for any ring $A$ the ring $W(A)$ has a structure of $\delta$-ring which functorial in $A$; it is given by the composition $W(A) \xrightarrow{\Delta} W(W(A)) \to W_1(W(A))$. According to a result of Joyal [10] (which will not be needed in the sequel) for any ring $A$ and any $\delta$-ring $B$ we have

$$\text{Hom}_{\text{rings}}(B', A) \simeq \text{Hom}_{\delta-\text{rings}}(B, W(A)),$$

where $!$ is the forgetful functor from $\delta$-rings to rings. More generally if $R$ is a $\delta$-ring by a $\delta$-ring over $R$ we shall mean a $\delta$-ring equipped with a $\delta$-ring homomorphism from $R$ into it. Similarly a ring over $R$ will mean an $R$-algebra. Then the above adjunction property implies that for any $\delta$-ring $B$ over $R$ and any ring $A$ over $R$,

$$\text{Hom}_{\text{rings}/R}(B', A) \simeq \text{Hom}_{\delta-\text{rings}/R}(B, W(A)),$$

where $W(A)$ is an $R$-algebra via $R \to W(R) \to W(A)$. Let now $R = W(k)$ with $k$ a perfect field of characteristic $p$. Recall that for any $R$-algebra we defined in the Introduction $R$-algebras $J^n(C)$ and $J^\infty(C)$. The set theoretic maps $\delta : J^n(C) \to J^{n+1}(C)$ and $\delta : J^\infty(C) \to J^\infty(C)$ constructed in the Introduction
are then $p$-derivations and we have the following adjunction property: for any $\delta$-ring $D$ over $R$ and any ring $C$ over $R$ we have
\[
\text{Hom}_{\text{rings}/R}(C, D^0) \simeq \text{Hom}_{\delta-\text{rings}/R}(J^\infty(C), D).
\]
Putting together the two adjunction properties above we get
\[
\text{Hom}_{\text{rings}/R}(J^\infty(C), A) \simeq \text{Hom}_{\delta-\text{rings}/R}(J^\infty(C), W(A)) \simeq \text{Hom}_{\text{rings}/R}(C, W(A)^0)
\]
for any rings $A$ and $C$ over $R$.

It is sometimes useful to use a universality property that is more refined than that of $J^\infty$. To that purpose let us define a prolongation sequence to be a sequence
\[
B^0 \xrightarrow{\varphi_0} B^1 \xrightarrow{\varphi_1} B^2 \xrightarrow{\varphi_2} \ldots
\]
of ring homomorphisms which is equipped with $p$-derivations
\[
B^0 \xrightarrow{\delta} B^1 \xrightarrow{\delta} B^2 \xrightarrow{\delta} \ldots
\]
such that $\varphi_n \circ \delta = \delta \circ \varphi_{n-1}$ for all $n$. We denote by $B^*$ a prolongation sequence as above. A morphism of prolongation sequences $B^* = (B^n)$ and $C^* = (C^n)$ is by definition a sequence of morphisms $B^n \to C^n$ that commute, in the obvious sense, with the $\varphi$s and the $\delta$s. Clearly, for any ring $C$, $J^*(C) := (J^n(C))$ is naturally a prolongation sequence. Moreover, for any prolongation sequence $D^* = (D^n)$ and any ring $C$ we have
\[
\text{Hom}_{\text{rings}/R}(C, D^0) \simeq \text{Hom}_{\text{prol.seq.}}(J^*(C), D^*).
\]
Finally consider the prolongation sequence $R^* = (R^n)$ where all $R^n$ are $R = W(k)$, $k$ a perfect field of characteristic $p$, and all $\varphi$ are the identity. By a prolongation sequence over $R$ we understand a morphism of prolongation sequences $R^* \to B^*$; we have a natural notion of morphism of prolongation sequences over $R^*$. Clearly, for any ring $C$ over $R$, $J^*(C) := (J^n(C))$ is naturally a prolongation sequence over $R$. Moreover, for any prolongation sequence $D^* = (D^n)$ over $R$ and any ring $C$ over $R$ we have
\[
\text{Hom}_{\text{rings}/R}(C, D^0) \simeq \text{Hom}_{\text{prol.seq./R}}(J^*(C), D^*).
\]
Note that for any ring $A$ over $R$ the morphisms $\Delta : W_{m}(A) \to W_{1}(W_{m-1}(A))$ induce $p$-derivations $\delta : W_{m}(A) \to W_{m-1}(A)$ which, for each $N$, fit into a prolongation sequence
\[
W_{N}(A) \xrightarrow{\delta} W_{N-1}(A) \xrightarrow{\delta} \ldots \xrightarrow{\delta} W_{0}(A) = A \to 0 \to 0 \to \ldots
\]
This is a prolongation sequence over $R$ because of the $\phi$-linearity of $F : W_{m}(R) \to W_{m-1}(R)$ and hence of $F : W_{m}(A) \to W_{m-1}(A)$; cf. Lemma 2.2. So by the universality property for prolongation sequences we have a natural (compatible) family of $R$-homomorphisms:
\[
(2.16) \quad s : J^n(W_{N}(A)) \to W_{N-n}(A)
\]
for $0 \leq n \leq N$. Note that for any $R$-algebra $A$ the $p$-derivation $\delta : W_{m}(A) \to W_{m-1}(A)$ sends $v_i$ into
\[
(2.17) \quad \delta v_i = \rho(v_{i-1}) - p^{i(p-1)}\rho(v_i)
\]
for \( i = 1, \ldots, m \), where \( v_0 = 1 \). Indeed it is enough to show this for \( A = R \) in which case this follows from Lemmas 2.1 and 2.2. Finally note that by the commutativity of (2.4) if \( m' \) is fixed and \( m \) varies the morphisms \( \Delta : W_{m+m'}(R) \to W_m(W_{m'}(R)) \) fit into a morphism of prolongation sequences. This induces commutative diagrams

\[
\begin{array}{ccc}
J^n(W_{m+m'}(R)) & \xrightarrow{J^n(\Delta)} & J^n(W_m(W_{m'}(R))) \\
\downarrow s & & \downarrow s \\
W_{m+m'-n}(R) & \xrightarrow{\Delta} & W_{m-n}(W_{m'}(R))
\end{array}
\] (2.18)

Remark 2.13. If the upper row of the diagram 2.18, for \( m, m', n \) variable, is viewed as the “Lie groupoid of the integers” (i.e. an arithmetic analogue of the Lie groupoid of the line) then one is tempted to view the bottom row of the above diagram as an analogue of a “subgroupoid” of that “Lie groupoid”. However this candidate for a “subgroupoid” is contained in the “complement of the identity component” of the “Lie groupoid of integers”; cf. Remark 4.7.

3. \( p \)-jets and \( p \)-triangular bases

Let \( R \) be any ring, and let \( C \) be a commutative unital \( R \)-algebra, equipped with an \( R \)-algebra homomorphism \( C \to R \). Let \( C^+ \) be the kernel of this homomorphism and assume \( C^+ \) is a free \( R \)-module of finite rank. Let \( \{ v_\alpha; \alpha \in \Omega \} \) be an \( R \)-basis of \( C^+ \) where \( \Omega \) is a finite set equipped with a total order \( \leq \). Write

\[
v_\alpha \cdot v_\beta = \sum_{\gamma \in \Omega} c_{\alpha\beta\gamma} v_\gamma
\]

for \( \alpha \leq \beta \), where \( c_{\alpha\beta\gamma} \in R \). Let \( x \) be a collection of variables \( x_\alpha \) indexed by \( \alpha \in \Omega \) and

\[
Q_{\alpha\beta} = x_\alpha x_\beta - \sum_{\gamma \in \Omega} c_{\alpha\beta\gamma} x_\gamma \in R[x].
\]

**Lemma 3.1.** The \( R \)-algebra map

\[
R[x]/(Q_{\alpha\beta}; \alpha \leq \beta) \to C
\]
sending \( x_\alpha \mapsto v_\alpha \) is an isomorphism.

**Proof.** Indeed the source is generated as an \( R \)-module by 1 and the classes of \( x_\alpha \) so the algebra map above is injective (because 1 and the \( v_\alpha \)'s are linearly independent) and surjective (because 1 and the \( v_\alpha \)'s generate \( C \)). \( \square \)

**Definition 3.2.** Let \( C \) and \( C^+ \) be as above and let \( p \) be a prime. Let us say that \( v_\alpha \) is a \( p \)-triangular basis of \( C^+ \) if for all \( \alpha \leq \beta \) and all \( \gamma \) the structure constants \( c_{\alpha\beta\gamma} \) satisfy the following conditions:

1) \( c_{\alpha\beta\gamma} = 0 \) for \( \gamma < \alpha \);
2) \( c_{\alpha\beta\gamma} \equiv 0 \mod p^2 \) for \( \gamma \neq \beta \);
3) \( c_{\alpha\beta\gamma} \equiv \epsilon_{\alpha\beta} \mod p^2 \) where \( \epsilon_{\alpha\beta} \in \{0, 1\} \).
For the rest of this section, we assume that $C^+$ possesses a $p$-triangular basis $v_\alpha$, $\alpha \in \Omega$. We also assume $R = W(k)$ for $k$ a perfect field of characteristic $p \geq 3$.

Let $A = R[x] = R[x, x', x'', \ldots]$ and $A^n = R[x, x', \ldots, x^{(n)}]$. We start by recalling some filtrations from [8]. Let

$$A^{(n)} = A^n + pA^{n+1} + p^2A^{n+2} + \ldots \subset A.$$ 

Also let

$$I = (x, x', x'', \ldots) \subset A.$$

Consider the ideal $I^{[p]} \subset A$ generated by all elements of the form $pf$ and $f^p$ where $f \in I$; equivalently $I^{[p]} \subset A$ is the ideal generated by all elements of the form $px_{(j)}$ and $(x_{(j)})^p$ where $\alpha \in \Omega$, $j \geq 0$. It is trivial to check (cf. [8]) that

$$\delta(A^{(n)}) \subset A^{(n+1)}, \quad \phi(A^{(n)}) \subset A^{(n)}; \quad \delta(I^{[p]}) \subset I^{[p]}.$$ 

$$\delta(p^{j+1}A^{(n)}) \subset p^jA^{(n)}, \quad \delta(p^{j+1}I) \subset p^jI.$$ 

For any set $S$ let us denote by $[S]$ an arbitrary element of $S$. In particular for our algebra $C$ and the $V$-basis $v_\alpha$ of $C^+$,

$$Q_{\alpha \beta} = x_\alpha x_\beta - p\epsilon_{\alpha \beta} x_\beta + p^2[A^0 \cap I]$$

and $Q_{\alpha \beta}$ depends only on the variables $x_\alpha$ with $\gamma \geq \alpha$.

Finally let $Q^{(n)} \subset A^n$ be the ideal generated by

$$\{\delta^r Q_{\alpha \beta}; \alpha \leq \beta, 0 \leq r \leq n\}.$$ 

Note that if $F, G \in A^n$ and $F \equiv G \mod Q^{(n-1)}A^n$ then $\delta F \equiv \delta G \mod Q^{(n)}A^{n+1}$. Here is our main computation in characteristic zero.

**Theorem 3.3.** Assume $C^+$ has a $p$-triangular basis and let $Q_{\alpha \beta}$ and $Q^{(n)} \subset A^n$ be as above. Then for $n \geq 1$ and $\alpha \leq \beta$ we have the congruences

$$\delta^n Q_{\alpha \beta} \equiv F_{\alpha \beta n} \mod Q^{(n-1)}A^n$$

in the ring $A^n$ where

$$F_{\alpha \beta n} = \begin{cases} px_{(n)}\phi(x_{(n)}) - p\epsilon_{\alpha \beta} x_{(n)} + p[A^0 \cap I], & n = 1 \\ (x_{(n)}^p)\phi(x_{(n)}) + (x_{(n)}')^p\phi(x_{(n)}) - (x_{(n)}x_{(n)})^p - \epsilon_{\alpha \beta} \phi(x_{(n)}) + [A^0 \cap I^{[p]}], & n = 2 \\ (x_{(n)}^p - \phi(x_{(n-1)})) + (x_{(n)}')^p - \phi(x_{(n-1)}) - \epsilon_{\alpha \beta} \phi(x_{(n-1)}) + [A^{(n-2)} \cap I^{[p]}], & n \geq 3, \end{cases}$$

and $F_{\alpha \beta n}$ only depends on the variables indexed by $\gamma \geq \alpha$. 

**Documenta Mathematica** 18 (2013) 971–996
Proof. Note that \(x^s_\alpha \equiv p^2[I] \mod Q(0)\) for \(s \geq 3\). We get the following congruences mod \(Q(0)A^1\):

\[
\delta Q_{\alpha \beta} = \delta(x_\alpha x_\beta) - \delta(\epsilon_{\alpha \beta}px_\beta) + \delta(p^2[A(0) \cap I]) + p[A(0) \cap I]
\]

\[
= x^p_\alpha x^p_\beta + px_\alpha x_\beta - \epsilon_{\alpha \beta}x^p_\beta \delta p - \epsilon_{\alpha \beta}px_\beta + p[A(0) \cap I]
\]

\[
\equiv p^2[A(1) \cap I] + px_\alpha x_\beta - \epsilon_{\alpha \beta}px_\beta + p[A(0) \cap I]
\]

\[
= px_\alpha x_\beta - \epsilon_{\alpha \beta}px_\beta + p[A(0) \cap I].
\]

Using the fact that \(\delta p \equiv 1 \mod p\) we get the following congruences mod \(Q(1)A^2\):

\[
\delta^2 Q_{\alpha \beta} = \delta(px^p_\alpha x^p_\beta) - \delta(p\epsilon_{\alpha \beta}px^p_\beta) + \delta(p[A(0) \cap I])
\]

\[
+ C_{p\epsilon}(px^p_\alpha x^p_\beta, -\epsilon_{\alpha \beta}x^p_\beta, \epsilon_{\alpha \beta}x_\beta) \cap [A(0) \cap I[p]]
\]

\[
= (x^p_\alpha x^p_\beta)\delta(p) + p\delta(px^p_\alpha x^p_\beta) - \delta(p\epsilon_{\alpha \beta})(x^p_\beta)p - \epsilon_{\alpha \beta}px^p_\beta + [A(0) \cap I[p]]
\]

\[
= (x^p_\alpha x^p_\beta) + p((x^p_\alpha)p)x^p_\beta + (x^p_\beta)p\epsilon_{\alpha \beta} + p[A(0) \cap I]
\]

\[
- \epsilon_{\alpha \beta}\phi(x^p_\beta) + [A(0) \cap I[p]]
\]

\[
= (x^p_\alpha)p\phi(x^p_\beta) + (x^p_\beta)p\phi(x^p_\alpha) - (x^p_\alpha x^p_\beta)p - \epsilon_{\alpha \beta}\phi(x^p_\beta) + [A(0) \cap I[p]].
\]

Using the fact that the 5 terms above are in \(A(1)^1 \cap I[p]\), the fact that \(\phi\delta = \delta\phi\), and the fact that \(\delta((x^p_\alpha)p), \delta((x^p_\beta)p) \in pA^2 \cap I[p] \subset A(1)^1 \cap I[p]\) we get the following congruence mod \(Q(2)A^3\):

\[
\delta^3 Q_{\alpha \beta} = \delta((x^p_\alpha)p\phi(x^p_\beta)) + \delta((x^p_\beta)p\phi(x^p_\alpha)) - \delta((x^p_\alpha x^p_\beta)p)
\]

\[
- \delta(\epsilon_{\alpha \beta}\phi(x^p_\beta)) + \delta([A(0) \cap I[p]] + C_{p\epsilon}(\epsilon_{\alpha \beta}x^p_\beta, \epsilon_{\alpha \beta}x_\beta)) \cap [A(0) \cap I[p]]
\]

\[
= \delta((x^p_\alpha)p\phi(x^p_\beta)) + \delta((x^p_\beta)p\phi(x^p_\alpha)) - \delta(\epsilon_{\alpha \beta}\phi(x^p_\beta)) + [A(1) \cap I[p]]
\]

\[
= (x^p_\alpha)p^2\delta(\phi(x^p_\beta)) + \phi^2(x^p_\beta)p\delta((x^p_\alpha)p) + (x^p_\beta)p^2\delta(\phi(x^p_\alpha))
\]

\[
+ \phi^2(x^p_\beta)p\delta((x^p_\alpha)p) - \epsilon_{\alpha \beta}\delta(\phi(x^p_\beta)) + [A(1) \cap I[p]]
\]

\[
= (x^p_\alpha)p^2\phi(x^p_\beta) + (x^p_\beta)p^2\phi(x^p_\alpha) - \epsilon_{\alpha \beta}\phi(x^p_\beta) + [A(1) \cap I[p]].
\]

Finally using the same kind of computation as for \(\delta Q_{\alpha \beta}\) one proves by induction on \(n\) that for \(n \geq 3\) we have the following congruence mod \(Q^{(n-1)}A^n\):

\[
\delta^n Q_{\alpha \beta} \equiv (x^p_\alpha)_{p^{n-1}}\phi(x^{(n-1)}_\beta) + (x^p_\beta)_{p^{n-1}}\phi(x^{(n-1)}_\alpha) - \epsilon_{\alpha \beta}\phi(x^{(n-1)}_\beta) + [A^{(n-2)} \cap I[p]],
\]

The fact that \(F_{\alpha \beta n}\) only depends on the variables indexed by \(\gamma \geq \alpha\) follows simply from the fact that for any pair \(\alpha \leq \beta\) we can make the computations above in the rings with variables indexed by indices \(\gamma \geq \alpha\). \(\square\)
Let $\overline{B} = B/pB$ for any ring $B$, also for $b \in B$ let $\overline{b} \in \overline{B}$ be the class of $b$. Set $k = R/pR$. Then the ideal $\overline{I}$ in $\overline{A} = k[x, x', x''...]$ is generated by the set $\{(x^{(r)})_\alpha^p; r \geq 0, \alpha \in \Omega\}$ and it is trivial to see that we have an equality of ideals $$\overline{A^{n-2}} \cap \overline{I}^{\overline{n}} = \overline{A^{n-2}} \cap \overline{I}^{\overline{n}} = \{(x^{(r)})^p_\alpha; 0 \leq r \leq n-2, \alpha \in \Omega\}$$ in the ring $\overline{A}^{n-2} = k[x, x', ..., x^{(n-2)}]$.

Set $F_{\alpha \beta \delta} = Q_{\alpha \beta}$. We get the following:

**Corollary 3.4.** The ideal $Q^{(\alpha)} \subset A^n$ is generated by the set $\{F_{\alpha \beta \delta}; \alpha \leq \beta, \ 0 \leq r \leq n\}$.

In particular:

$J^n(C) = A^n/(F_{\alpha \beta \delta}; \alpha \leq \beta, \ 0 \leq r \leq n)$,\\ $\overline{J^n(C)} = \overline{A^n}/(\overline{F_{\alpha \beta \delta}}; \alpha \leq \beta, \ 0 \leq r \leq n)$.

Since $\overline{Q^{(\alpha)}} \subset (x_\alpha x_\beta, (x^{(r)}_\alpha)^p; \alpha, \beta \in \Omega, 1 \leq r \leq n-1) \subset \overline{A^n}$ we get

**Corollary 3.5.** There is a natural surjection

$$\overline{J^n(C)} \rightarrow \overline{A^n} / (x_\alpha x_\beta, (x^{(r)}_\alpha)^p; \alpha, \beta \in \Omega, 1 \leq r \leq n-1).$$

In some important cases the above surjection is close to an isomorphism as we shall see next.

**Definition 3.7.** Assume $C^\ast$ has a $p$-triangular basis $v_\alpha$, $\alpha \in \Omega$. Let $\{\gamma \in \Omega; \epsilon_{\gamma \gamma} = 1\} = \{\gamma_1, ..., \gamma_N\}, \ \gamma_1 < ... < \gamma_N$.

We say that $v_\alpha$ is a non-degenerate $p$-triangular basis if $\min \Omega = \gamma_1$ and $\epsilon_{\gamma, \gamma} = 1$ for all $i = 1, ..., N$ and all $\gamma_i \leq \gamma < \gamma_{i+1}$. (Here and later we discard the condition $\gamma < \gamma_{i+1}$ if $i = N$.)
Remark 3.8. Let \( m \) be a non-negative integer. Let \( \Omega = \Omega_m = \{1, ..., m\} \) be equipped with the usual total order and consider the \( R \)-algebra \( C = W_m(R) \), the ghost homomorphism \( \omega_0 : W_m(R) \to R \), and its kernel \( W_m(R)^+ \). Then by Lemma 2.1 \( v_1, ..., v_m \) is a non-degenerate \( p \)-triangular basis of \( W_m(R)^+ \); indeed \( \{i \in \Omega_m; \epsilon_i = 1\} = \{1\} \) and \( \epsilon_{ij} = 1 \) for \( j = 1, ..., m \).

**Remark 3.9.** Let \( m, m' \) be non-negative integers. Consider the set
\[
\Omega := \Omega_{m,m'} = \{(i, i') \in \mathbb{Z} \times \mathbb{Z}; 0 \leq i \leq m, 0 \leq i' \leq m'\}\{(0, 0)\},
\]
ordered by the lexicographic order: \((i, i) \leq (j, j')\) if either \( i < j \) or \( i = j, i' \leq j' \). Consider the \( R \)-algebra \( W_m(W_{m'}(R)) \) (\( p \geq 5 \)) and the composition of ghost maps \( \omega_0 \circ \omega_0 : W_m(W_{m'}(R)) \to R \) with kernel \( W_m(W_{m'}(R))^+ \). Consider the \( R \)-basis \( v_{(i,i')} = v_{i,i'} \) of \( W_m(W_{m'}(R))^+ \) where \((i, i') \in \Omega_{m,m'}\); cf Lemma 2.9. Then \( v_{i,i'} \) is a non-degenerate \( p \)-triangular basis by Lemma 2.8; note that for \( m, m' \geq 1 \)
\[
\{(i, i') \in \Omega_{m,m'}; \epsilon_{(i,i')(i,i')} = 1\} = \{(0, 1), (1, 0)\}.
\]

Going back to our general \( C \), viewed with augmentation \( C \to C/C^+_p = R \), the following is a computation of the identity component of \( J^n(C) \) in the presence of non-degenerate \( p \)-triangular bases.

**Theorem 3.10.** Assume \( C^+ \) has a non-degenerate \( p \)-triangular basis \( v_{\alpha} \) and set
\[
\pi = \prod_{i=1}^{N}(\delta v_{\gamma_i} - 1) \in J^1(C).
\]
Then the image of \( \pi^n \) in \( J^2(C) \) is idempotent and for all \( n \geq 2 \) we have a natural isomorphism
\[
J^n(C)_\pi \cong \frac{\mathbb{A}^n}{(x_\alpha x_\beta, (x_\alpha^{(r)})^p; \alpha, \beta \in \Omega, 1 \leq r \leq n - 1)}
\]
sending the class of \( \delta^r v_{\alpha} \) into the class of \( x_\alpha^{(r)} \) for all \( r \geq 0 \) and all \( \alpha \).

By the theorem it follows that \( J^n(C)_\pi \), respectively \( J^n(C)_{1-\pi} \), are isomorphic to the identity component, respectively to the complement of the identity component, of \( J^n(C) \).

**Proof of Theorem 3.10.** Since the map (3.3) sends \( \pi \) into the class of the polynomial \( \Phi = \prod_{i=1}^{N}(x_{\gamma_i} - 1) \) (which is an invertible element) we get a surjective map from the left hand side of (3.4) to the right hand side of (3.4). In order to prove that the latter map is an isomorphism it is enough to show that the inclusion of ideals
\[
(\mathbb{Q}^{(n)})_\Phi \subset (x_\alpha x_\beta, (x_\alpha^{(r)})^p; \alpha, \beta \in \Omega, 1 \leq r \leq n - 1)_\Phi
\]
in the ring \( (\mathbb{A}^n)_\Phi \) is an equality. Clearly \( x_\alpha x_\beta \in (\mathbb{Q}^{(n)})_\Phi \). Next we show that for all \( i = 1, ..., N \), all \( \gamma_i \leq \gamma < \gamma_{i+1} \), and all \( 1 \leq r \leq n - 1 \) we have
\[(x^{(r)})^p \in \left(\mathbb{Q}(n)\right)_\Phi.\] We proceed by induction on \(N - i \geq 0.\) The proof of the case \(N - i = 0\) is similar to the proof of the induction step so we skip it. For the induction step assume the assertion is true for \(i + 1, i + 2, \ldots, N,\) for some index \(1 \leq i < N,\) and let us prove it for \(i.\) We proceed by induction on \(r \geq 1.\) To check the base case \(r = 1\) note that for all \(\gamma \in \Omega,\) we have
\[
F_{\gamma \gamma 2} = (x_\gamma^r - 1)^p(x_\gamma^r)^p + G_{\gamma \gamma 2} \in \left(\mathbb{Q}(n)\right)_\Phi.
\]
Since \(G_{\gamma \gamma 2} \in \left(\mathbb{Q}(n)\right)_\Phi\) we get \((x_\gamma^r)^p \in \left(\mathbb{Q}(n)\right)_\Phi.\) Now let \(3 \leq s \leq n\) and assume that for all \(1 \leq r \leq s - 2\) and all \(\gamma \in \Omega,\) we have \((x_\gamma^{(s)})^p \in \left(\mathbb{Q}(n)\right)_\Phi;\) we want to show that \((x_\gamma^{(s-1)})^p \in \left(\mathbb{Q}(n)\right)_\Phi.\) Note that
\[
F_{\gamma \gamma s} = (x_\gamma^r)^{p^{r-1}}(x_\gamma^{(s-1)})^p + (x_\gamma^r)^{p^{r-1}}(x_\gamma^{(s-1)})^p - (x_\gamma^{(s-1)})^p + G_{\gamma \gamma s} \in \left(\mathbb{Q}(n)\right)_\Phi.
\]
Recall that \(G_{\gamma \gamma s}\) is in the ideal generated by \((x_\mu^{(r)})^p\) with \(\mu \geq \gamma_i\) and \(0 \leq r \leq s - 2.\) By the induction hypotheses (and the fact that \(x_\alpha^p \in \left(\mathbb{Q}(n)\right)_\Phi\)) we have that
\[
(x_\gamma^r)^{p^{r-1}}, (x_\gamma^{(s-1)})^p, G_{\gamma \gamma s} \in \left(\mathbb{Q}(n)\right)_\Phi.
\]
It follows that \((x_\gamma^{(s-1)})^p \in \left(\mathbb{Q}(n)\right)_\Phi\) which ends the induction on \(r\) and hence the induction on \(i\) as well. To conclude the proof of the Theorem note that
\[
\gamma_{\gamma \gamma 2} - G_{\gamma \gamma 2} = (x_\gamma^r)^{2p} - (x_\gamma^{(s)})^p \in \mathbb{Q}(2)
\]
so the image of \((\delta v_\gamma)^p\) in \(J^2(C)\) is idempotent hence so is the image of \(\pi^p.\)

Finally we need the following

**Lemma 3.11.** For \(U_1, U_2 \in I \cap A^0\) and \(n \geq 0\) we have
\[
\delta^n(U_1 + pu_2) = \delta^n U_1 + [I[p] \cap A^{(n)}].
\]

**Proof.** We proceed by induction on \(n.\) The case \(n = 0\) is tautological. For the induction step,
\[
\delta^{n+1}(U_1 + pu_2) = \delta^{n+1} U_1 + \delta([I[p] \cap A^{(n)}])
\]
\[
+ C_p(\delta^n U_1, [I[p] \cap A^{(n)}])
\]
\[
= \delta^{n+1} U_1 + [I[p] \cap A^{(n+1)}]
\]
which ends the proof.

**Corollary 3.12.** Assume the notation in Theorem 3.10 and let \(u_1, u_2 \in C^+.\) Then for any \(i = 0, \ldots, n - 1\) we have
\[
\delta^i(u_1 + pu_2) = \delta^i(u_1)
\]
in the ring $J^n(C)$. In particular for all $u_1,\ldots,u_s,u \in C^+$ we have
\[ \delta(\sum_{t=1}^{s} u_t + pu) = \sum_{t=1}^{s} \delta(u_t). \]

Proof. Let $u_1, u_2$ be the classes of polynomials $U_1, U_2 \in I \cap A^0$. Then the first assertion follows from Lemma 3.11 because, for $i \leq n-1$, $I^{[n]} \cap A^{[r]}$ is contained in the denominator of the fraction in Equation 3.4. The second assertion follows from the fact that the difference between $\delta(\sum_{t=1}^{s} u_t)$ and $\sum_{t=1}^{s'} \delta(u_i)$ is in the square of the ideal $C^+$ and hence its reduction mod $p$ is zero. □

4. Applications to $p$-jets of Witt rings

In this section we continue to write $R = W(k)$ for $k$ a perfect field of characteristic $p \geq 5$. In view of Remark 3.8 the results in the previous section apply to the algebra $C = W_m(R)$ equipped with the homomorphism $u_0 : W_m(R) \to R$ and with the non-degenerate $p$-triangular basis $v_1,\ldots,v_m$ indexed by $\Omega_m = \{1,\ldots,n\}$. For $m = 1$ we have:

**Corollary 4.1.**
\[
J^n(W_1(R)) = \begin{cases} 
\frac{k[x,x']}{(x^1)}, & n = 1 \\
\frac{k[x,x',x^{(r)}]}{(x^2, (x^1)^2 - x^1)^p)}, & n = 2 \\
\frac{k[x,x',\ldots,x^{(n)}]}{(x^2, (x^1)^2 - x^1)^p, (2x^1 - 1)^{p-1}(x^r - 1)^p + \Omega_{1r}, 2 \leq r \leq n)}, & n \geq 3,
\end{cases}
\]

where $\Omega_{1r} \in ((x^s)^p; 0 \leq s \leq r - 2) \subset k[x,\ldots,x^{(r-2)}]$. In particular, since $2x^1 - 1$ is invertible in the ring $k[x]/((x^2)^2 - x^1)^p)$, the ring $J^n(W_1(R))$ is a polynomial ring in $m$ variables $x_1^{(n)},\ldots,x_m^{(n)}$ over an Artin ring. This Artin ring is a free module of rank $p^{n-1}$ over the ring $W_1(R) = k[x]/(x^2)$. Moreover the ring $J^{n}(W_1(R))$ is a flat integral extension of $W_1(R)$.

Set $\pi = 1 - \delta v_1$. For $m \geq 1$ and $n \geq 2$ we have a splitting
\[
J^n(W_m(R)) = \left( J^n(W_m(R)) \right)_\pi \times \left( J^n(W_m(R)) \right)_{1-\pi}.
\]

For the identity component we have the following direct consequence of Theorem 3.10:

**Corollary 4.2.** For $n \geq 2$
\[
\left( J^n(W_m(R)) \right)_\pi \simeq \frac{k[x^{(r)}; i \in \Omega_m, 0 \leq r \leq n]}{(x_i, x_j^{(r)}); i,j \in \Omega_m, 1 \leq r \leq n - 1},
\]

In particular the above ring is a polynomial ring in $m$ variables $x_1^{(n)},\ldots,x_m^{(n)}$ over a local Artin ring. This local Artin ring is a free module of rank $p^{m(n-1)}$. 

Documenta Mathematica 18 (2013) 971–996
over $W_m(R)$. Moreover the ring
\[
\left( J^\infty(W_m(R)) \right)_{\pi} = \frac{k[x^{(r)}_i; i \in \Omega_m, r \geq 0]}{(x_ix_j, (x^{(r)}_i)p; i, j \in \Omega_m, r \geq 1)}
\]
is local and is a flat integral extension of $W_m(R)$.

On the other hand the canonical lifts $J^n(W_m(R)) \to R$ of $w_1, ..., w_m$ send $\delta w_1 \mapsto \delta p \in R^\times$ so they factor through $J^n(W_m(R))_{1-\pi} \to R$. In particular the complement of the identity component, $\left( J^n(W_m(R)) \right)_{1-\pi}$, is non-zero.

Moreover for $m \geq 2$ this ring is, in some sense, more “degenerate” than the identity component. Indeed this ring is trivially seen to be the quotient of $(\mathbb{A}^2)_x$ by the ideal generated by the following elements:
1) $x_ix_j, \ 1 \leq i \leq j \leq m$,
2) $(x^p_i)_i, \ 2 \leq i \leq m$,
3) $(x^p_i)(x^p_j)_i, \ 2 \leq i \leq j \leq m$,
4) $(x^{(r-1)}_i)_i + \overline{G}_{ir}, \ 3 \leq r \leq n$,
5) $G_{ijr}, \ 2 \leq j \leq m, \ 3 \leq r \leq n$,
6) $G_{ijr}, \ 2 \leq i \leq j \leq m, \ 3 \leq r \leq n$.

Since the variables $x^{(n-1)}_2, ..., x^{(n-1)}_m, x^{(n)}_1, ..., x^{(n)}_m$ do not appear in any of the above generators, and since, as we saw, the ring $\left( J^\infty(W_m(R)) \right)_{1-\pi}$ is non-zero we get:

**Corollary 4.3.** For $n \geq 3$ and $m \geq 2$ the ring $\left( J^\infty(W_m(R)) \right)_{1-\pi}$ is isomorphic to a polynomial ring in $2m-1$ variables $x^{(n-1)}_2, ..., x^{(n-1)}_m, x^{(n)}_1, ..., x^{(n)}_m$ over some non-zero ring.

Assume from now on $p \geq 5$. In a similar way, in view of Remark 3.9, the results in the previous section apply to the $R$-algebra $C = W_m(W_m(R))$, $1 \leq m, m' < \infty$, equipped with the homomorphism $\omega_0 \circ \omega_0 : W_m(W_m'(R)) \to R$ and with the non-degenerate $p$-triangular basis $v_{i,i'}$ indexed by $\Omega = \Omega_{m,m'}$ in loc. cit. In particular if $\Pi = (1-\delta_{m,0})(1-\delta_{1,n})$ we have a splitting
\[
J^n(W_m(W_m'(R))) = \left( J^n(W_m(W_m'(R))) \right)_{1-\Pi} \times \left( J^n(W_m(W_m'(R))) \right)_{1-\Pi}.
\]

For the identity component we have the following direct consequence of Theorem 3.10:

**Corollary 4.4.** For $n \geq 2$ we have
\[
\left( J^n(W_m(W_m'(R))) \right)_{\Pi} \cong \frac{k[x^{(r)}_{ii'}; (i, i') \in \Omega, 0 \leq r \leq n]}{(x_{ii'}x_{jj'}, (x^{(r)}_{ii'})p; (i, i'), (j, j') \in \Omega, 1 \leq r \leq n-1)}.
\]

Next consider the comultiplication $R$-algebra map
\[
\Delta : W_{m+m'}(R) \to W_m(W_m'(R))
\]
and the induced $R$-algebra maps

$$J^n(\Delta) : J^n(W_{m+m'}(R)) \to J^n(W_m(W_{m'}(R))).$$

**Lemma 4.5.** For $n \geq 2$ the above $R$-algebra map induces an $R$-algebra map

$$J^n(\Delta) : J^n(W_{m+m'}(R)) \to J^n(W_m(W_{m'}(R))).$$

**Proof.** It is enough to show that $J^n(\Delta)$ in (4.1) sends $\pi = 1 - \delta v_1$ into an invertible element of $J^n(W_m(W_{m'}(R)))$. But by Lemmas 2.10 and 3.12 we have the following congruences mod $p$:

$$J^n(\Delta)(1 - \delta v_1) = 1 - \delta(\Delta v_1) \equiv 1 - \delta v_1,$$

and we conclude by the fact that the image of $(\delta v_1)(\delta v_0)$ in $J^n(W_m(W_{m'}(R)))$ is nilpotent; cf. Corollary 4.4.

□

By the identifications in Lemmas 4.2 and 4.4 we get that the $R$-algebra map in Lemma 4.5 induces an $R$-algebra map

$$J^\infty(\Delta) : k[x_i(r); i \in \Omega_{m+m'}, r \geq 0] \to k[x_{i,i'}(r); (i,i') \in \Omega_{m,m'}, r \geq 0].$$

**Corollary 4.6.** For $1 \leq i'' \leq m + m'$ we have that $J^\infty(\Delta)$ sends the class of $x_{i'}^{(r)}$ into the class of

$$\sum_{i + i' = i''} x_{i,i'}, \quad \text{if} \quad r = 0$$

and into the class of

$$\delta^{r-1} \left( \sum_{i + i' = i''} x_{i,i'} \right), \quad \text{if} \quad r \geq 1.$$

**Proof.** By Corollary 3.12 and Lemma 2.10 $J^\infty(\Delta)$ sends the class of $x_{i'}^{(r)}$ into the class of

$$\delta^r \left( \sum_{i + i' = i''} x_{i,i'} \right)$$

if $r \geq 0$. It is then enough to prove that

$$\delta^r \left( \sum_{i + i' = i''} x_{i,i'} \right) = \delta^{r-1} \left( \sum_{i + i' = i''} x_{i,i'} \right) + (I^2 \cap A^0)A + I^p$$

for $r \geq 1$ in $A = R[x,x',x'',...],$ where $I^2$ denotes as usual the square of the ideal $I$. The later follows by induction using the fact that

$$\delta((I^2 \cap A^0)A) \subset (I^2 \cap A^0)A + I^p.$$
Remark 4.7. It would be interesting to have an explicit understanding of the homomorphisms $s : J^n(W_m(R)) \to W_{m-n}(R)$ in (2.16), or at least of their reduction mod $p$. This involves understanding the iterates of formula 4.2. Note however that by formula 4.2 it follows that
\[ s(\pi) = 1 - (1 - p^{p-2}p(v_1)) \in pW_{m-1}(R). \]

In other words for $n \geq 2$
\[ \pi : J^n(W_m(R)) \to W_{m-n}(R) \]
factors through the complement of the identity component!
\[ J^n(W_m(R))_{1-\pi} \to W_{m-n}(R) \]
rather than through the identity component $J^n(W_m(R))_\pi$; this makes the problem more subtle.

References

ON ARTIN REPRESENTATIONS
AND NEARLY ORDINARY HECKE ALGEBRAS
OVER TOTALLY REAL FIELDS

SHU SASAKI

Received: May 29, 2012
Revised: May 19, 2013

Communicated by Don Blasius

Abstract. We prove many new cases of the strong Artin conjecture for two-dimensional, totally odd, insoluble (icosahedral) representations $\text{Gal}(\overline{F}/F) \to GL_2(\mathbb{C})$ of the absolute Galois group of a totally real field $F$.

2010 Mathematics Subject Classification: 11G80, 11F33, 11F41, 14G22, 14G35
Keywords and Phrases: Galois representations, $p$-adic modular forms, Hilbert modular varieties

1 Introduction

Let $K$ be a number field. Artin conjectured that the $L$-series of any continuous representation $\rho : \text{Gal}(\overline{K}/K) \to GL_n(\mathbb{C})$ of the absolute Galois group $\text{Gal}(\overline{K}/K)$ of $K$ is holomorphic except a possible pole at $s = 1$ when the trivial representation is a constituent of $\rho$.

A result of Brauer (See [36]) about finite groups immediately implies that $L(\rho, s)$ has meromorphic continuation and satisfies a certain functional equation relating the values at $s$ and $1 - s$. Any such complex representation is semi-simple, and because Artin showed that $L(\rho_1 + \rho_2) = L(\rho_1, s)L(\rho_2, s)$, the conjecture immediately follows from the case where $\rho$ is irreducible. In the case where $\rho$ is irreducible, the strong form of this conjecture, known as the strong Artin conjecture, asserts that there is a cuspidal automorphic representation $\pi$ of $GL_n(\mathbb{A}_K)$ such that $L(\pi, s) = L(\rho, s)$, and Artin conjecture follows from the strong Artin conjecture (See [22], Theorem 8.8 with its proof (p.286) attributed to Ramakrishnan).
When \( n = 2 \) and the image of the projective representation \( \text{proj} \rho : \text{Gal}(\overline{K}/K) \to PGL_2(\mathbb{C}) = GL_2(\mathbb{C})/C^* \) is dihedral (\( D_{2n} \) for some \( n \geq 2 \)), \( \rho \) is induced from a character \( \chi \) of the absolute Galois group \( \text{Gal}(K/M) \) of a quadratic extension \( M \) of \( \mathbb{Q} \), and Artin himself proved the conjecture (the holomorphy of \( L(\rho, s) = L(\text{Ind}^{\text{Gal}(\overline{K}/E)} \chi, s) = L(\chi, s) \) follows from earlier work of Hecke).

When \( n = 2 \) and the image of \( \text{proj} \rho \) is tetrahedral (\( A_4 \)) and when \( n = 2 \), \( K = \mathbb{Q} \), \( \rho \) odd, and the projective image of \( \rho \) is octahedral (\( S_4 \)), Langlands [23], using his theory of (cyclic) base change, “deduced” the strong Artin conjecture from the dihedral case. Tunnell, building on work of Langlands, completed the octahedral case \( n = 2 \) and general \( K \). In the octahedral case, in order to “descend” a cuspidal automorphic representation \( \Pi \) of \( GL_2(\mathbb{A}_E) \) such that \( L(\Pi, s) = L(\rho|_{\text{Gal}(\mathbb{Q}/F)}), s) \) to a cuspidal automorphic representation \( \pi \) of \( GL_2(\mathbb{A}_K) \), where \( E \) is the quadratic extension of \( K \) corresponding to the unique index 2 subgroup (\( \cong A_4 \)) of \( S_4 \), Langlands uses a theorem of Deligne-Serre (and therefore \( K = \mathbb{Q} \) and \( \rho \) should be necessarily odd) whilst Tunnell uses cubic base change to match up, for all but finitely many places \( \nu \) of \( K \), the restriction of \( \rho \) to the decomposition group at \( \nu \) and the local representation \( \pi_\nu \).

The icosahedral (\( A_5 \)) case had remained largely intractable until Buzzard-Dickinson-Shepherd-Barron-Taylor [4] proved many new cases of the strong Artin conjecture for odd \( \rho \): \( \text{Gal}(\mathbb{Q}/\mathbb{Q}) \to GL_2(\mathbb{C}) \).

[4] follows the program of Taylor ([37]), which may be succinctly described as an approach to deduce results about weight one forms from results about weight two forms (more specifically the idea of Wiles in [42]), and it is a culmination of a series of work: “\( R = T \) theorem for 2-adic ordinary finite flat representations” by Dickinson [10], “modularity of mod 2 icosahedral representations” by Shepherd-Barron and Taylor [33], and “modular lifting theorem for two-dimensional \( p \)-adic Artin representations unramified at \( p \) (for any prime \( p \))” by Buzzard and Taylor [5]. Buzzard [3] later extended [5] to treat almost all two-dimensional \( p \)-adic Artin representations potentially unramified at \( p \) (the image of the inertia group at \( p \) is finite) and subsequently it led to modularity of two-dimensional “5-adic” icosahedral Artin representations by Taylor [39]. The strong Artin conjecture for odd two-dimensional representations of \( \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \) is now completely proved by work of Khare-Wintenberger and Kisin on Serre’s conjecture for odd two-dimensional mod \( p \) representations of \( \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \).

In this paper, we push through Taylor’s program and generalise them to treat new cases of the strong Artin conjecture for two-dimensional, totally odd, icosahedral Artin representations of the absolute Galois group of a totally real field. More precisely, we prove the following theorems.

**Theorem 1** Let \( F \) be a totally real field. Suppose that 5 splits completely in \( F \). Suppose that \( \rho : \text{Gal}(\overline{F}/F) \to GL_2(\mathbb{C}) \) is a totally odd, irreducible, continuous representation satisfying the following conditions.

- The image of the projective representation \( \text{proj} \rho \) of \( \rho \) is \( A_5 \).
• The projective image of the decomposition group at every place of \( F \) above 5 has order 2.

Then \( \rho \) arises from a holomorphic cuspidal Hilbert modular eigenform of weight 1 and the Artin L-function \( L(\rho, s) \) is entire.

**Theorem 2** Let \( F \) be a totally real field. Suppose that 2 splits completely in \( F \) and that \( [F(\zeta_5) : F] = 4 \). Suppose that \( \rho: \text{Gal}(\overline{F}/F) \to GL_2(\mathbb{C}) \) is a totally odd, irreducible, continuous representation satisfying the following conditions.

• The image of the projective representation \( \text{proj} \rho \) of \( \rho \) is \( A_5 \).

• At every place \( p \) of \( F \) above 2, the projective representation of \( \rho \) is unramified, and the image of \( \text{Frob}_p \) has order 3.

Then \( \rho \) arises from a holomorphic cuspidal Hilbert modular eigenform of weight 1 and the Artin L-function \( L(\rho, s) \) is entire.

These are corollaries of the following theorems, first of which is about “if \( \overline{\rho}: \text{Gal}(\overline{F}/F) \to GL_2(\mathbb{F}_p) \) is modular, then \( \rho: \text{Gal}(\overline{F}/F) \to GL_2(\mathbb{Q}_p) \simeq GL_2(\mathbb{C}) \) is modular”:

**Theorem 3** Let \( p \) be a rational prime. Let \( K \) be a finite extension of \( \mathbb{Q}_p \) with ring \( \mathcal{O} \) of integers and maximal ideal \( \mathfrak{m} \). Let \( F \) be a totally real field. Suppose that \( p \) splits completely in \( F \). Let \( \rho: \text{Gal}(\overline{F}/F) \to GL_2(\mathcal{O}) \) be a continuous representation satisfying the following conditions.

• \( \rho \) ramifies at only finitely many primes.

• \( \overline{\rho} = (\rho \mod \mathfrak{m}) \) is absolutely irreducible when restricted to \( \text{Gal}(\overline{F}/F(\zeta_p)) \), and has a modular lifting which is potentially ordinary and potentially Barsotti-Tate at every prime of \( F \) above \( p \).

• For every prime \( p \) of \( F \) above \( p \), the restriction \( \rho|_{G_p} \) to the decomposition group \( G_p \) at \( p \) is the direct sum of 1-dimensional characters \( \chi_{p,1} \) and \( \chi_{p,2} \) of \( G_p \) such that the images of the inertia subgroup at \( p \) are finite and \((\chi_{p,1} \mod \mathfrak{m}) \neq (\chi_{p,2} \mod \mathfrak{m}) \).

If \( p = 2 \), assume moreover the following conditions.

• The image of the complex conjugation, with respect to every embedding of \( F \) into \( \mathbb{R} \), is not the identity matrix.

• \( \overline{\rho} \) has insoluble image.

• For every prime \( p \) of \( F \) above 2, \( \rho \) is unramified at \( p \).

Then there exists an embedding \( \iota : K \hookrightarrow \overline{\mathbb{Q}}_p \simeq \mathbb{C} \) and a classical holomorphic cuspidal Hilbert modular eigenform \( f \) of weight 1 such that \( \iota \circ \rho \) is isomorphic to the representation associated to \( f \) by Rogawski-Tunnell [28].
In proving the theorem, we shall firstly establish $R = T$ theorems for Hida
$p$-ordinary families over a finite soluble totally real extension $F_{\Sigma}$ of $F$ in which
$p \geq 2$ remains split completely—for lack of reference we shall prove them. Since
$\overline{\rho}$ has a potentially $p$-Barsotti-Tate and potentially $p$-ordinary modular lifting,
one can deduce $R = T$ in $p$-adic families from Kisin’s $R = T$ theorems in
the $p$-Barsotti-Tate case. Note that, unfortunately, it is not possible to make
appeal to Geraghty’s $R = T$ theorems in $p$-ordinary families as they assume
that $p > 2$ and that $\overline{\rho}$ is trivial at every prime of $F$ above $p$. This is because
one can not eliminate the possibility that, upon ‘soluble’ base-changing to $F_{\Sigma}$
to set $\rho_{| \text{Gal}(F/F_{\Sigma})}$ trivial at every prime of $F_{\Sigma}$ above $p$, $F_{\Sigma}$ may no longer be
split at $p$, which is crucial in constructing weight one forms in our approach.
In the light of [1], the condition about the existence of a potentially ordinary
Barsotti-Tate lifting of $\overline{\rho}$ can be weaker, more precisely, it suffices to assume
‘$\overline{\rho}$ is modular’. It is not necessary to make appeal to their results however.

The next two theorems are about modularity of $\overline{\rho}$.

**Theorem 4.** Let $F$ be a totally real field. Suppose that 5 is unramified in $F$.
Let $\overline{\rho} : \text{Gal}(\overline{F}/F) \to \text{GL}_2(\mathbb{F}_5)$ be a continuous representation of satisfying the
following conditions.

- $\overline{\rho}$ is totally odd.
- $\overline{\rho}$ has projective image $A_5$.

There exists a cuspidal Hilbert modular eigenform of weight 2 such that its
associated 5-adic Galois representation is potentially Barsotti-Tate and poten-
tially ordinary at every prime of $F$ above 5, and its associated mod 5 Galois
representation is isomorphic to $\overline{\rho}$.

The idea is exactly the same as that of Taylor—to prove modularity of $\overline{\rho}$, one
firstly finds an elliptic curve $E$ over a finite soluble totally real field extension
$F_{\Sigma}$ of $F$ such that the action of $\text{Gal}(\overline{F}/F_{\Sigma})$ on the 5-torsion points of $E$ is iso-
morphic to $\overline{\rho}_{| \text{Gal}(\overline{F}/F_{\Sigma})}$; secondly one proves $E$ modular, therefore $\overline{\rho}_{| \text{Gal}(\overline{F}/F_{\Sigma})}$ modular; and finally it follows from Khare-Wintenberger [18] and Kisin [20]
that $\overline{\rho}_{| \text{Gal}(\overline{F}/F_{\Sigma})}$ has a characteristic zero lifting which is modular. The ‘auto-
morphic descent’ works as in [39].

In proving $E$ is modular, we make some technical improvements on a ‘naive’
generalisation over totally real fields of the main theorem of Taylor in [39]
by making appeal to the main result of Kisin [20] rather than the main
result of Skinner-Wiles [35]. While Taylor/Skinner-Wiles requires the mod 3
representation $E[3](\overline{F}_{\Sigma})$ of $\text{Gal}(\overline{F}/F_{\Sigma})$ to be reducible with distinct characters
on the diagonal at every prime of $F_{\Sigma}$ above 3, we no longer requires this and
consequently remove the ‘3-distinguishedness condition’ in the main theorems
of [39]. The key observation is that the weight 2 specialisation $F_{\Sigma,2}$ of the Hida
family $F_{\Sigma}$, whose weight 1 specialisation $F_{\Sigma,1}$ renders $E[3](\overline{F}_{\Sigma})$ modular by

---

"Documenta Mathematica 18 (2013) 997–1038"
Langlands-Tunnell, does indeed render the 3-adic Barsotti-Tate representation $T_3E$ ’strongly residually modular’ in the sense of Kisin [20] if $E[3](\mathbb{F}_2)$ is unramified at every prime above 3.

As is clear from its proof, what we are proving indeed is modularity of general mod 5 representations $\text{Gal}(\mathbb{F}/F) \rightarrow \text{GL}_2(F_5)$, and this allows us to work with the prime 2–proving modularity of $\overline{\rho}_2 : \text{Gal}(\mathbb{F}/F) \rightarrow \text{SL}_2(F_4)$ with $\text{proj}\overline{\rho}_2 \simeq A_5$—instead of the prime 5, going back to the original approach of Buzzard-Dickinson-Shepherd-Barron-Taylor; in [4] one firstly finds an abelian surface $A$ over $F$ with real multiplication $\mathbb{Z}[(1 + \sqrt{5})/2]$ such that $A(\mathbb{F})[2] \simeq \rho_2$; secondly proves the mod 5 representation $\text{Gal}(\mathbb{F}/F) \rightarrow \text{GL}_2(A(\mathbb{F})[\sqrt{5}]) \simeq \text{GL}_2(F_5)$ is modular; and deduce $A$ is modular by a modular lifting theorem.

THEOREM 5 Let $F$ be a totally real field. Suppose that $[F(\zeta_5) : F] = 4$. Let $\overline{\rho} : \text{Gal}(\mathbb{F}/F) \rightarrow \text{SL}_2(F_4)$ be a continuous representation. Then there exists a cuspidal Hilbert modular eigenform of weight 2 such that its associated 2-adic Galois representation is potentially Barsotti-Tate and potentially unramified at every prime of $F$ above 2 and its associated mod 2 Galois representation is isomorphic to $\overline{\rho}$.

Lastly it might come in useful comparing our work and others. After the first draft of this paper was written in 2010, Kassaei announced a result proving an analogue of the main theorem 3 in the case when $p$ is odd, $p$ is unramified in $F$, and $\chi_{p,1}/\chi_{p,2}$ and $\chi_{p,2}/\chi_{p,1}$ are both unramified at every prime $p$ of $F$ above $p$. Pilloni, on the other hand, seems to have proved a slightly stronger analogue in which $p$ is allowed to ramify a little in $F$. The fundamental ideas in both works and ours are essentially the same and are due to Buzzard, more specifically to Buzzard’s Theorem 9.1 in [3]. In forthcoming joint work with Kassaei and Tian, we extend Kassaei’s work to the case where $\chi_{p,1}/\chi_{p,2}$ and $\chi_{p,2}/\chi_{p,1}$ are of conductor $p$ for every prime $p$ of $F$ above $p$ (unramified in $F$) and prove many new cases of the strong Artin conjecture for $\rho : \text{Gal}(\mathbb{F}/F) \rightarrow \text{GL}_2(\mathbb{C})$ in the insoluble case as above.

To prove an analogue of the main theorem 3 in the case where $\chi_{p,1}/\chi_{p,2}$ and $\chi_{p,2}/\chi_{p,1}$ are of conductor $p^r$ with $r > 1$ for every prime $p$ of $F$ above $p$, one needs to know precise geometry of Hilbert modular varieties of level $p^r$ and, unless $p$ splits completely in $F$ which we solve, this may not even be possible. Calculating $q$-expansions at cusps to glue weight one forms does not seem to depend on the ramification of $p$ in $F$ and, for that, this work is very useful in general. On the other hand, in order to prove the general case ($p$ ramifies arbitrarily in $F$), the author [30] considers new moduli spaces of Hilbert-Blumenthal abelian varieties; and he expects to make progress in the general case in his forthcoming work.

ACKNOWLEDGEMENTS The author would like to thank Kevin Buzzard, Fred Diamond, and Payman Kassaei for extremely helpful comments whilst the author was preparing this paper. He would also like to thank David Geraghty.
and Wansu Kim for their help. Finally, he would like to thank the referee for very helpful comments which led to many improvements in the paper.

The author was supported financially by EPSRC (UK) whilst he was preparing the manuscript. He would like to thank their support. He would also like to thank Institut Henri Poincaré (Paris, France) and King’s College London (London, UK) for the financial support and hospitality they offered.

2 Modularity of mod 5 icosahedral representations of $\text{Gal}(\overline{F}/F)$

**Lemma 6** Let $F$ be a totally real field. Suppose that 5 is unramified in $F$. Suppose that $\overline{\rho} : \text{Gal}(\overline{F}/F) \to \text{GL}_2(\mathbb{F}_5)$ is a continuous representation satisfying the following conditions.

• $\overline{\rho}$ is totally odd.

• $\overline{\rho}$ has projective image $A_5$.

Then there is a finite soluble totally real field extension $F_\Sigma$ of $F$ and an elliptic curve $E$ over $F_\Sigma$ such that

• $F(\sqrt{5}) \subset F_\Sigma \subset \overline{F}$, and $\sqrt{5}$ splits completely in $F_\Sigma$;

• $E$ has good ordinary reduction at every prime of $F$ above 3 and has potentially ordinary reduction at every prime of $F$ above 5;

• $\overline{\rho}_{E,5} : \text{Gal}(\overline{F}/F_\Sigma) \to \text{Aut}(E(\overline{F}_\Sigma)[5])$ is equivalent to a twist of $\overline{\rho}|_{\text{Gal}(\overline{F}/K_1)}$;

• $\overline{\rho}_{E,3}|_{\text{Gal}(\overline{F}/F_\Sigma(\zeta_3))} : \text{Gal}(\overline{F}/F_\Sigma(\zeta_3)) \to \text{Aut}(E(\overline{F}_\Sigma)[3])$ is absolutely irreducible.

**Proof.** Firstly, as in [39], find a biquadratic totally real extension $K_1 \subset \overline{F}$ of $F$, which is a quadratic totally real extension of $F(\sqrt{5})$ in which $\sqrt{5}$ splits completely, such that $\text{proj} \overline{\rho} : \text{Gal}(\overline{F}/K_1) \to PSL_2(\mathbb{F}_5) \simeq A_5$ lifts to a representation $\overline{\rho}_1 : \text{Gal}(\overline{F}/K_1) \to GL_2(\mathbb{F}_5)$ with determinant the mod 5 cyclotomic character $\epsilon$. Choose, by class field theory, a finite soluble totally real extension $K_2 \subset \overline{F}$ of $K_1$ such that $\overline{\rho}_1|_{\text{Gal}(\overline{F}/K_2)}$ is trivial when restricted to the decomposition group at every prime of $K_2$ above 3. Let $F_\Sigma$ denote the Galois closure of $K_2$ over $F$. Let $\overline{\rho}_{\Sigma}$ denote the restriction of $\overline{\rho}$ to $\text{Gal}(\overline{F}/F_\Sigma)$.

As in section 1 of [33], let $Y_{\pi_0}/F_\Sigma$ (resp. $X_{\pi_0}/F_\Sigma$) denote the twist of the (resp. compactified) modular curve $Y_5$ (resp. $X_5$) with full level 5 structure by the cohomology class in $H^1(\text{Gal}(\overline{F}/F_\Sigma), \text{Aut} X_5)$ defined by an isomorphism $\overline{\rho}_\Sigma \simeq (\mathbb{Z}/5\mathbb{Z}) \times \mu_5$ of the $\mathbb{F}_5$-vector spaces. As proved in Lemma 1.1 in [33], the ‘twist’ cohomology class is indeed trivial, and therefore $X_{\pi_0} \simeq X_5$ and $Y_{\pi_0}$ is isomorphic over $F_\Sigma$ to a Zariski open subset of the projective line $\mathbb{P}^1$. In particular, $Y_{\pi_0}$ has infinitely many rational points.

Let $Y_{\pi_0,0}(3)$ denote the degree 4 cover over $Y_{\pi_0}$ which parameterises isomorphism classes of elliptic curves $E$ equipped with an isomorphism $E[5] \simeq \overline{\rho}_\Sigma$ taking...
the Weil pairing on $E[5]$ to $\epsilon : \Lambda^2 T_2^e \to \mu_5$ and a finite flat subgroup scheme $C \subset E[3]$ of order 3.

Let $Y_{F_2,\text{split}}(3)$ denote the étale cover over $Y_{E_2}$ which parameterises isomorphism classes of elliptic curves $E$ equipped with an isomorphism $E[5] \simeq T_2^e$ taking the Weil pairing on $E[5]$ to $\epsilon : \Lambda^2 T_2^e \to \mu_5$ and an unordered pair, fixed by $\text{Gal}(\overline{F}/F_{2})$, of finite flat subgroup schemes $C, D \subset E[3]$ of order 3 which intersect trivially. Then it follows from Lemma 12 in [27] that $Y_{F_2,\text{split}}(3)$ and $Y_{F_2,0}(3)$ has only finitely many rational points.

For every prime $p$ of $F_{2}$ above 3, the elliptic curve $y^2 = x^3 + x^2 - x$ defines an element of $Y_{\Sigma_2}(F_{2,p})$ with good ordinary reduction, and we let $U_p \subset Y_{\Sigma_2}(F_{2,p})$ denote a (non-empty) open neighbourhood (for the 3-adic topology) of the point, consisting of elliptic curves with good ordinary reduction at $p$.

For every prime $p$ of $F_{2}$ above 5, we define a non-empty open subset (for the 5-adic topology) $U_p \subset Y_{\Sigma_2}(F_{2,p})$ as in the proof of Lemma 2.3 in [39].

By Hilbert irreducibility theorem (Theorem 1.3 in [11]; see also Theorem 3.5.7 in [32]), we may then find a rational point in $Y_{\Sigma_2}(F_{2})$ which lies in $U_p$ for every $p$ of $F_{2}$ above either 3 or 5 and does not lie in the images of $Y_{\Sigma_2,0}(3)(F_{2}) \to Y_{E_2}(F_{2})$ and $Y_{\Sigma_2,\text{split}}(3)(F_{2}) \to Y_{E_2}(F_{2})$. The elliptic curve over $F_{2}$ corresponding to the rational point is what we are looking for. □

**Theorem 7** Let $F$ be a totally real field. Suppose that 5 is unramified in $F$.

Let $\overline{\rho} : \text{Gal}(\overline{F}/F) \to \text{GL}_2(\overline{\mathbb{F}}_5)$ be a continuous representation of satisfying the following conditions.

- $\overline{\rho}$ is totally odd.
- $\overline{\rho}$ has projective image $A_5$.

Then there exists a cuspidal Hilbert modular eigenform of weight 2 such that its associated 5-adic Galois representation is potentially Barsotti-Tate and potentially ordinary at every prime of $F$ above 5 and its associated mod 5 Galois representation is isomorphic to $\overline{\rho}$.

**Proof.** Choose an elliptic curve over a finite soluble totally real extension $F_{2}$ of $F$ as in the lemma. Replace $F_{2}$ by its finite soluble totally real extension if necessary to assume that the mod 3 representation $\overline{\rho}_{E,3}$ of $\text{Gal}(\overline{F}/F_{2})$ on $E(F_{2})[3]$ is unramified when restricted to the decomposition subgroup at every prime of $F_{2}$ above 3. By the Langlands-Tunnell theorem, there exists a weight 1 holomorphic cuspidal Hilbert modular eigenform $f_1$ which gives rise to $\overline{\rho}_{E,3}$. By 3-adic Hida theory [14], we may find a holomorphic cuspidal Hilbert modular eigenform $f_2$ of weight 2 and level prime to 3, ordinary at every prime of $F_{2}$ above 3, which gives rise to $\overline{\rho}_{E,3}$. As $E$ is ordinary at 3, $\rho_{E,3}$ is strongly residually modular in the sense of Kisin [20] (3.5.4), and it follows from Theorem 3.5.5 in [20] that $T_3E$ is modular. By Faltings’ isogeny theorem, $E$ is therefore modular. As $\overline{\rho}_{E,5}$ is modular, $\overline{\rho}_{\text{Gal}(\overline{F}/F_{E})}$ is modular. Since $F_{2}$ is a soluble extension of $F$, $\overline{\rho}_{E}$ remains absolutely irreducible when restricted
to $\text{Gal}(\overline{F}/F_\Sigma(\zeta))$. Furthermore, since $5$ is unramified in $F$, the kernel of $\text{proj} \overline{\rho}_\Sigma$ does not fix $F_\Sigma(\zeta_5)$. It then follows from results of Khare-Wintenberger [18] and Kisin [20] that there exists a modular lifting of $\overline{\rho}_\Sigma$. The ‘soluble descent’ to $F$ is exactly as in [39]. □

**Remark.** In the forthcoming work with Kassaei and Tian, we remove the assumption that $5$ is unramified in $F$ in Lemma 6, and thereby in Theorem 7. Essentially the same argument works.

### 3 Modularity of mod 2 icosahedral representations of $\text{Gal}(\overline{F}/F)$

**Theorem 8** Let $F$ be a totally real field. Suppose that $[F(\zeta_5) : F] = 4$. Let $\overline{\rho} : \text{Gal}(\overline{F}/F) \to SL_2(F_4)$ be a continuous representation. Then there exists a cuspidal Hilbert modular eigenform of weight 2 such that its associated $2$-adic Galois representation is potentially Barsotti-Tate and potentially unramified at every prime of $F$ above 2 and its associated mod 2 Galois representation is isomorphic to $\overline{\rho}$.

**Proof.** By Theorem 3.4 in [33], there exists a principally polarised abelian surface $A$ over $F$ with real multiplication by $\mathbb{Z}[(1 + \sqrt{5})/2]$ compatible with the polarisation such that the action of $\text{Gal}(\overline{F}/F)$ on $A(\overline{F})[2] \cong F_4^2$ is equivalent to $\overline{\rho}$; and the action of $\text{Gal}(\overline{F}/F)$ on $A(\overline{F})[\sqrt{5}] \cong F_5^2$ is given via a homomorphism

$$\overline{\rho}_{A,\sqrt{5}} : \text{Gal}(\overline{F}/F) \to GL_2(F_5)$$

which is surjective and whose image contains $SL_2(F_5)$. It suffices to prove that $A$ is modular.

Firstly, the Weil pairing on $A(\overline{F})[\sqrt{5}]$ shows that $\det \overline{\rho}_{A,\sqrt{5}}$ is the mod 5 cyclotomic character. Since $[F(\zeta_5) : F] = 4$, the determinant is indeed surjective, and therefore $\overline{\rho}_{A,\sqrt{5}}$ is absolutely irreducible.

If $\overline{\rho}_{A,\sqrt{5}}$ is irreducible at some place of $F$ above 5, the absolute irreducibility of $\overline{\rho}_{A,\sqrt{5}}$ implies the absolute irreducibility of its restriction to $\text{Gal}(\overline{F}/F(\zeta_5))$. Otherwise, $\overline{\rho}_{A,\sqrt{5}}$ is reducible at every place of $F$ above 5; in which case, it is also equally easy to check that its restriction to $\text{Gal}(\overline{F}/F(\sqrt{5}))$ of $\overline{\rho}_{A,\sqrt{5}}$ is absolutely irreducible (See Proposition 7 in [27], for example). It follows from results of Khare-Wintenberger [16] and Kisin [20] that it is possible to construct a modular lifting of $\overline{\rho}_{A,\sqrt{5}}$; more precisely, $\overline{\rho}_{A,\sqrt{5}}$ is strongly residually modular. The modularity of $\rho_{A,\sqrt{5}}$ follows from Theorem 3.5.5 in [20] and [12]. □

### 4 Holomorphic Hilbert modular forms and Hida theory of modular Galois representations

Let $F$ be a totally real field. We let $\mathcal{O}_F$ denote the ring of integers, $\mathfrak{p}_F$ the different of $F$, $\mathcal{A}_F = \mathcal{A}_F^\infty \times F_\infty$, and $\mathcal{O}_F \otimes \mathbb{Z}^\times \subset \mathcal{A}_F^\infty$. Let $S_\infty$
denote the set of infinite places of $F$. For an ideal $n$ of $\mathcal{O}_F$, let $F_n$ denote the strict ray class field of conductor $nS_\infty$.

For an ideal $n$, let $U^1(n)$ (resp. $U_1(n)$) denote the open compact subgroup of $GL_2(\mathcal{O}_F^n)$ consisting of matrices which are congruent modulo $n\mathcal{O}_F^n$ to matrices with first column $(1,0)$ (resp. the second row $(0,1)$). Let $I_n$ denote $A_F^n/(F^n(A_F^n \cap U^1(n))^F_n\times)$.

For $k \in \mathbb{Z}$ and an open compact subgroup $U$ of $GL_2(\mathcal{O}_F^n)$, let $S_k(U)$ denote the space, $S_k(U)$ in the sense of Hida [14], of cuspidal holomorphic Hilbert modular forms $f$ of parallel weight $k$ and level $U$ with the Fourier coefficient $c(n,f) \in \mathbb{Z}$ for all ideals $n$ of $\mathcal{O}_F$. The spaces $S_k(U^1(n))$ and $S_k(U_1(n))$ for an ideal $n$ of $\mathcal{O}_F$ come equipped with an action of $I_n$ via the diamond operator $\langle \rangle$, and Hecke operators $T_q$ for every prime $q$ of $\mathcal{O}_F$ not dividing $n$ and $U_q$ for every prime of $q$ dividing $n$.

Let $h_k(n)$ denote the sub $\mathbb{Z}$-algebra of $\text{End}(S_k(U^1(n)))$ generated over $\mathbb{Z}$ by all these operators (See Proposition 2.3, Theorem 4.10, and Theorem 4.11 of [14]). For every prime $q$ not dividing $n$, let $S_q = (\mathbb{N}_F/q\mathbb{Q})^{k-2}(q) \in h_k(n)$; this corresponds to the action of the scalar matrix with a uniformiser of $\mathcal{O}_F$ at $q$ on the diagonal. Following [14], for every ideal $m$ of $\mathcal{O}_F$, we may define $T_m \in h_k(n)$.

Let $p$ be a rational prime and let $S_p$ denote the set of prime ideals of $\mathcal{O}_F$ dividing $p$. Fix an algebraic closure $\overline{\mathbb{Q}}_p$, an isomorphism $\overline{\mathbb{Q}}_p \simeq \mathbb{C}$, and an embedding $\overline{\mathbb{Q}} \rightarrow \overline{\mathbb{Q}}_p$.

For a ring $R \subset \overline{\mathbb{Q}}_p$, we shall let $S_k(U^1(n))_R$ denote $S_k(U_1(n)) \otimes \mathbb{Z} R$ and $h_k(n)_R$ denote $h_k(n) \otimes \mathbb{Z} R$; there is a pairing $\langle \ , \ \rangle : h_k(n)_R \times S_k(U_1(n))_R \rightarrow R$ defined by $(T, f) = c(\mathcal{O}_F, T, f)$.

For a ray class character $\psi : I_n \rightarrow \overline{\mathbb{Q}}_p^\times$ mod $nS_\infty$, let $S_{k, \psi}(U_1(n))_{\mathbb{Z}[\psi]}$ denote the submodule of $S_k(U_1(n))_{\mathbb{Z}[\psi]}$ consisting of cuspidal Hilbert modular forms of parallel weight $k$ and level $U_1(n)$ with central character $\psi$ and level $\mathbb{Z}$ acts via $\psi$ at $q$; the forms in $S_{k, \psi}(U_1(n))_{\mathbb{Z}[\psi]}$ may be thought of as $[I_n]$-tuples of classical Hilbert modular forms of ‘level $\Gamma_1(n)$’ on the $[I_n]$-copies of $GL_2(R)/\Gamma(SO_2(R))^{\text{Hom}(F, R)}$ with ‘Dirichlet character mod $n$’.

Fix an ideal $n$ of $\mathcal{O}_F$ coprime to $p$. For a finite extension $K$ of $\mathbb{Q}_p$ with ring $\mathcal{O}$ of integers, Hida [14] defines the ideompotent $e$ and we set $h^0_\mathcal{O}(n)$ to be the inverse limit with respect to $r \in \mathbb{Z}_{>1}$ of $h_2(n^r)_{\mathcal{O}_e} \otimes \mathcal{O}_e$. Let $I_{np}^\infty$ denote the inverse limit of the $I_{np}$ and the diamond operators $\langle \rangle : I_{np}^\infty \rightarrow e h_2(n^r)_{\mathcal{O}_e}$ induce

$$
\langle \rangle : I_{np}^\infty \rightarrow h^0_\mathcal{O}(n)^\times.
$$

One can also see $\langle \rangle$ as the action of $(\mathcal{O}_F/n)^\times \times (\mathcal{O}_F \otimes \mathbb{Z})^\times$ by the composite:

$$(\mathcal{O}_F/n)^\times \times (\mathcal{O}_F \otimes \mathbb{Z})^\times \rightarrow I_{np}^\infty \rightarrow h^0_\mathcal{O}(n).$$
Hida [14] proves that $h^0_\mathcal{O}(n)$ is a torsion free $\Lambda_K$-module and, for a character $\psi : I_{\mathfrak{p}^\infty} \to K$ which factors through $I_{\mathfrak{p}^r}$ for $r \in \mathbb{Z}_{\geq 1}$, $k \geq 2$, then $h^0_\mathcal{O}(n)_{\ker((\psi \circ \text{Art})^s - \psi)}$ is isomorphic to the subspace of $\epsilon S_k(U_1(np^s))_\mathcal{O}$ where $(\cdot) = \psi$ on $\mathbb{F}_{np^s}$.

We will let $e^{\text{cyclo}}$ denote the character
\[\text{Gal}(\overline{F}/F)\to \text{Gal}(\overline{F}/F)^{\text{ab}} \to I_{\mathfrak{p}^\infty} \leftrightarrow \mathcal{O}[[I_{\mathfrak{p}^\infty}]]^\times = \Lambda_K[\text{Tor}_{\mathfrak{p}^\infty}]^\times.\]

Note that $q \mapsto \mathbb{N}qS_q$ extends to $\mathbb{N}S : (\mathcal{O}_F/n)^\times \times (\mathcal{O}_F \otimes \mathbb{Z}_p)^\times \to I_{\mathfrak{p}^\infty} \to h^0_\mathcal{O}(n)^\times$. Let $\mathbb{N}S_{\Sigma}$ (resp. $\mathbb{N}S_{ap}^\times$) denote the $\Sigma$ (resp. the prime to $S_F$) part $\prod_{p \in \Sigma} \mathcal{O}_F^\times \to h^0_\mathcal{O}(n)$ (resp. $(\mathcal{O}_F/n)^\times \to h^0_\mathcal{O}(n)$) for $s$ subset $\Sigma$ of $S_F$.

If $m$ is a maximal ideal of $h^0_\mathcal{O}(n)$ with residue field $k_m$, there is a continuous representation
\[\overline{\rho}_m : G_F = \text{Gal}(\overline{F}/F) \to GL_2(k_m)\]
such that, for every prime ideal $q$ of $\mathcal{O}_F$ not dividing $np$, $\overline{\rho}_m$ is unramified at $q$ and $\text{tr}_{\mathcal{O}_m}((\text{Frob}_q)) = T_q$. Set $S^0_\mathcal{O}(n) = \text{Hom}_{\Lambda_K}(h^0_\mathcal{O}(n), \Lambda_K)$. For a finite field extension $L$ of the field $\text{Frac} \Lambda_K$ of fractions of $\Lambda_K$ with its integral closure $\mathcal{O}_L$ of $\Lambda_K$ in $L$, Buzzard-Taylor [5] calls a $\Lambda_K$-algebra homomorphism $F_{\mathfrak{p}^s} \in S^0_\mathcal{O}(n) \otimes_{\Lambda_K} L = \text{Hom}_{\Lambda_K}(h^0_\mathcal{O}, L)$ a $\Lambda$-adic eigenform (of level $n$).

If the unique maximal ideal $m$ above $\ker F_{\mathfrak{p}^s} \subset h^0_\mathcal{O}(n)$ is non-Eisenstein, i.e., $\overline{\rho}_m$ as above is absolutely irreducible, then there is a continuous representation
\[\rho_{F_{\mathfrak{p}^s}} : G_F \to GL_2(h^0_\mathcal{O}(n)_m) \cong GL_2(O_L),\]

which is unramified at every prime ideal $q$ of $\mathcal{O}_F$ not dividing $np$ and satisfies $\text{tr} \rho_{F_{\mathfrak{p}^s}}((\text{Frob}_q)) = T_q$ and $\det \rho_{F_{\mathfrak{p}^s}} = (NS) \circ e^{\text{cyclo}}$. Moreover, it is a result of Wiles [43] that, for every place $p$ of $F$ above $p$, the restriction to the decomposition group $G_{\mathfrak{p}}$ at $p$ of $\rho_{F_{\mathfrak{p}^s}}$ is of the form
\[\begin{pmatrix} \chi_{F_{\mathfrak{p}^s}, p, 2} & * \\ 0 & \chi_{F_{\mathfrak{p}^s}, p, 1} \end{pmatrix}\]
where $\chi_{F_{i,2},1}$ is an unramified character of $G_p$ such that $\chi_{F_{i,2},1}(\text{Frob}_p) = U_p$ and $\chi_{F_{i,2},1}\chi_{F_{i,2},2} = (F_{i,1} \otimes NS) \circ e^{-\text{cho}}|_{G_p}$.

**Definition.** Following [5], we call two $\Lambda$-adic eigenforms $F_{i,1}$ and $F_{i,2}$: $\mathcal{H}_G(n) \rightarrow \mathcal{O}_\mathfrak{L}$ of level $\mathfrak{n}$ $\Lambda$-adic companion form with respect to primes $\varphi_1$ and $\varphi_2$ of $\mathcal{O}_\mathfrak{L}$ which do not divide $p$, if there are embeddings $\iota_1 : \mathcal{O}_\mathfrak{L}/\varphi_1 \hookrightarrow \mathbb{Q}_p$ and $\iota_2 : \mathcal{O}_\mathfrak{L}/\varphi_2 \hookrightarrow \mathbb{Q}_p$ such that, for every ideal $\mathfrak{m}$ of $\mathcal{O}_F$ not dividing $p$, there exists a subset $\Sigma$ of $S_p$ such that $(F_{i,1}(T_{\mathfrak{m}}) \mod \varphi_2) = (F_{i,1}((T_{\mathfrak{m}}(\mathfrak{NS}_\mathfrak{m}))(\mathfrak{m}^{-1}) \mod \varphi_1))$; and such that, for every place $\mathfrak{p}$ in $\Sigma$, $(F_{i,1}(U_{\mathfrak{p}}) \mod \varphi_2) = (F_{i,1}(U_{\mathfrak{p}}(\mathfrak{NS}^{\mathfrak{m}})(\mathfrak{p})) \mod \varphi_1)$ while, for every $\mathfrak{p}$ in $S_p - \Sigma$, $(F_{i,2}(U_{\mathfrak{p}}) \mod \varphi_2) = (F_{i,1}(U_{\mathfrak{p}}) \mod \varphi_1)$.

5 Deformation rings and Hecke algebras

Let $F$ be a totally real field of even degree in which $p$ is unramified; if $p = 2$ assume furthermore that 2 splits completely in $F$. If $p$ is odd, suppose $p \geq 5$.

Let $D$ be the quaternion algebra over $F$ which ramifies exactly at a finite set $\Sigma$ of finite places of $F$ not dividing $p$ and the infinite places $S_\infty$ of $F$. Let $\mathcal{O}_D$ denote a maximal order and fix an isomorphism $\mathcal{O}_D \cong M_2(\mathcal{O}_F)$ for $\mathfrak{q}$ not in $\Sigma$. Let $S$ denote the disjoint union of $\Sigma$, the set $S_p$ of places of $F$ above $p$, and the infinite places of $F$.

For a topological $\mathbb{Z}_p$-algebra $R$, let $\psi : \mathbb{A}_F^{\infty}/F \rightarrow \mathbb{R}^\times$ be a continuous character such that $\psi|_{\mathbb{R}_p}$ is trivial for every place $\mathfrak{p}$ of $F$ above $p$, and, for an open compact subgroup $U = \prod_{\mathfrak{q}} U_{\mathfrak{q}} \subset \prod_{\mathfrak{q}} \mathcal{O}_D^\times$, let $S_{\mathfrak{q}}^{\psi}(U)_R$ denote the $R$-module of $R$-valued modular forms on $\mathbb{A}_F^{\infty}/(D \otimes F \mathbb{A}_F^{\infty})^\times$ of weight 2 and of level $U$ in the sense of Taylor [40].

Let $\n_\Sigma$ denote the square-free product of the primes in $\Sigma$ and define $U_\Sigma \subset (D \otimes F \mathbb{A}_F^{\infty})^\times$ by $U_{\Sigma,\mathfrak{q}} = GL_2(\mathcal{O}_{F_{\mathfrak{q}}})$ for $\mathfrak{q}$ not in $\Sigma$; and $U_{\Sigma,\mathfrak{q}} = \mathcal{O}_D^\times$ for $\mathfrak{q} \in \Sigma$.

We shall write $S_{\mathfrak{q}}^{\psi}(\n_\Sigma)$ for $S_{\mathfrak{q}}^{\psi}(U_\Sigma)$ and $h_{\mathfrak{q}}^{\psi}(\n_\Sigma)_R$ for the $R$-subalgebra of $\text{End}_R(S_{\mathfrak{q}}^{\psi}(\n_\Sigma)_R)$ generated by $T_{\mathfrak{q}}$ and $S_{\mathfrak{q}}$ for all $\mathfrak{q}$ not in $S$; and $T_{\mathfrak{q}}$ and $S_{\mathfrak{q}}$ for all $\mathfrak{p}$ in $S_p$.

Let $K$ be a finite extension of $\mathbb{Q}_p$ and $\mathcal{O}$ be the ring of integers with maximal ideal $\mathfrak{m}$ ad residue field $k$. Let $\rho : \text{Gal}(\overline{F}/F) \rightarrow GL_2(\mathcal{O})$

be a continuous representation such that

- $\overline{\rho} = (\rho \mod \mathfrak{m})$ is unramified outside $S_p$,
- $\overline{\rho}$ is not scalar at every place $\mathfrak{p}$ above $p$,
- if $p$ is odd, the restriction of $\overline{\rho}$ to $\text{Gal}(\overline{F}/F(\zeta_p))$ is absolutely irreducible; if $p = 2$, $\overline{\rho}$ has insoluble image,
there exists a holomorphic automorphic representation $\pi$ of $(D \otimes F A_F)^\times$ generated by a cusp form in $S^D_{2\Sigma}(\pi_S)$ such that $\pi_q$ is unramified for every $q$ not in $\Sigma \cup S_p$, $\pi_p$ is ordinary at every $p$ in $S_p$, for every $q \in \Sigma$, $\pi_q$ corresponds by the local Jacquet-Langlands correspondence to a special representation of conductor $q$, and such that $\overline{\rho}_q \simeq \overline{\rho}$.

- $\rho$ ramifies at $\Sigma$ and possibly at $S_p$; for every $p$ in $S_p$

\[ \rho|_{G_p} \sim \begin{pmatrix} * & * \\ 0 & \chi_p \end{pmatrix} \]

with $\chi_p$ unramified; and for $q \in \Sigma$

\[ \rho|_{G_q} \sim \begin{pmatrix} \epsilon \chi_q & * \\ 0 & \chi_q \end{pmatrix} \]

with $\chi_q$ unramified such that $\chi_q^2 = (\psi \circ \text{Art})|_{G_q}$.

Let $A_F^\times = A_F^{\times S} \times A_F^{\times S}$ for a finite subset $S$ of the places of $F$. Let $\psi$ be a character of $A_F^{\times S}$. For $p = 2$ let $\psi_{p, \pm}$ denote the $\mathbb{Z}_p$-linear extension of the norm $N : (O_F \otimes \mathbb{Z}_2)^\times \to \mathbb{Z}_2^\times$ followed by the character $\mathbb{Z}_2^\times \to \mathbb{Z}_2^\times$ whose restriction to $(\mathbb{Z}/4)^\times = \{ \pm 1 \}$ sends $-1$ to $\mp 1$ and whose restriction to $(1 + 4\mathbb{Z}_2)^\times$ is trivial. For $p$ odd, let $\psi_p$ denote the norm followed by the trivial character on $\mathbb{Z}_p^\times$.

5.1 (Framed) deformation rings $R$

$p|p$; if $p$ is odd, let $R_{p, \text{ord}}^{\square}$ (resp. $R_{p, \text{ord}}^{\square, \text{BT,ord}}$) denote the $O$-algebra which represents the $p$-ordinary (resp. Barsotti-Tate $p$-ordinary) framed deformations of $\overline{\rho}|_{G_p}$ of the form

\[ \begin{pmatrix} * & * \\ 0 & \chi_p^{ur} \end{pmatrix} \]

with an unramified lifting $\chi_p^{ur}$ of $\chi_p$ (resp. and its determinant is $\epsilon \psi_p$); if $p = 2$, we shall write $`\pm`$ in shorthand to mean two independent cases—`+` corresponds to the 2-old case while `-` corresponds to the 2-new case in the sense to be made precise below, and let $R_{p, \pm}^{\square, \text{ord}}$ (resp. $R_{p, \pm}^{\square, \text{BT,ord}}$) denote the complete local noetherian $O$-algebra which represents the $p$-ordinary (resp. Barsotti-Tate $p$-ordinary) liftings of $\overline{\rho}|_{G_p}$ of the form

\[ \begin{pmatrix} * & * \\ 0 & \chi_p^{ur} \end{pmatrix} \]

with an unramified lifting $\chi_p^{ur}$ of $\chi_p$, and with its determinant corresponding, by the local class field theory, to the norm $O_F^\times \hookrightarrow (O_F \otimes \mathbb{Z}_2)^\times \xrightarrow{N} \mathbb{Z}_2^\times$ followed by the character $\mathbb{Z}_2^\times \to \mathbb{Z}_2^\times$ whose restriction to $(\mathbb{Z}/4)^\times = \{ \pm 1 \}$ sends $-1$ to...
\(1\) (resp. with its determinant \(\epsilon \psi P,\pm\)).

Let \(R_{p,\text{ord}}^\square = \bigotimes_{P \in S_p} R_{p,\text{ord}}^\square\) (resp. \(R_{p,\text{BT,ord}}^\square = \bigotimes_{P \in S_p} R_{p,\text{BT,ord}}^\square\)) if \(p\) is odd; and \(R_{p,\text{ord}}^\square = \bigotimes_{P \in S_p} R_{p,\text{ord}}^\square\) (resp. \(R_{p,\text{BT,ord}}^\square = \bigotimes_{P \in S_p} R_{p,\text{BT,ord}}^\square\)) if \(p = 2\).

For \(q \in \Sigma\) let \(R_q^{\square,\psi}\) denote the domain (see 2.6 in [20], or Proposition 2.12 and 3.3.4 in [18]) parameterising liftings of \(\rho_{|G_q}\) of the form

\[
\begin{pmatrix}
\epsilon \chi_q & 0 \\
0 & \chi_q
\end{pmatrix}
\]

with \(\chi_q\) an unramified lifting of \(\chi_q\) such that \((\chi_q)^2 = (\psi \circ \text{Art}^{-1})|G_q\).

Let \(R_g^{\square,\psi}\) denote the completed tensor product \(\bigotimes_{q \in \Sigma} R_q^{\square,\psi}\).

For \(\tau|\infty\) let \(R_{\tau}^{\square,\text{odd}}\) denote the formally smooth ring which represents the liftings of \(\rho_{|G_{\tau}}\) which, if \(p\) is odd, are odd; and, if \(p = 2\), the image of complex conjugation in \(G_{\tau} \simeq \text{Gal}(\mathbb{C}/\mathbb{R})\) is not the identity matrix.

Let \(R_{\infty}^{\square,\text{odd}}\) denote the completed tensor product \(\bigotimes_{\tau|\infty} R_{\tau}^{\square,\text{odd}}\).

Fix a \(k\)-basis of \(\overline{\rho}\) and let

\[
\rho^{\square S} : G_F \to GL_2(R^{\square S})
\]

denote the \(S\)-framed universal deformation ring. Let \(R_S^{\square S}\) denote the completed tensor product of the local framed deformation rings at places in \(S\).

Let

\[
R_{S,\text{ord},\psi}^{\square} = R^{\square S} \otimes_{R_S^{\square S}} (R_{p,\text{ord}}^{\square} \otimes_{R_p^{\square}} R_{\Sigma,\psi}^{\square} \otimes_{R_{\Sigma}^{\square}} R_{\infty}^{\square,\text{odd}})
\]

\[
R_{S,\text{BT,ord},\psi}^{\square} = R^{\square S} \otimes_{R_S^{\square S}} (R_{p,\text{BT,ord}}^{\square} \otimes_{R_p^{\square}} R_{\Sigma,\psi}^{\square} \otimes_{R_{\Sigma}^{\square}} R_{\infty}^{\square,\text{odd}})
\]

if \(p\) is odd; and

\[
R_{S,\pm,\text{ord},\psi}^{\square} = R^{\square S} \otimes_{R_S^{\square S}} (R_{p,\pm,\text{ord}}^{\square} \otimes_{R_p^{\square}} R_{\Sigma,\psi}^{\square} \otimes_{R_{\Sigma}^{\square}} R_{\infty}^{\square,\text{odd}})
\]

\[
R_{S,\pm,\text{BT,ord},\psi}^{\square} = R^{\square S} \otimes_{R_S^{\square S}} (R_{p,\pm,\text{BT,ord}}^{\square} \otimes_{R_p^{\square}} R_{\Sigma,\psi}^{\square} \otimes_{R_{\Sigma}^{\square}} R_{\infty}^{\square,\text{odd}})
\]

if \(p = 2\).

Let \(R_{S,\text{ord},\psi}\) (resp. \(R_{S,\text{BT,ord},\psi}\)) denote the subring of \(R_{S,\text{ord},\psi}^{\square}\) (resp. \(R_{S,\text{BT,ord},\psi}^{\square}\)) generated by the traces of \(\rho^{\square S}\). Similarly define \(R_{S,\pm,\text{ord},\psi}\) and \(R_{S,\pm,\text{BT,ord},\psi}\).
5.2 Hecke Algebras

Since $\mathbf{7}$ arises from a holomorphic cusp form in $S^D_{2}(\mathbf{n}_\Sigma, \psi)_{\mathcal{O}}$ on the quaternion algebra $D$ over $F_\Sigma$ by assumption, there exists a maximal ideal $m^D \subset h^D_{2}(\mathbf{n}_\Sigma, \psi \psi_p)_{\mathcal{O}}$ if $p$ odd (resp. $m^D \subset h^D_{2}(\mathbf{n}_\Sigma, \psi \psi_{p,+})_{\mathcal{O}}$ if $p = 2$). It then follows that there exists a maximal ideal $m \subset h^D_{2}(\mathbf{n}_\Sigma p, \psi \psi_p)_{\mathcal{O}}$ such that

$$h^D_{2}(\mathbf{n}_\Sigma p, \psi)_{m} \simeq h^D_{2}(\mathbf{n}_\Sigma, \psi \psi_p)_{m^D}$$

if $p$ odd (resp. $m^+ \subset h^D_{2}(\mathbf{n}_\Sigma 2, \psi \psi_{p,+})_{\mathcal{O}}$ such that

$$h^D_{2}(\mathbf{n}_\Sigma 2, \psi \psi_{p,+})_{m^+} \simeq h^D_{2}(\mathbf{n}_\Sigma, \psi \psi_{p,+})_{m^D}$$

if $p = 2$). When $p = 2$, there also exists $m_- \subset h^D_{2}(\mathbf{n}_\Sigma 4, \psi \psi_{p,-})$ such that

$$h^D_{2}(\mathbf{n}_\Sigma 4, \psi \psi_{p,-})_{m_-} / (2) \simeq h^D_{2}(\mathbf{n}_\Sigma 2, \psi \psi_{p,+})_{m^+} / (2)$$

This can be proved exactly as in the proof of Lemma 3.2 in [4]; instead use the 0-dimensional Shimura variety corresponding to $D$ over $F_\Sigma$. For $p = 2$ define $e_{\text{BDST}, \pm}$ and let $h^0(n)_{\pm} = e_{\text{BDST}, \pm} h^0(n)$. Let

$$h^D_{2}(\mathbf{n}_\Sigma p, \psi \psi_p)_{m} = h^D_{2}(\mathbf{n}_\Sigma p, \psi \psi_p)_{m} \otimes_{R^S_{2+\text{ord}, \psi}} R^\square_{S, \text{ord}, \psi}$$

$$h^0(\mathbf{n}_\Sigma, \psi \psi_p)_{m} = h^0(\mathbf{n}_\Sigma, \psi \psi_p)_{m} \otimes_{R^S_{2+\text{ord}, \psi}} R^\square_{S, \text{ord}, \psi}$$

if $p$ is odd; and let

$$h^D_{2}(\mathbf{n}_\Sigma 4, \psi \psi_{p,-})_{m_-} = h^D_{2}(\mathbf{n}_\Sigma 4, \psi \psi_{p,-})_{m_-} \otimes_{R^S_{2+\text{ord}, \psi}} R^\square_{S, \text{ord}, \psi}$$

$$h^0(\mathbf{n}_\Sigma, \psi \psi_{p,-})_{m_-} = h^0(\mathbf{n}_\Sigma, \psi \psi_{p,-})_{m_-} \otimes_{R^S_{2+\text{ord}, \psi}} R^\square_{S, \text{ord}, \psi}$$

if $p = 2$. It then follow from results of Kisin and Khare-Wintenberger that there is a natural surjection

$$R^\square_{S, \text{ord}, \psi} \twoheadrightarrow h^D_{2}(\mathbf{n}_\Sigma p, \psi \psi_p)_{m}$$

if $p$ odd and

$$R^\square_{S, \text{ord}, \psi} \twoheadrightarrow h^D_{2}(\mathbf{n}_\Sigma 4, \psi \psi_{p,-})_{m_-}$$

if $p = 2$, which induce isomorphisms

$$R^\square_{S, \text{ord}, \psi} [1/p] \simeq h^D_{2}(\mathbf{n}_\Sigma p, \psi \psi_p)_{m} [1/p]$$

if $p$ odd and

$$R^\square_{S, \text{ord}, \psi} [1/2] \simeq h^D_{2}(\mathbf{n}_\Sigma 4, \psi \psi_{p,-})_{m_-} [1/2].$$

The determinant of $\rho^\square_S$ defines

$$\text{NS} : I_{n_{\Sigma} p} \rightarrow R^\text{ord,} \psi$$
and $R_{S,\varphi}^{\text{ord},\psi}/\ker(S - (\psi\varphi_1 \circ \text{cyclo})) \simeq R_{S}^{\text{BT,ord},\psi}$ if $p$ odd, and

$$NS : I_{n_{2},p} \rightarrow R_{S,-}^{\text{ord},\psi}$$

induces $R_{S,-}^{\text{ord},\psi}/\ker(S - (\psi\varphi_{1,-} \circ \text{cyclo})) \simeq R_{S,-}^{\text{BT,ord},\psi}$ if $p = 2$. On the other hand, $h^0(n_{\Sigma},\psi)_{m}/\ker(S - (\psi\varphi_{1} \circ \text{cyclo})) \simeq h_2(n_{\Sigma},\psi_{1})_{m}$ and $h^0(n_{\Sigma},\psi)_{-m,-}/\ker(S - (\psi\varphi_{1,-} \circ \text{cyclo})) \simeq h_2(n_{\Sigma},\psi_{1,-})_{m,-}$. Then the surjective $\Lambda$-algebra homomorphisms

$$R_{S}^{\square,\text{ord},\psi} \rightarrow h^0(n_{\Sigma},\psi)_{m}$$

if $p$ odd and

$$R_{S,-}^{\text{ord},\psi} \rightarrow h^0(n_{\Sigma},\psi)_{-m,-}$$

if $p = 2$ induce the isomorphisms

$$R_{S}^{\square,\text{ord},\psi}[1/p] \simeq h^0(n_{\Sigma},\psi)_{m}[1/p]$$

and

$$R_{S,-}^{\text{ord},\psi}[1/2] \simeq h^0(n_{\Sigma},\psi)_{-m,-}[1/2].$$

6 Companion forms mod $p$

Let $F$ be a totally real field and $p$ be a rational prime. Suppose that $[F(\zeta_p) : F] > 3$ if $p > 3$ and that $2$ splits completely in $F$ if $p = 2$. Let $f_2$ be a holomorphic cuspidal Hilbert eigenform of weight $2 \leq k_2 \leq p$ and of level prime to $p$. Assume that the associated $p$-adic representation $\rho_2$ of $\text{Gal}(\overline{F}/F)$ is crystalline and ordinary at every prime of $F$ above $p$. It is a well-known theorem of Wiles (Theorem 2.1.4 in [43]) that, for every prime $p$ of $F$ above $p$, the restriction $\rho|_{G_p}$ to the decomposition group $G_p$ at $p$ is of the form

$$\rho|_{G_p} \simeq \begin{pmatrix} \chi_{p,1}^{-1} & * \\ 0 & \chi_{p,2} \end{pmatrix}$$

where $\chi_{p,1}$ and $\chi_{p,2}$ are unramified characters of $G_p$, and $\chi_{p,2}(\text{Frob}_p)$ is a unit $U_p$-eigenvalue of the $p$-stabilised newform of $f_2$.

**Theorem 9** Let $f_2$ be a holomorphic cuspidal Hilbert eigenform of weight $2 \leq k_2 \leq p$ and of level prime to $p$ as above. Let $k_1 \overset{\text{def}}{=} p$ if $k_2 = p$ and $k_1 \overset{\text{def}}{=} p+1-k_2$ if $2 \leq k_2 < p$. Suppose that

- if $p > 2$, the associated mod $p$ representation $\overline{\rho}_2 : \text{Gal}(\overline{F}/F) \to GL_2(\overline{\mathbb{F}}_p)$ is absolutely irreducible when restricted to $\text{Gal}(\overline{F}/F(\zeta_p))$, and if $p = 2$, $\overline{\rho}_2 : \text{Gal}(\overline{F}/F) \to GL_2(\overline{\mathbb{F}}_2)$ has insoluble image;
- if $p > 2$ and if $\chi_{p,2}^{-1} \neq \chi_{p,1}$, the ramification index of $F_p$ is strictly less than $p-1$ for every prime $p$ of $F$ above $p$, and if $p = 2$, $\overline{\rho}_2$ is unramified at every prime of $F$ above $2$.
Then there exists a holomorphic cuspidal Hilbert eigenform of weight $2 \leq k_1 \leq p$ and of level prime to $p$ with its associated mod $p$ representation $\rho_1 : \text{Gal}(\overline{F}/F) \to GL_2(\mathbb{F}_p)$ satisfying $\overline{\pi}_1 \simeq \overline{\pi}_2 \otimes \overline{\rho}_1^{-1}$ if $p > 2$ and $\overline{\pi}_1 \simeq \overline{\pi}_2$ if $p = 2$, and the $U_p$-eigenvalue of the $p$-stabilised new form is a lifting of $\overline{\chi}_{p,1}$.

Proof. For $p > 2$, this is a result of Gee (Theorem 2.1 [13]). Let $p = 2$; thus $k_1 = k_2 = 2$. For clarity, let $\overline{\rho}$ denote $\overline{\rho}_2 \otimes \tau$ where $\tau$ is the mod 4 cyclotomic character. Clearly the twist of $\rho_2$ by the Teichmuller lift of $\tau$ defines a modular lifting of $\overline{\rho}$ potentially ordinary and potentially Barsotti-Tate at $p$. By class field theory, find a finite totally real soluble extension $F_{\Sigma} \subset \overline{F}$ of $F$ of even degree in which 2 remains split completely, and satisfies the following conditions:

• there exists a quaternion algebra $D$ over $F_{\Sigma}$ ramified exactly at a finite set $\Sigma$ of finite primes of $F_{\Sigma}$ not dividing 2;

• $\overline{\rho}|_{\text{Gal}(\overline{F}/F_{\Sigma})}$ is ramified exactly at $\Sigma$ and the infinite places, and , in particular, for every prime $q \in \Sigma$, $\overline{\rho}|_{\text{Gal}(\overline{F}/F_{\Sigma})}$ at $q$ is an extension of an unramified character by the twist of the character by $\epsilon$ at $q$;

• there exists a maximal open compact subgroup $U \subset (D \otimes_{F_{\Sigma}} \mathbb{A}_{F_{\Sigma}})_{\Sigma}^\times$ such that $U_q = GL_2(O_{F_{\Sigma}})$ for $q \notin S^D$ and $U_q = O_{D_q}^\times$ for $q \in S^D$, and a holomorphic cuspidal automorphic representation $\pi_2$ of $(D \otimes_{F_{\Sigma}} \mathbb{A}_{F_{\Sigma}})_{\Sigma}^\times$ with central character $\psi$ such that $\overline{\rho}|_{\text{Gal}(\overline{F}/F_{\Sigma})} \simeq \overline{\rho}_{\pi_2} : \text{Gal}(\overline{F}/F_{\Sigma}) \to GL_2(\mathbb{F}_p)$ and $\det \overline{\rho}|_{\text{Gal}(\overline{F}/F_{\Sigma})} = \overline{\psi} \epsilon$ and such that $\pi$ is unramified at every prime of $F_{\Sigma}$ above 2.

It then follows from work of Khare-Wintenberger (See Corollary 4.7 and Theorem 10.1 in [18]) that there is a lifting $\rho : \text{Gal}(\overline{F}/F_{\Sigma}) \to GL_2(\mathbb{Q}_p)$ of $\overline{\rho}|_{\text{Gal}(\overline{F}/F_{\Sigma})}$, unramified outside $S_{\Sigma,p} \coprod S_{\Sigma,\infty}$ with $\det \rho = \psi \epsilon$ such that, for every prime $p$ of $F_{\Sigma}$ above 2, $\rho$ is ordinary at $p$ and Barsotti-Tate and is of the form

$$
\begin{pmatrix}
\epsilon \chi_{p,2} & * \\
0 & \chi_{p,1}
\end{pmatrix}
$$

where $\chi_{p,1}$ and $\chi_{p,2}$ are unramified liftings of $\chi_{p,1}|_{\text{Gal}(\overline{F}/F_{\Sigma})}$ and $\chi_{p,2}|_{\text{Gal}(\overline{F}/F_{\Sigma})}$ respectively. It then follows from the main theorem of Kisin [19] and Khare-Wintenberer [18], and by soluble descent that there exists a holomorphic cuspidal Hilbert eigenform $f_1$ of weight $k_1 = 2$ and of level prime to 2 such that $\rho|_{f_1}|_{\text{Gal}(\overline{F}/F_{\Sigma})} \simeq \rho$. \(\square\)
7 $\Lambda$-adic companion forms

Theorem 10 Let $p$ be a rational prime. Let $F$ be a totally real field. Suppose that $p$ splits completely in $F$. Let $K$ be a finite extension of $\mathbb{Q}_p$ with ring of integers $\mathcal{O}$ and residue field $k = \mathcal{O}/m$. Suppose that

$$\rho : \text{Gal}(\overline{F}/F) \to GL_2(\mathcal{O})$$

is a continuous representation satisfying

- $\rho$ ramifies at only finite many primes;
- $\overline{\rho} = (\rho \mod m)$ is absolutely irreducible when restricted to $\text{Gal}(\overline{F}/F(\zeta_p))$, and has a modular lifting which is potentially ordinary and potentially Barsotti-Tate at every prime of $F$ above $p$;
- for every prime ideal $\mathfrak{p}$ of $F$ above $p$, the restriction $\rho |_{G_{\mathfrak{p}}}$ to the decomposition group $G_{\mathfrak{p}}$ at $\mathfrak{p}$ is the direct sum of characters $\chi_{p,1}$ and $\chi_{p,2} : G_{\mathfrak{p}} \to \mathcal{O}^\times$ such that the images of the inertia subgroup at $\mathfrak{p}$ are finite and $(\chi_{p,1} \mod m) \neq (\chi_{p,2} \mod m)$;

If $p = 2$, assume furthermore that

- the image of the complex conjugation with respect to every embedding of $F$ into $\mathbb{R}$ is not the identity matrix;
- $\overline{\rho}$ has insoluble image;
- for every $\mathfrak{p}$ of $F$ above $p$, $\rho$ is unramified at $\mathfrak{p}$.

Then there is a finite totally real soluble extension $F_\Sigma \subset \overline{F}$ of $F$ in which $p$ splits completely; a finite set $\Sigma$ of finite places of $F_\Sigma$ (at which $\rho |_{G_{\mathfrak{p}}}$, where $G_{\Sigma} \defeq \text{Gal}(\overline{F}/F_\Sigma)$ is ramified of conductor $n_\Sigma$); an ideal $n$ of $\mathcal{O}_{F_{\Sigma}}$ coprime to $p$ which $n_\Sigma$ divides; and, for every subset $P$ of the set $S_{\Sigma,P}$ of places of $F_{\Sigma}$ above $p$,

1. a character

$$\chi_P : G_\Sigma \to \mathcal{O}^\times$$

of finite order, unramified outside a finite set of places containing $S_{\Sigma,P}$, such that the restriction to the inertia subgroup of $G_\Sigma$ at $\mathfrak{p}$ of $\chi_P$ equals that of $\chi_{p,1}$ (resp. $\chi_{p,2}$) for all $\mathfrak{p}$ in $P$ (resp. $S_{\Sigma,P} - P$);

2. a finite extension $L$ of $\text{Frac} \Lambda_K$ and a $\Lambda$-adic form

$$F_{\text{Hida}, P} : h_0^0(n_\Sigma) \to L;$$

3. a homomorphism $f_P : h_0^0(n) \to \mathcal{O}$ if $p > 2$ while $f_P : h_0^0(n)_- \to \mathcal{O}$ if $p = 2$ satisfying
8 Models of Hilbert modular varieties

Let $F$ be a totally real field $F_\Sigma$ in the preceding section–of degree $d = [F : \mathbb{Q}]$ with ring of integers $\mathcal{O}_F$. Fix a rational prime $p$ and an ideal $\mathfrak{n}$ of $\mathcal{O}_F$ prime to $p$. For every integer $r \geq 1$, fix a $p^r$-th primitive root $\zeta_{p^r}$ of unity. For a prime $\mathfrak{p}$
Let \( F \) above \( p \), let \( F_p \) denote the completion of \( F \) with respect to the absolute value corresponding to \( p \), \( k_p \) the residue field of \( F_p \), \( f_p \) the residue class degree, and \( e_p \) the ramification index.

Fix embeddings \( Q \rightarrow \mathcal{O}_p \rightarrow \mathcal{O}_F \). Let \( K \) denote a finite extension of \( Q_p \) which contains the image of \( F \) by every embedding of \( F \) into \( \mathcal{O}_F \); and let \( \mathcal{O} \) denote its ring of integers and \( k \) denote the residue field.

For a fractional ideal \( I \) of \( F \) canonically ordered, let \( I^+ \) denote the totally positive elements. Fix a set \( T \) of representatives in \( \mathbf{A}_F^\times \) of the strict ideal class group \( \mathbf{A}_F^\times /(F^\times (\mathcal{O}_F \otimes \mathbf{Z}^\times ) \times F_\infty^\times ) \), and we shall let \( t \) also mean the fractional ideal \( \mathfrak{o}_F^t \) corresponding to a representative \( t \) in \( T \).

**Definition.** A \( t \)-polarised Hilbert-Blumenthal abelian variety (henceforth abbreviated as HBAV) with level \( \Gamma_1(\mathfrak{n}) \)-structure over a \( \mathcal{O} \)-scheme \( S \) is an abelian variety \( A \) over \( S \) of relative dimension \( d \) together with

- \( i : \mathcal{O}_F \rightarrow \text{End}(A/S) \);
- a homomorphism \( \lambda : (t, t^+) \rightarrow (\text{Sym}(A/S), \text{Pol}(A/S)) \) of ordered invertible \( \mathcal{O}_F \)-modules, where \( \text{Sym}(A/S) \) (resp. \( \text{Pol}(A/S) \)) denotes the invertible \( \mathcal{O}_F \)-module (via \( i \)) of symmetric homomorphisms (resp. polarisations), such that \( A \otimes_{\mathcal{O}_F} t \rightarrow A^\vee \), induced by \( \lambda \), is an isomorphism of HBAVs—it is shown in [41] that this is equivalent to the condition that there exists a prime-to-\( p \) polarisation \( A \rightarrow A^\vee \); and to the ‘determinant condition’ on \( \text{Lie}(A) \) in the sense of Kottwitz;
- an \( \mathcal{O}_F / \mathfrak{n} \)-module morphism \( \eta : (\mathcal{O}_F / \mathfrak{n})^\vee = (GL_1 \otimes \mathfrak{d}_F^{-1})[\mathfrak{n}] \rightarrow A[\mathfrak{n}] \).

**Definition.** Let \( Y_{\Gamma_1(n, t)} \) (resp. \( Y_{\Gamma_1(n, t) \cap \mathfrak{l}w} \)) denote the \( \mathcal{O} \)-scheme representing the functor which sends an \( \mathcal{O} \)-scheme \( S \) to the set of isomorphism classes \((A, i, \lambda, \eta)\) (resp. \((A, i, \lambda, \eta, C)\)) of \( t \)-polarised HBAVs with level \( \Gamma_1(\mathfrak{n}) \)-level structure (resp. and a finite flat subgroup scheme \( C \) of \( A[p] \)) with compatible \( \mathcal{O}_F \)-action locally free of rank \( \sum_p (\mathcal{O}_F/p) \).

It follows from [27] and [8] that if \( \mathfrak{n} \) does not divide 2, or 3, \( Y_{\Gamma_1(n, t)} \) is a smooth scheme over \( \mathcal{O} \) of relative dimension \([F : Q]\). If \( \mathfrak{n} \) does divide 2, or 3, we let \( Y_{\Gamma_1(n, t)} \) denote the \( \mathcal{O} \)-scheme

\[
(\Gamma_1(n, t)/\Gamma_1(m, t)) \setminus Y_{\Gamma_1(m, t)}
\]

for an auxiliary ideal \( \mathfrak{m} \) of \( \mathcal{O}_F \) such that \( \mathfrak{n} \mid \mathfrak{m} \) and \( \Gamma_1(\mathfrak{m}) \) small enough, i.e., torsion-free.

Let \( Y_{\Gamma_1(n, t)} \) denote the fibre over \( k \) of \( Y_{\Gamma_1(n, t)} \); and let \( Y_{\Gamma_1(n, t) \cap \mathfrak{l}w} \) denote the fibre over \( k \) of \( Y_{\Gamma_1(n, t) \cap \mathfrak{l}w} \).

It is a well-known result of Deligne-Ribet that the fibre \( Y_{\Gamma_1(n, t)} \) is irreducible (Corollary 4.6 in [9]). It is a result of local model theory by Pappas that \( Y_{\Gamma_1(n, t) \cap \mathfrak{l}w} \) is normal (Corollary 2.2.3 in [25]).
Suppose that $p$ splits completely in $F$. In which case, the $p$-divisible group of a HBAV over the ring of integers of a finite extension of $\mathbf{Q}_p$ decomposes as the product of $[F : \mathbf{Q}]$ one-dimensional $p$-divisible groups, one for each prime $\mathfrak{p}$ of $F$ above $p$, and this allows us to define ‘Katz-Mazur-Drinfeld’ higher level structures at $p$ by defining level structures at $\mathfrak{p}$ on the ‘$\mathfrak{p}$-divisible group’ for each $\mathfrak{p}$.

**Definition.** Let $r$ be an integer $\geq 1$. Define $Y_{T_1(n)} \times \Gamma_1(p^r) \times \Gamma_1(p^r)$ to be the $\mathcal{O}$-scheme representing the functor which sends an $\mathcal{O}$-scheme $S$ to the set of isomorphism classes of the sextuples $(A, i, \lambda, \eta, C, \eta_{KM})$ over $S$ where $(A, i, \lambda, \eta)$ is a $t$-polarised HBAV over $S$ with $\Gamma_1(n)$-level structure, $C$ is a finite flat subgroup scheme of $A[p^r]$ locally free of finite rank $\mathcal{O}_F/p^r$ with compatible action of $\mathcal{O}_F$, and an $\mathcal{O}_F$-linear group homomorphism

$$\eta_{KM} : \mathcal{O}_F/p^r \to \text{Mor}(S, C) \subset \text{Mor}(S, A[p^r])$$

such that the image of $\eta_{KM}$ defines a ‘full set of sections’ in the sense of Katz-Mazur [17] (See 1.10.5 and 1.10.10 in [17]).

**Definition.** For every prime $\mathfrak{p}$ of $F$ above $p$, let $Y_{T_1(n)} \times \Gamma_1(p^r) \times \Gamma_1(p^r)$ denote the fine moduli space over $K$ of the septuples $(A, i, \lambda, \eta, C, \eta_{KM}, D_\mathfrak{p})$ where the sextuple $(A, i, \lambda, \eta, C, \eta_{KM})$ defines a point of $Y_{T_1(n)} \times \Gamma_1(p^r) \times \text{Spec} \mathcal{O}_K$ Spec $K$, and $D_\mathfrak{p}$ is finite flat subgroup scheme of $A[p]$ of rank $\vert \mathcal{O}_F/p \vert$ which has trivial intersection with $C$.

9 Compactification

By an unramified cusp $C$ of $Y_{T_1(n)}$ over $R$, we shall mean a pair of fractional ideals $M_1, M_2$ of $F$ such that $M_1M_2^{-1} \simeq t$ which comes equipped with

- an $\mathcal{O}_F \otimes \mathbb{Z} R$-linear isomorphism $\lambda : M_1^{-1} \otimes \mathbb{Z} R \simeq \mathcal{O}_F \otimes \mathbb{Z} R$;
- an $\mathcal{O}_F$-linear embedding $\eta : \mathcal{O}_F/n \to n^{-1}M_2^{-1}/M_1^{-1}$.

For brevity, let $M = M_1M_2$, $M^\vee = \text{Hom}_{\mathbb{Z}}(M, \mathbb{Z}) = \text{Hom}_{\mathcal{O}_F}(M, \mathcal{O}_F^\vee) \simeq \epsilon M_2^{-1} \mathbb{Z}$, and $M^\vee+ \subset M^\vee$ of the totally positive elements in $(\epsilon M_2^{-1} \mathbb{Z})^+$. Choose a rational polyhedral cone decomposition $\Sigma_C$ of $(M^\vee+ \otimes \mathbb{Z} R) \cup \{0\}$. For some cone $\sigma \subset (M^\vee+ \otimes \mathbb{Z} R)$, we let $\sigma^\vee \subset M \otimes \mathbb{Z} R$ denote the dual cone.

Let $S_n = \text{Spec} \mathbb{Z}[q^{-1}M^\vee]$ and $S_n \hookrightarrow S_{n, \sigma}$ denote the affine torus embedding (see Theorem 2.5 in [6]) corresponding to the cone $\sigma$ and let $S_{n, \sigma}^{\infty} = \text{Spf} \mathbb{Z}[q^{-1}M^\vee \otimes \sigma^\vee]$ denote the formal completion of $S_{n, \sigma}$ along the boundary $S_{n, \sigma}^{\infty} = S_{n, \sigma} - S_n$.

Let $T_{n, \sigma} = \text{Spec} \mathbb{Z}[q^{-1}M^\vee \otimes \sigma^\vee]$ and $T_{n, \sigma}^{0} = T_{n, \sigma} - S_{n, \sigma}^{\infty} = \text{Spec} \mathbb{Z}[q^{-1}M^\vee \otimes \sigma^\vee]$. The henselisation of $(S_{n, \sigma}, S_{n, \sigma}^{\infty})$ projects onto an affine étale scheme $U_{n, \sigma}$ over $S_{n, \sigma}$ which approximates $S_{n, \sigma}^{\infty}$ in the sense of Artin, and let $U_{n, \sigma}^{0} = U_{n, \sigma} \times_{T_{n, \sigma}} T_{n, \sigma}^{0}$. 
The Mumford construction applied to the $\mathcal{O}_F$-linear ‘period’ map $q : M_2 \to GL_1(U_{n,\sigma}) \otimes \mathcal{O}_F 1^{-1} M_1^{-1}$ gives rise to a semi-abelian scheme

$$\text{Tate}_{M_1, M_2}(q) \overset{\text{def}}{=} (GL_1 \otimes \mathcal{O}_F 1^{-1} M_1^{-1}) / q M_2$$

over the complete ring $U_{n,\sigma}$ with action of $\mathcal{O}_F$, whose pull-back, which we shall denote by $\text{Tate}_{M_1, M_2}(q)$ to $U_{n,\sigma}$, is naturally a $HBAV$, $t$-polarised

$$\text{Tate}_{M_1, M_2}(q) \otimes \mathcal{O}_F 1 M_2^{-1} \simeq (GL_1 \otimes \mathcal{O}_F 1^{-1} M_2^{-1}) / q M_1$$

$$\text{Tate}_{M_2, M_1}(q) \simeq \text{Tate}_{M_1, M_2}(q)$$

with level $\Gamma_1(\mathfrak{n})$-structure, and which gives rise to a map

$$U_{n,\sigma} \times \text{Spec} \mathcal{O} \to \text{Y}_{\Gamma_1(n, t)}.$$

We glue $\prod_{T/\mathbb{Z}} \prod_{\sigma \in \Sigma C} U_{n,\sigma} \times \text{Spec} \mathcal{O}$ along the map to get a toroidal compactification $X_{\Gamma_1(n, t)}$ over $\text{O}$ of $Y_{\Gamma_1(n, t)}$ ([26]). Similarly, one can define a compactification $X_{\Gamma_1(n, t) \text{tw}}$ over $\text{Y}_{\Gamma_1(n, t) \text{tw}}$ with its choice of a rational cone decomposition compatible with that of $X_{\Gamma_1(n, t)}$.

Let

$$\text{Tate}_{M_1, M_2, S}(q) \overset{\text{def}}{=} \text{Tate}_{M_1, M_2}(q) \times \text{Spec} \mathcal{Z}[[q^M, q^{-M}]] S$$

for a $\mathcal{Z}[[q^M, q^{-M}]]$-scheme $S$; it is $t$-polarised. Let $S$ be a $\mathcal{O} \otimes \mathcal{Z}[[q^M, q^{-M}]]$-scheme. Then there is a ‘connected-étale’ exact sequence

$$0 \to (GL_1 \otimes \mathcal{O}_F 1^{-1} M_1^{-1})[p] \to \text{Tate}_{M_1, M_2, S}(q)[p] \to (1/p^r) M_2 / M_2 \to 0$$

of $(\mathcal{O}_F / p^r)$-modules schemes over $S$.

**Lemma 11** Fix an integer $r \geq 1$. Let $S$ be a connected $\mathcal{O} \otimes \mathcal{Z}[[q^M, q^{-M}]]$-scheme. Suppose that $C$ is an $\mathcal{O}_F$-stable finite flat subgroup scheme of $\text{Tate}_{M_1, M_2, S}(q)[p^r]$ of order $|\mathcal{O}_F / p^r|$. Then for every $\tau = \tau_2$, there exists a unique pair of non-negative integers $\rho_{\tau, 1}, \rho_{\tau, 2}$ such that $\rho_{\tau, 1} + \rho_{\tau, 2} = r$ and such that

$$C_p \cap (GL_1 \otimes \mathcal{O}_F 1^{-1} M_1^{-1})[p^r] \simeq (GL_1 \otimes \mathcal{O}_F 1^{-1} M_1^{-1})[p^{\rho_{\tau, 1}}]$$

and the image of $C_p$ in $(1/p^r) M_2 / M_2$ is isomorphic to $p^{-\rho_{\tau, 2}} M_2 / M_2$. 

**Proof.** This is essentially Proposition 13.6.2 in [17]. □

By a cusp of $C$ of $Y_{\Gamma_1(n, t) \Gamma_1(p^r)}$ over $R$, we shall mean a pair of fractional ideals $M_1, M_2$ of $F$ such that $M_1 M_2^{-1} \simeq t$ which comes equipped with

- an $\mathcal{O}_F \otimes \mathcal{Z} R$-linear isomorphism $\lambda : M_1^{-1} \otimes \mathcal{O}_F \otimes \mathcal{Z} R \simeq \mathcal{O}_F \otimes \mathcal{Z} R$;

- an $\mathcal{O}_F$-linear embedding $\eta : \mathcal{O}_F / n \to n^{-1} M_2^{-1} / M_2^{-1}$. 

\begin{document}
\begin{thebibliography}{99}
\item []
\end{thebibliography}
\end{document}
• an $\mathcal{O}_F$-linear isomorphism $\eta_{KM}: \mathcal{O}_F/p^r \simeq p^{-r}M_2/M_2$.

Let $M = M_1M_2$ as above. Fix an integer $r \geq 1$. Suppose that $S$ is an $\mathcal{O} \otimes \mathbb{Z} [q^M, q^{-M}]$-scheme.

**Definition** Let $\zeta_r$ denote the image of 1 by

$$\zeta_{KM, \tau} : (\mathcal{O}_F/p^r) \simeq \mathbb{Z}^{-1}/p^r \simeq GL_1[p^r] \simeq (GL_1 \otimes \mathbb{Z}^{-1})[p^r]$$

and $\zeta_{r, \tau}$ denote its $\tau = \tau_p$ component. We often allow $\zeta_r$ and $\zeta_{r, \tau}$ to mean their images in $(GL_1 \otimes \mathbb{Z}^{-1})(S)$ and Tate$_{M_1, M_2, S}(q)(S)$.

Let $\eta^{\ast}_{\tau}$ denote the image of 1 by

$$\eta^{\ast}_{KM, \tau} : (\mathcal{O}_F/p^r) \simeq p^{-r}M_2/M_2 \rightarrow q^{-r}M_2/M_2$$

defining a point of Tate$_{M_1, M_2}(q)$ of exact order $|\mathcal{O}_F/p^r|$. Let $\eta^{\ast}_{r, \tau}$ denote its $\tau = \tau_p$ component.

**Lemma 12** Fix an integer $r \geq 1$. Let $S$ be a connected $\mathcal{O} \otimes \mathbb{Z} [q^M, q^{-M}]$-scheme. Suppose that $C$ is an $\mathcal{O}_F$-stable finite flat subgroup scheme of Tate$_{M_1, M_2, S}(q)[p^r]$ of order $|\mathcal{O}_F/p^r|$. Suppose that $C$ is of type $\rho = (\rho_{r, 1}, \rho_{r, 2})_\tau$. Let $P_{KM} \in C(S)$ denote a point of exact order $|\mathcal{O}_F/p^r|$. Then for every $\tau = \tau_p$, $P_{KM, \tau}$ is of the form $\zeta_{r, \tau}^{\sigma_1} \eta^{\ast}_{r, \tau}$ for a pair of integers $0 \leq \sigma_1 \leq \rho_{r, 1}$ and $0 \leq \sigma_2 \leq \rho_{r, 2}$ such that both $\sigma_1$ and $\sigma_2$ are coprime to $p$.

**Proof.** This is essentially 13.6.3 in [17].

10 **Generic fibres**

With $n$ fixed, for every integer $r \geq 1$, let $\overline{U}_r$ denote the quotient group of the totally positive units of $F'$ by the subgroup of elements which are squares of elements in $\mathcal{O}_F$ which are congruent to 1 mod $np^r$. If $r = 0$, we simply write $\overline{U}$.

Let $Y_{\Gamma_1(n)} \times X_{\Gamma_1(n)}$, $Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)}$ respectively denote the disjoint unions, $t$ ranging over $T$, of $Y_{\Gamma_1(n)}$, $Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)}$.

Let $X_{\Gamma_1(n)} \times X_{\Gamma_1(n)}$ respectively denote the generic fibres over $K$ of the $K$-schemes $Y_{\Gamma_1(n)}$, $X_{\Gamma_1(n)}$.

Let $X_{\Gamma_1(n)} \times X_{\Gamma_1(n)} \times Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)}$ respectively denote the toroidal compactifications of the $K$-schemes $Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)}$, $X_{\Gamma_1(n)} \times Y_{\Gamma_1(n)} \times Y_{\Gamma_1(n)}$. Let

Documenta Mathematica 18 (2013) 997–1038
\[ Y_{\Gamma_1(n)^\wedge} \cup X_{\Gamma_1(n)^\wedge} \cup Y_{\Gamma_1(n)^\wedge} \cap Iw \cup X_{\Gamma_1(n)^\wedge} \cap Iw \]

denote their disjoint unions over \( T \).

Finally, let \( Y_{\Gamma_1(n)^\rig} \cup X_{\Gamma_1(n)^\rig} \cup Y_{\Gamma_1(n)^\rig} \cap Iw \cup X_{\Gamma_1(n)^\rig} \cap Iw \) respectively denote the Raynaud rigid generic fibres of \( Y_{\Gamma_1(n)^\wedge} \cup X_{\Gamma_1(n)^\wedge} \cup Y_{\Gamma_1(n)^\wedge} \cap Iw \cup X_{\Gamma_1(n)^\wedge} \cap Iw \).

11 \( p \)-adic classical Hilbert modular forms

Suppose that \( (k = \sum_{\tau \in \text{Hom}(F,K)} k_{\tau}, w = \sum_{\tau \in \text{Hom}(F,K)} w_{\tau}) \in \mathbb{Z}^{\text{Hom}(F,K)} \times \mathbb{Z}^{\text{Hom}(F,K)} \) is such that \( w = 2w_{\tau} - k_{\tau} \) is independent of \( \tau \) (this is Taylor’s \( \mu \) in [38]).

For \( S \in \{ Y_{\Gamma_1(n),K} \cup Y_{\Gamma_1(n)^\rig,K} \cup Y_{\Gamma_1(n)^\rig \cap Iw,K} \} \), let \( \text{Lie}^\vee(A/S) \) (resp. \( H^1_{\text{dR}}(A/S) \)) denote the pull-back by the identity section of the sheaf of relative differentials of the universal HBAV \( A \) over \( S \) (resp. the higher direct image of the relative de Rham complex). By the decomposition,

\[ \mathcal{O}_F \otimes_{\mathbb{Z}} \mathcal{O} \simeq \bigoplus_{\tau \in \text{Hom}(F,K)} \mathcal{O}_\tau \]

where \( \mathcal{O}_\tau \) is \( \mathcal{O} \) into which \( F \) embeds by \( \tau \), we have

\[ \text{Lie}^\vee(A/S) = \bigoplus_{\tau \in \text{Hom}(F,K)} \text{Lie}^\vee(A/S)_{\tau}, \quad H^1_{\text{dR}}(A/S) = \bigoplus_{\tau \in \text{Hom}(F,K)} H^1_{\text{dR}}(A/S)_{\tau} \]

where \( \text{Lie}^\vee(A/S) \) and \( H^1_{\text{dR}}(A/S) \) are locally free sheaves of \( \mathcal{O}_S \)-modules of rank 1 and 2 respectively. Following Hida [14], let

\[ L_{(k,w)} = \bigotimes_{\tau \in \text{Hom}(F,K)} (\bigwedge H^1_{\text{dR}}(A/S)_{\tau})^{\otimes_{\mathbb{Z}} w_{\tau}} \otimes_{\mathcal{O}_S} (\text{Lie}^\vee(A/S)_{\tau})^{\otimes_{\mathbb{Z}} k_{\tau}} \]

If \( k \) is parallel, more precisely, if \( (k, w) = ((k, \ldots, k), (k/2, \ldots, k/2)) \), we will often write \( L_k \) for \( L_{(k,w)} \). We shall also let \( L_{(k,w)} \) denote its extension to the compactification.

Let \( \pi_1 \) (resp. \( \pi_{2,p} \)) denote the degeneracy map

\[ X_{\Gamma_1(n) \cap \Gamma_1(p'), Iw, K} \to X_{\Gamma_1(n) \cap \Gamma_1(p'), K} \]

defined, on the non-cuspidal points, by

\[ (A, i, \lambda, \eta, C, \eta_{\text{KM}}, D_p) \mapsto (A, i, \lambda, \eta, C, \eta_{\text{KM}}) \]

(resp. \( (A/D_p, (i \mod D_p), (\lambda \mod D_p), (\eta \mod D_p), (\eta_{\text{KM}} \mod D_p)) \)).
12 Canonical subgroups for one-dimensional formal groups

Let $L$ be a finite extension of $K$, and let $\mathrm{val}_L$ be a valuation on $L$ normalised so that $\mathrm{val}_L(p) = 1$. Let $G$ be a one-dimensional principally polarised $p$-divisible/Barsotti-Tate group over $\mathcal{O}_L$.

**Definition.** The identity component $G^\wedge$ of $G$ is a one-dimensional formal group, and define $\mathrm{Ha}(G)$ to be $\mathrm{val}_L(a)$ for $a$ as defined in Proposition 3.6.6, [16] (see also [29]).

By definition, $G$ is ordinary if and only if $\mathrm{Ha}(G) = 0$.

Let $C$ be a finite flat subgroup scheme of $G[p]$ of order $p$.

**Definition.** Define deg$(G, C)$ to be $1 - \mathrm{val}_L(\mathrm{Ann}(\mathrm{coker}(\mathrm{Lie}^\vee(G/C) \to \mathrm{Lie}^\vee(G))))$.

It follows immediately from the definition that $\text{deg}(G, C)+\text{deg}(G/C, G[p]/C) = 1$.

Suppose that $\text{deg}(G, C) < p/(p + 1)$. Then there exists a canonical subgroup $H(G)$ of $G$. If $C = H(G)$, then $\text{deg}(G, C) = \mathrm{Ha}(G)$. To see this, note that $H(G)(L)$ consists of 0 and $p - 1$ points $P$ of the formal group $G^\wedge$ of valuation $(1 - \mathrm{Ha}(G))/(p - 1)$ (Theorem 3.10.7, [16]). Since $\text{deg}(G, C) = 1 - \prod_P \mathrm{val}(P)$ (Lemma 1.3 [24]), $\text{deg}(G, C) = \mathrm{Ha}(G)$.

**Lemma 13** Let $r$ be a rational number $< p/(p + 1)$. Suppose that $G$ is not ordinary. Then

$$\{(G, C) \mid \mathrm{Ha}(G) \leq r\}$$

divides into two disjoint subsets, namely

$$\{(G, C) \mid C = H(G) \text{ and } \text{deg}(G, C) \in (0, r]\}$$

and

$$\{(G, C) \mid C \neq H(G) \text{ and } \text{deg}(G, C) \in [1 - r/p, 1)\}.$$ 

On the other hand,

$$\{(G, C) \mid \mathrm{Ha}(G/C) \leq r\}$$

divides into two disjoint subsets, namely

$$\{(G, C) \mid \text{deg}(G, C) \in (0, r/p], C = H(G), \text{ and } \mathrm{Ha}(G) < 1/(p + 1) \}$$

$$\cup \{(G, C) \mid \text{deg}(G, C) \in (0, r/p], C \neq H(G), \text{ and } \mathrm{Ha}(G) < p/(p + 1) \}$$

and

$$\{(G, C) \mid \text{deg}(G, C) \in [1 - r, 1), C = H(G), 1/(p + 1) < \mathrm{Ha}(G) < p/(p + 1)\}$$

$$\cup \{(G, C) \mid \text{deg}(G, C) \in [1 - r, 1), C \neq H(G), \mathrm{Ha}(G) \geq p/(p + 1)\}.$$
Proof. This follows from canonical subgroup theorem in [29]. □

Fix an integer $n \geq 1$ and suppose furthermore that $\deg(G, C) \leq p^{1-n}/(p+1) < p/(p+1)$. Then define subgroup $H_n = H_n(G)$ of $G$ order $p^n$ inductively as follows: If $n = 1$, set $H_1 = D$. If $n > 1$, then let $H_n$ to be the pre-image by the map $G \to G/H(G)$ of $H_{n-1}(G/H(G)) \subset G/H(G)$.

**Proposition 14** Suppose that one-dimensional principally polarised $p$-divisible group $G$ over $\mathcal{O}_L$ has a subgroup $H_n(G)$ as defined above. Suppose that $m \geq 1$ is an integer. Suppose that $C_m$ is a subgroup of $G$ of order $p^m$ such that $H_n(G) \cap C_m = \{0\}$, and suppose that $D_m + n$ is a cyclic subgroup of $G$ of order $p^{m+n}$ such that $H_n(G) \subseteq D_{m+n}$. Then $\deg(G/C_m) < p^{1-(m+n)/(p+1)}$ and $G/C_m$ has the subgroup $H_{m+n}(G/C_m)$. Indeed, $H_{m+n}(G/C_m) = (D_{m+n} + C_m)/C_m$.

**Proof.** This can be proved as in Proposition 3.5 in [3]. □

**13 $p$-adic overconvergent Hilbert modular forms**

Let $X_{\Gamma_1(n), K}^{an}, X_{\Gamma_1(n), K}^{an, 1}$, respectively denote the rigid analytic spaces in the sense of Tate ([2]) associated to the $K$-schemes $X_{\Gamma_1(n), K}, X_{\Gamma_1(n), K}^{an, 1}$. \[X_{\Gamma_1(n), K}^{an} \supseteq X_{\Gamma_1(n), K}^{an, 1}.\]

Given a closed point of $Y_{\Gamma_1(n)}^{rig}$, it corresponds to a point $(A, \lambda, \eta)$ defined over the integer $\mathcal{O}_L$ of a finite extension $L$ of $K$. We then define $\deg_{\tau}(A)$, for $\tau = \tau_p$ for a place $p$ of $F$ above $p$, to be ‘$\deg$’ as in the previous section with the (one-dimension) Barsotti-Tate group of $p$-power torsions of $A$ in place of ‘$G’$.

The $\mathcal{O}$-scheme $X_{\Gamma_1(n)}$ is of finite type, hence $X_{\Gamma_1(n)}^{rig}$ is quasi-compact. There exists a finite many sufficiently small affine formal schemes $U^\wedge$ such that their generic fibres $U^{rig}$ form an admissible covering of $X_{\Gamma_1(n)}^{rig}$. Let $U^{\wedge}_{\text{good}}$ denote the smooth formal scheme $U^\wedge \cap Y_{\Gamma_1(n)}^{rig}$ and let $i : U^{\wedge}_{\text{good}} \hookrightarrow U^\wedge$. On each $U^{\wedge}_{\text{good}}$, there is a function whose corresponding rigid function has its valuation $\text{deg}$; indeed, apply the construction to the formal completion of the ‘universal’ semi-abelian scheme over $X_{\Gamma_1(n)}$ along the underlying scheme of $U^{\wedge}_{\text{good}}$. We may think of the function on $U^{\wedge}_{\text{good}}$ as a lift of the Hasse invariant at $p$, and it follows from Kocher’s principle that $i_*\mathcal{O}_{U^{\wedge}_{\text{good}}} = \mathcal{O}_{U^\wedge}$, i.e., the function extends to $U^\wedge$. The valuation of its induced function on the generic fibre $U^{rig}$ extends the function on $U^{rig}_{\text{good}}$. Glue these functions on $U^{rig}$'s, there is a rigid function on $X_{\Gamma_1(n)}^{rig} \simeq X_{\Gamma_1(n)}^{an}$ that defines $\text{deg}$.

**Definition.** If $I \subseteq [0, 1)$ is a closed, open, or half open interval with endpoint in $Q$, define the rigid space $X_{\Gamma_1(n), K}^{an, I} = \coprod_{r} X_{\Gamma_1(n, r), K}^{an}$ to be the admissible open set of points whose degrees are all in the range $I$. 
For every $t$, $X_{\Gamma_1(n,t),K}^\text{an} I$ is connected; this follows from the fact that $X_{\Gamma_1(n,t),K}^{\text{rig}}$ is connected (since $X_{\Gamma_1(n,t)}$ is irreducible) and its ordinary locus is open, dense, and connected.

Similarly, given a closed point of $Y_{\Gamma_1(n)\cap \Iw}^{\text{rig}}$, it corresponds to a point $(A,\lambda,\eta,C)$ defined over the integer $\mathcal{O}_L$ of a finite extension $L$ of $K$. Let $B = A/C$ and $S = \text{Spec} \mathcal{O}_L$; let $\text{val}_S$ denote the valuation on $L$ normalised such that $\text{val}_S(p) = 1$.

Then the $\mathcal{O}_F$-equivariant map of $\mathcal{O}_S$-modules $\text{Lie}^\vee(B/S) \rightarrow \text{Lie}^\vee(A/S)$ decomposes into $\text{Lie}^\vee(B/S)_\tau \rightarrow \text{Lie}^\vee(A/S)_\tau$ for every $\tau \in \text{Hom}(F,K)$, and, for the unique prime $p$ of $F$ above $p$ corresponding to $\tau$, let $\text{deg}_p((A,C))$ denote $1 - \text{val}_S(\text{Ann}(\text{Coker}(\text{Lie}^\vee(B/S)_\tau \rightarrow \text{Lie}^\vee(A/S)_\tau)))$. Applying the construction to the universal HBAV over $Y_{\Gamma_1(n)\cap \Iw}^{\text{rig}}$, we locally have functions on $X_{\Gamma_1(n)\cap \Iw}^\text{an}$ whose valuations define the degrees. As for $\text{deg}(A)$, Kocher’s principle allows us to extend the function to $X_{\Gamma_1(n)\cap \Iw}^{\text{rig}} \simeq X_{\Gamma_1(n)\cap \Iw}^\text{an}$.

**Definition.** If $S_1$ and $S_2$ are disjoint subsets of $\text{Hom}(F,K)$ and if $I, I_1, I_2 \subseteq [0,1]$ are closed, open, or half open intervals in $\mathbb{Q}$, define the rigid space $(X_{\Gamma_1(n)\cap \Iw,K}^\text{an} I)_{S_1 S_2}$ to be the admissible open set of points whose degree at $\tau \in \text{Hom}(F,K) - S_1 - S_2$ (resp. $S_1$, resp. $S_2$) is in the range $I$ (resp. $I_1$, resp. $I_2$).

**Definition.** Let $\pi_1$ (resp. $\pi_{2,p}$) denote the degeneracy map $X_{\Gamma_1(n)\cap \Iw,K}^\text{an} \rightarrow X_{\Gamma_1(n),K}^\text{an}$ which, on the non-cuspidal points, is defined by

$$(A, i, \lambda, \eta, C) \mapsto (A, i, \lambda, \eta)$$

(resp. $(A, i, \lambda, \eta, C) \mapsto (A/C_p, (i \mod C_p), (\lambda \mod C_p), (\eta \mod C_p))$)

**Definition.** Let $\pi$ denote the degeneracy map $X_{\Gamma_1(n)\cap \Gamma_{1'(p')},K}^\text{an} \rightarrow X_{\Gamma_1(n)\cap \Iw,K}^\text{an}$ which, on the non-cuspidal points, is defined by

$$(A, i, \lambda, \eta, \eta_{\text{KM}}) \mapsto (A/pP_{\text{BM}}(\eta_{\text{KM}}), (i \mod (pP_{\text{BM}})), (\lambda \mod (pP_{\text{BM}})), (\eta \mod (pP_{\text{BM}}))).$$

*Documenta Mathematica* 18 (2013) 997–1038
where by $P_{\eta_{K^M}}$, we mean the image of 1 by $\eta_{K^M}$.

**Definition.** Define $(X_{\Gamma_1(n)\cap \Gamma_1(p^r),K})_1 I_1 S_1 I_2 S_2$ to be the preimage by $\pi$ of $(X_{\Gamma_1(n)\cap \Gamma_1(p^r),K})_1 I_1 S_1 I_2 S_2$.

For $0 \leq r \leq p/(p + 1)$, it follows from the previous section that

$$\pi^{-1}_1(X_{\Gamma_1(n),K}[0,r]) \cong X_{\Gamma_1(n)\cap \Gamma_2,K}[0,r] \prod X_{\Gamma_1(n)\cap \Gamma_2,K}[1-r/p,1];$$

and for $\tau = \tau_p$

$$\pi^{-1}_{2,p}(X_{\Gamma_1(n),K}[0,r]) \cong (X_{\Gamma_1(n)\cap \Gamma_2,K}[0,r],0,r/p] \prod (X_{\Gamma_1(n)\cap \Gamma_2,K}[0,r],1-r,1].$$

The theory of canonical subgroups provides rigid sections:

$$\pi_1 : X_{\Gamma_1(n)\cap \Gamma_2,K}[0,r] \rightarrow X_{\Gamma_1(n),K}[0,r]$$

and

$$\pi_2 : X_{\Gamma_1(n)\cap \Gamma_2,K}[1-r,1] \rightarrow X_{\Gamma_1(n),K}[0,r].$$

On the other hand,

$$\pi_1 : X_{\Gamma_1(n)\cap \Gamma_2,K}[1-r/p,1] \rightarrow X_{\Gamma_1(n),K}[0,r]$$

is finite flat of degree $|O_F/p|$, and

$$\pi_2 : X_{\Gamma_1(n)\cap \Gamma_2,K}[0,r/p] \rightarrow X_{\Gamma_1(n),K}[0,r]$$

is finite flat of degree $|O_F/p|$. 

Hida [14] proves (Theorem 5.6 in [14]) that, for a character $\psi : Fr_{n^\infty} \rightarrow K^\times$ which factors through $I_{np^r}$ and $k \geq 2$, an element $F_H : h_0^0(n) \rightarrow L$ of $\mathcal{S}_{\mathcal{O}}^0(n) \otimes L$ defines, modulo($\ker(\epsilon \circ \text{Art})^{k-2}\psi$), a cusp eigenform of weight $k$ and level $\Gamma_1(np^r)$ which is an eigenform with its $T_m$-eigenvalue $F_H(T_m)$ mod $(\ker(\epsilon \circ \text{Art})^{k-2}\psi)$ and $S$ acting by $(\epsilon \circ \text{Art})^{k-2}\psi$. Indeed $I_{np^r}$-action defines the character of $F_H$ mod $(\ker(\epsilon \circ \text{Art})^{k-2}\psi)$, i.e.,

$$(F_H \mod (\ker(\epsilon \circ \text{Art})^{k-2}\psi))(\cdot)$$

$$= \psi_F \mod (\ker(\epsilon \circ \text{Art})^{k-2}\psi)(\psi_T \circ \epsilon^{2-k})$$

where $\psi_F$ is the composite $\text{Tor}_{np^r} \rightarrow I_{np^r} \rightarrow h_0^0(n)$ followed by $F_H : h_0^0(n) \rightarrow L$ and $\psi_T$ is the Teichmuller character’, the projection from $\mathbb{Z}_p^\times$ to its torsion subgroup of finite order. We shall prove that the specialisation $F_H \mod \ker(\epsilon \circ \text{Art})^{k-2}\psi$ defines a $p$-ordinary overconvergent eigenform of weight $k$ and of level $\Gamma_1(np^r)$ for any $k = 1$. 

**Documenta Mathematica** 18 (2013) 997–1038
For $\epsilon$ such that $0 \leq \epsilon < 1/(p^{r-2}(p + 1))$, the theory of canonical subgroups in [29] (see also Proposition 2.3.1 and 2.4.1 in [21]) shows that $U_p \overset{\text{def}}{=} \prod_p U_p$ defines a completely continuous endomorphism on $H^0(X_{\Gamma_1(n) \cap \Gamma_1((p'),K)[0,\epsilon],L_k(\text{cusps})})$ over $\mathbb{Q}$, where $X_{\Gamma_1(n) \cap \Gamma_1((p'),K)[0,\epsilon],L_k(\text{cusps})}$ is the pre-image by the forgetful morphism of $X_{\Gamma_1(n),K}[0,\epsilon]$. We remark that, when $F = \mathbb{Q}$, this is proved in [4] Lemma 2.3 as a result of calculations with $q$-expansions.

By Serre’s theory [31], there is an idempotent $e$ commuting with $U_p$ by which we may write

$$H^0(X_{\Gamma_1(n) \cap \Gamma_1((p'),K)[0,\epsilon],L_k(\text{cusps})}) = eH^0(X_{\Gamma_1(n) \cap \Gamma_1((p'),K)[0,\epsilon],L_k(\text{cusps})})$$

$$+ (1 - e)H^0(X_{\Gamma_1(n) \cap \Gamma_1((p'),K)[0,\epsilon],L_k(\text{cusps})})$$

where $eH^0(X_{\Gamma_1(n) \cap \Gamma_1((p'),K)[0,\epsilon],L_k(\text{cusps})})$ is finite-dimensional $K$-vector space and all the generalised eigenvalues of $U_p$ are units, while $U_p$ is topologically nilpotent on the complement. It is well-known that $e = eH^0(X_{\Gamma_1(n) \cap \Gamma_1((p'),K)[0,\epsilon],L_k(\text{cusps})})$.

**Lemma 15** For any integer $k$, the $p$-adic eigenform $F_H \mod (\ker(\epsilon \circ \text{Art})^{k-2})$ as above is overconvergent of weight $k$ and of level $\Gamma_1(np^r)$.

**Proof.** This can be proved as in Lemma 1 in [5]; replace the Eisenstein series ‘$E$’ of weight $(p - 1)$ therein by the pull-back to $X_{\Gamma_1(n) \cap \Gamma_1((p'),K)$ of a characteristic zero lifting of a sufficiently large power of the Hasse invariant. $\square$

It follows from the theorem in the previous section that, given a $p$-adic representation

$$\rho : \text{Gal}(\overline{F}/F) \to GL_2(\mathcal{O})$$

as in the main theorem, there are

1. a finite soluble totally real field extension $F_\Sigma \subset \overline{F}$ of $F$ in which $p$ splits completely,
2. a finite set $S = \Sigma \prod S_{\Sigma,p} \prod S_{\Sigma,\infty}$ of places in $F_\Sigma$, where $S_{\Sigma,p}$ denotes the set of places of $F$ above $p$ and $S_{\Sigma,\infty}$ denotes the set of infinite places of $F_\Sigma$,
3. an ideal $\mathfrak{n}$ of $\mathcal{O}_F$ divisible by $\mathfrak{n}_\Sigma = \prod_{\sigma \in \Sigma} \mathfrak{q}_\sigma$,
4. $2^{|S_{\Sigma,p}|}$ characters $\chi_P : \text{Gal}(\overline{F}/F_\Sigma) \to \mathcal{O}^\times$ of finite order and $2^{|S_{\Sigma,p}|}$ weight one $p$-ordinary overconvergent cuspidal Hilbert modular eigenforms $f_P$ of ‘tame level’ $\mathfrak{n}$, one for every subset $P$ of $S_{\Sigma,p}$, such that:
• $f_P$ is the weight one specialisation of the $A$-adic companion form $F_{\text{Hida}, P} : h^0_{\text{Iw}}(n) \to K$, with character $\psi_P = \psi_P^{\text{Art}} \psi_{P, S, \psi_P}$ of $(\mathcal{O}_{F_\psi}/n)^x \times (\mathcal{O}_{F_\psi}/p)^x$.

• The Galois representation $\rho_P$ associated to $f_P$ is $\rho| \text{Gal}(\overline{\mathbb{F}}_p/F_\psi) \otimes \chi_P^{-1}$, $\rho_P$ is unramified outside $S$ and ordinary at every place in $S_{\Sigma, P}$, and the $f_P$'s are 'in companion' in the sense that

- $c(\mathcal{O}_{F_\psi}, f_P) = 1$, and $c(m, f_P) = 0$ if $m$ is not coprime to $n$;
- $c(q, f_P) = \text{tr} \rho(Frob_q)/\chi_P(Frob_q)$ for every prime ideal $q$ not dividing $\eta_P$;
- For $p$ in $P$, $c(m, f_P)(\chi_P \circ \text{Art})(m) = c(m, f_{P-\{p\}})(\chi_{P-\{p\}} \circ \text{Art})(m)$ for every ideal $m$ coprime to $\eta_P$;
- For $p$ in $P$, the character of $f_P$ at $p$ is $\chi_{P-\{p\}}^{-1}$ while for $p \in S_{\Sigma, P} - P$, the character of $f_P$ at $p$ is $\chi_{P-\{p\}}^{-1}$;
- For $p$ in $P$, $(\chi_P^{S_{\Sigma, P}} \circ \text{Art})(p) = (\psi_P^{S_{\Sigma, P}} \circ \text{Art})(p)$;

• For a place $p$ of $P$, the $U_p$-eigenvalue of $f_P$ is $(\chi_P, 1^{-1})(\text{Frob}_p)$ while for $p$ in $S_{\Sigma, P} - P$, the $U_p$-eigenvalue of $f_P$ is $(\chi_P, 2^{-1})(\text{Frob}_p)$.

14 Analytic Continuation of Overconvergent Eigenforms

Fix $\tau = \tau_P$ throughout the section (except the last two assertions).

**Definition.** Fix $t$. For brevity, let $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau}$ denote $(X_{\Gamma_1(n, t) \cap \text{Lw}_t, K}[0, r])/[0,1])_\tau$; and for an integer $n \geq 0$, let $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau, n}$ denote $(X_{\Gamma_1(n, t) \cap \text{Lw}_t, K}[0, r])/[0,1 - 1/p^n(p + 1)]_\tau$.

Let $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau}$ (resp. $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau, n}$) denote the disjoint union over $T$ of $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau}$ (resp. $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau, n}$).

**Proposition 16** For every integer $n \geq 0$, $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau, n}$ is an admissible open subset of $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau}$, and the $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau, n}$ form an admissible covering of $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau}$. For every $t$ and every $n \in \mathbb{Z}_{\geq 0}$, $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau, n}$ is connected.

**Proof.** Clear. $\square$

**Definition.** Let $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau}$ (resp. $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau, n}$) denote the pre-image by the degeneracy morphism

$$\pi : X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau} \to X_{\Gamma_1(n, t) \cap \text{Lw}_t, K}$$

of $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau}$ (resp. $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau, n}$).

Let $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau}$ (resp. $X_{\Gamma_1(n) \cap \text{Lw}_t, \tau, n}$) denote the disjoint union over $T$ of $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau}$ (resp. $X_{\Gamma_1(n, t) \cap \text{Lw}_t, \tau, n}$).

**References**
PROPOSITION 17 For every integer $n \geq 0$, $X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau,n}$ is an admissible open subset of $X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau}$, and the $X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau,n}$ form an admissible covering of $X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau}$. For every $t$ and an integer $n \geq 0$, $X_{\Gamma_1(n,t)\cap \Gamma_1(p^r),\tau,n}$ is connected.

Proof. Analogous to the proposition above. □

COROLLARY 18 We have $\pi_{1}^{-1}(X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau,n+1}) \subset \pi_{2}^{-1}(X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau,n})$.

Proof. This follows from [29]. □

Let $(\text{Tate}_{M_1,M_2}(q) = (GL_1 \otimes \mathbb{Z} \mathcal{O}^{-1}M_1^{-1})/q^M_2, i, \lambda, \eta, \eta_{\text{KM}} : 1 \rightarrow \zeta) \otimes \mathbb{Z}(\mathfrak{q}^M_2))$ for the pair $M_1, M_2$ of the fractional ideals such that $M_1M_2^{-1} \cong \mathcal{O}$ be a family of HBVs around a cusp of $X_{\Gamma_1(n,t)\cap \Gamma_1(p^r),\tau}$. Choose (non-canonically) once for all a basis of the pull-back by Max $(\mathcal{O} \otimes \mathbb{Z}(\mathfrak{q}^M_2)) \rightarrow X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau}$ of the line bundle $\mathcal{L}_k$, since a subgroup of $\text{Tate}_{M_1,M_2}(q)[p]$ of order $|\mathcal{O}_F/p|$, disjoint from $\eta_{\text{r}}$, is of the form $\zeta \eta + q^M_2$ where $\zeta$ ranges over the $|\mathcal{O}_F/p|$ points of $(GL_1 \otimes \mathbb{Z} \mathcal{O}^{-1}M_1^{-1})/p^M_2 \otimes \mathbb{Z}$ and $\eta_{\text{r}}^{\mathcal{L}_k} \in q^M_2\mathcal{M}_2$. $U_p(f)(\text{Tate}_{M_1,M_2}(q), i, \lambda, \eta, \eta_{\text{KM}})$ is:

$$|\mathcal{O}_F/p|^{-1} \sum_{\zeta} f((GL_1 \otimes \mathbb{Z} \mathcal{O}^{-1}M_1^{-1})/q^M_2)/((\zeta \eta))$$

$$= |\mathcal{O}_F/p|^{-1} \sum_{\zeta} f((GL_1 \otimes \mathbb{Z} \mathcal{O}^{-1}M_1^{-1})/((\zeta \eta)q^M_2))$$

$$= |\mathcal{O}_F/p|^{-1} \sum_{\zeta} |\mathcal{O}_F/t|^{-1} \sum_{\nu \in (p-1)^+} c(pM^{-1} \nu, f)(\zeta \eta)^\nu$$

$$= |\mathcal{O}_F/p|^{-1} |\mathcal{O}_F/t|^{-1} \sum_{\nu \in (p-1)^+} \sum_{\zeta} c(pM^{-1} \nu, f)\eta^{\nu}$$

where $q_\eta$ denotes a representative in $q^M_2$ of the class $\eta \in q^M_2\mathcal{M}_2 = q^M_2\mathcal{M}_2\otimes \mathcal{M}_2$ defined earlier; and $t_p$ represents the class of $pt \simeq pM_2^{-1}$.

THEOREM 19 Suppose that $f \in H^0((X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau})_r[0,\varepsilon], L_k)$ is an eigenform for $U_p$ with non-zero eigenvalue, then $f$ extends to $X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau} = (X_{\Gamma_1(n)\cap \Gamma_1(p^r),\tau})_r[0,1]$.

DEFINITION. Let

$$X_{\Gamma_1(n,t)\cap \Gamma_1(p^r),\tau}^{[0]} \subset X_{\Gamma_1(n,t)\cap \Gamma_1(p^r),\tau}^{[1]} \subset \cdots \subset X_{\Gamma_1(n,t)\cap \Gamma_1(p^r),\tau}^{[r-1]}$$

$$X_{\Gamma_1(n,t)\cap \Gamma_1(p^r),\tau}^{[s]}$$

denote the admissible open subsets of $X_{\Gamma_1(n,t)\cap \Gamma_1(p^r),\tau}$ defined in such a way that the non-cuspidal $S$-points of $X_{\Gamma_1(n,t)\cap \Gamma_1(p^r),\tau}^{[s]}$ parameterises $(A/S, i, \lambda, \eta)$.
equipped with a point \( P_{\text{HCM}} \) of exact of order \( \sum_{\mathfrak{p}} |O_F/\mathfrak{p}^r| \) where \( A/S \) is either \( p \)-non-ordinary, or it is \( p \)-ordinary and \( H_{r-s}(A[p]) \) equals the subgroup generated by \( |O_F/p| P_{\text{HCM}} \).

For every \( 0 \leq s \leq r - 1 \), \( X_{\Gamma_1(n),\Gamma_1(p'),\tau}^{[s]} \) is connected since it is the pre-image of a closed subset of the union of irreducible components intersecting precisely at the \( p \)-non-ordinary locus of \( X_{\Gamma_1(n),\Gamma_1(p'),\tau} \).

**Theorem 20** If \( r \) is an integer \( \geq 2 \) and suppose that \( f \in H^0(X_{\Gamma_1(n),\Gamma_1(p'),\tau}, L_k) \) is an eigenform for \( U_\mathfrak{p} \) with non-zero eigenvalue. Then \( f \) extends to \( X_{\Gamma_1(n),\Gamma_1(p'),\tau}^{[r-1]} \).

**Proof.** This can be proved as Lemma 6.1 in [3]. \( \square \)

**Corollary 21** If \( f \in H^0(X_{\Gamma_1(n),\Gamma_1(p'),\tau}, L_k) \) for some \( 0 < \epsilon < 1 \) is an eigenform for every \( U_\mathfrak{p}, \mathfrak{p} \), with non-zero eigenvalue, then \( f \) extends to \( H^0(X_{\Gamma_1(n),\Gamma_1(p'),\tau}, L_k) \). Similarly, if \( f \in H^0(X_{\Gamma_1(n),\Gamma_1(p'),\tau}, L_k) \) for some \( 0 < \epsilon < 1 \) is an eigenform for every \( U_\mathfrak{p}, \mathfrak{p} \), with non-zero eigenvalue, then \( g \) extends to \( H^0(X_{\Gamma_1(n),\Gamma_1(p'),\tau}, L_k) \).

15 GLUING EIGENFORMS

15.1 THE IWAHORI CASE

**Definition.** For every subset \( P \) of the set of places of \( F \) above \( p \), let \( w_P \) denote the automorphism of \( X_{p_1(n),\Gamma_1(p',\tau)}^{\text{an}} \) defined by a composite (independent of ordering) of the \( w_\mathfrak{p} \) for all \( \mathfrak{p} \) in \( P \).

**Theorem 22** For every subset \( P \) of the set \( S = S_P \) of places of \( F \) above \( p \), suppose \( f_\mathfrak{p} \in H^0(X_{\Gamma_1(n),\tau}, L_k) \) is an overconvergent modular form of parallel weight \( k = \sum_{\tau \in \text{Hom}(F,K)} k_\tau \in \mathbb{Z} \) and of level \( \Gamma_1(n) \). Assume furthermore that

- the Fourier coefficient \( c(f_\mathfrak{p}, \mathcal{O}_F) = 1 \);
- for every place \( \mathfrak{p} \) of \( F \) above \( p \), there exist \( \alpha_\mathfrak{p}, \beta_\mathfrak{p} \in K \) such that \( \alpha_\mathfrak{p} \neq \beta_\mathfrak{p} \) and such that, for every \( P, f_\mathfrak{p} \) is an eigenform for \( U_\mathfrak{p} \) with eigenvalue \( \alpha_\mathfrak{p} \) if \( \mathfrak{p} \in P \) whilst with eigenvalue \( \beta_\mathfrak{p} \) if \( \mathfrak{p} \notin P \);
- for all ideal \( \mathfrak{m} \) of \( \mathcal{O}_F \) coprime to \( p \), \( c(\mathfrak{m}, f_\mathfrak{p}) \) are equal for every \( P \).

Then every \( f_\mathfrak{p} \) is a classical Hilbert modular eigenform of weight \( k \) and of level \( \Gamma_1(n) \cap \text{Iw} \).

**Proof.** By the isomorphism

\[
\pi_1: X_{\Gamma_1(n),\text{Iw},K}^{\text{an}}[0,r] \xrightarrow{\sim} X_{\Gamma_1(n),K}^{\text{an}}[0,r]
\]

for \( r < p/(p+1) \) given by the canonical subgroups theorem [29], we may think of \( f_\mathfrak{p} \) as an element of \( H^0(X_{\Gamma_1(n),\text{Iw},K}^{\text{an}}[0,r], L_k) \). It follows from results in [29]
that $\pi^*_1 F_P$ extends to a section over $X^{an}_{\Gamma_1(n)\cap \Iw, K}[0,1)$. For brevity, we shall only show that $f_P$, with $P$ the (full) set $S$ of places of $F$ above $p$, is classical; the general case follows by changing the roles of $\alpha_p$ and $\beta_p$.

Choose a rational number $r \in \mathbb{Q}$ with $1/2 < r < p/(p+1)$. Suppose that $f_S$ extends to a section of $L_k$ over $(X_{\Gamma_1(n)\cap \Iw, K}[0, r][0,1]_{S-P}$ for some $P \subseteq S$. Fix a prime $p \in P$. It suffices to show that $f_S$ extends to $(X_{\Gamma_1(n)\cap \Iw, K}[0, r][0,1]_{S-(p-P)})$.

For $f \in H^0(X_{\Gamma_1(n)\cap \Iw, K}[0, r], L_k)$ and for every subset $Q \subseteq S-P$, let $f^Q$ denote the restriction of $f$ to $(X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q[0, r]_{(S-P)-Q}$ by the map $\pi_1 \circ w_Q$ which defines an isomorphism

$$(X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q[0, r]_{(S-P)-Q} \simeq X_{\Gamma_1(n), K}[0, r]$$

The pre-image by $\pi_{2,p} \circ w_Q$ of $(X_{\Gamma_1(n), K}[0, r])(0, r)_p$ is the union of two components

$$(X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q(1-r,1)_p \coprod (X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q(0, r/p)_p$$

and it induces an isomorphism

$$(X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q(1-r,1)_p \simeq (X_{\Gamma_1(n), K}[0, r])(0, r)_p$$

on the one component and a finite flat morphism of degree $|\mathcal{O}_F/p|

(X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q(0, r/p)_p \longrightarrow (X_{\Gamma_1(n), K}[0, r])(0, r)_p$$

on the other.

We are going to glue $f_S$ and $f_{S-(p)}$: more precisely glue $f^Q_S$ and $f^Q_{S-(p)}$.

Let $F$ denote the section

$$(\alpha_p f^Q_S - \beta_p f^Q_{S-(p)})/(\alpha_p - \beta_p) \in H^0((X_{\Gamma_1(n), K}[0, r])(0, r)_p, L_k)$$

and $G$ denote the section

$|\mathcal{O}_F/p|(f^Q_S - f^Q_{S-(p)})/(\alpha_p - \beta_p) \in H^0((X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q(0, r/p)_p, L_k)$$

Since one can show readily the $q$-expansions of $\pi^*_2 F$ and $G$ are equal at around $C = (\text{Tate}(M_\mu(q),\ldots,\langle q \rangle))$, we shall glue $\pi^*_2 F$ and $G$ at $(X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q(0, r/p)_p$ to construct an extension $\tilde{F}^\prime$ of $F$ to a section over $(X_{\Gamma_1(n), K}[0, r)[0,1]_p$; this extension constructs an extension of $f^Q_S$ (and $f^Q_{S-(p)}$) to $(X_{\Gamma_1(n)\cap \Iw, K}[0, r][1-r,1]_Q[0, r]_{(S-P)-Q}[0,1]_p$ and therefore to $(X_{\Gamma_1(n)\cap \Iw, K}[0, r][0,1]_{S-(p-P)}$ (by assumption, there is an extension 'over

$[0,1]$ at $S-P$)
Gluing of \( \pi_{2,p}^* F \) and \( G \) is analogous to [3] since we have a commutative diagram

\[
\begin{array}{ccc}
(X_{\Gamma_1(n)\cap \text{Iw},[0,r]}[[1-r]Q(0,r/p)_p & \to & (X_{\Gamma_1(n)\cap \text{Iw},K}[0,r]_Q[[1-r]Q(0,1)_p) \\
(X_{\Gamma_1(n),K}[0,r])_p & \to & (X_{\Gamma_1(n),K}[0,r])_p \\
\end{array}
\]

where the vertical arrows are \( \pi_{2,p} \circ w_Q \) but of degree \( |O_F/p| \) on the left and \( 1 + |O_F/p| \) on the right. \( \square \)

15.2 The \( \Gamma_1(p) \) Case

For every \( t \) in \( T \) and for every subset \( P \) of the set \( S = S_P \) of places \( F \) above \( p \), we well let

\[
\pi_1^{-1}(X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{Iw},p,\text{K}][0,1])(0,1)_{S-P} \subset X_{\Gamma_1(n),t}^{\text{an}}(\Gamma_1(p),\text{Iw},p,\text{K})
\]

Let \( w_P \) denote the composite of the \( w_{Q_p} \) for all \( p \in P \). Note that

\[
(X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{Iw},p,\text{K}][0,1])(0,1)_{S-P} \not\supset w_P^{-1}X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{Iw},p,\text{K}][0,1]
\]

and each \( (X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{Iw},p,\text{K}][0,1])(0,1)_{S-P} \) is connected since it is isomorphic to \( X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{O},p,\text{K}][0,1] \) and the latter is connected since it is the pre-image of a connected component in the Zariski topology of the closed fibre. Let \( (X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{Iw},p,\text{K}][0,1])(0,1)_{S-P} \) denote the disjoint union over \( T \) of \( (X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{Iw},p,\text{K}][0,1])(0,1)_{S-P} \).

**Definition.** If \( f \) is a section of \( L_k \) over \( X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{K}][0,1] \), then \( w_{S-P}^{-1} f \) is a section of \( w_{S-P}^{-1} L_k \) over \( w_{S-P}^{-1} (X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{K}][0,1]) = (X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{K}][0,1])(0,1)_{S-P} \). Because \( p \) is inverted, the natural morphism of invertible sheaves

\[
L_k((X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{K}][0,1])(0,1)_{S-P} \not\supset w_{S-P}^{-1} (L_k|X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{K}][0,1])
\]

is an isomorphism and we let \( f|w_P \) denote the section of \( L_k \) over \( (X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{K}][0,1])(0,1)_{P-S} \) corresponding to \( w_{S-P}^{-1} f \) by the isomorphism.

**Theorem 23** For every subset \( P \) of \( S = S_P \), let \( f_P \in H^0(X_{\Gamma_1(n),t} \cap \Gamma_1(p),[\text{K}][0,1],L_k)^{\text{Iw}} \) be an overconvergent Hilbert modular form of parallel weight \( k = \sum_{r \in \text{Hom}(F,K)} k_r \in \mathbb{Z} \) and of level \( \Gamma_1(np) \). For every subset \( P \) of \( S \), suppose that \( f_P \) has a Hecke character

\[
\psi_P \overset{\text{def}}{=} \psi_P^S \psi_{S,P} : (\mathcal{O}_F/np)^\times \cong (\mathcal{O}_F/n)^\times \times (\mathcal{O}_F/p)^\times \to \mathcal{O}^\times;
\]

and that \( \psi_P^P(p) = \psi_P^{P-[p]}(p) \) for every \( p \) in \( P \). Suppose that
• the Fourier coefficients \( c(\mathcal{O}_F, f_P) = 1 \) and \( c(m, f_P) = 0 \) if \( m \) and \( n \) are not coprime,

• for every \( p \in P \), \( c(m, f_P) = \psi_{P,p}(m)c(m, f_{P-\{p\}}) \) for every ideal \( m \) coprime to \( np \), where by \( \psi_{P,p} \) we mean the \( p \)-component of \( \psi_{P,S} \) which we assume non-trivial,

• for every \( p \in S \), \( f_P \) is an \( U_p \)-eigenform with non-zero eigenvalue \( \alpha(p, f_P) \), and for every \( p \in P \), \( \alpha(p, f_P)\alpha(p, f_{P-\{p\}}) = \psi_{P,p}(p)|\mathcal{O}_F/p|^{k-1} = \psi_{P-\{p\}}(p)|\mathcal{O}_F/p|^{k-1} \).

Then \( f_P \) is a section of \( L_k \) over \( X_{\Gamma_1(n)\cap\Gamma_1(p),\mathcal{O}_K}^\an \).

Proof. For every subset \( P \) of \( S \), let \( g_P \) denote \( f_P|w_P \in H^0((X_{\Gamma_1(n)\cap\Gamma_1(p),\mathcal{O}_K}[0,1])(0,1)_{S-P},L_k) \). Clearly \( g_S = f_S \). We shall prove that \( f_S \) is classical.

Fix an integer \( 0 \leq n \leq |S| \) and suppose that the \( g_P \) with \( P \subseteq S \) such that \( |P| \geq n \) glue together to define sections, which will again be denoted by \( g_P \), over

\[
\bigcup_{P \subseteq S, |P| \geq n} (X_{\Gamma_1(n)\cap\Gamma_1(p),\mathcal{O}_K}[0,1])(0,1)_{S-P}.
\]

Fix a subset \( P \subseteq S \) with \( \#P = n \) and fix \( p \in P \). It suffice to show that \( g_P(\simeq w^*_{S-P}f_P) \) and (a constant multiple of) \( g_{P-\{p\}} (\simeq w^*_{S-(P-\{p\})}f_{P-\{p\}} = w^*_p w^*_{S-P}f_{P-\{p\}} \) glue.

Let \( \alpha_P \) (resp. \( \beta_P \)) denote the \( U_p \)-eigenvalue \( \alpha(p, f_P) \) (resp. \( \alpha(p, f_{P-\{p\}}) \)).

Fix a \( p \)-th root \( \zeta_1 \) of unity. Let \( (\text{Tate}_M, M_\infty(q), \ldots, \eta_{\text{KM}} : 1 \mapsto \zeta_1) \) be a point around a cusp \( C \). By abuse of notation, we call it \( C \).

There is a morphism

\[
\pi_1 : X_{\Gamma_1(n,t)\cap\Gamma_1(p),\mathcal{I}_w,p,K}^\an \longrightarrow X_{\Gamma_1(n,t)\cap\Gamma_1(p),\mathcal{O}_K}^\an
\]

defined, on the non-cuspidal points, by

\[
(A, i, \lambda, \eta, \eta_{\text{KM}}, D_p) \mapsto (A, i, \lambda, \eta, \eta_{\text{KM}})
\]

and, for \( \gamma \in (\mathcal{O}_F/p)^\times \),

\[
\pi_{2,\gamma} : X_{\Gamma_1(n,t)\cap\Gamma_1(p),\mathcal{I}_w,p,K}^\an \longrightarrow X_{\Gamma_1(n,t)\cap\Gamma_1(p),\mathcal{O}_K}^\an
\]

defined, on the non-cuspidal points, by

\[
(A, i, \lambda, \eta, \eta_{\text{KM}}, D_p) \mapsto (A/\gamma \eta_{\text{KM}}(1)p + D_p) \ldots (\eta_{\text{KM}} \bmod (\gamma \eta_{\text{KM}}(1)p + D_p)).
\]

To single out, let

\[
\pi_{2,p} : X_{\Gamma_1(n,t)\cap\Gamma_1(p),\mathcal{I}_w,p,K}^\an \longrightarrow X_{\Gamma_1(n,t)\cap\Gamma_1(p),\mathcal{O}_K}^\an
\]

Documenta Mathematica 18 (2013) 997–1038
 denote the morphism ‘\( \gamma = 0 \) in \( \mathcal{O}_F/p' \) which takes \((A, i, \lambda, \eta, \eta_{KM}, D_p)\) to \((A/D_p, \ldots, (\eta_{KM} \bmod D_p))\).

By abuse of notation, let \( C \) also denote the pre-image

\((\text{Tate}_{M_1, M_2}(q), \ldots, \eta_{KM} : 1 \mapsto \zeta_1, \langle \eta_1^{\nu} \rangle) \in X_{\Gamma_1(n, t)\cap \Gamma_1(p), Iw_p, K}\)

by \( \pi_1 \) above of \( C = (\text{Tate}_{M_1, M_2}(q), \ldots, \zeta_1) \) for \((M_1, M_2) = (\mathcal{O}_F, t^{-1})\) and \( M = M_1M_2 = t^{-1} ; \) and let \( C_P \in (X_{\Gamma_1(n, t)\cap \Gamma_1(p), Iw_p, K}[0, 1]) \) \((\mathbf{0}, 1)_{S-P} \) denote the cups \( w_{S-P}^*C \).

Then

\[(g_P|1)(C_P) = pr^*f(\text{Tate}_{M_1, M_2}(q), \ldots, \zeta_1) = |\mathcal{O}_F/t|^{-1} \sum_{\nu \in M^+} c(\nu M^{-1}, f_P)q_\nu^{\nu} \]

On the other hand, for \( \gamma \in (\mathcal{O}_F/p)^\times \),

\[
\begin{align*}
(g_P|\pi_2, \gamma)(C_P) &= (f_P|\pi_2, \gamma)(\text{Tate}_{M_1, M_2}(q), \ldots, \zeta_1) \\
&= pr^*f_P(\text{Tate}_{M_1, M_2}(q)/\langle \zeta, \eta \rangle, \ldots) \text{ where } \eta := \eta_{1, p}^\text{et} \text{ and } \zeta := \zeta_{1, p} \\
&= pr^*f_P((GL_1 \otimes \mathbb{Z} \delta^{-1}M_1^{-1})/(\zeta, \eta_{M_2})^{-1}, \ldots) \\
&= |\mathcal{O}_F/t_p|^{-1} \sum_{\nu \in (p^{-1}M_2)^+} c(\nu \delta M^{-1}, f_P)\zeta^{\nu} q_\nu^{\nu} 
\end{align*}
\]

where \( t_p \) is one of the (fixed) representatives of the narrow class group of \( F \) representing the class of \( tp \), and where \( q_\nu \) denote a representative in \( q^{p^{-1}M_2}/q^{M_2} \). Finally

\[
(g_P|\pi_2, p)(C_P) = pr^*(f_P|\pi_2, p)(\text{Tate}_{M_1, M_2}(q), \ldots, \zeta_p) = |\mathcal{O}_F/t_p|^{-1} \sum_{\nu \in (p^{-1}M)^+} c(\nu \delta M^{-1}, f_P)q_\nu^{\nu} 
\]

For brevity, let \( S \) denote the ‘Gauss sum’

\[ S \overset{\text{def}}{=} \sum_{\gamma \in (\mathcal{O}_F/p)^\times} \zeta^\gamma \psi_{p, p}(\gamma) \]

for \((M_1, M_2) = (\mathcal{O}_F, t^{-1}) \). Then for \( \nu \in (p^{-1}M)^+ \) such that \( \nu \delta M^{-1} \subset \mathcal{O}_F^+ \) is not divisible by \( p \),

\[
S = \sum_{\gamma \in (\mathcal{O}_F/p)^\times} \zeta^{\nu \delta M^{-1}} \psi_{p, p}(\gamma \nu \delta M^{-1}) = \psi_{p, p}(\nu \delta M^{-1}) \sum_{\gamma \in (\mathcal{O}_F/p)^\times} \zeta^{\nu} \psi_{p, p}(\gamma) .
\]
It then follows that

\[
\sum_{\gamma \in (O_F/p)} \psi_{p,p}(\gamma) (g_p|\pi_2,\gamma)(C_F) = \sum_{\gamma \in (O_F/p)} \psi_{p,p}(\gamma) |O_F/t_p|^{-1} \sum_{\nu \in (1,M)} c(\nu p M^{-1}, f_p) \zeta^{\nu} q_\eta^\nu
\]

\[
= \left|O_F/t_p\right|^{-1} \sum_{\nu \in (1,M)} c(\nu p M^{-1}, f_p) q_\eta^\nu \sum_{\gamma \in (O_F/p)} \psi_{p,p}(\gamma) (C_F)
\]

\[
= S|O_F/t_p|^{-1} \sum_{\nu \in (1,M), \nu \mid p M-1} c(\nu p M^{-1}, f_p) q_\eta^\nu
\]

\[
= S|O_F/t_p|^{-1} \sum_{\nu \in (1,M), \nu \mid p M-1} c(\nu p M^{-1}, f_p|p-\eta) q_\eta^\nu
\]

\[
= S|O_F/t_p|^{-1} \sum_{\nu \in (1,M), \nu \mid p M-1} c(\nu p M^{-1}, f_p|p-\eta) q_\eta^\nu
\]

By the connectedness of \((X_{\Gamma_1(n,\eta)}|\Gamma_1(p),K)[0,1])\) \((0,1]_{S-P}\).

\[
\sum_{\gamma \in (O_F/p)} \psi_{p,p}(\gamma) (g_p|\pi_2,\gamma) = S(g_p|p-\eta) = \beta_p g_p|p-\eta|\pi_1(C_F)
\]

on \((X_{\Gamma_1(n,\eta)}|\Gamma_1(p),K)[0,1])\) \((0,1]_{S-P}\).

Let \((A, i, \lambda, \eta, \eta_{KM}, D_p = \{Q_p\})\) be a non-cuspidal point of \(X_{\Gamma_1(n,\eta)}|\Gamma_1(p),l_{w_p},K\)
and let \(P = \eta_{KM}(1) = P^{\ast} \times P_p\) and \(Q = P^{\ast} \times Q_p\). Then

\[
|O_F/p|\alpha_p g_p(A,i,\lambda,\eta,Q) = pr^* g_p(A/(P_p),\ldots,\eta,\overline{Q})
\]

where \(\eta := \eta \mod \langle P_p \rangle\) and \(\overline{Q} := Q \mod \langle P_p \rangle\) is:

\[
= |O_F/p|S_{\nu_{\eta_{KM}}}(iA, i, \lambda, \eta, Q) = pr^* g_p(A/(P_p),\ldots,\eta, \overline{Q})
\]

\[
= \sum_{c_p \not\mid A, c_p \neq (P_p)(Q_p)} pr^* g_p(A/C_p, \ldots, (Q \mod C_p))
\]

\[
= \sum_{c_p = g_p(A/(P_p) \not\mid \eta)} pr^* g_p(A/(\gamma P_p + Q_p), \ldots, (Q \mod (\gamma P_p + Q_p)))
\]

\[
= \sum_{\gamma \in (O_F/p)} \psi_{p,p}(\gamma) pr^* g_p(A/(\gamma P_p + Q_p), \ldots, (Q \mod (\gamma P_p + Q_p)))
\]

\[
= \psi_{p}(\eta) S_{\nu_{\eta_{KM}}}(iA, i, \lambda, \eta, Q) = pr^* g_p(A/(P_p), \ldots, (Q \mod (\gamma P_p + Q_p)))
\]

\[
= \psi_{p}(\eta) S_{\nu_{\eta_{KM}}}(iA, i, \lambda, \eta, Q) = pr^* g_p(A/(P_p), \ldots, (Q \mod (\gamma P_p + Q_p)))
\]

\[
= \psi_{p}(\eta) S_{\nu_{\eta_{KM}}}(iA, i, \lambda, \eta, Q) = pr^* g_p(A/(P_p), \ldots, (Q \mod (\gamma P_p + Q_p)))
\]

\[
= \sum_{\gamma \in (O_F/p)} \psi_{p,p}(\gamma) g_p|\pi_2,\gamma)(C_F) - \beta_p g_p|p-\eta|\pi_1(C_F)
\]

Therefore

\[
(|O_F/p|\alpha_p g_p - S(g_p|p-\eta))|\pi_2,\gamma) = 0
\]

\[
= (|O_F/p|\alpha_p g_p - S(g_p|p-\eta)|w_\pi)^*(A/\pi, Q)
\]

\[
= (|O_F/p|\alpha_p g_p - S(g_p|p-\eta)|w_\pi)^*(A/\pi, Q)
\]

\[
= (|O_F/p|\alpha_p g_p - S(g_p|p-\eta)|w_\pi)^*(A/\pi, Q)
\]

\[
= (|O_F/p|\alpha_p g_p - S(g_p|p-\eta)|w_\pi)^*(A/\pi, Q)
\]

It suffices to show that \(|O_F/p|\alpha_p g_p - S(g_p|p-\eta)|w_\pi\) is identically zero; in which case, one can glue \(g_p\) and \((|O_F/p|\alpha_p g_p - S(g_p|p-\eta)|w_\pi)\) as desired. Showing that it is identically zero is exactly as in [3].
15.3 The $\Gamma_1(p^r)$, $r \geq 2$, case

Theorem 24 Let $S$ denote the set $S_P$ of places of $F$ above $p$. For any set $P \subseteq S$, let $f_P \in H^0(X_{\Gamma_1(n)\cap \Gamma_1(p^r)}, K[0,1], L_k)_{\mathbf{U}}$ be an overconvergent modular form of weight $k = \sum_{\tau \in \text{Hom}(F,K)} k_{\tau} \in \mathbb{Z}$ and of level $\Gamma_1(np^r)$. Suppose that, for every $P \subseteq S$, $f_P$ has a character $\psi_P, \psi_{S,P}$ of $(\mathcal{O}_F/np^r)^\times \simeq (\mathcal{O}_F/n)^\times \times (\mathcal{O}_F/p^r)^\times$. Suppose furthermore that $f_P$ is an eigenform for $U_p$ with non-zero eigenvalue for every $p \in S$. Suppose finally that, for every $P \subseteq S$,

- $c(\mathcal{O}_F, f_P) = 1$;
- $c(m, f_P) = 0$ if $m$ and $n$ are not coprime;
- for every $p \in P$, $c(m, f_P) = \psi_{p, P}(m)c(m, f_{P-\{p\}})$ for every ideal $m$ coprime to $np$, where $\psi_{p, P}$ is the $p$-component of $\psi_{S,P}$.

Then the $f_P$ are classical Hilbert modular forms in $H^0(X_{\Gamma_1(n)\cap \Gamma_1(p^r)}, K[0,1], L_k)_{\mathbf{U}}$.

Proof. As in the previous subsection, we shall prove the theorem by induction. For every subset $P$ of $S$, let $g_P$ denote $f_P|_{\mathcal{O}_F} \in H^0(X_{\Gamma_1(n)\cap \Gamma_1(p^r)}, K[0,1], L_k)$. We shall prove that $g_S$ is classical. Fix an integer $0 \leq n \leq |S|$ and suppose that the $g_P$ with $P \subseteq S$ such that $|P| \geq n$ glue together to define sections, which will again be denoted by $g_P$, over

$$\bigcup_{P \subseteq S, |P| \geq n} X_{\Gamma_1(n)\cap \Gamma_1(p^r), K[0,1], L_k}.$$ 

Fix a subset $P \subseteq S$ with $#P = n$, and fix $p \in P$. It suffice to show that $g_P$ and (a constant multiple of) $g_{P-\{p\}}$ glue. Let $C$ denote a point $(\text{Tate}_{M_1, M_2}(q), \iota, \lambda, \eta, P)$ around a cusp $(M_1, M_2) = (\mathcal{O}_F, t^{-1})$ where

$$P = \eta_{\text{CM}}(1) = P^p \times P_p \in \text{Tate}_{M_1, M_2}(q)(\mathcal{O}_F \otimes \mathbb{Z}((q^{M}))$$

where $P^p \overset{\text{def}}{=} \prod_{p|\eta \neq p} \zeta_{\eta_{\text{CM}}}^\eta$ and $P_p \overset{\text{def}}{=} \eta_{\text{CM}}^\eta P_p$. For brevity, let $\mu$ denote $\zeta_{\text{CM}}^\eta$, and $\mu_p$ its $p$-component.

We shall compute $q$-expansions of $g_P$ and $g_{P-\{p\}}$ at the cusp $C_P \overset{\text{def}}{=} w^p_{S-P} C$. Let $\alpha_p$ denote the $U_p$-eigenvalue of $f_P$. 

$$\frac{\langle \mathcal{O}_F/p \eta_{\text{CM}} g_P(C_P) \rangle}{\langle \mathcal{O}_F/p \eta_{\text{CM}} g_P(C_P) \rangle} = \sum_{\gamma \in \text{Tate}_{M_1}(q)/C_\gamma \setminus \{\gamma\}} \psi_{\gamma, P}(\text{Tate}_{M_1, M_2}(q)/C_\gamma, \ldots, (P \mod C_\gamma))$$

for every $P \subseteq S$, let $f_P \in H^0(X_{\Gamma_1(n)\cap \Gamma_1(p^r)}, K[0,1], L_k)_{\mathbf{U}}$ be an overconvergent modular form of weight $k = \sum_{\tau \in \text{Hom}(F,K)} k_{\tau} \in \mathbb{Z}$ and of level $\Gamma_1(np^r)$. Suppose that, for every $P \subseteq S$, $f_P$ has a character $\psi_P, \psi_{S,P}$ of $(\mathcal{O}_F/np^r)^\times \simeq (\mathcal{O}_F/n)^\times \times (\mathcal{O}_F/p^r)^\times$. Suppose furthermore that $f_P$ is an eigenform for $U_p$ with non-zero eigenvalue for every $p \in S$. Suppose finally that, for every $P \subseteq S$,
We know that $\psi_{S,p}$ has a conductor $p^r$, and hence $\psi_{S,p}((1+p^{r-1})pM^{-1}) = \mu_p^{r\gamma}$
for some integer $0 < \nu_1 < p$ (not that $1 + p^{r-1}$ is thought of as an element of
$p^{-1}M_2$). $M_1 = M_1/p^rM_1 \simeq \mathcal{O}_F/p^r$; it therefore follows that $\psi_{p,p}((1 + p^{r-1})pM^{-1}) = \mu_p^{r\gamma}$. In particular, $\psi_{p,p}((1 - p^{r-1}\gamma)pM^{-1}) = \mu_p^{-\gamma\nu_1}$. Hence
\[
\sum_{\gamma} \psi_{p,p}((1 - p^{r-1}\gamma)pM^{-1})\mu_p^{\nu} = \sum_{\gamma} \mu_p^{(\nu - \nu_1)} = \begin{cases} |\mathcal{O}_F/p| & \text{if } p|\nu - \nu_1 \\ 0 & \text{otherwise} \end{cases}
\]
Therefore,
\[
g_p(C_P) = (\alpha_p|\mathcal{O}_F/p|)^{-1}|\mathcal{O}_F/p| \sum_{\nu \in (p^{-1}M)^+} C(\nu\mathcal{F}^{-1}, f_P)q_\nu^{\nu}.
\]
We now calculate the $q$-expansion of $g_{P-\{p\}}|\omega_{CS}$ at $C_P$. Firstly, note that
\[
(g_{P-\{p\}}|\omega_{CS})(C_P) = p^{r\gamma}g_{P-\{p\}}(\text{Tate}_{M_1,p^r\{\gamma\}, p}(q, \{\zeta_{r, p}\eta_1^{p^r}, \ldots, M_1^p \times Q_p)
\]
where $Q_p$ is defined by $(\zeta_{r, p}\eta_1^{p^r}, Q_p) = \zeta_p$. Tensoring over $\mathcal{O}_F$ with $p^{r-1}$ on
$GL_1 \otimes \delta^{-1}M_1^{-1}$ induces an isomorphism
\[
(GL_1 \otimes \delta^{-1}M_1^{-1})/(\langle q^{M_2}, \zeta_{r, p}\eta_1^{p^r} \rangle)
\simeq (GL_1 \otimes \delta^{-1}M_1^{-1})/(\langle q^{p^{r-1}M_2}, \zeta_{r, (r-1), p}q^{p^{r-2}M_2} \rangle)
\]
\[
\simeq (GL_1 \otimes \delta^{-1}M_1^{-1})/(\zeta_{1, p}\eta_1^{p^r})^{p^{r-2}M_2}
\]
The HBAV
\[
\text{Tate}_{M_1,p^{r-2}M_2}(\zeta_{1, p}\eta_1^{p^r}) = (GL_1 \otimes \delta^{-1}M_1^{-1})/(\zeta_{1, p}\eta_1^{p^r})^{p^{r-2}M_2}
\]
is naturally $p^{2-r}(\simeq (p^{r-2}M_2)^{-1}M_1)$-polarised, and comes equipped with the
level structure $(p)^{r-1}\eta$ and the point $P \times \eta_1^{p^r}$ of order $|\mathcal{O}_F/p^r|$; it defines a
point of $X^{[r-1]}(\Gamma_1(n, t) \cap \Gamma_1(t), \tau, (P \mod C_P))$. For a point $(A, i, \lambda, \eta, P)$ of $X^{[r-1]}(\Gamma_1(n, t) \cap \Gamma_1(t), \tau, (P \mod C_P))$
\[
\begin{align*}
& (|\mathcal{O}_F/p|\beta_P)^{r-1}g_{P-\{p\}}(A, i, \lambda, \eta, P) \\
& = (|\mathcal{O}_F/p|\beta_P)^{r-1}g_{P-\{p\}}(A, i, \lambda, \eta, P) \\
& = \sum_{C_p \subset A_p, |C_p| = |\mathcal{O}_F/p^{r-1}|} 1_{P \cap P = \{P\}} P^{r-1}g_{P-\{p\}}(A/C_P, \ldots, (P \mod C_P))
\end{align*}
\]
In which case, observe that $(A/C_P, \ldots, (P \mod C_P)) \in X^{[0]}(\Gamma_1(n, t) \cap \Gamma_1(t), \tau)$
and that it allows one to extend finite slope $U_p$-eigenforms over $X^{[0]}(\Gamma_1(n, t) \cap \Gamma_1(t), \tau)$ to
$U_p$-eigenforms over $X^{[r-1]}(\Gamma_1(n, t) \cap \Gamma_1(t), \tau)$ by ‘analytic continuation’. If we let
\[
(A, i, \lambda, \eta, P) = (\text{Tate}_{M_1,p^{r-2}M_2}(\zeta_{1, p}\eta_1^{p^r}), \ldots, (p)^{r-1}\eta, P \times \eta_1^{p^r})
\]
then the cyclic subgroups $C_p$ of order $|\mathcal{O}_F/p^{r-1}|$, disjoint from the
subgroup of order $|\mathcal{O}_F/p|$ generated by $\eta_1^{p^r}$, are of the form
where \( q_\eta \) is a representative in \( q^{p-1}M_z \) of \( q_\eta^e \in q^{p-1}M_z / q^M_z \), is naturally \((p^{-1}M)^{-1}M_1 \simeq \mathfrak{pt}\)-polarised. Then there exists a non-zero constant \( \kappa_1 \) such that

\[
\left( \psi_{p^{-1-1}} \right)^{(p)}|\nu_1(S_p) \equiv \sum_{\nu_1-1} p^{*}\psi_{p\nu_1-1}((GL_1 \otimes \mathcal{Z} \oplus s^{-1}1_M) / \langle \psi_{1-p}\nu_1-1, \psi_{p\nu_1-1} \rangle) = \psi_{p\nu_1-1}((GL_1 \otimes \mathcal{Z} \oplus s^{-1}1_M) / \langle \psi_{1-p}\nu_1-1, \psi_{p\nu_1-1} \rangle)
\]

where \( \nu_1 \) ranges over \((\mathcal{O}_F/p^{r-1}). \) For brevity, let \( S_{\nu} \overset{\text{def}}{=} \sum_{\nu_1-1} \nu_1 \psi_{p\nu_1-1}((1+p\nu_1-1)p^{-1}M_1)^{-1} \psi_{p\nu_1-1}. \) As in the proof of Theorem 11.1 in [3], one can deduce that

\[
S_{\nu} = \mu_{\nu_1} S_{\nu_1}
\]

where \( \nu_1 \) is defined by \( \psi_{p\nu_1-1}((1+p\nu_1-1)p^{-1}M_1)^{-1} = \mu_{\nu_1} \), and therefore \( S_{\nu} = 0 \) unless \( p | (\nu - \nu_1) \); one also deduces that, for \( \nu \in (p^{-1}M)^-1 \) such that \( p | (\nu - \nu_1) \),

\[
S_{\nu} = \psi_{p\nu_1-1}(\nu\nu_1M^{-1})S_{\nu_1}
\]

for \( \nu' \in (p^{-1}M)^-1 \) such that \( \nu'\nu_1M^{-1} = 1 \mod p \) and therefore \( S_{\nu'}/\psi_{p\nu_1-1}(\nu'\nu_1M^{-1})) = S_{\nu}/\psi_{p\nu_1-1}(\nu\nu_1M^{-1}). \) Consequently, there is a non-zero constant \( \kappa_2 \) such that

\[
\left( g_{p^{-1}} \right)^{(p)} \nu_1(S_p) = \sum_{\nu_1 \in (p^{-1}M)^+} \psi_{\nu_1}(\nu_1\nu_1M^{-1})S_{\nu}(q_\eta)
\]

Therefore \( \alpha_p g_{p^{-1}} \) and \( \kappa_2^{-1}g_{p^{-1}} \) agree at \( C_p \) and hence \( f_p \) and \( (\alpha_p\kappa_2^{-1})g_{p^{-1}} \) glue together. \( \square \)

References


Documenta Mathematica 18 (2013) 997–1038


Shu Sasaki
Universität Duisburg-Essen
Fakultät für Mathematik
Thea-Leymann-Str. 9
45127 Essen, Germany
s.sasaki.03@cantabgold.net
EXTENDING SELF-MAPS
TO PROJECTIVE SPACE OVER FINITE FIELDS

BJORN POONEN

Received: May 21 2012
Communicated by Stephen Lichtenbaum

Abstract. Using the closed point sieve, we extend to finite fields the following theorem proved by A. Bhatnagar and L. Szpiro over infinite fields: if \( X \) is a closed subscheme of \( \mathbb{P}^n \) over a field, and \( \phi: X \to X \) satisfies \( \phi^* \mathcal{O}_X(1) \simeq \mathcal{O}_X(d) \) for some \( d \geq 2 \), then there exists \( r \geq 1 \) such that \( \phi^r \) extends to a morphism \( \mathbb{P}^n \to \mathbb{P}^n \).

2010 Mathematics Subject Classification: Primary 37P25; Secondary 37P55
Keywords and Phrases: self-map, closed point sieve

1 Introduction

Let \( k \) be a field. Given a closed subscheme \( X \subseteq \mathbb{P}^n \) over \( k \), and given a self-map (i.e., \( k \)-scheme endomorphism) \( \phi: X \to X \), does \( \phi \) extend to a self-map \( \psi: \mathbb{P}^n \to \mathbb{P}^n \)? Such questions have applications in arithmetic dynamics; for instance, [Fak03, Corollary 2.4] uses a positive answer to a variant of this to show that the Morton–Silverman uniform boundedness conjecture for preperiodic points of a self-map of projective space over a number field [MS94, p. 100] implies the uniform boundedness conjecture for torsion points on abelian varieties over a number field.

If the extension \( \psi \) exists, then \( \psi^* \mathcal{O}(1) \simeq \mathcal{O}(d) \) for some integer \( d \), and then \( \phi^* \mathcal{O}_X(1) \simeq \mathcal{O}_X(d) \). But A. Bhatnagar and L. Szpiro [BS12, Proposition 2.3] gave an example showing that the existence of \( d \) such that \( \phi^* \mathcal{O}_X(1) \simeq \mathcal{O}_X(d) \) is not sufficient for the extension \( \psi \) to exist.

To obtain an extension theorem, one can relax the requirements. Two ways of doing this lead to the following questions:
Question 1.1 (Changing the embedding). Let $X$ be a projective $k$-scheme. Let $\mathcal{L}$ be an ample line bundle on $X$. Let $\phi: X \to X$ be a morphism such that $\phi^*\mathcal{L} \simeq \mathcal{L} \otimes \mathcal{O}_X(d)$ for some $d \geq 1$. Does there exist a closed immersion $X \hookrightarrow \mathbb{P}^n$ such that $\phi$ extends to a morphism $\mathbb{P}^n \to \mathbb{P}^n$?

Question 1.2 (Replacing the self-map by a power). Let $X$ be a closed subscheme of $\mathbb{P}^n$ over $k$. Let $\phi: X \to X$ be a morphism such that $\phi^*\mathcal{O}_X(1) \simeq \mathcal{O}_X(d)$ for some $d \geq 2$. Then there exists $r \geq 1$ such that $\phi^r$ extends to a morphism $\mathbb{P}^n \to \mathbb{P}^n$.

Remark 1.3. Section 4 explains why we cannot allow $d = 1$ in Question 1.2.

Suppose that $k$ is infinite. Then the answer to both questions is yes: see [Fak03, Corollary 2.3] and [BS12, Theorem 2.1], respectively (in the proof of the latter, one should replace the prime avoidance lemma there by the lemma used in [Fak03], that a finite union of proper subspaces in a vector space over an infinite field cannot cover the whole space). A positive answer to Question 1.2 is also an immediate consequence of [Fak03, Proposition 2.1] if one notices that the statement and proof there remain valid if hypothesis (1) is imposed only for $n = d$ instead of all $n \geq 0$. (The word “variety” in [Fak03] and [BS12] may be read as “scheme of finite type”, so there is no difference between “projective variety” and “projective scheme”.)

Our main result is the following:

Theorem 1.4. Question 1.2 has a positive answer over any field $k$.

In the case where $k$ is finite, the general position arguments in [Fak03] and [BS12] fail, so a new idea is needed. To prove Theorem 1.4, we use the closed point sieve introduced in [Poo04] to show that a random choice leads to an extension of $\phi$, even though we cannot exhibit one explicitly. As far as we know, this is the first time that sieve techniques have been applied to a problem in dynamics.

Remark 1.5. See [MZMS13, Theorem 3] for an analogous statement on self-maps of equicharacteristic complete local rings.

Remark 1.6. We still do not know if Question 1.1 has a positive answer when $k$ is finite.

2 Extending morphisms to projective space

The finite field case of Theorem 1.4 will be proved with the aid of the following quantitative theorem, involving a zeta function $\zeta_U(s)$ defined as in [Poo04]:

Theorem 2.1. Let $k$ be a finite field $\mathbb{F}_q$. Fix a closed subscheme $X$ of $\mathbb{P}^n = \text{Proj} \, S$ over $k$. Let $U := \mathbb{P}^n - X$. Let $I = \bigoplus_{d \geq 0} I_d \subseteq S = \bigoplus_{d \geq 0} S_d$ be the homogeneous ideal of $X \subseteq \mathbb{P}^n$. Let $N \geq n$. Fix $f_0, \ldots, f_N \in S_d$. Then if
Extending self-maps

\[ g_0, \ldots, g_N \] are chosen independently and uniformly at random from the finite set \( I_d \).

\[ \text{Prob}(f_0 + g_0, \ldots, f_N + g_N \text{ have no common zeros on } U) = \zeta_U(N+1)^{-1} + o(1), \]

where the \( o(1) \) is bounded by a function of \( k, X, n, N \), and \( d \) that tends to 0 as \( d \to \infty \) while \( k, X, n, \) and \( N \) are fixed.

Theorem 2.1 will be proved in Section 3. For now, we show how it implies Theorem 1.4, through the following:

**Theorem 2.2.** Fix a closed subscheme \( X \) of \( \mathbb{P}^n \) over a field \( k \). If \( d \) is sufficiently large and \( N \geq n \), then any morphism \( \phi : X \to \mathbb{P}^N \) such that \( \phi^* \mathcal{O}(1) \cong \mathcal{O}_X(d) \) extends to a morphism \( \mathbb{P}^n \to \mathbb{P}^N \).

**Proof.** Let \( z_0, \ldots, z_N \) be the homogeneous coordinates on \( \mathbb{P}^N \). For sufficiently large \( d \), the restriction map \( S_d = \Gamma(\mathbb{P}^n, \mathcal{O}(d)) \to \Gamma(X, \mathcal{O}_X(d)) \) is surjective. So each \( \phi^*(z_i) \) is the restriction of some \( f_i \in S_d \).

If \( k \) is infinite, the proof of [Fak03, Proposition 2.1] applies for any \( d \) that is moreover large enough that \( X \) is cut out in \( \mathbb{P}^n \) by homogeneous polynomials of degree at most \( d \).

If \( k \) is finite, Theorem 2.1 implies that for sufficiently large \( d \), there exist \( g_0, \ldots, g_N \in I_d \) such that \( f_0 + g_0, \ldots, f_N + g_N \) have no common zeros in \( \mathbb{P}^n - X \). On the other hand, restricted to \( X \), they define the same map \( \phi \) as \( f_0, \ldots, f_N \) do, so they have no common zeros on \( X \) either. Thus \( f_0 + g_0, \ldots, f_N + g_N \) define a morphism \( \mathbb{P}^n \to \mathbb{P}^N \) extending \( \phi \).

**Proof of Theorem 1.4.** Apply Theorem 2.2 with \( N = n \) and with \( \phi \) equal to a sufficiently large power of the \( \phi \) given in Theorem 1.4.

3 Proof of Theorem 2.1

The idea of the proof of Theorem 2.1, borrowed from [Poo04], is to sieve out, for each closed point \( P \in U \), the \((g_0, \ldots, g_N)\) for which \( f_0 + g_0, \ldots, f_N + g_N \) have a common zero at \( P \). Heuristically, the probability that a given \( f_i + g_i \) vanishes at \( P \) is \( q^{-\deg P} \), so, assuming independence, the probability that \( f_0 + g_0, \ldots, f_N + g_N \) have no common zeros on \( U \) should be

\[ \prod_{\text{closed } P \in U} \left( 1 - q^{-(N+1)\deg P} \right) = \zeta_U(N+1)^{-1}. \]

But independence holds only for finitely many \( P \), so to make this rigorous, we impose the conditions only for \( P \) of degree up to some bound \( \rho \), and then prove that the number of \((g_0, \ldots, g_N)\) sieved out by higher-degree closed points is negligible.
3.1 Points of low degree

Let \( f = (f_0, \ldots, f_N) \) and \( g = (g_0, \ldots, g_N) \). Let \( V(f + g) \) be the common zero locus of the \( f_i + g_i \). Given \( \rho \in \mathbb{Z}_{>0} \) and a \( k \)-scheme \( Z \), let \( Z_{<\rho} \) be the set of closed points of \( Z \) of degree less than \( \rho \), and define \( Z_{>\rho} \) similarly.

**Lemma 3.1 (Points of low degree).** For fixed \( \rho \), if \( d \) is sufficiently large, then 

\[
\text{Prob}(V(f + g) \cap U_{<\rho} = \emptyset) = \prod_{P \in U_{<\rho}} (1 - q^{-(N+1)\deg P}).
\]

**Proof.** Let \( \mathcal{I} \) be the ideal sheaf of \( X \subseteq \mathbb{P}^n \). View \( U_{<\rho} \) as a 0-dimensional closed subscheme of \( \mathbb{P}^n \). By [Poo08, Lemma 2.1], if \( d \) is sufficiently large, then the restriction map \( I_d \to \Gamma(U_{<\rho}, \mathcal{I} : \mathcal{O}_{U_{<\rho}}(d)) \) is surjective. In particular, for each \( i \), the tuple of "values" \( ((f_i + g_i)(P))_{P \in U_{<\rho}} \) is equidistributed. The residue field at \( P \) has size \( q^\deg P \), so the probability that \( f + g \) vanishes at \( P \) is \( q^{-(N+1)\deg P}, \) and the probability that \( f + g \) is nonvanishing at all \( P \in U_{<\rho} \) is 

\[
\prod_{P \in U_{<\rho}} (1 - q^{-(N+1)\deg P}) .
\]

\( \square \)

3.2 Points of medium degree

Let \( U_{a\leq \gamma \leq b} \) be the set of closed points of \( U \) of degree between \( a \) and \( b \). As in [Poo08, Section 2], fix \( c \) so that \( S_1 I_m = I_{m+1} \) for all \( m \geq c \).

**Lemma 3.2 (Points of medium degree).** If \( d \) is sufficiently large, then

\[
\text{Prob}(V(f + g) \cap U_{\rho \leq \gamma \leq d-c} = \emptyset) = O(q^{-\rho}).
\]

**Proof.** By [Poo08, Lemma 2.2], the fraction of \( h \in I_d \) vanishing at a closed point \( P \) of degree \( e \in [\rho, d-c] \) is at most \( q^{-(d-c)\rho} \). The set of \( g_i \in I_d \) such that \( f_i + g_i \) vanishes at \( P \) is either empty or a coset of this set of polynomials \( h \), so \( \text{Prob}(f_i + g_i \text{ vanishes at } P) \leq q^{-e} \). Hence \( \text{Prob}(f + g \text{ vanishes at } P) \leq q^{-(N+1)e} \). Summing over all \( P \in U_{\rho \leq \gamma \leq d-c} \) and using the trivial bound that \( U \) contains \( O(q^{Nc}) \) closed points of degree \( c \) yields

\[
\sum_{c = \rho}^{d-c} O(q^{Nc})q^{-(N+1)e} = O(q^{-\rho}).
\]

\( \square \)

3.3 Points of high degree

**Lemma 3.3.** Given a closed subvariety \( Z \subset \mathbb{P}^n \) such that \( \dim Z \cap U > 0 \), the probability that a random \( h \in I_d \) vanishes identically on \( Z \) is at most \( q^{-(d-c)} \).

**Proof.** Choose \( P \in (Z \cap U)_{>d-c} \). If \( h \) vanishes on \( Z \), it vanishes at \( P \). By [Poo08, Lemma 4.1], \( \text{Prob}(h(P) = 0) \leq q^{-(d-c)} \).

\( \square \)
LEMMA 3.4 (Points of high degree). We have
\[ \text{Prob}(V(f + g) \cap U_{d-c} = \emptyset) = 1 - o(1) \]
as \(d \to \infty\).

Proof. Let \( W_{-1} = \mathbb{P}^n \). For \( i = 0, \ldots, N \), let \( W_i \) be the common zero locus of \( f_0 + g_0, \ldots, f_i + g_i \). We pick \( g_0, \ldots, g_N \) randomly one at a time.

Claim 1: For \( i = -1, \ldots, n - 2 \), conditioned on a choice of \( g_0, \ldots, g_i \) for which \( \dim W_i \cap U = n - i - 1 \), the probability that \( \dim W_{i+1} \cap U = n - i - 2 \) is \( 1 - o(1) \) as \( d \to \infty \).

Proof of Claim 1: We have \( \dim W_{i+1} \cap U = n - i - 2 \) if \( f_{i+1} + g_{i+1} \) does not vanish identically on any irreducible component of \( W_i \cap U \). The number of such components is at most the number of components of \( W_i \), which, by Bézout’s theorem as in [Ful84, p. 10], is at most \( O(d^{i+1}) \). For each component \( Z \) meeting \( U \), the set of \( g_{i+1} \) such that \( f_{i+1} + g_{i+1} \) vanishes identically on \( Z \) is either empty or a coset of the subspace of \( h \in I_d \) vanishing identically on \( Z \), and the probability that \( h \) vanishes on \( Z \) is at most \( q^{-(d-c)} \), by Lemma 3.3. Thus the desired probability is at least \( 1 - O(d^{i+1})q^{-(d-c)} = 1 - o(1) \).

Claim 2: Conditioned on a choice of \( g_0, \ldots, g_{n-1} \) for which \( \dim W_{n-1} \cap U \) is finite, \( \text{Prob}(W_n \cap U_{d-c} = \emptyset) = 1 - o(1) \) as \( d \to \infty \).

Proof of Claim 2: By Bézout’s theorem again, \( \#(W_{n-1} \cap U) = O(d^n) \). For each \( P \in W_{n-1} \cap U \), the set of \( g_n \in I_d \) such that \( f_n + g_n \) vanishes at \( P \) is either empty or a coset of the subspace of \( h \in I_d \) vanishing at \( P \). If, moreover, \( \deg P > d - c \), then \( \text{Prob}(h(P) = 0) \leq q^{-(d-c)} \) by [Poo08, Lemma 4.1]. Thus the desired probability is at least \( 1 - O(d^n)q^{-(d-c)} = 1 - o(1) \) as \( d \to \infty \).

Applying Claim 1 inductively and finally Claim 2 shows that with probability \( 1 - o(1) \), we have \( W_n \cap U_{d-c} = \emptyset \) and hence also \( V(f + g) \cap U_{d-c} = \emptyset \) since \( V(f + g) \subseteq W_n \).

3.4 END OF PROOF

Combining Lemmas 3.1, 3.2, and 3.4 shows that for any \( \rho \in \mathbb{Z}_{>0} \),
\[ \text{Prob}(V(f + g) \cap U = \emptyset) = \prod_{P \in U_{<\rho}} \left( 1 - q^{-(N+1)\deg P} \right) - O(q^{-\rho}) - o(1) \]
as \( d \to \infty \). Applying this to larger and larger \( \rho \) completes the proof of Theorem 2.1.

4 A COUNTEREXAMPLE

Here we show that Question 1.2 has a negative answer if we allow \( d = 1 \), even for projective integral varieties over \( k = \mathbb{C} \). Our counterexample is inspired by [BS12, Proposition 2.3].

Documenta Mathematica 18 (2013) 1039–1044
Let $k = \mathbb{C}$. Let $X$ be the image of the morphism $\mathbb{P}^1 \to \mathbb{P}^3$ given by $(x : y) \mapsto (x^4 : x^3y : xy^3 : y^4)$. Let $\phi : X \to X$ correspond under $X \simeq \mathbb{P}^1$ to the automorphism of $\mathbb{P}^1$ given by $(t \mapsto t^2)$. For $r \geq 1$, the self-map $\phi^r$ corresponds to $(t \mapsto t^{2r})$. But this does not preserve the span of $\{x^4, x^3y, xy^3, y^4\}$, since the coefficient of $x^2y^2$ in $(x + ry)^4$ is nonzero. Thus $\phi^r$ cannot be the restriction of an automorphism of $\mathbb{P}^3$.

**Acknowledgements**

I thank Lucien Szpiro for comments. This research was supported by the Guggenheim Foundation and National Science Foundation grants DMS-0841321 and DMS-1069236.

**References**


The Virtual Haken Conjecture

Dedicated to Mike Freedman
on the occasion of his 60th birthday

Ian Agol

with an appendix by

Ian Agol, Daniel Groves, and Jason Manning

Received: April 16, 2012
Revised: July 14, 2013

Communicated by Ursula Hamenstädt

Abstract. We prove that cubulated hyperbolic groups are virtually special. The proof relies on results of Haglund and Wise which also imply that they are linear groups, and quasi-convex subgroups are separable. A consequence is that closed hyperbolic 3-manifolds have finite-sheeted Haken covers, which resolves the virtual Haken question of Waldhausen and Thurston’s virtual fibering question. An appendix to this paper by Agol, Groves, and Manning proves a generalization of the main result of [1].

2010 Mathematics Subject Classification: 57M, 20F67
Keywords and Phrases: hyperbolic, group, Haken

1 Introduction

In this paper, we will be interested in word-hyperbolic fundamental groups of compact non-positively curved (NPC) cube complexes. These notions were introduced in the seminal paper of Gromov [20].
Theorem 1.1. [24, Problem 11.7] [50, Conjecture 19.5] Let $G$ be a word-hyperbolic group acting properly and cocompactly on a CAT(0) cube complex $X$. Then $G$ has a finite index subgroup $F$ acting faithfully and specially on $X$.

Remark 1.2. The conclusion of the theorem means that the cube complex $X/F$ satisfies the special conditions of Haglund-Wise, namely that the hyperplanes are embedded, and there are no self-osculating or inter-osculating hyperplanes [24, Definition 3.2]. We will not delve into the definition in this paper since we are interested in the implications of specialness, but will not be using the hyperplane conditions directly. Suffice it to say that fundamental groups of compact special cube complexes are precisely the convex subgroups of right-angled Artin groups, and therefore inherit many of the nice properties of these groups.

Corollary 1.3. Let $G$ be a non-elementary word-hyperbolic group acting properly and cocompactly on a CAT(0) cube complex $X$. Then $G$ is linear, large, and quasi-convex subgroups are separable.

Proof. Since $G$ is virtually special by Theorem 1.1, linearity follows from [24, Theorem 4.4] and quasi-convex separability follows from [24, Theorem 1.3]. The implication of largeness is well-known, see e.g. [50, Theorem 14.10]. □

Remark 1.4. The condition here that $G$ is word-hyperbolic is necessary, since there are examples of simple groups acting properly cocompactly on a product of trees [11]. By [24, Proposition 7.2], a quasi-convex subgroup $H \leq G$ is represented by a convex subcomplex $Y \subset X$ such that such that $H$ acts cocompactly on $Y$. A group $G$ is large if it has a finite-index subgroup $G' \leq G$ surjecting $\mathbb{Z} \ast \mathbb{Z}$.

Theorem 9.1. Let $M$ be a closed aspherical 3-manifold. Then there is a finite-sheeted cover $\tilde{M} \to M$ such that $\tilde{M}$ is Haken.

Theorem 9.2. Let $M$ be a closed hyperbolic 3-manifold. Then there is a finite-sheeted cover $\tilde{M} \to M$ such that $\tilde{M}$ fibers over the circle. Moreover, $\pi_1(M)$ is LERF and large.

Theorem 9.1 resolves a question of Waldhausen [49]. Moreover, Theorem 9.2 resolves [30, Problems 3.50-51] from Kirby’s problem list, as well as [48, Questions 15-18].

There has been much work on the virtual Haken conjecture before for certain classes of manifolds. These include manifolds in the Snappea census [17], surgeries on various classes of cusped hyperbolic manifolds [2, 3, 4, 9, 13, 14, 15, 31, 38, 39], certain arithmetic hyperbolic 3-manifolds (see [46] and references therein), and manifolds satisfying various group-theoretic criteria [32, 33, 35]. A key breakthrough was made by Kahn and Markovic who proved that every closed hyperbolic 3-manifold contains an immersed quasi-fuchsian surface [28] (see also Lackenby [34] who had previously resolved the arithmetic case).
approach in this paper uses techniques from geometric group theory, and as such does not specifically rely on 3-manifold techniques, although some of the arguments (such as word-hyperbolic Dehn surgery and hierarchies) are inspired by 3-manifold techniques.

Here is a short summary of the approach to the proof of Theorem 1.1. In Section 4 we use a weak separability result (Theorem A.1) to find an infinite-sheeted regular cover $X$ of $X/G$ which has a collection of embedded compact 2-sided walls $W$. This covering space has a finite hierarchy obtained by labeling the walls $W$ with finitely many numbers (which we think of as colors), so that walls with the same color do not intersect, and cutting successively along the walls ordered by their labels to get an infinite collection of “cubical polyhedra”, giving a “hierarchy” for $X$. The goal is to construct a finite-sheeted cover which is “modeled” on this hierarchy for $X$. We first construct a measure on the space of colorings of the wall graph of $X$ in Section 5. We then refine the colors to reflect how each wall is cut up by previous stages of the hierarchy in Section 6. We use the measure to find a solution to certain gluing equations on the colored cubical polyhedra defined by the refined colorings, and use solutions to these equations to get the base case of the hierarchy in Section 7. We glue up successively each stage of the hierarchy, using a gluing theorem 3.1 to glue at each stage after passing to a finite-sheeted cover. The inductive hypotheses and inductive step of the proof of Theorem 1.1 are given in section 8.

Theorem A.1 generalizes the main result of [1], and is proved in the appendix which is joint work with Groves and Manning. The proof of Theorem A.1 relies on the Malnormal Special Quotient Theorem A.10 which is a result of Wise [50, Theorem 12.3].

Acknowledgements: We thank Nathan Dunfield and Dani Wise for helpful conversations, and Martin Bridson, Frédéric Haglund, Yi Liu, Eduardo Martínez-Pedroza, and Henry Wilton for comments on an earlier draft. We also thank the referee for many comments.

2 Definitions

We expect the reader to be familiar with non-positively curved (NPC) cube complexes [10], special cube complexes [24], and hyperbolic groups [20].

Definition 2.1. A flag simplicial complex is a complex determined by its 1-skeleton: for every clique (complete subgraph) of the 1-skeleton, there is a simplex with 1-skeleton equal to that subgraph. A non-positively curved (NPC) cube complex is a cube complex $X$ such that for each vertex $v \in X$, the link $\text{link}_X(v)$ is a flag simplicial complex. If $X$ is simply-connected, then $X$ is CAT(0) [10]. More generally, an NPC cube orbicomplex or orbihedron is a pair $(G, X)$, where each component of $X$ is a CAT(0) cube complex and $G \to \text{Aut}(X)$ is a proper cocompact effective action. We will also call such pairs cubulated groups when $X$ is connected. If $G$ is torsion-free, then $X/G$ is...
an NPC cube complex. When $G$ has torsion, we may also think of the quotient $X/G$ as an orbi-space in the sense of Haefliger [22, 23]. The orbihedra we will consider in this paper will have covering spaces which are cube complexes, so they are developable, in which case we can ignore subtleties arising in the theory of general orbi-spaces.

Gluing the cubes isometrically out of unit Euclidean cubes gives a canonical metric on an NPC cube complex.

**Definition 2.2.** Given an NPC cube complex $X$, the wall of $X$ is an immersed NPC cube complex $W$ (possibly disconnected). For each $n$-cube $[-1, 1]^n \cong C \subset X$, take the $n-1$-cubes obtained by cutting the cube in half (setting one coordinate $= 0$), called the hyperplanes of $C$. If a $k$-cube $D$ is a face of an $n$-cube $C$, then there is a corresponding embedding of the hyperplanes of $D$ as faces of the hyperplanes of $C$. Take the cube complex $W$ with cubes given by hyperplanes of the cubes of $X$, and gluings given by inclusion of cube hyperplanes. This cube complex immerses into the cube complex $X$. We will call this immersed cube complex the wall complex of $X$. There is a natural line bundle over $W$ obtained by piecing together the normal bundles of hyperplanes in each cube. If this line bundle is non-orientable, then the wall $W$ is one-sided. Otherwise, it is 2-sided or co-orientable, and there are two possible co-orientations.

**Definition 2.3.** Let $X$ be an NPC cube complex. A subcomplex $Y \subset X$ is locally convex if the embedding $Y \to X$ is a local isometry. Similarly, a combinatorial map $Y \leftrightarrow X$ between NPC cube complexes is called locally convex if it is a local isometry.

The condition of being a local isometry is equivalent to saying that $Y$ is NPC, and for each vertex $v \in Y$, $\text{link}_Y(v) \subset \text{link}_X(v)$ is a very full subcomplex, which means that for any two vertices of $\text{link}_Y(v)$ which are joined by an edge in $\text{link}_X(v)$, they are also joined by an edge in $\text{link}_Y(v)$. For example, an embedded cube in an NPC cube complex is a locally convex subcomplex.

**Definition 2.4 (Almost malnormal Collection).** A collection of subgroups $H_1, \ldots, H_g$ of $G$ is almost malnormal provided that $|H_i \cap H_j| < \infty$ unless $i = j$ and $g \in H_i$.

For example, finite collections of non-conjugate maximal elementary subgroups of a torsion-free hyperbolic group form an almost malnormal collection.

**Definition 2.5.** Let $X$ be an NPC cube complex, $Y \subset X$ a locally convex subcomplex. We say that $Y$ is acylindrical if any map $(S^1 \times [0, 1], S^1 \times \{0, 1\}) \to (X, Y)$ which is injective on $\pi_1$ is relatively homotopic to a map $(S^1 \times [0, 1], S^1 \times \{0, 1\}) \to (Y, Y)$.

In particular, if $Y \subset X$ is acylindrical, then the collection of subgroups of the fundamental groups of its components form a malnormal collection of subgroups of $\pi_1(X)$. 
Definition 2.6. ([50, Definition 11.5]) Let $QVH$ denote the smallest class of hyperbolic groups that is closed under the following operations.

1. $1 \in QVH$
2. If $G = A \ast_B C$ and $A, C \in QVH$ and $B$ is finitely generated and embeds by a quasi-isometry in $G$, then $G$ is in $QVH$.
3. If $G = A \ast_B$ and $A \in QVH$ and $B$ is f.g. and embeds by a quasi-isometry, then $G$ is in $QVH$.
4. Let $H < G$ with $[G : H] < \infty$. If $H \in QVH$ then $G \in QVH$ (in particular with (1), any finite group $K \in QVH$).

The notation $QVH$ is meant to be an abbreviation for “quasi-convex virtual hierarchy”. A hierarchy for a group is a sequence of decompositions as amalgamated products and HNN extensions, which would hold if item (4) were eliminated. In particular, items (2) and (3) may be replaced by $G$ is a graph of groups, with vertex groups in $QVH$ and edge groups quasi-convex and f.g. in $G$. A motivating example of a group in $QVH$ is a 1-relator group with torsion [50, Theorem 18.1], or a closed-hyperbolic 3-manifold group which contains an embedded quasi-fuchsian surface [50, Section 14].

Remark 2.7. Let $M$ be a connected closed oriented hyperbolic 3-manifold, such that there is an embedded surface $S \subset M$ which is quasi-fuchsian. Then $\pi_1 S < \pi_1 M$ is a quasi-convex subgroup. The subgroups of components of $\pi_1 (M \setminus S)$ are geometrically finite Kleinian groups with non-trivial domain of discontinuity, and we may obtain $G = \pi_1 M$ as an amalgamated product or HNN extension from these groups. Then fundamental groups of 3-manifolds with boundary are Haken and therefore admit a finite-stage hierarchy. The rest of the surfaces in the hierarchy of $M$ will be finitely generated subgroups of these groups, and therefore will also be geometrically finite and quasi-convex in $\pi_1 M$, by a theorem of Thurston [12, Theorem 2.1].

Definition 2.8. ([50, Definition 11.1]) Let $MQH$ denote the smallest class of hyperbolic groups that is closed under the following operations.

1. If $|G| < \infty$, then $G \in MQH$
2. If $G = A \ast_B C$ and $A, C \in MQH$ and $B$ is finitely generated, almost malnormal, and embeds by a quasi-isometry in $G$, then $G$ is in $MQH$.
3. If $G = A \ast_B$ and $A \in MQH$ and $B$ is f.g., almost malnormal, and embeds by a quasi-isometry, then $G$ is in $MQH$.

We will not define special cube complexes in this paper (see [24]). However, we note that a special cube complex with hyperbolic fundamental group has embedded components of the wall subcomplex, and therefore its fundamental group is in the class $QVH$. Moreover, we have
Theorem 2.9. [50, Theorem 13.3] A torsion-free hyperbolic group is in $\mathcal{QVH}$ if and only if it is the fundamental group of a compact virtually special cube complex.

The reader not familiar with virtually special cube complexes or their fundamental groups may therefore take this theorem as the defining property for a virtually special group which will be used in this paper. Also, we note that using subgroup separability of quasi-convex subgroups, Wise’s theorem implies that a group is in $\mathcal{QVH}$ if an only if it is virtually in $\mathcal{MQH}$. We will make use of this form of the theorem, by proving that a cubulated hyperbolic group is virtually in $\mathcal{MQH}$ in the proof of Theorem 1.1.

We state here a lemma which will be used in the case that $G$ has torsion.

Lemma 2.10. Let $G \in \mathcal{QVH}$, and suppose that $G'$ is an extension of $G$ by a finite group $K < G'$, so there is a homomorphism $\varphi : G' \to G$ such that $\ker(\varphi) = K$. Then $G' \in \mathcal{QVH}$.

Proof. First, we remark that we need only check this for central extensions of $G$, since the kernel of the homomorphism $G' \to \text{Aut}(K)$ given by the conjugacy action will be a finite central extension of a finite-index subgroup of $G$. However, this observation does not seem to simplify the argument.

We prove this by induction on the defining properties of $G \in \mathcal{QVH}$ (essentially, on the length of the quasiconvex hierarchy defining the quotient group $G$). Consider the set of all extensions of groups in $\mathcal{QVH}$ by the finite group $K$. We prove that this class lies in $\mathcal{QVH}$ by showing that these groups are closed under the four operations defining $\mathcal{QVH}$.

1. The extension of 1 by $K$ is in $\mathcal{QVH}$ by property (4), so $K \in \mathcal{QVH}$.

2. Suppose $G = A \ast_B C$ and $A, C \in \mathcal{QVH}$ and $B$ is finitely generated and embeds by a quasi-isometry in $G$.

   We see that for $A' = \varphi^{-1}(A), B' = \varphi^{-1}(B), C' = \varphi^{-1}(C)$, then $\varphi$ is a quasi-isometry, so $B' < G'$ embeds quasi-isometrically in $G'$, and is finitely generated since $B$ is. Also, $A', B', C'$ are finite extensions of $A, B, C$ by $K$ respectively. If $A', B', C' \in \mathcal{QVH}$, then $G' \in \mathcal{QVH}$ by condition (2).

3. Suppose $G = A \ast_B$ and $A \in \mathcal{QVH}$ and $B$ is f.g. and embeds by a quasi-isometry.

   We see that for $A' = \varphi^{-1}(A), B' = \varphi^{-1}(B)$, then $\varphi$ is a quasi-isometry, so $B' < G'$ embeds quasi-isometrically in $G'$, and is finitely generated since $B$ is. Also, $A', B'$ are finite extensions of $A, B$ by $K$ respectively. If $A', B' \in \mathcal{QVH}$, then $G' \in \mathcal{QVH}$ by condition (3).
4. Suppose \( H < G \) with \( [G : H] < \infty \) and \( H \in \QVH \).

Then for \( H' = \varphi^{-1}(H) \), we have \( [G' : H'] = [G : H] < \infty \), and \( H' \) is a finite extension of \( H \) by \( K \). So if \( H' \in \QVH \), then \( G' \in \QVH \) by condition (4).

Thus, we see that finite extensions of elements of \( \QVH \) by \( K \) are in \( \QVH \) by induction, so \( G' \in \QVH \). \( \Box \)

3 Virtual gluing

In this section, we introduce a technical theorem which will be used in the proof of Theorem 1.1. We call this a “virtual gluing theorem”, because it will allow us to glue together certain subcomplexes of a cube complex which may agree only up to finite index, by first passing to a finite cover in which they match exactly by an isometry.

**Theorem 3.1.** Let \( X \) be a compact cube complex which is virtually special and \( \pi_1(X) \) hyperbolic (for each component of \( X \) if \( X \) is disconnected). Let \( Y \subset X \) be an embedded locally convex acylindrical subcomplex such that there is an NPC cube orbi-complex \( Y_0 \) and a finite-sheeted cover \( \pi : Y \rightarrow Y_0 \). Then there exists a regular cover \( \overline{X} \rightarrow X \) such that the preimage of \( Y \leftarrow \overline{Y} \subset \overline{X} \) is a regular orbi-cover \( \overline{Y} \rightarrow Y_0 \). In other words, \( \pi_1(\overline{Y}) \triangleleft \pi_1(Y_0) \) for each component, and for each component of \( Y_0 \), the components of \( \overline{Y} \) covering this component are equivalent.

**Remark 3.2.** Keep in mind that all of the complexes in the statement of this theorem may be disconnected.

**Proof.** We construct a bipartite graph of groups \( G \) with vertex groups given by the fundamental groups of components of \( X \) and \( Y_0 \), and edge groups given by the fundamental groups of components of \( Y \). The inclusion \( Y \subset X \) and the covering space \( \pi : Y \rightarrow Y_0 \) gives the inclusion maps of the edge groups into the vertex groups to give the graph of groups. By the Bestvina-Feighn combination theorem [6, Corollary 7], the fundamental group of this graph of groups is a hyperbolic group. Since the subcomplex \( Y \) is acylindrical in \( X \) and \( Y \) is a finite-sheeted cover of \( Y_0 \), the graph of groups \( G \) is 2-acylindrical (cylinders between edge groups have length at most 1). By [29, Theorem 1.1], each edge group is quasi-convex in \( G \). Thus, \( G \) is in \( \QVH \), and therefore is virtually special by Theorem 2.9. Let \( Y' \rightarrow Y_0 \) be a regular finite-sheeted covering space which factors through \( Y \rightarrow Y_0 \). By separability of quasi-convex subgroups [24, Theorem 1.3], there is a finite index subgroup \( G' \triangleleft G \) such that the induced cover of \( Y_0 \) factors through \( Y' \). The induced cover \( \overline{X} \) of \( X \) has the desired property, since it induces a regular orbifold cover \( \overline{Y} \rightarrow Y_0 \). \( \Box \)

**Remark 3.3.** It is possible to give a proof of this theorem using the techniques of [25, Theorem 6.1] rather than citing Theorem 2.9 which depends on [50].
In this section, we will be setting up for a proof of Theorem 1.1 by induction on dimension. As in the hypothesis of Theorem 1.1, let $X$ be a CAT(0) cube complex, $G$ a hyperbolic group acting properly and cocompactly on $X$. Recall that we say that $(G, X)$ is a cubulated hyperbolic group. Since the action of $G$ is proper and cocompact, the cube complex $X$ is finite dimensional, locally finite, and quasi-isometric to $G$. To prove that $G$ is virtually special, we may assume that $G$ acts faithfully on $X$, since properness implies that the subgroup $K$ acting trivially must be finite. If we can show that $G/K$ is in $QVH$, then by Lemma 2.10, we can conclude that $G$ is also in $QVH$. So we will assume that the action of $G$ on $X$ is faithful as well as proper and cocompact. The quotient $X/G$ may be interpreted as an orbihedron [22, 23] if $G$ has torsion. Moreover, there are finitely many orbits of walls $W \subset X$ since the action of $G$ is cocompact. The stabilizer $G_W$ of a wall $W$ acts properly and cocompactly on $W$, and therefore is quasi-isometric to $W$. Therefore $G_W$ is a quasi-convex subgroup of $G$ since $W$ is totally geodesic and therefore convex in $X$, and quasi-convexity is preserved by quasi-isometry of $\delta$-hyperbolic metric spaces. Since the action of $G_W$ on $W$ is proper, the kernel of the action $G_W \to \text{Aut}(W)$ induced by restriction is finite. So there is a finite-index quotient group $G_W \to \overline{G_W}$ which acts faithfully on $W$.

Let $\{W_1, \ldots, W_m\}$ be orbit representatives for the walls of $X$ under the action of $G$. By induction on the maximal dimension of a cube, we may assume that $G_{W_i}$ is virtually special for $1 \leq i \leq m$. Therefore, by Lemma 2.10, $G_{W_i}$ is also virtually special. In particular, for each $i$, there is a finite index torsion-free normal subgroup $G_i' \leq G_{W_i}$, such that $W_i/G_i'$ is a special cube complex. There exists $R > 0$ such that if two walls $W, W' \subset X$ have the property that $d(W, W') > R$, then $|G_W \cap G_{W'}| < \infty$. This follows because the subgroups $G_W, G_{W'}$ are quasi-convex, so their intersection is also quasi-convex [10, Proposition 3.9], and therefore a hyperbolic subgroup [10, Proposition 3.7]. If $G_W \cap G_{W'}$ is infinite, then there exists an infinite order element $g \in G_W \cap G_{W'}$. Since $g$ preserves both $W$ and $W'$, there is an axis for $g$ in both totally geodesic subsets. But since the space $X$ is $\delta$-hyperbolic for some $\delta$, this means that the axes must be distance $\leq 5\delta$ apart [10, Lemma 3.3(2)], and therefore $d(W, W') \leq 5\delta$.

For each $1 \leq i \leq m$, let

$$A_i = \{G_W, gG_W, |d(g(W_i), W_i) \leq 2R\} - \{G_{W_i}\}.$$ 

Then $A_i$ is finite for all $i$, as follows. Let $g \in G$ be an element such that $d(W_i, g(W_i)) \leq R$. Let $\gamma$ be a geodesic of length $\leq R$ connecting $W_i$ and $g(W_i)$. Let $D \subset W_i$ be a fundamental domain for the action of $G_{W_i}$ on $W_i$. Let $w \in G_{W_i}$ be an element such that $w(\gamma) \cap D \neq \emptyset$. Then $d(wg(W_i), D) \leq R$. By local compactness, there are finitely many such translates of $W_i$ of distance $\leq R$ from $D$. Any other $k \in G$ such that $k(W_i) = wg(W_i)$ has the property that $k \in wgG_{W_i}$. Thus, $k \in G_{W_i}gG_{W_i}$. So we see that there are finitely many
double cosets $G_W, gG_W$, such that $d(W_i, g(W_i)) \leq R$.

The following lemma will give us a CAT(0) cube complex $X$ with compact embedded wall components $W$, which is a regular cover of the orbifold complex $X/G$. The construction of this cover will be crucial for the rest of the proof, in that we will model a hierarchy for a finite-sheeted cover of $X/G$ on hierarchies for $X$.

**Lemma 4.1.** We may find a quotient group homomorphism $\phi : G \to \mathcal{G}$ such that for all $1 \leq i \leq m$ and for all $G_W, gG_W \in \mathcal{A}_i$, $\phi(g) \notin \phi(G_W)$ and $\phi(G_W)$ is finite for all $j$. Moreover, we may assume that the action of $G_W \cap \ker(\phi)$ does not exchange the sides of $W_i$ (preserves the co-orientation), and that $\ker(\phi)$ is torsion-free and $X^{(1)}/\ker(\phi)$ contains no closed loops.

**Proof.** For each $W_i$, the set of double cosets $\mathcal{A}_i = \{G_W, gG_W, d(g(W_i), W_i) \leq 2R \} - \{G_W\}$ is finite. Fix an element $g$ such that $G_W, gG_W \in \mathcal{A}_i$. Choose elements $g_1, \ldots, g_m$ such that $g_i = 1$ and $H = (G_W^{g_1}, \ldots, G_W^{g_m}) \cong G_W^{g_1} \ast \cdots \ast G_W^{g_m}$ and $g \notin H$, and $H < G$ is quasiconvex. This may be arranged by standard ping-pong arguments. Then $H$ is virtually special since it is a free product of virtually special groups by the induction hypothesis (see the discussion in the first paragraph of this section). By Theorem A.1, we may find a quotient $\phi_g : G \to \mathcal{G}$ such that $\phi_g(g) \notin \phi(H)$ and $\phi_g(H)$ is finite. Clearly then $\phi_g(G_W)$ is finite for all $j$. Moreover, we may assume that $\ker(\phi_g) \cap G_W$ is contained in the subgroup preserving the orientation on the normal bundle to $W_i$. Let $\mathcal{A}$ be the finitely many double coset representatives for $\cup_i \mathcal{A}_i$ we use in this construction.

Let $\mathcal{T} \subset G$ be a finite set of representatives for each conjugacy class of torsion elements of $G$, such that $\mathcal{T} \cap G_W = \emptyset$ for all $j$, and conjugacy class representatives of any group elements identifying endpoints of edges of $X^{(1)}$. We may also apply the same technique to find for each $g \in \mathcal{T}$ a homomorphism $\psi_g : G \to \mathcal{G}$ such that $\psi_g(g) \neq 1$ and $\psi_g(G_W)$ is finite for all $i$.

Define $\phi$ by $\ker(\phi) = \cap_{g \in \mathcal{A}} \ker(\phi_g) \cap_{g \in \mathcal{T}} \ker(\psi_g)$, then $\phi : G \to \mathcal{G}$ is $G/\ker(\phi)$ has the desired properties. \hfill $\square$

Let $K = \ker(\phi)$, where $\phi$ comes from the previous lemma, and let $X = X/K$.

Let $\mathcal{W}$ be the collection of walls of $X$, so that each wall $W \in \mathcal{W}$ is a compact, embedded, 2-sided component of the walls of $X$. Then $X$ is an NPC cube complex, and for each $i$, the quotient $N_R(W_i)/(G_W \cap K)$ embeds in $X$ under the natural covering map, where $N_R(W_i)$ is the neighborhood of radius $R$ about $W_i$. Recall from the above discussion that $R$ could be taken to be $10\delta$, where $\delta$ is the hyperbolicity constant of $X$.

**Definition 4.2.** Form a graph $\Gamma(X)$, with vertices $V(\Gamma(X))$ consisting of the wall components of $W \subset X$, and edges $E(\Gamma(X))$ consisting of pairs of walls $(W_1, W_2)$ in $X$ such that $d(W_1, W_2) \leq R$. We have a natural action of $\mathcal{G}$ on $\Gamma(X)$.
5 Invariant coloring measures

Let $\Gamma$ be a (simplicial) graph of bounded valence $\leq k$, and let $G$ be a group acting cocompactly on $\Gamma$. Note that the quotient graph $\Gamma/G$ may have loops and multi-edges, so in particular may not be simplicial. We will denote the vertices of $\Gamma$ by $V(\Gamma)$, and the edges by $E(\Gamma) \subseteq V(\Gamma) \times V(\Gamma)$ consisting of the symmetric relation of pairs of adjacent vertices in $\Gamma$, so that $\Gamma$ is defined by the pair $\Gamma = (V(\Gamma), E(\Gamma))$. Since $\Gamma$ is simplicial, it has no loops, and therefore $E(\Gamma)$ does not meet the diagonal of $V(\Gamma) \times V(\Gamma)$.

Definition 5.1. An $n$-coloring of $\Gamma$ is a map $c: V(\Gamma) \to \{1, \ldots, n\} = [n]$ such that for every edge $(u, v) \in E(\Gamma)$, we have $c(u) \neq c(v)$. Let $[n]^{V(\Gamma)}$ be the space of all $n$-colorings of the trivial graph $(V(\Gamma), \emptyset)$, and endow this with the product topology to make it a compact space (Cantor set). Then the space of $n$-colorings of $\Gamma$ is naturally a closed $G$-invariant subspace of $[n]^{V(\Gamma)}$ which we will denote $C_n(\Gamma)$. The set $M(C_n(\Gamma))$ of probability measures on $C_n(\Gamma)$ endowed with the weak* topology is a convex compact metrizable set. Let $M_G(C_n(\Gamma)) \subset M(C_n(\Gamma)) \subset M([n]^{V(\Gamma)})$ denote the $G$-invariant measures.

Since we have assumed that the degree of every vertex of $\Gamma$ is $\leq k$, then clearly $C_{k+1}(\Gamma)$ is non-trivial: order the vertices, and color each vertex inductively by one of $k + 1$ colors not already used by one of its $\leq k$ neighbors.

Theorem 5.2. The set $M_G(C_{k+1}(\Gamma))$ is non-empty, that is, there exists a $G$-invariant probability measure on the space of $k + 1$-colorings of the graph $\Gamma$.

Proof. For a $G$-invariant measure $\nu \in M_G([n]^{V(\Gamma)})$, we want to define a quantity which measures how far $\nu$ is from giving a $G$-invariant coloring measure in $M_G(C_n(\Gamma))$. For an edge $e = (u, v) \in E(\Gamma)$, let $B_e = \{f \in [n]^{V(\Gamma)} \mid f(u) = f(v)\}$. This is the subset of colorings of $V(\Gamma)$ which violate the coloring condition for $\Gamma$ at the edge $e$, so that $C_n(\Gamma) = \cap_{e \in E(\Gamma)} B_e$. Let $\{e_1, \ldots, e_m\} \subseteq E(\Gamma)$ be a complete set of representatives of the orbits of the action of $G$ on $E(\Gamma)$, which exists because we have assumed that the action of $G$ on $\Gamma$ is co-compact. For $\nu \in M_G([n]^{V(\Gamma)})$ define $\text{weight}(\nu) = \sum_{i=1}^{m} \nu(B_{e_i})$. If $\nu$ is a $G$-invariant coloring measure of $\Gamma$, then regarding $\nu \in M_G(C_n(\Gamma)) \subset M_G([n]^{V(\Gamma)})$, we have $\text{weight}(\nu) = 0$. Conversely, if $\text{weight}(\nu) = 0$ for $\nu \in M_G([n]^{V(\Gamma)})$, then $\nu \in M_G(C_n(\Gamma))$. To see this, let $\text{supp}(\nu) \subset [n]^{V(\Gamma)}$ be the support of $\nu$, which is $\cap_{G \text{-compact}, \nu(C) = 1} C$. Let $e \in E(\Gamma)$, then $\nu(B_e) = 0$, since there exists $e_1, g \in G$ such that $e = g(e_1)$, so $\nu(B_e) = \nu(B_{g(e_1)}) = \nu(B_{e_1}) = 0$ by $G$-invariance of $\nu$ and $\text{weight}(\nu) = 0$. Therefore $\text{supp}(\nu) \subset B_e^c$ for all $e \in E(\Gamma)$, and therefore $\text{supp}(\nu) \cap_{e \in E(\Gamma)} B_e^c = C_n(\Gamma)$. So $\nu \in M_G(C_n(\Gamma))$.

Take the uniform measure $\mu_n$ on $[n]^{V(\Gamma)}$, which is the product of $V(\Gamma)$ copies of the uniform measure on $[n]$, so $\mu_n \in M([n]^{V(\Gamma)})$. Clearly $\mu_n$ is $G$-invariant under the action of $G$ on $V(\Gamma)$, $\mu_n \in M_G([n]^{V(\Gamma)})$, since $G$ permutes the uniform measures on $[n]$. We note that for the uniform measure $\mu_n$, we have $\mu_n(B_e) = 1/n$. Then we see that $\text{weight}(\mu_n) = m/n$. 

Documenta Mathematica 18 (2013) 1045–1087
For $n > k + 1$, we define a map $p_n : [n]^{V(\Gamma)} \to [n-1]^{V(\Gamma)}$ which depends on $\Gamma$ and which is $G$-equivariant. For $c \in [n]^{V(\Gamma)}$ and $v \in V(\Gamma)$, define $p_n(c)(v) = c(v)$ if $c(v) < n$, and if $c(v) = n$, then $p_n(c)(v) = \min\{\{1,\ldots,n-1\} - \{c(u)|(u, v) \in E(\Gamma))\}$. Since the degree of $v$ is $\leq k$, this set is non-empty, and has a well-defined minimum which is $\leq k + 1$. In other words, $p_n(c)$ assigns to each vertex colored $n$ the smallest color not used by its neighbors, and otherwise does not change the color. Then $p_n(c)$ has the property that for any two vertices $u, v \in V(\Gamma)$ with $p_n(c)(u) = p_n(c)(v)$, then $c(u) = c(v)$. In particular, if $c$ is an $n$-coloring of $\Gamma$, then $p_n(c)$ is an $n - 1$-coloring of $\Gamma$. This implies that for all measures $\nu \in M_G([n]^{V(\Gamma)})$, \textit{weight}(p_n(\nu)) \leq \textit{weight}(\nu)$, where $p_n(\nu)$ is the push-forward measure. Notice that the map $p_n$ is continuous, since its definition is local, so that the push-forward is well-defined.

This gives a map $P_n : [n]^{V(\Gamma)} \to [k+1]^{V(\Gamma)}$ defined by $P_n = p_{k+1} \circ p_{k+2} \circ \cdots \circ p_n$. We get induced a map $P_n : M([n]^{V(\Gamma)}) \to M([k+1]^{V(\Gamma)})$ by push-forward of measures, and induces a map by restriction $P_n : M_G([n]^{V(\Gamma)}) \to M_G([k+1]^{V(\Gamma)})$ because the maps $p_n$ are $G$-equivariant.

Finally, we have $\textit{weight}(P_n(\mu)) \leq \textit{weight}(\mu)$ for any $\mu \in M_G([n]^{V(\Gamma)})$. In particular, $\textit{weight}(P_n(\mu_\infty)) \leq \textit{weight}(\mu_\infty) = m/n$. Take a subsequence of $\{P_n(\mu_\infty)\}$ converging to a $G$-invariant measure $\mu_\infty \in M_G([k+1]^{V(\Gamma)})$. Then $\textit{weight}(\mu_\infty) = 0$, which implies that $\mu_\infty \in M_G(C_{k+1}(\Gamma))$. □

6 Cube complexes with boundary patterns

In this section, we introduce the cubical barycentric subdivision of a cube complex, cubical polyhedra, and cube complexes with boundary pattern. We then introduce equivalence classes of colored walls, faces, and polyhedra of the cube complex $X$. These will be essential notions in the proof of Theorem 1.1.

Definition 6.1. Given a locally finite cube complex $X$, subdivide each $n$-cube into $2^n$ cubes of half the size to get a cube complex $\hat{X}$ (Figure 1(b)). This is called the cubical barycentric subdivision, and is analogous to the barycentric subdivision of a complex, in that one inserts new vertices in the barycenter of each cube, and connects each new barycenter vertex of each cube to the barycenter vertex of each cube of one higher dimension containing it, then filling in cubes using the flag condition (the difference with the usual barycentric subdivision is that one does not connect barycenter vertices of cubes to cubes of more than one dimension higher containing it). We may then regard the union of the hyperplanes $W \subset X$ as the union of the new topological codimension-one cubes of $\hat{X}$, which is the locally convex immersed complex $\hat{W} \subset X$ spanned by the barycenter vertices of $X$. Consider splitting $X$ along the hyperplanes $W$. By this, we mean remove each hyperplane, getting a disconnected complex, then put in $2^k$ copies of each codimension-$k$ cube that is removed to get a complex $X \setminus W = \hat{X} \setminus \hat{W}$ (see Figure 1(c)). We will think of this as a cube complex “with boundary”, where the boundary consists of the new cubes that were
attached at the missing hyperplanes. What remains are stars of the vertices of \( X \).

![Figure 1: Subdividing and splitting a cube complex](image)

**Definition 6.2.** A *cubical polyhedron* \( \mathcal{P} \) is a CAT(0) cube complex with a distinguished vertex \( v \in \mathcal{P} \) which is contained in every maximal cube. The star of \( v \) is combinatorially equivalent to \( \mathcal{P} \), and \( \mathcal{P} \) is determined by link(\( v \)).

The stars of vertices in an NPC cube complex are cubical polyhedra, and if we split \( \dot{X} \) along all of its hyperplanes, we get a union of stars of vertices and therefore cubical polyhedra.

**Definition 6.3.** A cube complex with boundary pattern is a cube complex \( X \) of bounded dimension together with locally convex subcomplexes \( \{\partial_1 X, \ldots, \partial_n X\} \), \( \partial_i X \subset X \) satisfying the following inductive definition (induct on the maximal dimension cube):

- For each \( i \), there is an isometrically embedded open product neighborhood \( \partial_i X = \partial_i X \times 0 \subset \partial_i X \times [0, 1) \subset X \). In particular, the dimension of each maximal cube of \( \partial_i X \) is one less than the dimension of a cube of \( X \) containing it. For each cube \( C \subset X \), the intersection \( C \cap X_{i_1} \cap \cdots \cap X_{i_j} \) is a (possibly empty) face of \( C \) for all \( \{i_1, \ldots, i_j\} \subset \{1, \ldots, n\} \).

- For each \( i \), the subcomplex \( \partial_i X \) forms a cube complex with boundary pattern \( \{\partial_j X \cap \partial_i X| j \neq i\} \), with induced collar neighborhoods \( (\partial_j X \cap \partial_i X) \times [0, 1) = (\partial_j X \times [0, 1)) \cap \partial_i X \).

What one may keep in mind for this definition is the analogy of a boundary pattern for a hierarchy of a 3-manifold, arising in the work of Haken [26].
If $X$ is a cube complex with boundary pattern $\{\partial_1 X, \ldots, \partial_n X\}$, then each $\partial_i X$ gets a co-orientation of the collar neighborhood $\partial_i X \times [0, 1)$, pointing into $X$ from 0 to 1 (into the cube complex).

**Examples:** Take a graph $X$, and let $\partial_0 X \subset X$ be the vertices of $X$ which have degree 1, then $X$ is a cube complex with boundary pattern $\partial_0 X$.

Take a cubical polyhedron $P(\Gamma)$ associated to a simplicial graph $\Gamma$. For each vertex $v \in \Gamma$, consider the subcomplex defined by $\text{link}(v) \subset \Gamma$, $P(\text{link}(v)) \subset P(\Gamma)$. Then $[0, 1] \times P(\text{link}(v)) \subset P(\Gamma)$. The collection $\{\{1\} \times P(\text{link}(v))| v \in V(\Gamma)\}$ forms a boundary pattern of $P(\Gamma)$. We will denote the union of this collection as $\partial P(\Gamma)$. We will call components $\{1\} \times P(\text{link}(v))$ the facets of $P(\Gamma)$. If a collection of facets of $P(\Gamma)$ have non-trivial intersection, then their intersection is a convex subcomplex we’ll call a face. The minimal faces will be points, which we will call vertices of $P(\Gamma)$.

**Remark 6.4.** The choice of terminology cubical polyhedron is meant to evoke a polyhedron. When $X$ is PL equivalent to a manifold, then each component of $X \setminus W$ is homeomorphic to a ball, with the boundary pattern corresponding to the facets of the polyhedron.

**Definition 6.5.** Let $X$ be a cube complex with boundary pattern $\{\partial_1 X, \ldots, \partial_n X\}$. Suppose there is an isometric involution $\tau : \partial_i X \to \partial_i X$ without fixed points, and with the property that $\tau(\partial_i X \cap \partial_n X) = \partial_i X \cap \partial_n X$ for $i < n$. Then we may form the quotient complex $X/\tau$, where for each cube $c \subset \partial_i X$, $c \neq \tau(c)$, since $\tau$ is fixed-point free, we amalgamate the cubes $c \times [-1,0]$ (changing the collar parameter of $c$ from $[0,1]$ to $[-1,0]$) and $\tau(c) \times [0,1]$ into a single cube isometric to $c \times [-1,1]$. We obtain an induced boundary pattern $\{\partial_i X/\tau|\partial_i X \cap \partial_n X\} | i < n\}$. This operation on $X$ is called gluing a cube complex with boundary pattern.

For the complex $X$ with walls $W$ constructed at the end of Section 4, let $P(X)$ be the set of cubical polyhedra which are stars of vertices, and let $P_1, \ldots, P_p$ be orbit representatives under the action of $\mathcal{G}$ of the cubical polyhedra of $X \setminus W$ which are vertex stars (we will think of these as the polyhedra obtained by splitting $X$ along its walls). Similarly, let $W_1, \ldots, W_w$ be orbit reps. of the walls $W$ under the action of $\mathcal{G}$. Let $\mathcal{F}(X)$ denote the set of all cubical polyhedra of the walls $W$. These are the stars of midpoints of edges of $X$ in $W$. Let $\mathcal{F} = \{F_1, \ldots, F_j\}$ be orbit representatives of the action of $\mathcal{G}$ on $\mathcal{F}(X)$ (we will assume that each $F_i \subset W_j$ for some $j$). There is a canonical map $\text{wall} : \mathcal{F}(X) \to V(\Gamma(X)) = W$ defined by $\text{wall}(F) = W$ if $F \subset W \in V(\Gamma(X))$. Notice that there is a one-to-one correspondence between $P_i$ and the vertices of $X/G$, and between $F_i$ and the edges of $X/G$ since $X/G = X/\mathcal{G}$.

Let $k = \maxdegree(\Gamma(X))$, then $C_{k+1}(\Gamma(X)) \neq \emptyset$. We define an equivalence relation $\simeq$ on $V(\Gamma(X)) \times C_{k+1}(\Gamma(X))$.

**Definition 6.6.** First, for $(v,c), (w,d) \in V(\Gamma(X)) \times C_{k+1}(\Gamma(X))$, if $(v,c) \simeq (w,d)$, then we must have $v = w$ and $c(v) = d(w)$ (so the partition respects

Documenta Mathematica 18 (2013) 1045–1087
the colored vertex type). For each \( v \in V(\Gamma(X)) \), we define the partition of \( \{v\} \times C_k(X) \) by induction on the color of \( v \).

1. If \( c(v) = d(v) = 1 \), then \( (v, c) \simeq (v, d) \).

2. If \( c(v) = d(v) = 2 \) and for all \( w \) such that \( (w, v) \in E(\Gamma(X)) \), we have \( c(w) = 1 \iff d(w) = 1 \), then \( (v, c) \simeq (v, d) \).

\( j \) The \( j \)th inductive step of the definition is given by: If \( c(v) = d(v) = j \) with \( 2 \leq j \leq k+1 \), and if for all \( w \) such that \( (w, v) \in E(\Gamma(X)) \), we have \( (w, c) \simeq (w, d) \) if \( c(w) < j \) or \( d(w) < j \), then \( (v, c) \simeq (v, d) \).

**Remarks on the equivalence relation \( \simeq \):** A coloring determines a hierarchy for \( X \), and an induced hierarchy on each wall of \( X \). The equivalence relation captures how each wall is cut up by previous stages of the hierarchy. This refinement is important for when we reconstruct the hierarchy to make sure after gluing up the \( j \)th level of the hierarchy that the \( j-1 \)st levels and lower are still matching up to finite index.

Notice that the equivalence class of \( (v, c) \) where \( c(v) = j \) depends only on \( c \) restricted to the ball of radius \( j-1 \) about \( v \) in \( \Gamma(X) \). This implies that the equivalence classes are clopen sets as subsets of \( \{v\} \times C_k(X) \). In fact, if we think of the coloring \( c \) as a Morse function on the vertices \( V(\Gamma(X)) \), then the equivalence class of \( (v, c) \) depends only on the “descending subgraph” of \( v \), consisting of the union of all paths in \( \Gamma(X) \) starting at \( v \) in which the values of \( c \) are decreasing.

We now want to define an equivalence relation \( \simeq_F \) on the set \( \mathcal{F}(X) \times C_k(X) \).

**Definition 6.7.** For \( E \in \mathcal{F}(X) \), we decree \( (E, c) \simeq (E, d) \) if \( (wall(E), c) \simeq (wall(E), d) \).

We define an equivalence relation \( \simeq_P \) on \( \mathcal{P}(X) \times C_k(X) \).

**Definition 6.8.** For \( P \in \mathcal{P}(X) \), \( (P, c) \simeq_P (P, d) \) if for every facet \( F \subset \partial P \), \( (F, c) \simeq_F (F, d) \). In particular, the colors \( c(F) \) of the facets \( F \subset \partial P \) depend only on the \( \simeq_P \) equivalence class of \( (P, c) \).

We have an action of \( G \) on each of these equivalence relations, by the action for \( g \in G \) given by \( g \cdot (v, c) = (g \cdot v, c \circ g^{-1}) \), for \( (v, c) \in \mathcal{W} \times C_k(X) \), and a similar formula for the action on faces and polyhedra. There are finitely many \( G \)-orbits of equivalence classes under the action of \( G \), and we may find representatives among \( \{W_1, \ldots, W_w\} \times C_k(X) \), \( \{F_1, \ldots, F_f\} \times C_k(X) \), and \( \{P_1, \ldots, P_p\} \times C_k(X) \).
7 Gluing equations

We will consider weights on equivalence classes of polyhedra \( \omega : \mathcal{P}(\mathcal{X}) \times C_{k+1}(\Gamma(\mathcal{X})) / \simeq_P \to \mathbb{R} \) which are invariant under the action of \( \mathcal{G} \), so that \( \omega(g \cdot (P, c)) = \omega(P, c) \), for all \( g \in \mathcal{G} \) and satisfying the polyhedral gluing equations. A weight \( \omega \) will be determined by its values on a \( \simeq_P \)-equivalence class \( [(P, j)]_\mathcal{P} \), \( 1 \leq j \leq p, c \in C_{k+1}(\Gamma(\mathcal{X})) \), and therefore is determined by finitely many variables. Given polyhedra \( P, P' \subset \mathcal{X} \) sharing a facet \( F \subset \partial P, F \subset \partial P' \), we get an equation on the weights for each equivalence class of \( \{F\} \times C_{k+1}(\Gamma(\mathcal{X})) / \simeq_F \). For each equivalence class \( [(F, c)]_\mathcal{X} \in \{F\} \times C_{k+1}(\Gamma(\mathcal{X})) / \simeq_F \), we have the equation

\[
\sum_{[(P, d)]_\mathcal{P} | (F, d) \simeq_F (F, c)} \omega([(P, d)]_\mathcal{P}) = \sum_{[(P', d)]_\mathcal{P} | (F, d) \simeq_F (F, c)} \omega([(P', d)]_\mathcal{P}).
\]

The polyhedral gluing equations on the polyhedra equivalence class weights are the equations obtained for each equivalence class \( [(F, c)]_\mathcal{X} \). These equations are also \( \mathcal{G} \)-equivariant, so in particular are determined by the equations for equivalence classes \( [(F_i, c)]_\mathcal{X}, 1 \leq i \leq f, c \in C_{k+1}(\Gamma(\mathcal{X})) \). Thus, we have finitely many equations determined by equivalence classes \( [(F, c)]_\mathcal{X}, 1 \leq i \leq f \) on finitely many variables \( \omega([(P, c)]_\mathcal{P}), 1 \leq j \leq p, \) together with the equations determined by \( \mathcal{G} \)-invariance.

For a measure \( \mu \in M_\mathcal{G}(C_{k+1}(\Gamma(\mathcal{X}))) \), we get non-negative polyhedral weights \( \mu([(P, c)]_\mathcal{P}) = \mu(\{d \in C_{k+1}(\Gamma(\mathcal{X})) | (P, c) \simeq_P (P, d)\}) \) (and \( \mu([(F, c)]_\mathcal{X}) \) is similarly defined for each facet \( F \)). These weights satisfy the polyhedral gluing equations. Consider a facet \( F = \partial P \cap \partial P' \), and an equivalence class \( [(F, c)]_\mathcal{X} \) which defines a gluing equation. Then using the additivity property of \( \mu \), we have

\[
\sum_{[(P, d)]_\mathcal{P} | (F, d) \simeq_F (F, c)} \mu([(P, d)]_\mathcal{P}) = \mu(\{d | (F, d) \simeq_F (F, c)\}) = \\
\mu([(F, c)]_\mathcal{X}) = \sum_{[(P', d)]_\mathcal{P} | (F, d) \simeq_F (F, c)} \mu([(P', d)]_\mathcal{P}).
\]

So \( \mu \) gives a non-negative real solution to the polyhedral gluing equations.

Since these equations are defined by finitely many linear equations with integral coefficients, there is a non-negative non-zero integral weight function satisfying the polyhedral gluing equations, \( \Omega : \mathcal{P}(\mathcal{X}) \times C_{k+1}(\Gamma(\mathcal{X})) / \simeq_P \to \mathbb{Z}_{\geq 0} \). In the next section we will use \( \Omega \) to create a tower hierarchy which gives a finite-sheeted cover of \( X/G \) in \( \mathcal{QVH} \).

8 Virtually gluing up the hierarchy

Let \( (w, c) \in V(\Gamma(\mathcal{X})) \times C_{k+1}(\Gamma(\mathcal{X})) \). Let \( W^c_j = \cup\{w \in V(\Gamma(\mathcal{X})) | c(w) = j\} \subset W, 1 \leq j \leq k+1 \), the union of walls colored \( j \) by \( c \). Suppose that
1.1. there is a locally convex combinatorial immersion \( \nu \) into \( \hat{X} \). We may think of \( w \) as being immersed in the wall \( \hat{w} \). Then define inductively immersed complexes in \( \hat{w} \) by \( \hat{w}_i = \hat{w}_{i-1} \setminus (\hat{w}_{i-1} \cap W_i) \), for \( 2 \leq i \leq j \). We don’t split \( w \) along \( \hat{w} \) since \( w \subset W_j \). We will use the notation \( w_{i-1} \) is \( \hat{w} \), since the \( j = c(w) \) is implicitly determined (if \( j = 1 \), then \( \hat{w} = w \)). The complex \( \hat{w} \) has a boundary pattern, given by \( \partial_i(\hat{w}) = w \cap W_i, 1 \leq i \leq j-1 \).

**Claim:** If \((w, c) \simeq (w, d)\), then \( \hat{w} = \hat{d} \) (as cube complexes with boundary pattern). In other words, \( \hat{w} \) depends only on the \( \simeq \)-equivalence class \((w, c)\).

This follows because \( \hat{w} \) is determined by \( w \cap W_i, 1 \leq i \leq j-1 \), which depends only on the equivalence class of \((w, c)\) since \( v \) is a component of \( W_i \) with \( w \cap v \neq \emptyset \), then \((w, v) \in E(\Gamma(X))\).

Consider now the symmetries of \( \hat{w} \) which preserve the equivalence class. That is, consider \( Stab(\hat{w}) \leq \mathcal{G} \), given by \( g \in \mathcal{G} \) such that \( g(\hat{w}) = \hat{w} \) (in particular, \( g(w) = w \)) and \( (w, c \circ g^{-1}) = (g(w), c \circ g^{-1}) \simeq (w, c) \). Now define \( \hat{w}_i = \hat{w} / Stab(\hat{w}) \), with its corresponding boundary pattern \( \partial_i(\hat{w}_i) = \partial_i(\hat{w}) / Stab(\hat{w}) \), \( 1 \leq i \leq j-1 \). In general, \( \hat{w}_i \) will be an orbihedron with boundary pattern.

For each \( j, 1 \leq j \leq k + 1 \), let \( \mathcal{Y}_j = \sqcup_{[(w, c) : c(w) = j]} \hat{w}_j \) (where we take precisely one \( \mathcal{G} \)-orbit representative of the equivalence relation \( \simeq \) so that there are only finitely many equivalence classes \([w, c]\) up to the action of \( \mathcal{G} \), and therefore \( \mathcal{Y}_j \) is a compact orbicomplex). The orbicomplex \( \mathcal{Y}_j \) has the property that for each \( \mathcal{G} \)-orbit of equivalence class \([F, c] \) with \( c(F) = j \), there is a unique representative of \((F, c) \) in the complex \( \mathcal{Y}_j \). At two extremes, we have \( \mathcal{Y}_1 = \sqcup \{ W_1 / Stab(W_1), \ldots, W_u / Stab(W_u) \} \), since the equivalence class depends only on the orbit of the walls under the action of \( \mathcal{G} \). We have \( \mathcal{Y}_k+1 = \sqcup \{ (F, c) / Stab([F, c], x) / c(wall(F)) = k + 1 \} \), with boundary pattern \( \partial(F, c) / Stab([F, c], x) = (F \cap W_i) / Stab([F, c], x), 1 \leq i \leq k \).

**proof of Theorem 1.1.** The idea of the proof is to reverse-engineer a (malnormal) quasi-convex hierarchy for a finite-sheeted cover of \( X/G \), proving that \( G \) is in \( \mathcal{M}Q\mathcal{H} \). This hierarchy will be realized as a sequence of graphs of acylindrical spaces. This will imply that \( G \) is virtually special. The hierarchy is in some sense an approximation to hierarchies of \( X \), in that the hierarchies induced on the walls will be covers of hierarchies of walls associated to colorings of \( X \).

We will construct a sequence of (usually disconnected) finite cube complexes \( \mathcal{V}_j, k+1 \geq j \geq 0 \), with boundary pattern \( \{ \partial_1(\mathcal{V}_j), \ldots, \partial_j(\mathcal{V}_j) \} \) which have the following properties:

1. there is a locally convex combinatorial immersion \( \nu : \mathcal{V}_j \rightarrow \hat{X}/G = \hat{X}/\hat{G} \)

2. \( \mathcal{V}_j \) is glued together from copies of \( \mathcal{G} \)-orbits of equivalence classes of polyhedra \( \mathcal{P} (X) \times \mathcal{C}_{k+1} (\Gamma(X)) / \simeq \mathcal{P} \) in such a way that if polyhedra \( \mathcal{P}, \mathcal{P}' \subset \mathcal{V}_j \) share a facet \( F \), then the induced equivalence class of \( F \) is the same. More formally, there is a decomposition of \( \mathcal{V}_j \) into cubical polyhedra \( \{ \mathcal{P}_h \} \), such
that there is a lift \( P_h \to \tilde{X} \) to a polyhedron of \( \tilde{X} \) (well-defined up to the action of \( \mathcal{G} \)) which projects to the map \( \nu_j : P_h \to \tilde{X}/\mathcal{G} \). Moreover, there is a coloring \( c_h \in C_{k+1}(V(\Gamma(\mathcal{X}))) \), with a well-defined equivalence class associated to the lift \( P_h \to \tilde{X} \). If \( P_g, P_h \) share a facet \( F \), so that \( F = \partial P_g \cap \partial P_h \subset \mathcal{V}_j \), then there is a lift \( P_g \cup_F P_h \to \tilde{X} \) which projects to the map \( \nu_j : P_g \cup_F P_h \to \tilde{X}/\mathcal{G} \). We want the colorings to be compatible, in the sense that \( (F, c_g) \sim_{\mathcal{X}} (F, c_h) \). Thus, there is a well-defined map \( c_j : \mathcal{F}(\mathcal{V}_j) \to [k+1] \).

3. The boundary of \( \mathcal{V}_j \) is the union of all facets \( F \) contained in precisely one polyhedron \( \partial P_g \subset \mathcal{V}_j \). Moreover, the boundary pattern \( \partial_i \mathcal{V}_j = \bigcup_{F \in \mathcal{F}(\mathcal{V}_j)} c_j(F) = i, 1 \leq i \leq j \). Thus, a facet \( F \) is an interior facet (contained in the boundary of two polyhedra) if and only if \( c_j(F) > j \).

4. The multiplicities of \( \mathcal{G} \)-orbits of equivalence classes of colored polyhedra making up \( \mathcal{V}_j \) satisfy the polyhedral gluing equations. In particular, for each equivalence class \( [(F, c)]_\mathcal{X}, F = \partial P \cap \partial P' \), the number of lifts \( P_g \to P \) with coloring \( c_g \) which induce equivalent colorings \( (F, c_g) \sim_{\mathcal{X}} (F, c) \) on \( F \) is equal to the number of lifts \( P_h \to P' \) which induce equivalent colorings \( (F, c_h) \sim_{\mathcal{X}} (F, c) \).

5. The complex \( \mathcal{V}_j \) will admit a malnormal quasi-convex hierarchy, realized geometrically by cutting along the walls colored \( j, \ldots, k+1 \). So the components of \( \mathcal{V}_j \) will have fundamental group in \( MQH \), and therefore will be virtually special by Theorem 2.9.

The base case \( \mathcal{V}_{k+1} \) is the collection of equivalence classes of polyhedra given by the solution to the polyhedral gluing equations \( \Omega \) found in the previous section. Recall we proved the existence of \( \Omega : \mathcal{P}(\mathcal{X}) \times C_{k+1}(\Gamma(\mathcal{X}))/\simeq_p \to Z_{\geq 0} \) satisfying the polyhedral gluing equations. For each equivalence class \( [(P_j, c)]_\mathcal{X} \), take \( \Omega(P_j, c) \) copies of \( P_j \), \( 1 \leq j \leq p \), keeping track of the coloring \( c \) associated to each copy of \( P_j \), and take the disjoint union of these to get \( \mathcal{V}_{k+1} \). Each polyhedron has a locally convex map to \( \tilde{X}/\mathcal{G} \), so condition (1) holds. These have the empty gluing, each component of \( \mathcal{V}_{k+1} \) has a lift to \( \tilde{X} \) and coloring determined by the polyhedral equivalence class, so condition (2) holds. Every facet of \( \mathcal{V}_{k+1} \) is a subset of \( \partial \mathcal{V}_{k+1} \), so there are no restrictions on the facets and condition (3) holds. Property (4) holds trivially since \( \Omega \) is a solution to the polyhedral gluing equations.

Now, suppose we have constructed \( \mathcal{V}_j \) with these properties, for \( 1 \leq j \leq k+1 \). Let’s prove the existence of \( \mathcal{V}_{j-1} \). The way that we will do this is to prove that \( \partial_i \mathcal{V}_j \) covers components of \( \mathcal{V}_j \) with degree zero. By degree zero, we mean that for each facet of \( \mathcal{V}_j \), the number of facets of \( \partial \mathcal{V}_j \) which cover the facet and have one co-orientation is equal to the number with the opposite co-orientation, where the co-orientation points into the adjacent polyhedron. Then we will appeal to Theorem 3.1 to take a cover \( \mathcal{V}_j \) of \( \mathcal{V}_j \) which may be glued along \( \partial_i \mathcal{V}_j \) to form \( \mathcal{V}_{j-1} \). We must further check that it satisfies the inductive hypotheses.
CLAIM: $\partial_j V_j$ covers components of $V_j$ with degree zero.

First, note that condition (1) implies that each facet $F$ of $V_j$ is contained in at most two polyhedra of $V_j$, because the map $\nu_j : V_j \to \hat{X}/G$ is locally convex. In particular, the map is injective on links of vertices lifted to $\hat{X}$, and therefore is also injective on links of facets lifted to $\hat{X}$. Since facets of $\hat{X}$ are contained in at most two polyhedra, the same holds for facets of $V_j$. So the gluing given in condition (2) identifies facets of the polyhedra in pairs. As described in condition (3), the facets contained in exactly one polyhedron form the boundary of $V_j$, and therefore a facet of $V_j$ which is not in the boundary of $V_j$ must be contained in precisely two polyhedra of $V_j$. Also, because the map $V_j \to \hat{X}/G$ is locally convex, the link of each polyhedron vertex of $V_j$ is the link of a product of open intervals and half-open intervals. This implies that any path in $\partial_j V_j$ may be deformed to lie in a sequence of adjacent facets of $\partial V_j$, meeting in codimension-one facets of $\partial V_j$. In fact, from the inductive construction, $V_j$ will have a hierarchy of length $k + 1 - j$ that induces such a hierarchy on each boundary component as well.

Consider a polyhedral facet $F$ involved in the boundary pattern $\partial_j V_j$, which by hypothesis (2) has a lift $\hat{F} \to \hat{X}$ and an associated equivalence class $[(F,c)]_\partial$, some $c \in C_{k+1}(\Gamma(X))$. The adjacent polyhedron $\partial \hat{P} \supset \hat{F}$ has an equivalence class $[(\hat{P},d)]_\partial$ that is a polyhedron of $\hat{V}_j$ by property (2) such that $(\hat{F},d) \simeq_\partial (F,c)$. For a facet $F'$ of $\partial \hat{P}$ adjacent to $\hat{F}$ with color $d(F') > j$, there must be an adjacent polyhedron $\hat{P}' \subset V_j$ containing $F' \subset \partial \hat{P}'$, since this facet cannot occur as part of the boundary pattern of $V_j$ by condition (3). Then there is a unique facet $\hat{F}' \subset \partial \hat{P}'$ meeting $F'$ such that $\hat{F}' \cap \hat{F} = \hat{F}'' \cap \hat{F}'$ and by condition (2) $\hat{F} \cup_{\hat{F}' \cap \hat{F}''} \hat{F}'' \subset P \cup_{F'} F' \to X$ is some lift of the map $\nu_j$ (the restriction of the map $\nu_j$ from condition (1)) such that $\hat{F} \cup \hat{F}'' \subset \text{wall}(\hat{F}) \subset \text{wall}(\hat{F}')$, see Figure 2. Let $[(F',d')]_\partial$ be the equivalence class associated to $F'$ (which exists by condition (2)). Then $(F',d') \simeq_\partial (F',d)$ by the condition (2). We have $(\text{wall}(\hat{F}), \text{wall}(\hat{F}')) \in E(\Gamma(X))$. Also, $d(\text{wall}(\hat{F})) < d(\text{wall}(\hat{F}')) < d'(\text{wall}(\hat{F}'))$, by the inductive hypothesis on $V_j$. Since $(F',d) \simeq_\partial (F',d')$, and therefore $(\text{wall}(\hat{F}),d) \simeq (\text{wall}(F'),d')$, we have $(\text{wall}(F),d) = (\text{wall}(\hat{F}'),d) \simeq (\text{wall}(\hat{F}''),d)$ by one of the conditions of the equivalence relation $\simeq$. Also, the lift $\hat{F} \cup_{\hat{F}' \cap \hat{F}''} \hat{F}'' \to \text{wall}(\hat{F})^d \to \text{wall}(\hat{F})$, since $d(F') > j$ by condition (j) of the equivalence relation 6.6.

Take a path $\alpha : I \to \partial_j V_j$ starting at $F$, and going through a sequence of facets $F = F_0, F_1, F_2, \ldots, F_m$, such that $F_i$ is associated to a coloring $d_i$. We may assume each of these facets intersects its neighbors in codimension-one facets of $\partial_j V_j$, by the observation above. We see that once we choose a lift $F \to \text{wall}(F)^d \subset X$, we get a lift $\hat{\alpha} : I \to \text{wall}(F)^d$, and corresponding lifts $F_i \to \text{wall}(F)^d$. Moreover, $(\text{wall}(F),d_0) \simeq (\text{wall}(F),d_i)$. If $\alpha$ is a closed path so that $F_k = F$, then the lift $F_k \to \text{wall}(F)^d$ induces an equivalent coloring of $\text{wall}(F)$. Thus, we see that the lift $F \to \text{wall}(F)^d$ is well-defined up to the action of $\text{Stab}(\text{wall}(F)^d)$, so we get a well-defined lift of the component $Z$ of

\[\text{Documenta Mathematica 18 (2013) 1045–1087}\]
The Virtual Haken Conjecture

\[ \partial_j \mathcal{V}_j \] contains \( F \) to a map \( Z \rightarrow \text{wall}(F)^d/\text{Stab}(\text{wall}(F)^d) = \text{wall}(F)^d_\mathcal{G}. \)

Conversely, if a facet \( F' \subset \partial P \) adjacent to \( F \) is colored \( d(\text{wall}(F')) = i < j = d(\text{wall}(F)) \), then \( F' \) must be part of the boundary pattern \( \partial_j \mathcal{V}_j \) by condition (3). Then \( F' \cap F \subset \partial(i(\partial_j \mathcal{V}_j)) \). Thus, we have a map \( Z \rightarrow \text{wall}(Z)^\mathcal{G} \) which is a covering projection onto the component of its image. The condition (4) ensures that the map \( \partial_j \mathcal{V}_j \rightarrow \mathcal{V}_j \) is degree zero, since for each facet equivalence class \( [(F, c)]_\mathcal{F} \) with \( c(F) = j \), there is a unique representative of the \( \mathcal{G} \)-orbit of \( (F, c) \) in the complex \( \mathcal{V}_j \). Thus, the number of representatives of \( [(F, c)]_\mathcal{F} \) in \( \partial_j \mathcal{V}_j \) with one co-orientation will cancel with the other co-orientation by the gluing equation for the class \( [(F, c)]_\mathcal{F} \). This finishes the proof of the claim that \( \partial_j \mathcal{V}_j \) covers components of \( \mathcal{V}_j \) with degree zero.

Next, we need to show that \( \partial_j \mathcal{V}_j \) is acylindrical in \( \mathcal{V}_j \) in order to apply Theorem 3.1. Suppose that there is an essential cylinder \( (S^1 \times [0, 1], S^1 \times \{0, 1\}) \rightarrow (\mathcal{V}_j, \partial_j \mathcal{V}_j) \). By induction hypothesis (1), there is a map \( \nu_j : \mathcal{V}_j \rightarrow \hat{\mathcal{X}}/G \), inducing a map of the cylinder \( S^1 \times [0, 1] \rightarrow \hat{\mathcal{X}}/G \). We may therefore choose a compatible elevation \( (S^1 \times [0, 1], S^1 \times \{0, 1\}) \rightarrow (\hat{\mathcal{X}}, \mathcal{Y}_0, \mathcal{Y}_q) \), where \( \mathcal{Y}_0, \mathcal{Y}_q \in V(\Gamma(\mathcal{X})) \), which must also be an essential cylinder. Moreover, since \( \mathcal{Y}_0, \mathcal{Y}_q \) are convex in \( \mathcal{X} \), we may assume that the maps \( S^1 \times \{0\} \rightarrow \mathcal{Y}_0, S^1 \times \{1\} \rightarrow \mathcal{Y}_1 \) are geodesic. Since parallel geodesics are distance \( \leq R = 2\delta \) apart in \( \mathcal{X} \) (where \( \delta \) is the hyperbolicity constant of \( \mathcal{X} \)), we may assume that for some \( z \in S^1, z \times [0, 1] \) is a geodesic of length \( \leq R \) connecting \( \mathcal{Y}_0 \) and \( \mathcal{Y}_q \) in \( \mathcal{X} \). Therefore \( (\mathcal{Y}_0, \mathcal{Y}_q) \in E(\Gamma(\mathcal{X})) \), so \( \mathcal{Y}_0 \) and \( \mathcal{Y}_q \) must have distinct colors in any coloring \( c \in C_{k+1}(\Gamma(\mathcal{X})) \), so \( c(\mathcal{Y}_0) \neq c(\mathcal{Y}_q) \). However, there must be a sequence of walls \( \mathcal{Y}_0, \mathcal{Y}_1, \ldots, \mathcal{Y}_q \) such that the geodesic \( z \times [0, 1] \) intersects this sequence of walls. There will also be a sequence of facets \( F_0, F_1, \ldots, F_q, F_i \subset \mathcal{Y}_i \) that the geodesic meets, and sequence of polyhedra \( P_1, \ldots, P_q, \) with \( F_{i-1} \cup F_i \subset \partial P_i, i = 1, \ldots, q \) (see Figure 3). Moreover, \( (\mathcal{Y}_i, \mathcal{Y}_j) \in E(\Gamma(\mathcal{X})) \) for all \( 0 \leq i < j \leq q \) since their distance is \( \leq R \). Associated to each \( P_i \) is an equivalence class of colorings.
[(P_i,d_i)]_p by (2), and since the facets F_1,\ldots, F_{q-1} are interior to V_j, we must have \((F_i,d_i) \simeq (F_i,d_i)\) by (2). In particular, \(d_{i-1}(Y_0) = d_i(Y_0), d_{i-1}(Y_q) = d_i(Y_q), i = 1,\ldots,q\), since \(d_i(Y_i) \neq i\) since \(F_i\) is interior to \(V_j\) by induction hypothesis (3). But then \(d_0(Y_0) = j, d_0(Y_q) = d_q(Y_q) = j\), which contradicts the fact that \(d_q(Y_0) \neq d_0(Y_q)\) since \((Y_0,Y_q) \in E(\Gamma(V))\). Thus, we conclude that the cylinder does not exist, and therefore \(\partial_j V_j\) is acylindrical in \(V_j\).

To recap, we have an acylindrical subcomplex \(\partial_j V_j \subset V_j\). Moreover, the components \(Z\) of \(\partial_j V_j\) are partitioned into equivalence classes determined by the equivalence relation of the equivalence class of \(wall(Z)\) together with coloring. Each component covers a component of \(\text{wall}(Z)^{c_j}\) for some \(\simeq\) equivalence class \([\text{wall}(Z),c_j]\). Thus, there is a union of components \(Z_j \subseteq V_j\) such that there is a cover \(\partial_j V_j \to Z_j\). Moreover, the cover is degree 0 with respect to the co-orientation. We split \(\partial_j V_j = \partial_j V_j^{+} \sqcup \partial_j V_j^{-} \sqcup \partial_j V_j^{0}\), determined on each component by whether the cover of the corresponding component of \(Z_j\) preserves or reverses co-orientation, unless \(\text{Stab}(\text{wall}(Z)^{c_j})\) exchanges the sides of \(\text{wall}(Z)\), in which case we may ignore the orientation and it lies in \(\partial_j V_j^{0}\).

By Theorem 3.1, there is a regular covering space of constant degree \(\tilde{V}_j \to V_j\), with boundary pattern \(\{\partial_1 \tilde{V}_j, \ldots, \partial_j \tilde{V}_j\}\) given by the preimages of \(\partial_j V_j\), such that the induced covering space \(\partial_j \tilde{V}_j \to Z_j\) is regular. Since the degree of the cover is zero, we must have that the covers \(\partial_j \tilde{V}_j \to Z_j\) and \(\partial_j \tilde{V}_j^{\pm} \to Z_j\) are common covers. After gluing the co-oriented components of \(\partial_j \tilde{V}_j\), we may take two copies of the resulting complex, and glue the non-co-oriented components \(\partial_j \tilde{V}_j^{0}\) by co-orientation reversing isometries which exchange the sides in pairs (we’ll rename the 2-fold cover \(\tilde{V}_j\) for simplicity). Thus, there is an isometric involution \(\tau_j : \partial_j V_j \leftrightarrow \partial_j \tilde{V}_j\). We may form the quotient space \(V_{j-1} = V_j/\tau_j\) by gluing the boundary pattern by \(\tau_j\). We need to check that the inductive hypotheses are satisfied for \(V_{j-1}\).

Since the involution \(\tau_j\) reverses co-orientation, we can see that the combinatorial immersion \(\tilde{V}_j \to V_j \to \tilde{X}/\tilde{G}\) extends to an immersion \(V_{j-1} \to \tilde{X}/\tilde{G}\).

Moreover, since \(\tau_j\) is an involution of the boundary pattern, we see that \(V_{j-1}\) has locally convex boundary since \(\partial_j V_j \subset V_j\) has a collar neighborhood, and therefore the map to \(\tilde{X}/\tilde{G}\) is locally convex, so condition (1) is satisfied. The boundary pattern \(\partial_i(\partial_j V_j)\) is preserved by the involution \(\tau_j\), since the color-

---

**Figure 3:** Sequence of walls meeting the geodesic.
ing of the boundary pattern is locally determined by the equivalence classes of walls being glued together for colors $i < j$. We define the boundary pattern of $V_{j-1}$ by $\partial_i V_{j-1} = \partial_i \tilde{V}_j / \tau_j$ (where $\tau_j$ is the restriction of $\tau_j$ to $\partial_i \tilde{V}_j$). The interior facets will all have color $> j - 1$, and the boundary facets will have color $\leq j - 1$. So condition (3) is satisfied. Since we have glued $V_{j-1}$ out of copies of colored polyhedra in a way consistent with the gluing equations, and taking regular covers preserves the gluing equations, conditions (2) and (4) are satisfied. Since by induction, each component of $V_j$ has fundamental group in $\mathcal{MQH}$, so does $\tilde{V}_j$. Each component of $V_j$ is obtained from components of $\tilde{V}_j$ by a graph of spaces with acylindrical edge spaces, which implies that the fundamental group is a graph of groups with vertex groups in $\mathcal{MQH}$ and edge groups malnormal. Thus, the fundamental group of each component of $V_j$ is in $\mathcal{MQH}$, so property (5) holds.

So all of the inductive hypotheses are satisfied.

The complex $V_0$ has trivial boundary pattern, and a locally convex map $V_0 \to \hat{X}/G$. Therefore, this map is a finite-sheeted covering space. Moreover, by property (5) $\pi_1(V_0) \in \mathcal{MQH}$. By [50, Theorem 13.3 or Theorem 11.2] (see also Theorem 2.9), $V_0$ has a finite-sheeted special cover, and thus $X/G$ does. This finishes the proof of Theorem 1.1. □

9 Conclusion

Recall that a Haken 3-manifold is a compact irreducible orientable 3-manifold containing an embedded $\pi_1$-injective surface.

**Theorem 9.1 (Virtual Haken conjecture [49]).** Let $M$ be a closed aspherical 3-manifold. Then there is a finite-sheeted cover $\tilde{M} \to M$ such that $\tilde{M}$ is Haken.

**Theorem 9.2 (Virtual fibering conjecture, Question 18 [48]).** Let $M$ be a closed hyperbolic 3-manifold. Then there is a finite-sheeted cover $\tilde{M} \to M$ such that $\tilde{M}$ fibers over the circle. Moreover, $\pi_1(M)$ is LERF and large.

**Proof of Theorems 9.1 and 9.2.** From the geometrization theorem [44, 43, 40], it is well-known that the virtual Haken conjecture reduces to the case that $M$ is a closed hyperbolic 3-manifold. For a closed hyperbolic 3-manifold, we have the following result of Bergeron-Wise based on work of Kahn-Markovic [28] (and making use of seminal results of Sageev on cubulating groups containing codimension-one subgroups [45]).

**Theorem 9.3.** [5, Theorem 5.3] Let $M$ be a closed hyperbolic 3-manifold. Then $\pi_1 M$ acts freely and cocompactly on a $\text{CAT}(0)$ cube complex.

Now, by Theorem 1.1, $\pi_1(M)$ is virtually special. This implies that $\pi_1 M$ is LERF and large following from the virtual specialness by Cor. 1.3. Therefore
$M$ is virtually Haken, and in fact $M$ is also virtually fibered by [50, Corollary 14.3]. □

We also have the following corollary, resolving a question of Thurston.

**Corollary 9.4 (Question 15 [48]).** Kleinian groups are LERF.

**Proof.** This follows combining 9.2 which proves that compact hyperbolic 3-manifold groups are LERF, together with the implication that therefore all finite-covolume Kleinian groups are LERF by [36, Proposition 5.3]. It is well known that any Kleinian group embeds in a finite covolume Kleinian group [41]. □

### A Appendix: Filling Virtually Special Subgroups

**by Ian Agol, Daniel Groves, and Jason Manning**

This appendix will be devoted to proving the following theorem, which may be regarded as a generalization of the main theorem of [1], with the extra ingredient of the malnormal virtually special quotient Theorem A.10 [50, Theorem 12.3].

**Theorem A.1.** Let $G$ be a hyperbolic group, let $H \leq G$ be a quasi-convex virtually special subgroup. For any $g \in G - H$, there is a hyperbolic group $\mathcal{G}$ and a homomorphism $\phi : G \to \mathcal{G}$ such that $\phi(g) \notin \phi(H)$ and $\phi(H)$ is finite.

**Remark A.2.** The conclusion of this theorem may be regarded as a weak version of subgroup separability. Under the hypotheses of the theorem, $H$ is subgroup separable in $G$ if one may also assume that the quotient group $\mathcal{G}$ is finite.

**Remark A.3.** It ought to be possible to prove this result using the techniques used by Wise to prove [50, Theorem 12.1]. However, we have decided to provide an alternative argument which gives a geometric perspective on the notion of height, and uses hyperbolic Dehn filling arguments from the literature instead of the small-cancellation theory developed in [50].

**Notation A.4.** In this appendix, we will sometimes use the notation $A \leq B$ to indicate that $A$ is a finite-index subgroup of $B$.

The height of a subgroup $H < G$ measures how far away $H$ is from being (almost) malnormal.

**Definition A.5.** Let $H < G$. The height of $H$ in $G$ is the maximum number $n \geq 0$ so that there are distinct cosets $\{g_1H, \ldots, g_nH\}$ with $\bigcap_{i=1}^{n} g_iHg_i^{-1}$ infinite.
Thus a finite subgroup has height 0, an infinite almost malnormal subgroup has height 1, and so on. Let $H$ be a quasiconvex subgroup of a hyperbolic group $G$. That $H$ has finite height was proved in [19] (we give a new proof in Corollary A.40 below). Our proof of Theorem A.1 is by an induction on the height of $H$ in $G$. In order to reduce the height of $H$, we first use $H$ to produce an associated peripheral structure consisting of commensurators of infinite intersections of $H$ with its conjugates. We then perform a Dehn filling on this peripheral structure.

We recall some more facts about quasiconvex subgroups of hyperbolic groups we’ll need to define the peripheral structure:

**Proposition A.6.** Let $G$ be a hyperbolic group.

1. [19, Lemma 2.7] Finite intersections of quasiconvex subgroups of $G$ are quasiconvex.

2. [19, Lemma 2.9] Quasiconvex subgroups of $G$ are finite index in their commensurators.

**Definition A.7.** We now define the malnormal core of $H$ and peripheral system induced by $H$ on $G$. Let $h$ be the height of $H$ in $G$. By [1, Corollary 3.5], there are finitely many $H$-conjugacy classes of minimal infinite subgroups of the form $H \cap H^{g_1} \cap H^{g_2} \cap \cdots \cap H^{g_j}$, where $1 \leq j \leq h$ and $\{g_1 = 1, g_2, \ldots, g_j\}$ are essentially distinct, in the sense that $g_i H = g_i' H$ if and only if $i = i'$. (Here minimality is with respect to inclusion.) These intersections are quasiconvex by Proposition A.6.(1).

Choose one $H$-conjugacy class of each such subgroup in $H$, and replace it with its commensurator in $H$ to obtain a collection of quasi-convex subgroups $D_0$ of $H$. Eliminating redundant entries which are $H$-conjugate, we obtain a collection $D$, which we will call the *malnormal core* of $H$ in $G$. The collection $D$ gives rise to a peripheral system of subgroups $\mathcal{P}$ in $G$ in two steps:

1. Change $D$ to $D'$ by replacing each $D \in D$ with $D' < G$ its commensurator in $G$.

2. Eliminate redundant entries of $D'$ to obtain $\mathcal{P} \subset D'$ which contains no two elements which are conjugate in $G$.

Call $\mathcal{P}$ the *peripheral structure on $G$ induced by $H$*. This peripheral structure is only well-defined up to replacement of some elements of $\mathcal{P}$ by conjugates. On the other hand, replacing $H$ by a commensurable subgroup of $G$ does not affect the induced peripheral structure. We consider two peripheral structures on a group to be the same if the same group elements are parabolic (i.e. conjugate into $\bigcup \mathcal{P}$) in the two structures.

**Remark A.8.** From Proposition A.6.(2) it follows that each of the elements of $D$ or $\mathcal{P}$ contains some such $H \cap H^{g_1} \cap \cdots \cap H^{g_j}$ as a finite index subgroup.
Example A.9. Suppose that \( H < G \) has height 3. If \( U = H \cap H^b \cap H^{ba} \) is a minimal infinite intersection of \( G \)-conjugates of \( H \), then \( U \) contributes at most 3 subgroups to the collection \( \mathcal{D} \), but only one subgroup to \( \mathcal{P} \). Suppose, for example, \( G = \langle a, b \rangle \) is free, and \( H = \langle a, bab^{-1}, b^2ab^{-2} \rangle \). Then \( \mathcal{D} \) consists of three cyclic groups \( \langle a \rangle, \langle bab^{-1} \rangle, \langle b^2ab^{-2} \rangle \). These subgroups are not conjugate in \( H \), but of course they are conjugate in \( G \), so \( \mathcal{P} = \{ \langle a \rangle \} \).

For a second example with slightly different behavior, take \( H = \langle a, ba^2b^{-1} \rangle \). Then \( U = \langle a^2 \rangle \), and we have to take commensurators in \( H \) to obtain \( \mathcal{D} = \{ \langle a \rangle, \langle ba^2b^{-1} \rangle \} \). Once again, \( \mathcal{P} = \{ \langle a \rangle \} \) consists of a single cyclic group.

Wise’s malnormal virtually special quotient theorem [50, Theorem 12.3] is a key tool in our argument:

Theorem A.10. Let \( G \) be a virtually special hyperbolic group. Let \( \{ H_1, \ldots, H_m \} \) be an almost malnormal collection of quasiconvex subgroups. Then there exists finite-index subgroups \( H_i \trianglelefteq H_i, i = 1, \ldots, m \) such that for any further finite index subgroups \( H_i' \trianglelefteq H_i \), the quotient \( G/ \langle H_1', \ldots, H_m' \rangle \) is virtually special.

In the proof of Theorem A.1 (Section A.5), we will combine Theorem A.10 with the Dehn filling results A.43 and A.22 to find a hyperbolic quotient of \( G \) in which the image of \( H \) is virtually special, quasiconvex, and has smaller height than \( H \).

A.1 Relative quasiconvexity and \( H \)-filling

In this subsection, we recall definitions of relative quasiconvexity and \( H \)-filling and state the main new Dehn filling result.

Our arguments use the cusped space \( X(G, \mathcal{P}, S) \) associated to a group \( G \), a system of subgroups \( \mathcal{P} \) and a finite generating set \( S \) for \( G \), assumed to satisfy the compatibility condition that \( S \cap \mathcal{P} \) generates \( P \) for each \( P \in \mathcal{P} \). A definition can be found in [21], but we recall and slightly modify the definition here for convenience. The cusped space is built from a Cayley graph \( \Gamma(G, S) \) for \( G \) by equivariantly attaching combinatorial horoballs to subgraphs of \( \Gamma(G, S) \) corresponding to left cosets of elements of \( \mathcal{P} \). It is convenient in this appendix to modify the definition of these combinatorial horoballs as follows:

Definition A.11. (Combinatorial horball) Let \( W \subset G \) be of the form \( tP \) for \( t \in G, P \in \mathcal{P} \). Let \( S_0 = S \cap P \setminus \{1\} \), and for \( n > 0 \), let \( S_n = S_{n-1} \cup \{ s_1s_2 \neq 1 \mid s_1, s_2 \in S_{n-1} \} \). Thus for each \( n \), \( S_n \) is a generating set for \( P \). Define \( H(W) \) to be a 1–complex with vertex set \( V = W \times \mathbb{Z}_{\geq 0} \), and edges as follows:

1. (vertical) For each \((w, n) \in V\), there is an edge from \((w, n)\) to \((w, n + 1)\).
2. (horizontal) For each \((w, n) \in V\), and each \( s \in S_n \) there is an edge from \((w, n)\) to \((ws, n)\). In particular, if \( s \) has order 2, then there is a pair of edges between \((w, n)\) and \((ws, n)\).
Appendix: Filling Virtually Special Subgroups

Definition A.12. (Cusped space) Note that the full subgraph of \( \Gamma(G, S) \) on \( W \) is the same as the full subgraph of \( H(W) \) on \( W \times \{ 0 \} \), so there is a natural way to glue \( H(W) \) to \( W \subset \Gamma(G, S) \). The cusped space \( X(G, \mathcal{P}, S) \) is obtained from \( \Gamma(G, S) \) by gluing on \( H(W) \) for all cosets \( tP \) with \( t \in G \) and \( P \in \mathcal{P} \).

Remark A.13. Using the modified definition A.11 to construct \( X(G, \mathcal{P}, S) \) gives a cusped space with the same distance function on vertices as the one constructed in [21]. In particular, the coarse geometry is unchanged, and all the results we will apply from [21, 1, 36, 27] apply in the same way to the cusped space we are using here. In particular, the pair \( (G, \mathcal{P}) \) is relatively hyperbolic if and only if \( X(G, \mathcal{P}, S) \) is Gromov hyperbolic. The advantage of the current construction is that the obvious \( G \)-action on \( X(G, \mathcal{P}, S) \) is free, even if \( G \) contains torsion elements.

Definition A.14. By a horoball of a cusped space \( X(G, \mathcal{P}, S) \), we will always mean either one of the glued-on graphs \( H(W) \), or a full subgraph of such an \( H(W) \) on the vertices \( W \times \mathbb{Z}_{\geq R} \) for some positive integer \( R \).

Definition A.15. (See [1, Section 3]) Let \( G \) be a hyperbolic group and \( H \) a quasi-convex subgroup, and let \( \mathcal{P} \) and \( D \) be the induced peripheral structures on \( G \) and \( H \) described above. Let \( X \) be the cusped space of \( (G, \mathcal{P}) \) and \( Y \) the cusped space of \( (H, D) \) (with respect to choices of generating sets). The inclusion \( \phi: H \to G \) sends peripheral subgroups in \( D \) into (conjugates of) peripheral subgroups in \( \mathcal{P} \), and so induces a proper \( H \)-equivariant Lipschitz map \( \tilde{\phi}: Y \to X \). We say that \( (H, D) \) is \( C \)-relatively quasiconvex in \( (G, \mathcal{P}) \) if \( \tilde{\phi} \) is \( C \)-Lipschitz and has \( C \)-quasiconvex image in \( X \).

In [36, Appendix A] it is explained that the above definition agrees with other notions of relative quasiconvexity, such as those in [27].

The following is proved in [1] under the assumption that \( G \) is torsion-free. It was extended to the general setting in [37].

Proposition A.16. [1, Proposition 3.12],[37, Corollary 1.9] The pairs \( (H, D) \) and \( (G, \mathcal{P}) \) are both relatively hyperbolic and with these peripheral structures \( (H, D) \) is a relatively quasi-convex subgroup of \( (G, \mathcal{P}) \).

Definition A.17. Let \( (H, D) \) be a relatively quasi-convex subgroup of \( (G, \mathcal{P}) \), where \( \mathcal{P} = \{ P_1, \ldots, P_m \} \). Let \( \{ N_i \triangleleft P_i \} \) be given. The quotient

\[
G(N_1, \ldots, N_m) := G/ \triangleleft N_1 \cup \cdots \cup N_m
\]

is a filling of \( (G, \mathcal{P}) \). It is an \( H \)-filling if \( N_i^\mathcal{P} \subset P_i^\mathcal{P} \cap H \) whenever \( H \cap P_i^\mathcal{P} \) is infinite.

Remark A.18. The current definition of \( H \)-filling agrees with the one in [1] only in case \( G \) is torsion-free. As explained in [36, Appendix B], Definition A.17 is the correct extension in case there is torsion.
Remark A.19. As explained in [1, Definition 3.2], an $H$-filling $G(N_1, \ldots, N_m)$ induces a filling $H(K_1, \ldots, K_n)$ of $H$: For each $D_i \in \mathcal{D}$, there is a $c_i \in G$ and $P_i \in \mathcal{P}$ so that $D_i \subseteq c_i P_i c_i^{-1}$. Then $K_i = c_i N_j c_i^{-1} \cap D_i$. The inclusion $H \hookrightarrow G$ induces a homomorphism $H(K_1, \ldots, K_n) \rightarrow G(N_1, \ldots, N_m)$.

Definition A.20. Let $(G, \mathcal{P})$ be a relatively hyperbolic group. We say that a statement $S$ about fillings $G(N_1, \ldots, N_m)$ holds for all sufficiently long fillings if there is a finite set $B \subset \bigcup \mathcal{P}$ so that whenever $G(N_1, \ldots, N_m)$ is a filling so that $\bigcup_{i=1}^m N_i$ does not intersect $B$, then $S$ holds.

Similarly, if $(H, \mathcal{D})$ is a relatively quasiconvex subgroup of $(G, \mathcal{P})$, a statement $S$ holds for all sufficiently long $H$-fillings if there is a finite set $B \subset \bigcup \mathcal{P}$ so that the statement $S$ holds for all $H$-fillings $G(N_1, \ldots, N_m)$ so that $\bigcup_{i=1}^m N_i$ does not intersect $B$.

Obviously if $S$ holds for all sufficiently long fillings, then $S$ holds for all sufficiently long $H$-fillings. The fundamental theorem of relatively hyperbolic Dehn filling can be stated:

Theorem A.21. [42, 16] (cf. [21] in the torsion-free case) Let $G$ be a group and $\mathcal{P} = \{P_1, \ldots, P_m\}$ a collection of subgroups so that $(G, \mathcal{P})$ is relatively hyperbolic, and let $F \subset G$ be finite. Then for all sufficiently long fillings $\phi: G \rightarrow \mathring{G} := G(N_1, \ldots, N_m)$,

1. $\ker(\phi|_{P_i}) = N_i$ for each $P_i \in \mathcal{P}$;
2. $(\mathring{G}, \{\phi(P_1), \ldots, \phi(P_m)\})$ is relatively hyperbolic; and
3. $\phi|_{F}$ is injective.

Our chief new Dehn filling result in this appendix is the following:

Theorem A.22. Let $G$ be hyperbolic, and let $H$ be height $k \geq 1$ and quasiconvex in $G$. Suppose that $\mathcal{D}$ and $\mathcal{P} = \{P_1, \ldots, P_m\}$ are as in Definition A.7. Then for all sufficiently long $H$-fillings

$$\phi: G \rightarrow \mathring{G} := G(N_1, \ldots, N_m)$$

with $N_i \lhd P_i$ finite index for all $i$, the subgroup $\phi(H)$ is quasi-convex of height strictly less than $k$ in the hyperbolic group $\mathring{G}$.

Remark A.23. We proved Theorem A.22 in [1] under the assumption that $G$ was torsion-free. Much of the proof from [1] still works without that assumption, but our argument that height is reduced in the quotient depended on the machinery of Part 2 of [21], in which torsion-freeness is assumed. Our main innovation in this appendix is a completely different proof that height decreases under Dehn filling.
A.2 Geometric finiteness

Geometric finiteness is a dynamical condition. We recall the relevant definitions, which are mostly standard. They originate in the study of Kleinian groups [18]; the fact that hyperbolic and relatively hyperbolic groups can be understood in this way is due to Bowditch [8] and Yaman [51]. The only non-standard terminology in this section is that of weak geometric finiteness, in which finite parabolic subgroups are explicitly allowed.

We also make the following standing assumption (natural in our context): $G$ is finitely generated, and $\mathcal{P}$ is a finite collection of subgroups of $G$.

**Definition A.24.** Let $M$ be a compact metrizable space with at least $3$ points, and let $\Theta(M)$ be the set of unordered distinct triples of points in $M$. Any action of $G$ on $M$ induces an action on $\Theta(M)$. The action of $G$ on $M$ is said to be a convergence group action if the induced action on $\Theta(M)$ is properly discontinuous.

**Definition A.25.** Suppose $G \curvearrowright M$ is a convergence group action. A point $p \in M$ is a conical limit point if there is a sequence $\{g_i\}_{i \in \mathbb{N}}$ and a pair of points $a, b$ so that $g_ip \to b$ but for every $x \in M \setminus \{p\}$, we have $g_ix \to a$.

A point $p$ is parabolic if $\text{Stab}_G(p)$ is infinite but there is no infinite order $g \in G$ and $q \neq p \in M$ so that $\text{Fix}(g) = \{p, q\}$.

A parabolic point $p$ is called bounded parabolic if $\text{Stab}_G(p)$ acts cocompactly on $M \setminus \{p\}$.

**Definition A.26.** The action $G \curvearrowright M$ is geometrically finite if every point in $M$ is a conical limit point or a bounded parabolic point. Say that $(G, \mathcal{P})$ acts geometrically finitely on $M$ if all of the following hold:

1. $G \curvearrowright M$ is a geometrically finite convergence action.
2. Each $P \in \mathcal{P}$ is equal to $\text{Stab}_G(p)$ for some bounded parabolic point $p$.
3. For any bounded parabolic point $p$, the stabilizer $\text{Stab}_G(p)$ is conjugate to exactly one element of $\mathcal{P}$.

Let $X$ be a $\delta$–hyperbolic $G$–space, so that $(G, \mathcal{P})$ acts geometrically finitely on $\partial X$. Then we say that $(G, \mathcal{P})$ acts geometrically finitely on $X$.

It is useful when talking about Dehn filling to allow parabolic subgroups to be finite. We will use the following definitions:

**Definition A.27.** Let $G \curvearrowright M$ be a convergence action, and say that $p \in M$ is a finite parabolic point if $p$ is isolated and has finite stabilizer.

For $\mathcal{P}$ a finite collection of subgroups of $G$, write $\mathcal{P}_\infty$ for the subcollection of infinite subgroups, and $\mathcal{P}_f$ for the subcollection of finite subgroups. Suppose
$G$ acts on the compact metrizable space $M$. Let $M'$ be obtained from $M$ by removing all isolated points. Say that $(G, P)$ acts \emph{weakly geometrically finitely} on $M$ (or that the action is \emph{WGF}) if all of the following occur:

1. $(G, P_\infty)$ acts geometrically finitely on $M'$.
2. Each $P \in P_f$ is equal to $\text{Stab}_G(p)$ for some $p \in M \setminus M'$.
3. Every $p \in M \setminus M'$ is a finite parabolic point, with stabilizer conjugate to exactly one element of $P_f$.

Finally, if $X$ is a $\delta$–hyperbolic $G$–space, then we say that the action of $(G, P)$ on $X$ is \emph{WGF} whenever the action of $(G, P)$ on $\partial X$ is WGF.

\textbf{Proposition A.28.} Let $(G, P)$ be relatively hyperbolic. Then the action of $(G, P)$ on its cusped space is WGF.

Conversely, if $(G, P)$ has a WGF action on a space $M$, then $(G, P)$ is relatively hyperbolic.

\textbf{Proof.} We first handle the case that $P = P_\infty$. In \cite[Theorem 3.25]{21} it is shown that the current definition of relative hyperbolicity is equivalent to several others, in particular to Gromov’s original definition of relative hyperbolicity. The papers of Bowditch \cite{7} and Yaman \cite{51} combine to show that, for $P = P_\infty$, the pair $(G, P)$ is relatively hyperbolic according to Gromov’s definition if and only if $(G, P)$ acts geometrically finitely on a compact perfect metrizable $M$.

The pair $(G, P)$ is relatively hyperbolic if and only if $(G, P_\infty)$ is relatively hyperbolic. Indeed, for any generating set $S$ of $G$, the cusped space $X_\infty = X(G, P_\infty, S)$ quasi-isometrically embeds into $X = X(G, P, S)$. The complement $X \setminus X_\infty$ is composed of combinatorial horoballs based on finite graphs. Thus $X$ is quasi-isometric to $X_\infty$ with rays attached coarsely equivariantly to the Cayley graph of $G$.

Suppose that $(G, P)$ is relatively hyperbolic, so that $X$ is Gromov hyperbolic. Then $X_\infty$ is also Gromov hyperbolic, and $(\partial X)'$ can be canonically identified with $\partial X_\infty$. Thus $G$ acts geometrically finitely on $(\partial X)'$. Moreover, the isolated points of $\partial X$ are in one to one correspondence with the left cosets of elements of $P_f$; the point corresponding to $tP$ has (finite) stabilizer equal to $tP t^{-1}$. Thus $(G, P)$ acts weakly geometrically finitely on $X$.

Conversely, if $(G, P)$ has a WGF action on $M$, then $(G, P_\infty)$ has a geometrically finite action on $M'$, so $(G, P_\infty)$ is relatively hyperbolic. Since $P \setminus P_\infty$ is composed of finite subgroups of $G$, the pair $(G, P)$ is also relatively hyperbolic.

\textbf{Remark A.29.} Given that $(G, P)$ is relatively hyperbolic if and only if $(G, P_\infty)$ is relatively hyperbolic, it is often convenient to simply ignore the possibility of finite parabolics, as for example in \cite{27}. In the present setting it is important...
to keep track of them, as otherwise we would not get uniform control of the geometry of cusped spaces of quotients, as in Theorem A.43 below.

Suppose that $X$ and $Y$ are $\delta$-hyperbolic $G$-spaces. A $G$-equivariant quasi-isometry from $X$ to $Y$ induces a $G$-equivariant homeomorphism from $\partial X$ to $\partial Y$. Since the property of being a (weakly) geometrically finite action on a $\delta$-hyperbolic space is defined in terms of the boundary, we have the following result.

**Lemma A.30.** Suppose that $(G, \mathcal{P})$ admits a WGF action on a $\delta$-hyperbolic space $X$ (as in Definition A.26) and that $f : X \to Y$ is a $G$-equivariant quasi-isometry to another $\delta$-hyperbolic $G$-space. Then the action of $G$ on $Y$ is WGF.

### A.3 Height from multiplicity

**Definition A.31.** $(H, \mathcal{D}) < (G, \mathcal{P})$ is fully quasiconvex if it is relatively quasiconvex and whenever $gDg^{-1} \cap P$ is infinite, for $D \in \mathcal{D}, P \in \mathcal{P}$, then $[P : gDg^{-1}] < \infty$.

In this section, $(G, \mathcal{P})$ is relatively hyperbolic, and $(H, \mathcal{D})$ is a fully quasiconvex subgroup. We allow the possibility that $\mathcal{P}$ and $\mathcal{D}$ are empty.

If $\mathcal{P}$ (and therefore $\mathcal{D}$) is empty, we take $\Gamma$ to be any graph on which $G$ acts freely and cocompactly, and choose $\tilde{*}$ to be some arbitrary vertex. Otherwise, we take $\Gamma$ to be the 1–skeleton of the cusped space $X(G, \mathcal{P}, S)$ (see Definition A.12). In this case $\Gamma$ contains a Cayley graph for $G$, and we take $\tilde{*} = 1 \in G \subset \Gamma$.

**Definition A.32.** Let $R \geq 0$. An $R$-hull for $H$ acting on $\Gamma$ is a connected $H$-invariant sub-graph $\tilde{Z} \subset \Gamma$ so that all of the following hold.

1. $\tilde{*} \in \tilde{Z}$.

2. If $\gamma$ is a geodesic in $\Gamma$ with endpoints in the limit set of $H$, then the $R$–neighborhood of $\gamma$ is contained in $\tilde{Z}$.

3. If $\mathcal{P}$ is nonempty and $B$ is a horoball of $\Gamma$ whose stabilizer in $H$ is infinite, then $B' \subset \tilde{Z}$ where $B'$ is some horoball nested in $B$.

4. The action of $(H, \mathcal{D})$ on $\tilde{Z}$ is WGF.

---

*Below, when applying Theorem A.43 to a quotient of $\Gamma$ by $G$, we will consider a graph which is the cusped space with some extra loops attached (in an equivariant way) to some vertices. It is straightforward to check that our arguments work as written for this slightly different space. In fact, with only a little extra work, one can take $\Gamma$ to be any graph with a free WGF $G$-action, but we decided to stick with the more restrictive setting in the interests of brevity.*
Remark A.33. For $\Gamma$ the Cayley graph of a hyperbolic group $G$, and $H$ a $\lambda$--quasiconvex subgroup, it is easy to show that the $((R+\lambda+5\delta)\text{--}1)$--neighborhood of $H$ is an $R$--hull for $H$ acting on $\Gamma$. We observe in any case that the $R$--neighborhood of a $0$--hull is an $R$--hull. The existence of $R$--hulls for $H$ acting on a cusped space is proved in Lemma A.41.

Example A.34. Let $G<\text{PSL}_2\mathbb{C}$ be a nonuniform lattice, for example a hyperbolic knot group. Let $P$ be a collection of representatives of conjugacy classes of maximal parabolic subgroups. If $\Gamma$ is a cusped space for $(G,P)$, then there is an equivariant quasi-isometry $\psi: \Gamma \to \mathbb{H}^3$. Let $H$ be a geometrically finite subgroup of $G$ which is fully quasiconvex, and let $C_H$ be the convex hull of the limit set of $H$ in $\mathbb{H}^3$. Then, for a sufficiently large neighborhood $N$ of $C_H$, $\psi^{-1}(N)$ has exactly one unbounded $H$--equivariant component $D$. Depending on the quality of the quasi-isometry $\psi$, $D$ may not even be a $0$--hull for $H$, but it can be shown that any sufficiently large neighborhood of $D$ is an $R$--hull for $H$. In general the behavior of the convex hull $C_H$ in $\mathbb{H}^3$ is what we are trying to capture with the $R$--hull in $\Gamma$.

Let $\tilde{Z}$ be an $R$--hull for $H$ acting on $\Gamma$, let $Z = \tilde{Z}/H$ be the quotient of $\tilde{Z}$ by the $H$--action, and let $Y = \Gamma/G$ be the quotient of $\Gamma$ by the $G$--action. If we let $*_H \in Z$ and $* \in Y$ be the images of $\tilde{*}$, we obtain canonical surjections $s: \pi_1(Z,*_H) \to H$ and $s: \pi_1(Y,*) \to G$. Moreover the canonical map

$$i: Z \to Y$$

which is the composition of the inclusion $Z \hookrightarrow \Gamma/H$ with the quotient map $\Gamma/H \to Y = \Gamma/G$ induces the inclusion of $H$ into $G$, in the sense that the diagram

$$\begin{array}{ccc}
\pi_1(Z,*_H) & \xrightarrow{i_*} & \pi_1(Y,*) \\
\downarrow{s} & & \downarrow{s} \\
H & \to & G
\end{array}$$

commutes.

Definition A.35. Let $n > 0$, and define the following subset of $Z^n$:

$$S_n = \{(z_1,\ldots,z_n) \mid i(z_1) = \cdots = i(z_n)\} \setminus \Delta \quad (2)$$

where $\Delta = \{(z_1,\ldots,z_n) \mid z_i = z_j \text{ for some } i \neq j\}$ is the “fat diagonal” of $Z^n$.

Let $s: \pi_1(Z,*_H) \to H$ be the canonical surjection. Let $\varpi_1,\ldots,\varpi_n$ be the $n$ projections of $S_n$ to $Z$.

(In Stallings’ language [47], $S_n$ is that part of the pullback of $n$ copies of $i: Z \to Y$ which lies outside $\Delta$.)

Let $C$ be a component of $S_n$, with a choice of basepoint $p = (p_1,\ldots,p_n)$. For $i \in \{1,\ldots,n\}$ define maps $\tau_{C,i}: \pi_1(C) \to H$ as follows:
Choose a maximal tree $T$ in $Z$. For each vertex $v$ of $Z$, the tree gives a canonical path $\sigma_v$ from $*_H$ to $v$, allowing the fundamental groups of $Z$ at different basepoints to be identified. To simplify notation define $\sigma_i = \sigma_{p_i}$. Now, the map $\varpi_i: C \to Z$ induces a well-defined map $(\varpi_i)_*: \pi_1(C,p) \to \pi_1(Z,*_H)$, taking a loop $\gamma$ based at $p \in C$ to the loop $\sigma_i\varpi_i(\gamma)\sigma_i$ (concatenating from left to right). We define $\tau_{C,i} = s \circ (\varpi_i)_*: \pi_1(C,p) \to H$. Since $H$ acts on $\hat{Z}$ by covering translations, the map $s$ can be seen by lifting paths starting and finishing at the basepoint in the usual way. Once we’ve used the path in the maximal tree to make based loops in $C$ map to paths in $Z$ starting and finishing at the basepoint $*_H$, the same is true of the maps $\tau_{C,i}$. 

**Definition A.36.** The multiplicity of $Z \to Y$ is the largest $n$ so that $S_n$ contains a component $C$ so that for all $i \in 1, \ldots, n$ the group 

$$\tau_{C,i}(\pi_1(C))$$

is an infinite subgroup of $H$.

**Lemma A.37.** For a fixed component $C$ of $S_n$, the groups

$$A_i = \tau_{C,i}(\pi_1(C,p)) < H$$

are conjugate in $G$. Specifically, if $\sigma_i$ are defined as above, and $g_{i,j}$ is represented by the loop $i \circ \sigma_i \cdot i \circ \sigma_j$, then $g_{i,j}A_jg_{i,j}^{-1} = A_i$.

**Proof.** As in the above discussion, the basepoint of $C$ is $p = \{p_1, \ldots, p_n\}$, and for each $i$ there is a canonical path $\sigma_i$ in $T \subset Z$ connecting the basepoint $*_H$ of $Z$ to $p_i$. We also recall the map $i: Z \to Y$ takes $*_H$ to $*$ and induces the inclusion $H < G$ in the sense that the diagram (1) above commutes.

Let $q = i(p_1) = \cdots = i(p_n)$. The paths $i \circ \sigma_i$ all begin at $*$ and end at $q$, so any concatenation of two of them gives a loop in $Y$ representing an element $g_{i,j}$ of $G$ conjugating one of the images of $\pi_1(C,p)$ to another. Precisely, for $i, j \in \{1, \ldots, n\}$ we get an element $g_{i,j}$ represented by $i \circ \sigma_i \cdot i \circ \sigma_j$ so that

$$g_{i,j}\tau_{C,j}(\alpha)g_{i,j}^{-1} = \tau_{C,i}(\alpha), \forall \alpha \in \pi_1(C,p).$$

\[\square\]

We aim in this section for the following:

**Theorem A.38.** Let $R$ be bigger than the quasi-geodesic stability constant $D = D(\delta)$ specified in the proof below. With the above notation, the height of $H$ in $G$ is equal to the multiplicity of $Z \to Y$.

**Remark A.39.** It is instructive to contemplate the proof of this theorem when $G$ is a Kleinian group, and $H$ a geometrically finite subgroup. Then it is not hard to verify that the multiplicity of a convex core of $H$ is equal to the height of $H$. In fact, the arguments in this section are motivated by carrying this geometric argument over to the broader category of hyperbolic groups.
Before doing the proof, we state and prove a corollary.

**Corollary A.40.** [19] The height of a quasiconvex subgroup of a hyperbolic group is finite.

**Proof.** Suppose $H$ is quasiconvex in $G$. Let $\Gamma$ be a Cayley graph for $G$, so that $H \subset \Gamma$ is $\lambda$-quasiconvex. Let $\tilde{Z}$ be the $(R + \lambda + 5\delta)$-neighborhood of $H$ in $\Gamma$. As noted in Remark A.33, $\tilde{Z}$ is an $R$-hull for $H$. Since $G$ (resp. $H$) acts cocompactly on $\Gamma$ (resp. $\tilde{Z}$), the complexes $Z$ and $Y$ are both finite. Thus $S_n$ is empty for large $n$. $\square$

**Proof of Theorem A.38.** We first bound multiplicity from below by height, and then conversely.

(Multiplicity $\geq$ Height): Suppose that $H$ has height $\geq n$. There are then $(H, g_2H, \ldots, g_nH)$ all distinct so that $J = H \cap H^{g_2} \cap \cdots \cap H^{g_n}$ is infinite. Since $(G, \mathcal{P})$ is relatively hyperbolic, every infinite subgroup of $G$ either contains a hyperbolic element or is conjugate into some $P \in \mathcal{P}$. This follows immediately from the classification of isometries of $\delta$-hyperbolic spaces (see [20, Section 8.2, p. 211]) and from the definition of WGF action. The proof therefore breaks up naturally into these two cases.

**Case 1.** The intersection $J$ contains a hyperbolic element $a$.

By replacing $a$ by a power we may suppose that $a$ has a $(K, C)$-quasi-geodesic axis $\tilde{\gamma}_a$, where $K$ and $C$ depend only on $\delta$. Quasi-geodesic stability implies that $\tilde{\gamma}_a$ lies Hausdorff distance at most $D$ from a geodesic, where $D$ depends only on $\delta$. So as long as $R > D$, the geodesic $\gamma_a$ lies in $\tilde{Z} \cap g_2\tilde{Z} \cap \cdots \cap g_n\tilde{Z}$.

Let $\pi_Z$ be the natural projection from $\tilde{Z}$ to $Z$. For $t \in \mathbb{R}$, define $\gamma_a : \mathbb{R} \to Z^n$ as follows:

$$
\gamma_a(t) = (\pi_Z(\tilde{\gamma}_a(t)), \pi_Z(g_2^{-1}\tilde{\gamma}_a(t)), \ldots, \pi_Z(g_n^{-1}\tilde{\gamma}_a(t)))
$$

Since $G$ acts freely and the $g_i$ are essentially distinct, $\gamma_a$ misses the diagonal. Since its coordinates differ only by elements of $G$, $\gamma_a$ has image in $S_n$. Moreover projection of $\gamma_a$ to any component gives a loop of infinite order in $H$. Thus we’ve shown that a component of $S_n$ has an element with infinite order projection to $G$, and therefore the multiplicity of $H$ is $\geq n$.

**Case 2.** The intersection $J$ is conjugate into $P \in \mathcal{P}$.

In this case $J$ preserves some horoball $B$ of $\Gamma$. By point (3) in the definition of $R$-hull, there is a horoball $B'$ nested inside $B$ so that $B' \subset \tilde{Z}$. By possibly replacing $B'$ with a horoball nested further inside, we have $B' \subset \tilde{Z} \cap g_2\tilde{Z} \cap \cdots \cap g_n\tilde{Z}$. 

*Documenta Mathematica 18 (2013) 1045–1087*
Appendix: Filling Virtually Special Subgroups

It follows that
\[
A = \{ (\pi_Z(b), \pi_Z(g_2^{-1}(b)), \ldots, \pi_Z(g_n^{-1}(b))) \mid b \in B' \}
\]
lies in some component \( C \) of \( S_n \). Moreover, each \( \tau_{C,i}(\pi_1(A)) < \tau_{C,i}(\pi_1(C)) \) is conjugate to \( J \), hence infinite.

\textit{(height \( \geq \) multiplicity): Suppose the multiplicity of \( Z \to Y \) is \( n \). Let \( C \subset S_n \) be a component with infinite fundamental group, and let \( p = (p_1, \ldots, p_n) \in C \). We define the paths \( \sigma_i \) from \(*_H\) to \( p_i \) as in the discussion before Definition A.36. Recall the homomorphisms \( \tau_{C,i} : \pi_1(C, p) \to H < G \) are defined by \( \tau_{C,i}([\gamma]) = [i \circ (\sigma_i \cdot \gamma_i \cdot \bar{\sigma}_i)] \), for any loop \( \gamma = (\gamma_1, \ldots, \gamma_n) \) based at \( p \) in \( C \). According to Lemma A.37, if we let \( A_i = \tau_{C,i}(\pi_1(C, p)) \), and \( g_{i,j} = [i \circ \sigma_i \cdot i \circ \bar{\sigma}_j] \), then \( A_{[0,1]} = A_i \).

In particular, writing \( g_i = g_{1,i} \), we have
\[
H \cap H^{g_2} \cdots \cap H^{g_n} \supseteq A_1
\]
is infinite. To establish the height of \( H \) is at least \( n \), we need to show that \((1, g_2, \ldots, g_n)\) are essentially distinct.

Let \( \tilde{T} \) be the lift of \( T \) to \( \Gamma \) which includes the point \( \tilde{x} \), and let \( \gamma = (\gamma_1, \ldots, \gamma_i) \) be a loop in \( C \) based at \( p \). For each \( i \), the path \( \sigma_i \) has a unique lift to \( \tilde{T} \). Let \( \tilde{\gamma}_i \) be the unique lift of \( \gamma_i \) starting at the terminus of \( \bar{\sigma}_i \). Then \( g_i(\gamma_i) = \gamma_i \).

Since \( p \in C \) lies outside the fat diagonal of \( Z^n \), the paths \( \gamma_1, \ldots, \gamma_n \) are all distinct. In particular, the lifts \( \tilde{\gamma}_1, \ldots, \tilde{\gamma}_n \) are also distinct.

Suppose that \((1, g_2, \ldots, g_n)\) are not essentially distinct. Then we would have \((\text{writing } g_1 = 1) \ g_{i,j} = g_i h \) for some \( 1 \leq i < j \leq n \) and some \( h \in H \). But then \( \gamma_j = h^{-1} \gamma_i \). Projecting back to \( Z \) we have \( \gamma_j = \gamma_i \), contradicting the fact that \( \gamma \) misses the fat diagonal of \( Z^n \). \( \square \)

A.4 Height decreases

Again, we have the setup: \((H, D)\) is fully quasiconvex in \((G, P)\). We now assume that \( D \) and \( P \) are nonempty, and that \( \Gamma = X(G, P, S) \) for some finite generating set \( S \) for \( G \). The graph \( \Gamma \) is acted on weakly geometrically finitely by \((G, P)\). Moreover given any finite generating set \( T \) for \( H \), there is a \( \lambda > 0 \) and an \( H \)-equivariant, proper, \( \lambda \)-Lipschitz map (defined in [1, Section 3])
\[
i : X(H, D, T) \to \Gamma
\]
\(X(H, D, T)\) with \( \lambda \)-quasiconvex image. We briefly recall the definition of this map on the vertex set of \(X(H, D, T)\). Vertices are labeled either by elements
There is some \( h \in H \) or by triples \((hD, k, n)\) so that \( D \in \mathcal{D}, k \in hD, \) and \( n \in \mathbb{N}. \) For \( h \in H \subset X(H, \mathcal{D}, T), \) \( i(h) = h. \) For each \( D \in \mathcal{D} \) choose \( e_D \in G \) and \( P_D \in \mathcal{P} \) so that \( D < e_D P_D e_D^{-1}. \) For a vertex of the form \((hD, k, n)\), define \( i(hD, k, n) \) to be the vertex of \( X(G, \mathcal{P}, S) \) labeled by \((h e_D P_D h e_D, n).\)

The quasi-convexity of the image of \( i \) is the definition of relative quasi-convexity in [1]. This definition is shown to be equivalent to the usual ones in [36, Appendix A].

Recall that the cusped space is built by attaching combinatorial horoballs to a Cayley graph, so there are canonical inclusions \( H \hookrightarrow X(H, \mathcal{D}, T) \) and \( G \hookrightarrow \Gamma. \) The map \( i \) extends the natural inclusion map of \( H \) into \( G. \)

**Lemma A.41.** There is some \( N \) so that the \( N \)-neighborhood of the image of \( i \) is a \( 0 \)-ball for the action of \( H \) on \( \Gamma. \)

*Proof.* There are four conditions to check. We will prove that each of them hold for any large enough value of \( N, \) and then take the maximum of the four lower bounds.

For a number \( M \geq 0 \) and a subset \( A \subset \Gamma, \) let \( N_M(A) \) denote the closed \( M \)-neighborhood of \( A \) in \( \Gamma. \) Let \( Y_M = N_M (i(X(H, \mathcal{D}, T))). \)

**Condition (1):** Since \( \tilde{i} = 1 \in H \subset G, \) we have \( \tilde{i} \in Y_M \) for any \( M \geq 0. \)

**Condition (2):** Suppose that \( \xi_1, \xi_2 \in \Lambda H, \) the limit set of \( H \) in \( \partial \Gamma, \) and suppose that \( l \) is a geodesic between \( \xi_1 \) and \( \xi_2. \) To satisfy the second condition of Definition A.32 (with \( R = 0 \)), we need \( l \) to be contained in \( Y_M \) for large enough \( M. \) The points \( \xi_1 \) and \( \xi_2 \) are limits of elements of \( H. \) Since \( Y_0 \) is \( \lambda \)-quasi-convex, a geodesic between any two elements of \( H \) is contained in \( Y_\lambda. \) It is now straightforward to see that \( l \) is contained in \( Y_{\lambda+\delta}. \)

**Condition (3):** Suppose that \( B \) is a horoball of \( \Gamma \) whose stabilizer in \( H \) is infinite. Algebraically, this gives peripheral subgroups \( D \) of \( H \) and \( P \) of \( G, \) and \( g \in G \) so that \( g D g^{-1} \cap P \) is infinite. The condition from Definition A.31 ensures that \( [P: g D g^{-1}] < \infty. \) This implies that there is some \( M_0 \) so that \( B \subset Y_{M_0}. \) Since there are only finitely many such \( D, P, \) and \( g, \) up to the action of \( G, \) the number \( M_0 \) may be taken to work for all such horoballs \( B. \)

**Condition (4):** The final condition from Definition A.32 is that \( H \) acts weakly geometrically finitely on \( Y_N. \) This is true for any \( N > 0. \) Note that \( Y_N \) is quasi-convex and quasi-isometric to \( Y_0, \) the image of \( i. \) Because the peripheral subgroups of \( H \) are finite index in maximal parabolic subgroups of \( G, \) the map \( i \) is a quasi-isometric embedding. (The proof is similar to the proof that a quasiconvex subgroup of a hyperbolic group is quasi-isometrically embedded.) The map \( i \) therefore gives an \( H \)-equivariant quasi-isometry between \( X(H, \mathcal{D}, T) \) and \( Y_N. \) Since \((H, \mathcal{D})\) acts weakly geometrically finitely on \( X(H, \mathcal{D}, T), \) it follows from Lemma A.30 that \((H, \mathcal{D})\) acts weakly geometrically finitely on \( Y_N. \) \( \square \)

**Definition A.42.** Let \( \tilde{Z}_0 \) be the \( N \)-neighborhood of \( \text{Im}(i) \) for \( N \) sufficiently
large that $\tilde{Z}_0$ is a 0-hull. For $R > 0$, let $\tilde{Z}_R$ be the $R$–neighborhood of $\tilde{Z}_0$. Clearly $\tilde{Z}_R$ is an $R$–hull. Let $Z_R$ be the quotient of $\tilde{Z}_R$ by the $H$–action, and let $Y$ be the quotient of $\Gamma$ by the $G$–action, as in the previous section.

**Theorem A.43.** Let $G$ be hyperbolic, $H < G$ quasiconvex, and let $g \in G \setminus H$. Let $(G,\mathcal{P})$, $(H,\mathcal{D})$ be the relatively hyperbolic structures from Definition A.7. Let $\Gamma = X(G,\mathcal{P},S)$ be a cusped space for $G$. Let $A$ be a finite set in $G$.

Then for all sufficiently long $H$–fillings $\phi$: $G \to \tilde{G} = G(N_1,\ldots,N_m)$:

1. If $K = \ker(\phi)$, then $\Gamma := \Gamma/K$ is $\delta'$–hyperbolic for some $\delta'$ independent of the filling, and is (except for trivial loops) equal to the cusped space for $(G,\mathcal{P})$. In particular $(G,\mathcal{P})$ is relatively hyperbolic.

2. Let $g: \Gamma \to \tilde{\Gamma}$ be the quotient by the $K$–action. For some $\lambda'$ independent of the filling, Image$(g \circ i)$ is $\lambda'$–quasiconvex. Thus the induced filling $(H,\mathcal{D})$ is relatively quasiconvex in $(G,\mathcal{P})$.

3. The induced map $H(K_1,\ldots,K_n) \to G(N_1,\ldots,N_m)$ (described in Remark A.19) is injective.

4. $\phi(g) \notin \phi(H)$.

5. $\phi|_A$ is injective.

**Proof.** We argue the conclusions in order, beginning with (1). It is straightforward to see that $\Gamma/K$ is almost the cusped space for $(G,\mathcal{P})$ as advertised, and we leave the details to the reader. The extra loops come from horizontal edges in horoballs between elements in the same $K$–orbit. Osin in [42] proves that, for all sufficiently long fillings, the constant in the linear isoperimetric inequality of $(G,\mathcal{P},S)$ is at most three times that of $(G,\mathcal{P},S)$. In [1, Proposition 2.3], it is explained how to use this bound to obtain a bound on the hyperbolicity constant of $X(G,\mathcal{P},S)$. That proposition is stated with the hypothesis that $G$ is torsion-free, but this hypothesis is not really necessary; we outline the argument. It is proved in [1, Lemma 2.2] that the linear isoperimetric constant of the coned-off Cayley complex can be estimated in terms of the linear relative isoperimetric constant for $(G,\mathcal{P},S)$. This in turn gives an estimate for the linear isoperimetric constant of the cusped space, with respect to a natural system of 2–cells [21, Theorem 3.24]. From this estimate one obtains an estimate for the hyperbolicity constant $\delta'$ as in [10, Theorem III.H.2.9].

In the torsion-free setting, Claims (2)–(4) are the same as [1, Propositions 4.3–4.5]. As explained in [36, Appendix B], the proof of these claims reduces to the extension of certain technical lemmas [1, 4.1 and 4.2] to the case in which $G$ may contain torsion. The necessary modifications to the proof are explained in [36, Appendix B]. These lemmas together imply the following. Let $K_H$ be the kernel of the filling of $H$ induced by a sufficiently long $H$–filling of $G$. Let $h \in H$ and $h = \phi(h)$. If $d(1,h) = d(1,K_H h)$, and $\gamma$ is a geodesic from 1 to
For any \( h \), then there is a \( 10\delta' \)-local geodesic (in particular a uniform quasi-geodesic) from 1 to \( h \) which is contained in a \( 2 \)–neighborhood of \( q(\gamma) \), and coincides with \( q(\gamma) \) within \( 10\delta + 2d(1,g) \) of the Cayley graph of \( \tilde{G} \).

Given this statement, we sketch the proof of the statement (2) about uniform quasiconvexity; full details are in the proof of [1, Proposition 4.3]. It suffices to consider a geodesic \( \sigma \) from 1 \( \in \tilde{G} \) to \( h = \phi(h) \) for \( h \in H \), and show it lies in a uniform neighborhood of \( \text{Image}(q \circ i) \). We may assume that no element \( h' \) of \( H \cap \phi^{-1}(\tilde{h}) \) is closer to 1 in \( \Gamma \) than \( h \). Choose a geodesic \( \gamma \) from 1 to \( h \) in \( \Gamma \). Using the italicized statement above, there is a \( 10\delta' \)-local geodesic from 1 to \( h \) contained in a \( 2 \)–neighborhood of \( q(\gamma) \). Quasi-geodesic stability then tells us that \( \sigma \) is contained in a uniform neighborhood of the image of \( \gamma \). Since \( H \) was relatively quasiconvex before filling, this implies that \( \sigma \) is contained in a uniform neighborhood of \( \text{Image}(q \circ i) \).

Next we sketch the proof of statement (3); full details are in the proof of [1, Proposition 4.4]. By way of contradiction, we let \( h \in H \) be in the kernel of \( \phi \), but not in \( K_H \), the kernel of the induced filling of \( H \). We may suppose that \( d(1,h) = d(1,K_Hh) \). Let \( \gamma \) be a geodesic from 1 to \( h \). The italicized statement implies that there is a \( 10\delta' \)-local geodesic loop in a \( 2 \)–neighborhood of \( q(\gamma) \), and coinciding with \( q(\gamma) \) in a \( 10\delta' \)-neighborhood of the Cayley graph. There can be no such loop, and we obtain a contradiction.

Next we sketch statement (4); details can be found in the proof of [1, Proposition 4.5]. If \( \phi(g) \in \phi(H) \), we choose \( h \in H \cap \phi^{-1}(\phi(g)) \) minimizing \( d(1,h) \), and let \( \gamma \) be a geodesic from 1 to \( h \). The italicized statement gives us a \( 10\delta' \)-local geodesic from 1 to \( \phi(g) \) whose length is at least \( 20\delta' + 4d(1,g) \). An easy computation gives a lower bound for \( d(1,\phi(g)) > \frac{\delta'}{4}d(1,g) \), which is impossible.

Finally, (5) is part of [42, Theorem 1.1].

**Proposition A.44.** For any \( R' > 0 \), there is an \( R \) so that \( q(\tilde{Z}_R) \) is an \( R' \)-hull for the action of \( \tilde{H} \) on \( \Gamma \), for all sufficiently long fillings. (In particular, \( R \) does not depend on the choice of long filling.)

**Proof.** Let \( \gamma \) be a geodesic joining limit points of \( \tilde{H} \). The \( \lambda' \)-neighborhood of \( W := q(\text{Im}(i)) \) contains \( \gamma \), by quasi-convexity. Thus the \( R \)-neighborhood of \( \gamma \) is contained in the \( R + \lambda' \)-neighborhood of \( W \), hence in the image of \( Z_{R+\lambda'} \).

The other conditions follow from the relative quasiconvexity of \( H \) in \( \tilde{G} \).

Let \( G \to \tilde{G} \) be a sufficiently long filling to satisfy the conclusions of Theorem A.43, so that \( \tilde{\Gamma} \) is \( \delta' \)-hyperbolic, \( \text{Im}(q \circ i) \) is \( \lambda' \)-quasiconvex, and so on. Fix \( R' \) bigger than the constant \( D(\delta') \) from Theorem A.38. Then \( q(\tilde{Z}_R) \) detects height in any sufficiently large filling, in a sense which we will describe below.

**Lemma A.45.** For all sufficiently long fillings \( \phi : G \to G(N_1,\ldots,N_m) \), if \( K = \ker(\phi) \), \( K_H = K \cap H \) and \( k \in K \setminus K_H \), then \( k\tilde{Z}_R \cap \tilde{Z}_R = \emptyset \).
Appendix: Filling Virtually Special Subgroups

Proof. The set $A = \{ g \in G \mid g\bar{Z}_R \cap \bar{Z}_R \neq \emptyset \}$ is a finite union of left cosets of $H$.

$$A = \bigsqcup_{i=0}^{t} g_i H, \ g_0 = 1.$$  

Applying Theorem A.43 for $g = g_1, \ldots, g = g_t$, we conclude that for all sufficiently long fillings, $\phi(g_i) \notin \phi(H)$ for $i > 0$. Equivalently $g_i h \notin K$ for any $h \in H$, and any $i > 0$. Thus for $k \in K \setminus K_H$, we have $k \notin A$, and so $k\bar{Z}_R \cap \bar{Z}_R = \emptyset$.  

Let $\tilde{Z}_R$ be the quotient of $\hat{Z}_R$ by $K_H$, and let $\hat{Z}_R = Z_R$ be the quotient of $\tilde{Z}_R$ by the action of $H$. By Lemma A.45, $\tilde{Z}_R$ embeds in $\hat{\Gamma}$. Now we have a commutative diagram,

$$
\begin{array}{cccc}
  & H & \ar@{^{(}->}[r] & G \\
  \downarrow & \downarrow & \downarrow & \downarrow \\
  \tilde{Z}_R & \ar[r] & \hat{Z}_R & \ar[r] & \hat{\Gamma} \\
  \downarrow & \downarrow & \downarrow & \downarrow \\
  H/K_H & \ar[r] & Z_R & \ar[r] & \bar{\Gamma} \\
  \end{array}
$$

where the horizontal maps are inclusions and the vertical maps are quotients by $K_H$ and $K$ respectively. After taking quotients by the relevant groups we get the diagram,

$$
\begin{array}{cccc}
  & Z_R & \ar[r]^i & Y \\
  \downarrow & \downarrow & \downarrow & \downarrow \\
  \bar{Z}_R & \ar[r] & \bar{Y} \\
  \end{array}
\tag{4}
$$

where the vertical maps are homeomorphisms, and the horizontal maps are immersions inducing the inclusions $H \rightarrow G$ and $\hat{H} \rightarrow \bar{\Gamma}$. Let

$$S_n = \{(z_1, \ldots, z_n) \in Z^n_R \mid i(z_1) = \cdots = i(z_n)\} \setminus \Delta$$

and

$$\bar{S}_n = \{(z_1, \ldots, z_n) \in \bar{Z}^n_R \mid \bar{i}(z_1) = \cdots = \bar{i}(z_n)\} \setminus \Delta.$$  

The maps in diagram (4) induce a bijection of $S_n$ with $\bar{S}_n$. For each $i \in \{1, \ldots, n\}$ and each component $C$ of $S_n$ the projections of $Z^n_R$ to its factors induce maps

$$\tau_{C,i} : \pi_1(C) \rightarrow H,$$

and

$$\bar{\tau}_{C,i} : \pi_1(C) \rightarrow \bar{H}.$$
Since the quotient $Z_R = \mathbb{Z}_R/H$ can also be thought of as $\mathbb{Z}_R/(H/K_H)$, we see that the homomorphisms $\tau_{C,i}$ all factor as $\tau_{C,i} = \phi|_H \circ \tau_{C,i}$, where $\phi$ is the filling map.

In particular, if $\gamma$ is a loop in $S_n$ so that $\tau_{C,i}(\gamma) = \gamma$ is infinite for each $i \in \{1, \ldots, n\}$ then it must be that $\tau_{C,i}(\gamma)$ is already infinite for each $i$. Therefore we have the following result.

**Corollary A.46.** The height of $\bar{H}$ in $\bar{G}$ is at most the height of $H$ in $G$.

We now specialize to the case that $(H, \mathcal{D}) < (G, \mathcal{P})$ comes from a quasiconvex subgroup $H$ of a hyperbolic group $G$, so that $\mathcal{D}$ is the malnormal core of $H$ and $\mathcal{P}$ the induced peripheral structure on $G$.

**Theorem A.47.** Assume $\mathcal{P}$ is the peripheral structure induced on $G$ by the quasiconvex subgroup $H$, and let $G \to G(N_1, \ldots, N_m)$ be a sufficiently long $H$-filling. In case every filling kernel $N_i$ has finite index in $P_i$, the height of $\bar{H}$ in $\bar{G}$ is strictly less than that of $H$ in $G$.

**Proof.** Suppose that $H$ has height $n$ in $G$ and that, contrary to the conclusion, $\bar{H}$ has height $n$ in $\bar{G}$.

Fix $R' > 0$. By Proposition A.44, there is an $R$ so that for any long enough filling the set $q(\mathbb{Z}_R)$ is an $R'$-hull for the action of $H$ in $\bar{G}$. We choose $R'$ large enough so that it satisfies the hypotheses of Theorem A.38. Specifically, we make sure $R' > D(\delta')$ for the universal constant of hyperbolicity $\delta'$ from Theorem A.43.

By Theorem A.38, the multiplicity of the map $\bar{i}: \mathbb{Z}_R \to \bar{Y}$ is $n$. Let $C$ be a component of $S_n$ (with basepoint $\bar{p} = (\bar{p}_1, \ldots, \bar{p}_n)$) so that each of the subgroups $\bar{A}_i = \bar{\tau}_{C,i}(\bar{\pi}_1(C, \bar{p}))$ are infinite.\footnote{We add a bar to our notation in the obvious way in the quotient.} By Lemma A.37, the groups $\bar{A}_i$ are all conjugate in $\bar{H}$, and so there are $\bar{g}_{i,j} \in \bar{G}$ so that $\bar{g}_{i,j} \bar{A}_j \bar{g}_{i,j}^{-1} = \bar{A}_i$. Since $G$ is a hyperbolic group, any infinite subgroup contains an infinite order element, so let $\bar{a} \in A_1$ be such an infinite order element, and suppose that $\gamma_0$ is a loop in $C$ based at $\bar{p}$ so that $\tau_{C,1}(\gamma_0) = \bar{a}$.

Now, consider the diagram (4). The vertical maps are homeomorphisms, and so (as in the above discussion), induce a homeomorphism between $S_n$ and $S_n$. Let $C$ be the component of $S_n$ corresponding to $\bar{C}$, let $p$ be the associated basepoint, and let $\gamma_0$ be the loop in $C$ associated to $\gamma_0$. As in the discussion above, the image of each $\tau_{C,i}(\gamma_0)$ is infinite in $H$. This shows that $a = \tau_{C,1}(\gamma_0)$ is an element of infinite order in the intersection in $G$ of $n$ essentially distinct conjugates of $H$. Thus $a$ lies in a conjugate of an element of $\mathcal{D}$. But each element of $\mathcal{D}$ has finite image in $\bar{G}$, contradicting the assumption that $\bar{a}$ has infinite order. This completes the proof. \(\Box\)

We now prove Theorem A.22.
Appendix: Filling Virtually Special Subgroups

Proof of Theorem A.22. We have $G$ hyperbolic, $H < G$ quasiconvex and height $k \geq 1$. We then have $(H, \mathcal{D})$ relatively quasiconvex in $(G, \mathcal{P})$ where $\mathcal{D}$ is the mahnornal core of $H$, and $\mathcal{P} = \{P_1, \ldots, P_m\}$ is the peripheral structure induced on $G$.

Let $\phi: G \to G(N_1, \ldots, N_m)$ be a sufficiently long $H$–filling, that the conclusions of Theorem A.43 and Theorem A.47 both hold, and suppose $N_i \not\leq P_i$ for each $i$.

By Theorem A.43, $\bar{G} = G(N_1, \ldots, N_m)$ is hyperbolic relative to $\bar{\mathcal{P}} = \{P_1/N_1, \ldots, P_m/N_m\}$, and $(\bar{H}, \bar{\mathcal{D}})$ is relatively quasiconvex in $(\bar{G}, \bar{\mathcal{P}})$. Since all the peripheral subgroups are finite, $\bar{G}$ is hyperbolic, and $\bar{H}$ is a quasiconvex subgroup of $\bar{G}$. By Theorem A.47, the height of $\bar{H}$ in $\bar{G}$ is at most $k - 1$. □

### A.5 Proof of main result

In this subsection we prove the main result of this appendix, first recalling the statement:

**Theorem A.1.** Let $G$ be a hyperbolic group, let $H \leq G$ be a quasi-convex virtually special subgroup. For any $g \in G - H$, there is a hyperbolic group $\mathcal{G}$ and a homomorphism $\phi: G \to \mathcal{G}$ such that $\phi(g) \not\in \phi(H)$ and $\phi(H)$ is finite.

**Proof.** Let $H \leq G$ be quasiconvex and virtually special, and let $g \in G \setminus H$. Let $h$ be the height of $H$ in $G$. We will induct on the height, noting that the height zero ($H$ finite) case holds trivially.

Let $\mathcal{P} = \{P_1, \ldots, P_m\}$ be the peripheral system associated to $H \leq G$, and $\mathcal{D}$ the peripheral system of $H$ from Definition A.7. By Theorem A.10, there are finite-index subgroups $D_j \leq D_j$ for each $D_j \in \mathcal{D}$ such that for any further finite-index subgroups $D_j' \leq D_j$, the quotient $H(D_1', \ldots, D_m') := H/\langle \bigcup_j D_j' \rangle$ is virtually special.

For each $D_j \in \mathcal{D}$, there is some unique $P_j$ and some $g_j$ so that

$$g_j^{-1}D_jg_j \leq P_j.$$

The element $g_j$ is not unique, but if $g_j'$ is another such element, then $g_j^{-1}g_j' \in P_j$. In particular, different $G$–conjugates of $D_j$ in $P_j$ are actually $P_j$– conjugates, so there are only finitely many of them.

Let $P_i \in \mathcal{P}$, and let

$$S_i = \{D_j' \mid D_j \in \mathcal{D}, g \in G, D_j' \leq P_i\}.$$

By the way $\mathcal{D}$ and $\mathcal{P}$ are defined, $S_i$ is never empty. By the argument in the last paragraph, $S_i$ is a finite collection, so $I_i := \bigcap S_i \leq P_i$.

Theorems A.43 and A.22 imply that there is a finite subset $B \subset \bigcup \mathcal{P}$ so that whenever $\phi: G \to G(N_1, \ldots, N_m)$ is an $H$–filling (see Definition A.17) satisfying $(\bigcup N_i) \cap B = \emptyset$ and $N_i \not\leq P_i$, then:
1. The image $\phi(H)$ is quasiconvex of height $< h$ in the hyperbolic group $G(N_1, \ldots, N_m)$. (Theorem A.22)

2. $\phi(H) \cong H(K_1, \ldots, K_n)$, where $H(K_1, \ldots, K_n)$ is the induced filling of $H$, described in Remark A.19. (Theorem A.43(3))

3. $\phi(g) \notin \phi(H)$. (Theorem A.43(4))

Since $H$ is virtually special, it is residually finite. Each $P_i$ is a finite extension of a subgroup of $H$, so each $P_i$ is residually finite. Hence there are normal subgroups $N_i \trianglelefteq P_i$ so that $(\bigcup N_i) \cap B = \emptyset$. These normal subgroups need not define an $H$–filling, but we can instead consider the subgroups

$$N_i' = N_i \cap I_i.$$

Then $\phi: G \to G(N_1', \ldots, N_m')$ is an $H$–filling inducing a filling $H \to H(K_1, \ldots, K_n)$ satisfying the hypotheses of Theorem A.10. In particular, the image $\bar{H}$ of $H$ in $\bar{G} := G(N_1', \ldots, N_m')$ is virtually special. By Theorem A.22, $\bar{G}$ is hyperbolic and $\bar{H} \leq \bar{G}$ is quasiconvex, of height $< h$. Moreover, Theorem A.43 implies $\phi(g) \notin \phi(H)$.

By induction, there is a quotient $\bar{\phi}: \bar{G} \to \bar{G}$ so that $\bar{\phi}(\phi(g)) \notin \bar{\phi}(\phi(H))$ and $\bar{\phi}(\phi(H))$ is finite. \qed

References


Appendix: Filling Virtually Special Subgroups


I. Agol, D. Groves, J. Manning


Appendix: Filling Virtually Special Subgroups


Ian Agol
University of California Berkeley
970 Evans Hall #3840
Berkeley, CA, 94720-3840
USA
ianagol@math.berkeley.edu

Daniel Groves
Department of Mathematics
Statistics, and Computer Science
University of Illinois at Chicago
322 Science and Engineering Offices (M/C 249)
851 S. Morgan St.
Chicago, IL 60607-7045
USA
groves@math.uic.edu

Jason Manning
244 Mathematics Building
Dept. of Mathematics
University at Buffalo
Buffalo, NY 14260-2900
USA
j399m@buffalo.edu

Documenta Mathematica 18 (2013) 1045–1087
Burniat Surfaces III: Deformations of Automorphisms and Extended Burniat Surfaces

I. Bauer and F. Catanese

Received: September 26, 2012
Revised: April 22, 2013

Communicated by Thomas Peternell

Abstract. We continue our investigation of the connected components of the moduli space of surfaces of general type containing the Burniat surfaces, correcting a mistake in part II. We define the family of extended Burniat surfaces with $K_S^2 = 4$, resp. 3, and prove that they are a deformation of the family of nodal Burniat surfaces with $K_S^2 = 4$, resp. 3. We show that the extended Burniat surfaces together with the nodal Burniat surfaces with $K_S^2 = 4$ form a connected component of the moduli space. We prove that the extended Burniat surfaces together with the nodal Burniat surfaces with $K_S^2 = 3$ form an irreducible open set in the moduli space. Finally we point out an interesting pathology of the moduli space of surfaces of general type given together with a group of automorphisms $G$. In fact, we show that for the minimal model $S$ of a nodal Burniat surface ($G = (\mathbb{Z}/2\mathbb{Z})^2$) we have $\text{Def}(S, G) \neq \text{Def}(S)$, whereas for the canonical model $X$ it holds $\text{Def}(X, G) = \text{Def}(X)$. All deformations of $S$ have a $G$-action, but there are different deformation types for the pairs $(S, G)$ of the minimal models $S$ together with the $G$-action, while the pairs $(X, G)$ have a unique deformation type.

2010 Mathematics Subject Classification: 14J29, 14J25, 14J10, 14D22, 14H30, 32G05

Dedicated to David Mumford.
I. Bauer and F. Catanese

Introduction

In the present article we continue our investigation, begun in [BC09a] and [BC10], of the connected components of the moduli space (of minimal surfaces \( S \) of general type) which contain the Burniat surfaces. We also correct an error in [BC10].

Recall that Burniat surfaces have \( K_S^2 = 6 - m \), \( m = 0, \ldots, 4 \), and the case \( m = 2 \) bifurcates in the subcases: the one of non-nodal Burniat surfaces, and the one of nodal Burniat surfaces. For \( m = 3 \) Burniat surfaces are three-nodal (this means that their canonical model has three nodes).

The main goals that we achieve in this paper are the following:

1. We define the family of extended Burniat surfaces for \( K_S^2 = 3 \), resp. \( 4 \), and prove that they are a deformation of the family of nodal Burniat surfaces with \( K_S^2 = 3 \), resp. \( 4 \).

2. We show that the extended Burniat surfaces with \( K_S^2 = 4 \), together with the nodal Burniat surfaces with \( K_S^2 = 4 \), form a set \( \mathcal{NEB}_4 \) which is a connected component of the moduli space: thereby we correct Theorem 1.1 of [BC10] and simultaneously we answer a question posed on page 562 of [BC10].

3. We show that the extended Burniat surfaces with \( K_S^2 = 3 \), together with the nodal Burniat surfaces with \( K_S^2 = 3 \) form an irreducible open set \( \mathcal{NEB}_3 \) of the moduli space, whose closure \( \overline{\mathcal{NEB}_3} \) consists of bidouble covers of normal cubic surfaces in \( \mathbb{P}^3 \) and is shown in Section 7 to be strictly larger than \( \mathcal{NEB}_3 \).

4. We point out a truly interesting pathology of the moduli space of varieties with a group \( G \) of automorphisms, which is the reason of our mistake mentioned above (Murphy’s law applies then, but in a different way than foreseen).

For nodal Burniat surfaces \( S \), we have a group \( G \cong (\mathbb{Z}/2\mathbb{Z})^2 \) of automorphisms, which is also the group of automorphisms of the canonical model \( X \). But whereas \( \text{Def}(X) = \text{Def}(X, G) \), i.e., all deformations of \( X \) carry along a deformation of the \( G \)-action, \( \text{Def}(S) \neq \text{Def}(S, G) \): thus even if all deformations of \( S \) have a \( G \)-action, the local moduli space \( \text{Def}(S, G) \) for the pairs yields a proper subvariety in the smooth germ \( \text{Def}(S) \).

We refer to [BC09a] and [BC10] for more details concerning investigation of the connected components of the moduli space containing the Burniat surfaces with \( K_S^2 = 6, 5, 4, 2 \), which is fully achieved thanks to the results of the present article.

What remains to be done in order to finish the investigation of Burniat surfaces is to decide, in the case \( K_S^2 = 3 \) of tertiary Burniat surfaces, whether the irreducible component mentioned above is also a connected component, describing in detail all the surfaces which are in the closure and their local deformations.

\(^1\)Namely, the integer \( m \geq 2 \) in Theorem 1.1 is indeed \( +\infty \), and the local moduli space of nodal Burniat surfaces is smooth.
The description of the closure of the irreducible component of the moduli space given by the extended nodal Burniat surfaces with $K_S^2 = 3$ has been carried out by Y. Chen in his Ph. D. thesis (cf. [Ch12]). In [BC10] we proved that 3 of the 4 irreducible families of Burniat surfaces with $K_S^2 \geq 4$, i.e., of primary and secondary Burniat surfaces, are a connected component of the moduli space of surfaces of general type.

In this paper we consider only nodal Burniat surfaces with $K_S^2 = 4$, showing that a general deformation of a nodal Burniat surface with $K_S^2 = 4$, resp. with $K_S^2 = 3$, is an extended Burniat surface, still a bidouble cover (through the bicanonical map) of a normal Del Pezzo surface of degree 4 with one ordinary double point, resp. of a cubic surface with three nodes.

The main results of the present paper are the following:

**Theorem 0.1.**

1) The subset $NEB_4$ of the moduli space of canonical surfaces of general type $M_{1,4}$ given by the union of the open set corresponding to extended Burniat surfaces with $K_S^2 = 4$ with the irreducible closed set parametrizing nodal Burniat surfaces with $K_S^2 = 4$ is a three dimensional irreducible connected component, normal and unirational. Moreover the base of the Kuranishi family of deformations of any such a minimal model $S$ is smooth.

2) The subset $NEB_3$ of the moduli space of canonical surfaces of general type $M_{1,3}$ corresponding to extended and nodal Burniat surfaces with $K_S^2 = 3$ is an irreducible open set, normal, unirational of dimension 4. Moreover the base of the Kuranishi family of deformations of any such a minimal model $S$ is smooth.

A very surprising and new phenomenon occurs for nodal surfaces, confirming Vakil’s ‘Murphy’s law’ philosophy ([Va06]).

To explain what happens for the moduli spaces of extended and nodal Burniat surfaces, let us recall again an old result due to Burns and Wahl (cf. [BW74]).

Let $S$ be a minimal surface of general type and let $X$ be its canonical model. Denote by $Def(S)$, resp. $Def(X)$, the base of the Kuranishi family of $S$, resp. of $X$.

Their result explains the relation between $Def(S)$ and $Def(X)$.

**Theorem (Burns-Wahl).**

Assume that $K_S$ is not ample and let $p : S \to X$ be the canonical morphism. Denote by $L_X$ the space of local deformations of the singularities of $X$ and by $L_S$ the space of deformations of a neighbourhood of the exceptional curves of $p$. Then $Def(S)$ is realized as the fibre product associated to the Cartesian diagram

$$
\begin{array}{ccc}
Def(S) & \longrightarrow & L_S \cong \mathbb{C}^\nu \\
\downarrow \quad & & \quad \downarrow \text{\textlambda} \\
Def(X) & \longrightarrow & L_X \cong \mathbb{C}^\nu,
\end{array}
$$
where $\nu$ is the number of rational $(-2)$-curves in $S$, and $\lambda$ is a Galois covering with Galois group $W := \oplus_{i=1}^r W_i$, the direct sum of the Weyl groups $W_i$ of the singular points of $X$.

An immediate consequence is the following

**Corollary (Burns-Wahl).**

1) $\psi : \text{Def}(S) \to \text{Def}(X)$ is a finite morphism, in particular, $\psi$ is surjective.
2) If $\text{Def}(X) \to \mathcal{L}_X$ is not surjective (i.e., the singularities of $X$ cannot be smoothened independently by deformations of $X$), then $\text{Def}(S)$ is singular.

Assume now that we have $1 \neq G \leq \text{Aut}(S) = \text{Aut}(X)$.

Then we can consider the space $\text{Def}(S,G)$ of local deformations of $S$ together with the $G$-action (by [Cat88] this is the space of $G$-invariant local deformations of $S$), and similarly consider the space $\text{Def}(X,G)$ of local deformations of $X$ and its $G$-action; we have then a natural map $\text{Def}(S,G) \to \text{Def}(X,G)$.

We indeed show here that, unlike the case for the corresponding morphism of local deformation spaces of the surfaces, this map needs not to be surjective; and, as far as we know, the following result gives the first global example of such a phenomenon.

**Theorem 0.2.** The deformations of nodal Burniat surfaces with $K_S^2 = 4,3$ to extended Burniat surfaces with $K_S^2 = 4,3$ yield examples where $\text{Def}(S, (\mathbb{Z}/2\mathbb{Z})^2) \to \text{Def}(X, (\mathbb{Z}/2\mathbb{Z})^2)$ is not surjective.

Moreover, $\text{Def}(S, (\mathbb{Z}/2\mathbb{Z})^2) \subset \text{Def}(S)$, whereas for the canonical model we have: $\text{Def}(X, (\mathbb{Z}/2\mathbb{Z})^2) = \text{Def}(X)$.

The moduli space of pairs, of an extended (or nodal) Burniat surface with $K_S^2 = 4,3$ and a $(\mathbb{Z}/2\mathbb{Z})^2$-action, is disconnected; but its image in the moduli space is a connected open set.

The above phenomenon can already be seen locally around a node, as it will be explained in Section 2. Our results show that the local pathology does indeed globalize.

Our paper is organized as follows: in Section 1 we give the definition of extended Burniat surfaces and describe the respective branch loci of the bidouble covers yielding nodal Burniat surfaces, respectively extended Burniat surfaces.

In the second chapter we analyse bidouble covers of a nodal surface singularity, explaining the phenomenon of Theorem 0.2 locally.

In the third section we show that nodal Burniat surfaces with $K_S^2 = 4,3$ deform to extended Burniat surfaces with $K_S^2 = 4,3$.

Section 4 is instead devoted to the calculation of $H^1(S, \Theta_S)$ for nodal and extended Burniat surfaces, and its eigenspaces for the $G = (\mathbb{Z}/2\mathbb{Z})^2$ action.

In the course of doing this we need to amend a small mistake in [BC10] Lemma 2.10; this is done in an appendix, where we actually generalize this lemma substantially in order to make it appropriate for our present purposes and also applicable in other situations.
In the end we succeed to prove that the subset $\mathcal{NEB}_4$ of the moduli space of canonical surfaces of general type $\mathcal{M}_{1,4}$ corresponding to nodal and extended Burniat surfaces with $K_S^2 = 4$ is an irreducible open set, normal, unirational of dimension 3 (similarly we show that the subset $\mathcal{NEB}_3$ of the moduli space of canonical surfaces of general type $\mathcal{M}_{1,3}$ corresponding to nodal and extended Burniat surfaces with $K_S^2 = 3$ is an irreducible open set, normal, unirational of dimension 4).

Section 5 is dedicated to the study of one-parameter limits of extended Burniat surfaces with $K_S^2 = 4$, showing that the subset of the moduli space of canonical surfaces of general type $\mathcal{M}_{1,4}$ corresponding to nodal and extended Burniat surfaces with $K_S^2 = 4$ is closed.

In Section 6 we give examples of other surfaces which lie in the closure of the family of extended Burniat surfaces with $K_S^2 = 3$.

In another appendix we give an alternative proof of three of the four assertions of Proposition 4.1, by other methods which could be of independent interest.

1. Definition of extended and nodal Burniat surfaces

Burniat surfaces are minimal surfaces of general type with $K^2 = 6, 5, 4, 3, 2$ and $p_g = 0$, which were constructed in [Bu66] as minimal resolutions of singular bidouble covers (that is, Galois covers with group $(\mathbb{Z}/2\mathbb{Z})^2$) of the projective plane branched on 9 lines.

We refer the reader to [BC10] for their construction, and we shall adhere to the notation introduced there.

Let $P_1, P_2, P_3 \in \mathbb{P}^2$ be three non-collinear points (which we assume to be the points $(1: 0: 0), (0: 1: 0)$ and $(0: 0: 1)$), and let $P_2, \ldots, P_{3+m}, \quad m = 2, 3$, be further (distinct) points not lying on the sides of the triangle with vertices $P_1, P_2, P_3$.

We make the further assumptions:

- for $m = 2$, the points $P_1, P_4, P_5$ are collinear, while,
- for $m = 3$, we assume moreover that also $P_2, P_4, P_6$ and $P_3, P_5, P_6$ are collinear (in particular, no four points are collinear); we may also use the notation $P'_1 := P_6, P'_2 := P_5, P'_3 := P_4$, so that $P_i, P'_i, P'_{i+1}, P'_{i+2}$ are collinear, where we use the convention $i \in \{1, 2, 3\}$ mod 3.

Let’s denote by $\tilde{Y} := \tilde{\mathbb{P}}^2(P_1, P_2, \ldots, P_{3+m})$ the weak Del Pezzo surface of degree $6 - m$, obtained blowing up $\tilde{\mathbb{P}}^2$ in the points $P_1, P_2, \ldots, P_{3+m}$.

Saying that $\tilde{Y}$ is a weak Del Pezzo surface means that the anticanonical divisor $-K_{\tilde{Y}}$ is nef and big; in our case it is not ample, because of the existence of $(\mathcal{O}_\tilde{Y}(2))-\text{curves}$, i.e. curves $N_i \cong \mathbb{P}^1$, with $N_i \cdot K_{\tilde{Y}} = 0$: for $m = 3$ $N_i$ is the strict transform of the line passing through $P_i, P'_{i+1}, P'_{i+2}$.

Contracting the $(\mathcal{O}_\tilde{Y}(2))-\text{curves}$ $N_i$ we obtain a normal singular Del Pezzo surface $Y'$ with $-K_{Y'}$ very ample.

We denote by $L$ the divisor on $\tilde{Y}$ which is the total transform of a general line in $\mathbb{P}^2$, by $E_i$ the exceptional curve lying over $P_i$, by $E'_i$ the exceptional curve
lying over $P_i'$ (hence $E_i' = E_{7-i}$) and by $D_{i,1}$ the strict transform of the line $y_{i-1} = 0$, side of the triangle joining the points $P_i, P_{i+1}$; that is, the unique effective divisor in $|L - E_i - E_{i+1}|$, where $[i+1]$ represents the residue class of $i + 1 \mod 3$, an element of $\{1, 2, 3\}$.

For $m = 2$ we have only one (-2)-curve $N_1$, such that $\{N_1\} = |L - E_1 - E_4 - E_5|$, whereas $D_2, D_3$ are divisors such that $\{D_2\} = |L - E_2 - E_4| + |L - E_2 - E_5| + E_1$, $\{D_3\} = |L - E_3 - E_4| + |L - E_3 - E_5| + E_2$.

Therefore the anticanonical image of $\tilde{Y}$ is a normal surface $Y' \subset \mathbb{P}^{6-m}$ of degree $6 - m$, whose singularities are one node $\nu_1$ (an $A_1$ singularity) in the case $m = 2$, and three nodes $\nu_1, \nu_2, \nu_3$ in the case $m = 3$ (the (-2)-curve $N_1$ is the total transform of the point $\nu_i$).

In order to improve readability we separate the definitions for $m = 2$ and $m = 3$.

1.1. NODAL AND EXTENDED BURNIAT SURFACES WITH $K_3^2 = 4$.

**Definition 1.1.** 1) Define the Burniat divisors for $m = 2$ as follows:

$$D_1 \in |L - E_1| + |L - E_1 - E_2| + |L - E_1 - E_4 - E_5| + E_3,$$

i.e., $D_1 = D_{1,1} + N_1 + C_1 + E_3$, where $C_1 \in |L - E_1|$ is assumed to be irreducible, whereas $D_2, D_3$ are divisors such that

$$\{D_2\} = |L - E_2 - E_3| + |L - E_2 - E_4| + |L - E_2 - E_5| + E_1,$$

$$\{D_3\} = |L - E_3 - E_1| + |L - E_3 - E_4| + |L - E_3 - E_5| + E_2.$$

2) The extended Burniat divisors for $m = 2$ are given as follows:

$$\Delta_1 \in |L - E_1| + |L - E_1 - E_2| + E_3,$$

i.e., $\Delta_1 = D_{1,1} + C_1 + E_3$, where $C_1 \in |L - E_1|$ is assumed to be irreducible, and
\[ \Delta_2 \in |L - E_2 - E_4| + |L - E_2 - E_5| + |2L - E_2 - E_3 - E_4 - E_5|, \]

where we assume the divisor \( \Gamma_2 \in |2L - E_2 - E_3 - E_4 - E_5| \) to be irreducible; and \( \Delta_3 \) is the divisor such that

\[ \{ \Delta_3 \} = (|L - E_3 - E_4| + |L - E_3 - E_5|) + (|L - E_3 - E_1| + E_2) + |L - E_1 - E_4 - E_5| \]

\[ =: \Delta_{3,1} + \Delta_{3,2} + N_1. \]

Remark 1.2. 1) Observe that \((D_1 + D_2 + D_3) \in |-3K_Y|\) is a reduced normal crossing divisor.

2) Similarly, \((\Delta_1 + \Delta_2 + \Delta_3) \in |-3K_Y + N_1|\) is a reduced normal crossing divisor.

3) On the normal Del Pezzo surface \(Y'\)

- \(D_1\) yields a conic plus two lines, and \(\Delta_1\) does the same (indeed \(D_1 = \Delta_1 + N_1\) and \(N_1\) is the ‘nodal’ exceptional curve).
- \(D_2\) yields four lines, \(\Delta_2\) yields a conic plus two lines (indeed \(\Delta_2 \equiv D_2 + N_1\))
- \(D_3\) yields four lines, the same does \(\Delta_3\) (indeed \(\Delta_3 = D_3 + N_1\)).

In particular, if the conic corresponding to \(\Delta_2\) specializes to contain the line corresponding to \(E_1\), we obtain then \(D_2\) subtracting the divisor \(N_1 \equiv L - E_1 - E_4 - E_5\).

Finally, the four lines of \(\Delta_3\) divide into two groups, i.e., we can write \(\Delta_3 = \Delta_{3,1} + \Delta_{3,2} + N_1\) so that, setting \(\Gamma_1 := C_1\) and writing \(\Delta_i = \Gamma_i + \Delta_{i}'\), for \(i = 1, 2\), then

\[ (*) : \Delta_{i}' + \Delta_{3,i} \equiv -K_Y \]
I. Bauer and F. Catanese

\[ (**) : \Gamma_1 + \Gamma_2 \equiv -K_{\tilde{Y}}. \]

4) The divisors \( D_i \) enjoy the property (cf. [BC10]) that there are divisor classes \( L_i \) such that \( D_{[i-1]} + D_{[i+1]} \equiv 2L_i \).

Hence, in particular, \( \Delta_{[i-1]} + \Delta_{[i+1]} \equiv 2\Lambda_i \), where, \( \Lambda_1 := L_1 + N_1 \) and \( \Lambda_j := L_j \) for \( j = 2,3 \).

5) Assume now that the conic corresponding to \( \Gamma_2 \) becomes reducible: if the conic passes through \( P_1 \), then necessarily \( \Gamma_2 \) splits as \( N_1 + E_1 + |L - E_2 - E_3| \), hence the conic is the union of two lines. If the conic is the union of two lines in another fashion, then necessarily either \( |L - E_2 - E_3| \) or \( |L - E_2 - E_4| \) is a component of \( \Gamma_2 \), hence \( \Delta_2 \) is not reduced.

We can now consider (cf. [Cat84b], [Cat99]) the associated bidouble covers \( S \to \tilde{Y} \) with branching divisors the Burniat divisors, respectively the extended Burniat divisors.

**Definition 1.3.** A secondary nodal Burniat surface is a bidouble cover \( S \to \tilde{Y} \) with branching divisors three Burniat divisors for \( m = 2 \). \( S \) is then a minimal surface of general type with \( p_g(S) = q(S) = 0 \), \( K_S^2 = 6 - m = 4 \) (cf. [BC10]).

If we let the three branch divisors be extended Burniat divisors, then we obtain a non minimal surface \( \tilde{S} \) whose minimal model \( S \) is called a secondary extended Burniat surface.

1.2. Nodal and extended Burniat surfaces with \( K_S^2 = 3 \).

**Definition 1.4.** 1) The Burniat divisors for \( m = 3 \) are defined to be the divisors \( D_1, D_2, D_3 \) such that

\[ D_i := D_{i,1} + N_i + E_{[i-1]} + |L - E_i - E' |. \]

More in detail,

\[ \{D_1\} = |L - E_1 - E_2| + |L - E_1 - E_4 - E_5| + |L - E_1 - E_6| + E_3, \]

\[ \{D_2\} = |L - E_2 - E_3| + |L - E_2 - E_4 - E_6| + |L - E_2 - E_5| + E_1, \]

\[ \{D_3\} = |L - E_3 - E_4| + |L - E_3 - E_5 - E_6| + |L - E_3 - E_4| + E_2. \]
2) The strictly extended Burniat divisors for $m = 3$ are defined as follows:

$$\Delta_i := |L - E_i - E_i'| + N_{i+1} + \Gamma_i,$$

where we assume $\Gamma_i \in |2L - E_i - E_{i+1} - E_i' - E_{i+1}'|$ to be the strict transform of an irreducible conic.

More in detail,

$$\Delta_1 \in |L - E_1 - E_6| + |2L - E_1 - E_2 - E_5 - E_6| + |L - E_2 - E_4 - E_6|,$$

$$\Delta_2 \in |L - E_2 - E_5| + |2L - E_2 - E_3 - E_4 - E_5| + |L - E_3 - E_5 - E_6|,$$

$$\Delta_3 \in |L - E_3 - E_4| + |2L - E_1 - E_3 - E_4 - E_6| + |L - E_1 - E_4 - E_5|.$$
3) Define $L_i := \frac{1}{2}(D_{[i-1]} + D_{[i+1]})$ and observe that it is an integral divisor; define also $\Lambda_i := L_i + N_i$.

Remark 1.5. 1) Observe that $(D_1 + D_2 + D_3) \in |-3K_{\tilde{Y}}|$ is a reduced normal crossing divisor.
2) Similarly, $(\Delta_1 + \Delta_2 + \Delta_3) \in |-3K_{\tilde{Y}} + \sum N_i|$ is a reduced normal crossing divisor.
3) On the normal Del Pezzo surface $Y'$, for $m = 3$, $\Delta_j$ yields a conic and one line, $D_j$ yields three lines (indeed $\Delta_j \equiv D_j - N_j + N_{j-1} + N_{j+1}$).

In particular, if the conic corresponding to $\Delta_j$ specializes to contain the line corresponding to $E_{[j-1]}$ (here as before $[j-1] \in \mathbb{Z}/3\mathbb{Z}$), we obtain $D_2$ subtracting the divisor $N_{j-1} + N_{j+1}$ and adding the divisor $N_j$.
4) The divisors $D_i$ enjoy the property (cf. [BC10]) that there are divisor classes $L_i$ such that $D_{[i-1]} + D_{[i+1]} \equiv 2L_i$.

Hence, in particular, $\Delta_{[i-1]} + \Delta_{[i+1]} \equiv 2\Lambda_i$, recalling that, for $m = 3$, $\Lambda_i := L_i + N_i$.
5) Assume that one or more of these conics become reducible. E.g., assume that the conic corresponding to $\Gamma_2$ becomes reducible, and observe that this will be the case if the conic passes through $P_1$ or $P_6$. We disregard this degeneration if the corresponding divisor $\Delta_2$ will be non reduced. The only possibility left
over is that $\Gamma_2$ splits as before, $N_1 + E_1 + |L - E_2 - E_3|$. This degeneration will be considered admissible. This motivates the following definition:

**Definition 1.6.** Assume $m = 3$ and that one or two of these conics $\Gamma_j$ become reducible in the admissible way $\Gamma_j = N_{j-1} + E_{j-1} + |L - E_j - E_{j+1}|$ (here, as usual, $j \in \mathbb{Z}/3\mathbb{Z}$).

In this case we define the (not strictly) extended Burniat divisors by subtracting to $\Gamma_j$ the nodal divisor $N_{j-1}$ it contains. Moreover, we replace

- $\Delta_{j+1}$ by $(\Delta_{j+1} - N_{j-1})$, and
- $\Delta_{j-1}$ by $(\Delta_{j-1} + N_{j-1})$.

For the convenience of the reader we have drawn the non strictly extended Burniat divisors in the case, where only $\Gamma_2$ degenerates.

**Remark 1.7.** If all three conics $\Gamma_j$ become reducible in the admissible way and we define in the same way as in the previous definition the three divisors by subtracting to $\Gamma_j$ the nodal divisor $N_{j-1}$ it contains, by subtracting again the nodal divisor $N_{j-1}$ from $\Delta_{j+1}$ and adding it to $\Delta_{j-1}$, we get the Burniat divisors from Definition 1.4.

We can now consider (cf. [Cat84b], [Cat99]) the associated bidouble covers $S \to \tilde{Y}$ with branching divisors the Burniat divisors, respectively the extended Burniat divisors.
Definition 1.8. A tertiary (three-)nodal Burniat surface $S$ is a bidouble cover $S \rightarrow \tilde{Y}$ with branch divisors three Burniat divisors for $m = 3$. $S$ is then a minimal surface of general type with $p_g(S) = q(S) = 0$, $K_S^2 = 6 - m = 3$ (cf. [BC10]).

If we let the three branch divisors be extended Burniat divisors (i.e., either strictly extended or non strictly extended!), then we obtain a non minimal surface $\tilde{S}$ whose minimal model $S$ is called a tertiary extended Burniat surface.

Remark 1.9. 1) In the nodal Burniat case the surface $S$ does not have an ample canonical divisor $K_S$, due to the existence of $(-2)$-curves, which are exactly the inverse images of the $(-2)$-curves $N_i \subset \tilde{Y}$.

For this reason we call the above Burniat surfaces of nodal type. We denote their canonical model by $X$, and observe that $X$ is a finite bidouble cover of the normal Del Pezzo surface $Y'$.

If $m = 2$, then $X$ has precisely one node (an $A_1$-singularity, corresponding to the contraction of the $(-2)$-curve) as singularity. While, for $m = 3$, the canonical model $X$ has exactly three nodes as singularities.

2) In the extended Burniat case $\tilde{S}$ is not minimal. In the strictly extended Burniat case the inverse image of each $N_i$ splits as the union of two disjoint $(-1)$-curves. $S$ has ample canonical divisor (hence $S = X$) exactly in the strictly extended case.

3) In all cases, the morphism $X \rightarrow Y'$ is exactly the bicanonical map of $X$ (see [BC10]).

4) Nodal Burniat surfaces are parametrized by a family with smooth base of dimension 2 for $m = 2$, of dimension 1 for $m = 3$. Strictly extended Burniat surfaces are parametrized by a family with smooth base of dimension 3 for $m = 2$, of dimension 4 for $m = 3$.

The key feature is that, both for nodal Burniat surfaces, and for extended Burniat surfaces, the canonical model $X$ is a finite bidouble cover of a singular Del Pezzo surface $Y'$, which has one node in the case $m = 2$, and three nodes for $m = 3$ (in this case $Y'$ is a cubic surface in $\mathbb{P}^3$).

In this case the direct image $p_* (\mathcal{O}_X)$ splits as a direct sum of four reflexive character sheaves of generic rank 1.

In the next section we shall describe how the covering behaves in the neighbourhood of a node in the two respective cases, and how these local coverings deform to each other (the Burniat case deforms to the extended Burniat case).

2. Local calculations around the nodes

In this section we consider finite bidouble covers of a node which are of Du Val type, i.e., yielding singularities which are at worst RDP’s (rational double points).

We obtain a classification which is a subset of the one made in [Cat87], classifying quotients of RDP’s by actions of $\mathbb{Z}/2\mathbb{Z}$ or of $G = (\mathbb{Z}/2\mathbb{Z})^2$.  

Documenta Mathematica 18 (2013) 1089–1136
We only need to look at Table 2, page 90, and Table 3, page 93, ibidem, to see which quotients of a rational double point by an involution, or by a pair of commuting involutions, yield an $A_1$ singularity, i.e., a node. There are six cases for such coverings of Du Val type of a node $Y$, which in local holomorphic coordinates is given by

$$xy - z^2 = 0.$$ 

In order to be more informative in our description, we denote by $\tilde{Y}$ the resolution of $Y$, which is the total space of a line bundle on $N \cong \mathbb{P}^1$ of degree $-2$ (hence $N^2 = -2$). Denoting the bidouble cover of $Y$ by $X$, we shall obtain, through the normalization of the fibre product, a finite bidouble cover of $\tilde{Y}$, for which we shall give the three corresponding branch divisors.

In the case where $X$ is not irreducible, we shall describe a connected component $X'$ of $X$.

1. $X' = Y$ (the covering is étale).
2. $X' = \mathbb{C}^2$, $X$ has two components and the covering morphism is given by

$$(u, v) \mapsto (x = u^2, y = v^2, z = uv).$$

The branch divisor on $\tilde{Y}$ is just the $(-2)$-curve $N$.
3. $X' = \{u^4 = xy\}$, $X$ has two components and the covering morphism is given by

$$(x, y, w) \mapsto (x, y, z = w^2).$$

The branch divisor on $\tilde{Y}$ consists of the $(-2)$-curve $N$ plus two fibres; the double cover of $\tilde{Y}$ has two nodes and resolving them we get the minimal resolution of the $A_3$ singularity $X'$.
4. $X = \{w^2 = uv\}$ and the covering morphism is given by

$$(u, v) \mapsto (x = u^2, y = v^2, z = w^2).$$

The three intermediate $\mathbb{Z}/2\mathbb{Z}$ covers are the two double covers $(2), (3)$ described above, plus the intermediate cover (here $a := uw, b := vw$)

$$\{(x, y, z, a, b) | \text{ Rank } \begin{pmatrix} x & a & z & b \\ a & z & b & y \end{pmatrix} = 1\},$$

which is the cone over a rational normal quartic (set $x = t_0^4, a = t_0^3t_1, z = t_0^3t_1, z = t_0t_1^2, z = t_1^4$).

The branch divisors on $\tilde{Y}$ are two: the $(-2)$-curve $N$ and the divisor $D$ formed by two fibres. The three intermediate double covers depend on the choice of the branch locus: $N$, respectively $N + D$, respectively $D$.
5. $X' = \{z^2 = (w^2 + y^{k+1})y\}$, $X$ has two components having a singularity of type $D_{k+3}$, and the covering morphism is given by

$$(y, z, w) \mapsto (x = w^2 + y^{k+1}, y, z).$$
The branch divisor on $\tilde{Y}$ is the total transform of the divisor $C := \{x = y^{k+1}, z^2 = y^{k+2}\}$ which is irreducible with a cusp for $k$ odd, else it is reducible with a $\frac{k}{2}$-tacnode for $k$ even.

In particular, $N$ is part of the branch locus.

(6) $X = \{ w^2 = (u - v^{k+1})(u + v^{k+1}) \} = \{ w^2 = u^2 - v^{2k+2} \}$ and the covering morphism is given by

$$(u, v, w) \mapsto (x = u^2, y = v^2, z = uv).$$

$X$ is a singularity of type $A_{2k+1}$ and, in order to treat a new case, we make the assumption $k \geq 1$.

The three intermediate $\mathbb{Z}/2\mathbb{Z}$ covers are the smooth double cover (2), the double cover (5) $\{ w^2 = x - y^{k+1} \}$, and a third singularity which we omit to describe.

The branch divisors on $\tilde{Y}$ are two: the (-2)-curve $N$ and the total transform of the divisor $C$ above.

The three intermediate covers depend on the choice of the branch locus: $N$, or $N + C'$, or $C'$, where $C'$ is the strict transform of $C$.

Letting $p : X \rightarrow Y$ be the finite bidouble cover, the direct image sheaf $p_* \mathcal{O}_X$ splits as

$$\mathcal{O}_Y \oplus \left( \bigoplus_{i=1,2,3} \mathcal{L}_i \right),$$

where in the first case the reflexive sheaves $\mathcal{L}_i$ are locally free.

To describe the other cases we use the reflexive sheaf $\mathcal{F}$ generated by $u, v$ as $\mathcal{O}_Y$-module, with relations

$$yu - zv = 0, zu - xv = 0.$$

We get

(2) $X' = \mathbb{C}^2$, $(u, v) \mapsto (x = u^2, y = v^2, z = uv)$,

$$p_* \mathcal{O}_X = (\mathcal{O}_Y \oplus \mathcal{F}) \oplus 2,$$

(3) $X' = \{ w^4 = xy \}$, $(x, y, w) \mapsto (x, y, z = w^2)$

$$p_* \mathcal{O}_X = (\mathcal{O}_Y \oplus \mathcal{O}_Y) \oplus 2,$$

(4) $X = \{ w^2 = uv \}$, $(u, v) \mapsto (x = u^2, y = v^2, z = w^2)$

$$p_* \mathcal{O}_X = (\mathcal{O}_Y \oplus \mathcal{F}) \oplus 2,$$

with generators $1, \{ u, v \}, w, \{ a = uw, b = vw \}$.

(5) $X' = \{ w^2 = x - y^{k+1} \}$, $(y, z, w) \mapsto (x = w^2 + y^{k+1}, y, z)$

$$p_* \mathcal{O}_X = (\mathcal{O}_Y \oplus \mathcal{O}_Y) \oplus 2.$$

(6) $X = \{ w^2 = u^2 - v^{2k+2} \}$, $(u, v, w) \mapsto (x = u^2, y = v^2, z = uv)$

$$p_* \mathcal{O}_X = (\mathcal{O}_Y \oplus \mathcal{F}) \oplus 2.$$
Remark 2.1. Cases 1, 3 and 5 are the case where we have a flat bidouble cover, i.e., $p_1, O_X$ is locally free. In cases 2, 4 and 6 we have non-flat bidouble covers, but with the same character sheaves. We shall soon show how case 4 deforms to case 2.

Proposition 2.2. In case 2) $X = \text{Spec}((O_Y \oplus F) \oplus (O_Y \oplus F))$, where the two addenda are orthogonal, and the algebra structure is determined by the nondegenerate pairing $F \times F \to O_Y$.

In case 4) $X = \text{Spec}((O_Y \oplus O_Y) \oplus w(O_Y \oplus F))$, and the algebra structure is determined by the nondegenerate pairing $F \times F \to O_Y$, together with the assignment $w^2 = z$.

We omit the simple proof.

Case 4) deforms now to case 2) by changing the assignment $w^2 = z$ to $w^2 = z + t$, $t \neq 0$, so that $w$ becomes then a local unit at the origin.

We can relate the resulting picture with the local semiuniversal deformation of a node.

Proposition 2.3. Let $t \in \mathbb{C}$, and consider the action of $G := (\mathbb{Z}/2\mathbb{Z})^2$ on $\mathbb{C}^3$ generated by $\sigma_1(u, v, w) = (u, v, -w)$, $\sigma_2(u, v, w) = (-u, -v, w)$. Then the hypersurfaces $X_t = \{(u, v, w) \mid w^2 = uv + t\}$ are $G$-invariant, and the quotient $X_t/G$ is the hypersurface

$$Y_t = Y_0 = \{(x, y, z) \mid z^2 = xy\},$$

which has a nodal singularity at the point $x = y = z = 0$.

$X_t \to Y_t$ is a bidouble covering of type 2 for $t \neq 0$, and of type 4 for $t = 0$. We get in this way a flat family of (non flat) bidouble covers.

Proof. The invariants for the action of $G$ on $\mathbb{C}^3 \times \mathbb{C}$ are:

$$x := u^2, y := v^2, z := uv, s := w^2, t.$$

Hence the family $X$ of the hypersurfaces $X_t$ is the inverse image of the family of hypersurfaces $s = z + t$ on the product

$$Y' \times \mathbb{C}^2 = \{x, y, z, s, t \mid xy = z^2\}.$$

Hence the quotient of $X_t$ is isomorphic to $Y'$.

The rest was already explained before. □

Remark 2.4. i) The simplest way to view $X_t$ is to see $\mathbb{C}^2$ as a double cover of $Y'$ branched only at the origin, and then $X_t$ as a family of double covers of $\mathbb{C}^2$ branched on the curve $uv + t = 0$, which acquires a double point for $t = 0$.

ii) The involution $\sigma_3(u, v, w) = (-u, -v, -w)$ has only the origin as fixed point, which lies on $X_0$. Whereas $\sigma_3$ acts freely on $X_t$, for $t \neq 0$.

$\text{Fix}(\sigma_1) = \{w = 0\}$, and $\{w = 0\} \cap X_t = \{uv + t = w = 0\}$.

Finally, $\text{Fix}(\sigma_2) = \{u = v = 0\}$, and $\{u = v = 0\} \cap X_t = \{u = v = 0, w^2 = t\}$, which consists of two points for $t \neq 0$, one for $t = 0$.

The corresponding branch loci are the origin, for $t = 0$, the divisor $z = 0$, and the point $x = y = z = -t = 0$. 

Documenta Mathematica 18 (2013) 1089–1136
iii) If we pull back the bidouble cover $X_t$ to $\tilde{Y}$, and we normalize it, we can see that

- $D_3$ is, for $t = 0$, the nodal curve $N$, and is the empty divisor for $t \neq 0$;
- $D_1$ is, for $t \neq 0$, the inverse image of the curve $z + t = 0$; while, for $t = 0$, it is only its strict transform, i.e. the divisor $D$ considered previously, made up of two fibres;
- $D_2$ is an empty divisor for $t = 0$, and the nodal curve $N$ for $t \neq 0$.

**Remark 2.5.** Part iii) of the previous remark shows that, as $t \to 0$, one subtracts the nodal divisor $N$ to $D_2$, and adds it to $D_3$; while for $D_1$, it specializes to $D + N$, and then we subtract $N$.

This is precisely the algorithm which applies when passing from extended Burniat to Burniat divisors.

The really interesting part of the story comes now: the family $X_t$ admits a simultaneous resolution only after that we perform a base change $t = \tau^2$ and the equation of $X_t$ becomes

$$X_{\tau} = \{ w^2 - \tau^2 = uv \}.$$ 

**Definition 2.6.** Let $X \to T'$ be the family where

$$X = \{ (u, v, w, \tau) | w^2 - \tau^2 = uv \}$$

and $T'$ is the affine line with coordinate $\tau$.

Define $S \subset X \times \mathbb{P}^1$ to be one of the small resolutions of $X$, and $S'$ to be the other one, namely:

$$S : \{ (u, v, w, \tau) (\xi) \in X \times \mathbb{P}^1 | \frac{w - \tau}{u} = \frac{v}{w + \tau} = \xi \}$$

$$S' : \{ (u, v, w, \tau) (\eta) \in X \times \mathbb{P}^1 | \frac{w + \tau}{u} = \frac{v}{w - \tau} = \eta \}.$$ 

Let $G$ be the group $G \cong (\mathbb{Z}/2\mathbb{Z})^2$ acting on $X$ trivially on the variable $\tau$, and else as in Proposition 2.3. Let further $\sigma_4$ act by $\sigma_4(u, v, w, \tau) = (u, v, w, -\tau)$, let $G' \cong (\mathbb{Z}/2\mathbb{Z})^3$ be the group generated by $G$ and $\sigma_4$, and let $H \cong (\mathbb{Z}/2\mathbb{Z})^2$ be the subgroup $\{ \text{Id}, \sigma_2, \sigma_1\sigma_4, \sigma_3\sigma_4 \}$.

The following is a rephrasing and a generalization of a discovery of Atiyah in our context: we omit the simple proof. For more details and a discussion of how these examples fit into the general theory of moduli spaces, see the ‘working guide’ written by the second author in [Cat11].

**Lemma 2.7.** The biregular action of $G'$ on $X$ lifts only to a birational action on $S$, respectively $S'$. The subgroup $H$ acts on $S$, respectively $S'$, as a group of biregular automorphisms.

The elements of $G' \setminus H = \{ \sigma_1, \sigma_3, \sigma_4, \sigma_2\sigma_4 \}$ yield isomorphisms between $S$ and $S'$.

The group $G$ acts on the punctured family $S \setminus S_0$, in particular it acts on each fibre $S_\tau$. 
Since \( \sigma_4 \) acts trivially on \( S_0 \), the group \( G' \) acts on \( S_0 \) through its direct summand \( G \).

The biregular actions of \( G \) on \( S \setminus S_0 \) and on \( S_0 \) do not patch together to a biregular action on \( S \), in particular \( \sigma_1 \) and \( \sigma_3 \) yield birational maps which are not biregular: they are called Atiyah flops (cf. \([58]\)).

3. Nodal Burniat surfaces deform to extended Burniat surfaces

In this section we will show that

- the canonical models \( X \) of nodal Burniat surfaces with \( K_X^2 = 4 \), together with the extended Burniat surfaces with \( K_X^2 = 4 \) are parametrized by a family with smooth connected base of dimension 3, which maps to the moduli space via a finite morphism;
- the canonical models \( X \) of nodal Burniat surfaces with \( K_X^2 = 3 \), together with the extended Burniat surfaces with \( K_X^2 = 3 \) are parametrized by a family with smooth connected base of respective dimension 4, which maps to the moduli space via a finite morphism.

We shall treat first the easier case

3.1. (Extended) Burniat surfaces with \( K^2 = 4 \).

**Proposition 3.1.** There exists a family, with connected smooth 3-dimensional base

\[
B = \{ (P_5, C_1, \Gamma_2) | C_1 \in |L - E_1|, \Gamma_2 \in |2L - E_2 - E_3 - E_4 - E_5| \}
\]

parametrizing a flat family of canonical models, including exactly all the nodal Burniat surfaces and the extended Burniat surfaces with \( K_X^2 = 4 \). The family maps to the moduli space via a quasi-finite morphism.

Here, \( P_1, P_2, P_3, P_4 \) are the standard projective basis in \( \mathbb{P}^2 \), the point \( P_5 \) belongs to the line \( P_1P_4 \) and, in the blow up of the plane in the given five points \( P_j, j = 1, \ldots, 5 \), \( C_1, \Gamma_2 \) are as in Definition 1.1 (\( C_1 \) is irreducible and either \( \Gamma_2 \) is irreducible, or splits as \( N_1 + E_1 + |L - E_2 - E_3| \)).

**Proof.** Recall that in this case \( D_1 + D_3 = \Delta_1 + \Delta_3 \), and that \( N_1 \) is a connected component of the above divisor \( D_1 + D_3 = \Delta_1 + \Delta_3 \).

We can therefore construct a family of double covers

\[
\tilde{W}_b \to \tilde{Y}
\]

such that the inverse image of \( N_1 \) is a (-1)-curve. Blowing down this (-1)-curve we get a family of finite double covers \( W_b' \to Y' \), which are nodal and equisingular.

Consider the pull back of the divisors \( \Delta_2 \) in the case where \( \Gamma_2 \) is irreducible, and of the divisors \( D_2 \) in the case where \( \Gamma_2 \) is reducible.

Since \( \Delta_2 \equiv D_2 + N_1 \), and the divisor \( N_1 \) becomes trivial on \( W_b' \), since it contracts to a smooth point, it follows that all these divisors are linearly equivalent, and we have a family of divisors on the family \( W_b' \).
We consider then the family of double covers $X_b \to W_b'$ branched on these divisors, and on the nodes of $W_b'$.

Finally, assume that two surfaces $S_1, S_2$ in the above family are isomorphic, equivalently that their canonical models $X_1, X_2$ are isomorphic. Then this isomorphism would yield an isomorphism of the bicanonical morphisms of each $X_j$, hence we would have isomorphisms of the image normal Del Pezzo surfaces $Y'_1, Y'_2$, sending the respective triple of branch curves for $X_1$ to the ones for $X_2$.

$Y'_j$ has exactly one node and contains exactly 12 lines (cf. Proposition 3.7 of [BC10]). $Y'_j$ is obtained blowing down the $(-2)$ curve $N_1$ and conversely, blowing up the singular point of $Y'_j$ and five disjoint $(-1)$-curves we obtain an isomorphism with the plane. Since these $(-1)$-curves correspond exactly to the lines in $Y'_j$, we have only a finite number of such birational contractions to the plane, and each of them determines a triple of branch curves, and five points in the plane.

Therefore the number of surfaces $S_2$ in the family which are isomorphic to $S_1$ is finite.

\[ \square \]

3.2. (Extended) Burniat surfaces with $K^2_X = 3$. We argue similarly, but it may be useful to make right away a simple geometrical observation.

Let $P_1, P_2, P_3, P_4$ be the standard projective basis in $\mathbb{P}^2$, and consider a line $L'$ with $P_3 \in L'$, different from the coordinate lines: then the line configuration of a ternary Burniat surface is completely determined by the line $L'$, since then $P_5 := L' \cap P_1P_4$, $P_6 := L' \cap P_2P_4$.

Proposition 3.2. There exists a family, with connected smooth 4-dimensional base

\[ T \subset \{(L', \Gamma_1, \Gamma_2, \Gamma_3)\} \]

where $L'$ is as above and $\Gamma_1, \Gamma_2, \Gamma_3$ are as in Definitions 1.4 and 1.6 parametrizing a flat family of canonical models, including exactly all the nodal Burniat surfaces and the extended Burniat surfaces with $K^2_X = 3$. The family maps to the moduli space via a quasi-finite morphism.

Proof. Given a triple $(\Gamma_1, \Gamma_2, \Gamma_3)$, according to the reducibility of each $\Gamma_i$, there corresponds either a Burniat divisor, or an extended Burniat divisor. We take the corresponding bidouble cover of $\tilde{Y}$, hence we construct four families of smooth surfaces, which are not necessarily minimal. We take now the corresponding canonical models, which are finite bidouble covers of the normal Del Pezzo surface $Y'$.

Observe that, given $p' : \tilde{S} \to \tilde{Y}$, and $\pi : \tilde{Y} \to Y'$,

\[ X = \text{Spec}(\pi_* (p')_* \mathcal{O}_{\tilde{S}}) = \text{Spec}(\mathcal{O}_{Y'} \bigoplus (\mathcal{E}^3_{i=1} F_i)) \]

Now the reflexive sheaves $F_i$ correspond to Weil divisors on $Y'$, and they are independent of $t \in T$ by virtue of 4) of remark 1.6.
The multiplication maps correspond to a family of Weil divisors on $Y'$: whence we get a flat family on $Y' \setminus \text{Sing}(Y')$. Locally around the nodes the structure of the deformation is as described in the previous section, therefore the family is flat everywhere.

The assertion that for a given surface $S_1$ in the family parametrized by $T$ the number of surfaces $S_2$ in the family which isomorphic to $S_1$ is finite follows as in the case $K^2 = 4$, using that the three nodal cubic surface $Y'$ contains only a finite number of lines.

\[ \square \]

Observe that proof given for the case $K^2 = 3$ works also in the case $K^2 = 4$.

4. Local deformations of the extended Burniat surfaces

The aim of this section is to calculate the dimension of $H^1(S, \Theta_S)$ for

- nodal Burniat surfaces with $K_S^2 = 4$,
- extended Burniat surfaces with $K_S^2 = 4$,
- nodal Burniat surfaces with $K_S^2 = 3$,
- extended Burniat surfaces with $K_S^2 = 3$.

The main results of this paragraph are the following (note that we use the notation introduced in Section 1):

**Proposition 4.1.**

1) Assume that $S$ is a nodal Burniat surface with $K_S^2 = 4$ ($m = 2$). Then the dimension of the vector space

\[ H^0(\Omega^1_Y(\log(D_i))(K_Y + L_i)) = H^0(\Omega^1_Y(\log(D_i))(E_i - E_{i+2})) \]

is 1 for $i = 3$, else it is 0.

2) Consider instead extended Burniat divisors for $m = 2$, and the corresponding vector spaces

\[ H^0(\Omega^1_Y(\log(\Delta_i))(K_Y + \Lambda_i)). \]

Then their dimensions are the same as in the Burniat case, namely, 1 for $i = 3$, else 0.

3) Assume that $S$ is a Burniat surface with $K_S^2 = 3$ ($m = 3$). Then each vector space

\[ H^0(\Omega^1_Y(\log(D_i))(K_Y + L_i)) = H^0(\Omega^1_Y(\log(D_i))(E_i - E_{i+2})) \]

is equal to 0.

4) In the case of (strictly or not strictly) extended Burniat divisors for $m = 3$ we have \forall:\n
\[ H^0(\Omega^1_Y(\log(\Delta_i))(K_Y + \Lambda_i)) = 0. \]

The proof of Proposition 4.1 is to be found in the second appendix.

Using Proposition 4.1 we can explicitly determine the several $G$-character spaces of $H^1(S, \Theta_S)$ and their dimensions (here $G = (\mathbb{Z}/2\mathbb{Z})^2 = \{1, g_1, g_2, g_3\}$). In the following, given a $G$-space $V$, we denote by $V^i$, for $i \in 1, 2, 3$, the eigenspace corresponding to the character whose kernel consists of $\{1, g_i\}$.
Proposition 4.2. 1) Let $S$ be the minimal model of a Burniat surface. Then the dimensions of the eigenspaces of the cohomology groups of the tangent sheaf $\Theta_S$ (for the natural $(\mathbb{Z}/2\mathbb{Z})^2$-action) are as follows.

1) $K_S^2 = 4$ of nodal type:

- $h^1(S, \Theta_S)^{inv} = 2$, $h^2(S, \Theta_S)^{inv} = 0$,
- $h^1(S, \Theta_S)^3 = 1 = h^2(S, \Theta_S)^3$,
- $h^2(S, \Theta_S)^4 = 0$, for $i \in \{1, 2\}$, $\forall j$;

2) $K_S^2 = 3$:

- $h^1(S, \Theta_S)^{inv} = 1$, $h^2(S, \Theta_S)^{inv} = 0$,
- $h^1(S, \Theta_S)^4 = 1 = h^2(S, \Theta_S)^4$, for $i \in \{1, 2, 3\}$.

2) Let $S$ be a minimal model of an extended Burniat surface with $K_S^2 = 4$. Then the dimensions of the eigenspaces of the cohomology groups of the tangent sheaf $\Theta_S$ (for the natural $(\mathbb{Z}/2\mathbb{Z})^2$-action) are as follows.

- $h^1(S, \Theta_S)^{inv} = 3$, $h^2(S, \Theta_S)^{inv} = 0$,
- $h^1(S, \Theta_S)^4 = 0 = h^2(S, \Theta_S)^4$, for $i \in \{1, 2\}$,
- $h^1(S, \Theta_S)^3 = 1 = h^2(S, \Theta_S)^3$.

3) Let $S$ be the minimal model of an extended Burniat surface with $K_S^2 = 3$. Then the dimensions of the eigenspaces of the cohomology groups of the tangent sheaf $\Theta_S$ (for the natural $(\mathbb{Z}/2\mathbb{Z})^2$-action) are as follows:

1) strictly extended case:

- $h^1(S, \Theta_S)^{inv} = 4$, $h^2(S, \Theta_S)^{inv} = 0$,
- $h^1(S, \Theta_S)^4 = 0$, for $i \in \{1, 2, 3\}$;

2) case where exactly one conic, w.l.o.g. the conic $\Gamma_1$, because of symmetry, degenerates to two lines:

- $h^1(S, \Theta_S)^{inv} = 3$, $h^2(S, \Theta_S)^{inv} = 0$,
- $h^1(S, \Theta_S)^4 = 0 = h^2(S, \Theta_S)^4$, for $i \in \{1, 3\}$,
- $h^1(S, \Theta_S)^2 = 1 = h^2(S, \Theta_S)^2 = 0$;

3) case where exactly two conics, w.l.o.g. $\Gamma_1, \Gamma_2$, degenerate to two lines each:

- $h^1(S, \Theta_S)^{inv} = 2$, $h^2(S, \Theta_S)^{inv} = 0$,
- $h^1(S, \Theta_S)^4 = 0 = h^2(S, \Theta_S)^4$,
- $h^1(S, \Theta_S)^4 = 1 = h^2(S, \Theta_S)^4$, for $i \in \{1, 3\}$.

For the reader’s convenience the above dimensions can also be found in Table I.

We begin with an easy but useful observation

Lemma 4.3. Assume that $N$ is a connected component of a smooth divisor $D \subset Y$, where $Y$ is a smooth projective surface. Moreover, let $M$ be a divisor on $Y$. Then

$$H^0(\Omega^1_Y(\log(D - N))(N + M)) = H^0(\Omega^1_Y(\log(D))(M))$$

provided $(K_Y + 2N + M) \cdot N < 0$.

Proof. The cokernel of $\Omega^1_Y(\log(D))(M) \rightarrow \Omega^1_Y(\log(D - N))(N + M)$ is supported on $N$ and equal to $\Omega^1_Y(N + M) = \mathcal{O}_N(K_Y + 2N + M)$.
The lemma will be applied several times in the case where \( N \cong \mathbb{P}^1 \) and \( N^2 < 0 \). Another useful lemma which will be crucial in some calculation is the following

**Lemma 4.4.** Assume that we have three linearly independent linear forms on \( \mathbb{P}^2 \), \( l_1 := x_1, l_2 := x_2, l_3 := x_3 \). Then

1. \( H^0(\Omega^1_{\mathbb{P}^2}(2)) \) has as basis the three 1-forms, for \( j < i \),
   \[
   \eta_{ji} := x_j dx_i - x_i dx_j = -\eta_{ij}.
   \]

2. \( H^0(\Omega^1_{\mathbb{P}^2}(\mathrm{dlog} l_1, \mathrm{dlog} l_2, \mathrm{dlog} l_3)(1)) \) has as basis the six 1-forms
   \[
   \omega_{ij} := \frac{x_j dx_i - x_i dx_j}{x_i}.
   \]

3. \( H^0(\Omega^1_{\mathbb{P}^2}(\mathrm{dlog} l_1, \mathrm{dlog} l_2, \mathrm{dlog} l_3)(2)) \) has as basis the three 1-forms \( \eta_{ji} \), for \( j < i \), plus the six 1-forms \( x_j \omega_{ij} \) and the three 1-forms \( x_1 \omega_{23}, x_2 \omega_{31}, x_3 \omega_{12} \).

**Proof.**

1) is well known and follows from the Euler sequence.

2) Take the chart \( x_i \neq 0 \Leftrightarrow x_i = 1 \); then in this chart \( \omega_{ij} := -dx_j \) is a regular 1-form.

In the chart \( x_j = 1 \) we have \( \omega_{ij} := \frac{dx_i}{x_i} \), while in the chart \( x_h = 1 \) we have \( \omega_{ij} := x_j \frac{dx_i}{x_i} - dx_j \).

Hence \( \omega_{ij} \) has logarithmic poles on \( x_i = 0 \), and the coefficient of the logarithmic term vanishes for \( x_i = x_j = 0 \), and is equal to 1 in \( x_i = x_h = 0 \).

The above observation shows the linear independence of the above 6 forms.

Moreover, \( \omega_{ij} \) is an eigenvector with character \( \lambda \) for the \( \mathbb{C}^* \)-action \( x_i \mapsto \lambda x_i \), hence \( \omega_{ij} \in H^0(\Omega^1_{\mathbb{P}^2}(\mathrm{dlog} l_1, \mathrm{dlog} l_2, \mathrm{dlog} l_3)(1)) \).

It suffices to show that this space has vector dimension equal to 6.

This follows however from the exact sequence

\[
0 \to \Omega^1_{\mathbb{P}^2}(1) \to \Omega^3_{\mathbb{P}^2}(\mathrm{dlog} l_1, \mathrm{dlog} l_2, \mathrm{dlog} l_3)(1) \to \bigoplus_{i=1}^3 \mathcal{O}_l(1) \to 0
\]

and the vanishing of \( H^j(\Omega^1_{\mathbb{P}^2}(1)) \) for \( j = 0, 1 \).

3) Observe that \( \omega_{ij} = \frac{1}{x_i} \eta_{ji} \), so that \( x_i \omega_{ij} = \eta_{ji} = -\eta_{ij} = x_j \omega_{ji} \).

Moreover, if \( h \neq i,j \), \( x_h \omega_{ij} = x_j \omega_{ih} = \eta_{jh} \), so that the products \( x_r \omega_{ij} \) generate a subspace of dimension at most 12.

By the exact sequence

\[
0 \to \Omega^1_{\mathbb{P}^2}(2) \to \Omega^3_{\mathbb{P}^2}(\mathrm{dlog} l_1, \mathrm{dlog} l_2, \mathrm{dlog} l_3)(2) \to \bigoplus_{i=1}^3 \mathcal{O}_l(2) \to 0
\]

and since \( H^1(\Omega^1_{\mathbb{P}^2}(2)) = 0, h^0(\mathcal{O}_l(2)) = 3 \) we infer that the dimension is indeed 12.

Since \( H^0(\mathcal{O}_l(2)) \) is generated by \( H^0(\mathcal{O}_{\mathbb{P}^2}(1)) \otimes_{\mathbb{C}} H^0(\mathcal{O}_l(1)) \) we conclude that the twelve 1-forms are a basis. \( \square \)

**Lemma 4.5.** Assume that we have two linearly independent linear forms on \( \mathbb{P}^2 \), \( l_1 := x_1, l_2 := x_2 \).
(1) $H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2)(1))$ has as basis the 4 forms
\[ \omega_{ij} := \frac{x_j dx_i - x_i dx_j}{x_i}, \quad 1 \leq i, j \leq 3, \quad i \neq 3. \]

(2) $H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2)(2))$ has as basis the 3 forms $\eta_{ji}$, for $j < i$, plus the 6 forms $x_2 x_{12}, x_1 x_{13}, x_3 x_{23}, x_2 x_{12}, x_1 x_{13}, x_1 x_{13}$.

Proof. Follows from Lemma 4.4 observing that $H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2)(i))$ is a subspace of $H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2, \log l_3)(i))$. The above two sets of vectors are linearly independent and the dimensions are 4, resp. 9. \hfill \square

Corollary 4.6. 1) Let $\omega \in H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2)(1))$. Then there are complex numbers $a_{ij}$ such that:
\[ \omega = a_{12} \omega_{12} + a_{21} \omega_{21} + a_{13} \omega_{13} + a_{23} \omega_{23} = \frac{dx_1}{x_1} (a_{12} x_2 - a_{21} x_1 + a_{13} x_3) + \]
\[ + \frac{dx_2}{x_2} (-a_{12} x_2 + a_{21} x_1 + a_{23} x_3) + dx_3 (-a_{13} - a_{23}). \]

2) Let $\omega \in H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2)(2))$: then we can write
\[ \omega = a_{12} \eta_{12} + a_{13} \eta_{13} + a_{23} \eta_{23} + a_{21} x_2 \omega_{12} + a_{121} x_1 \omega_{21} + a_{313} x_3 \omega_{13} + \]
\[ + a_{323} x_3 \omega_{23} + a_{213} x_2 \omega_{13} + a_{123} x_1 \omega_{23} = \]
\[ = \frac{dx_1}{x_1} (-a_{12} x_2 - a_{13} x_3 x_1 + a_{21} x_1 x_2 - a_{23} x_2 x_1 + a_{31} x_3 x_1 + a_{32} x_2 x_1) + \]
\[ + \frac{dx_2}{x_2} (a_{12} x_2 - a_{23} x_3 x_1 + a_{13} x_3 x_1 - a_{21} x_1 x_2 + a_{31} x_3 x_1 - a_{32} x_2 x_1) + \]
\[ + dx_3 (-a_{13} x_3 + a_{313} x_3 + a_{312} x_2 - a_{123} x_1). \]

3) Any $\omega \in H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2, \log l_3)(1))$ can be written as:
\[ \omega = a_{12} \omega_{12} + a_{13} \omega_{13} + a_{23} \omega_{23} + a_{21} \omega_{21} + a_{31} \omega_{31} + a_{32} \omega_{32} = \]
\[ = \frac{dx_1}{x_1} (a_{12} x_2 - a_{21} x_1 + a_{13} x_3 - a_{31} x_1) + \]
\[ + \frac{dx_2}{x_2} (-a_{12} x_2 + a_{21} x_1 + a_{23} x_3 - a_{32} x_2) + \]
\[ + \frac{dx_3}{x_3} (-a_{13} x_3 + a_{31} x_3 + a_{32} x_2 - a_{23} x_3). \]

4) Any $\omega \in H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2, \log l_3)(2))$ can be written as:
\[ \omega = a_{12} \eta_{12} + a_{13} \eta_{13} + a_{23} \eta_{23} + a_{21} x_2 \omega_{12} + a_{313} x_3 \omega_{13} + a_{323} x_3 \omega_{23} + a_{121} x_1 \omega_{21} + a_{131} x_1 \omega_{31} + a_{232} x_2 \omega_{32} + a_{123} x_1 \omega_{23} + a_{231} x_2 \omega_{31} + a_{312} x_3 \omega_{12} = \]
\[ \frac{dx_1}{x_1} \left( -a_{12} x_1 x_2 - a_{13} x_3 x_1 + a_{212} x_2^2 + a_{313} x_3^2 - a_{121} x_1^2 - a_{131} x_1^2 - a_{231} x_2 x_1 + a_{312} x_3 x_2 + a_{121} x_1^2 - a_{131} x_1^2 - a_{231} x_2 x_1 + a_{312} x_3 x_2 \right) + \frac{dx_2}{x_2} (a_{121} x_1 x_2 - a_{231} x_3 x_2 - a_{212} x_2^2 + a_{121} x_1^2 + a_{323} x_2^2 + a_{123} x_1 x_3 - a_{232} x_2^2 - a_{312} x_3 x_2) + \frac{dx_3}{x_3} (a_{131} x_1 x_3 + a_{232} x_2 x_3 - a_{231} x_3^2 - a_{323} x_3^2 + a_{131} x_1^2 + a_{232} x_2^2 - a_{231} x_1 x_3 + a_{231} x_1 x_2). \]

**Proof.** This is an easy verification. \(\square\)

Now we have prepared everything for the proof of Proposition 4.2. Since the proof is long and technical we prefer to put it in a separate section (cf. Appendix B).

**Proof of Proposition 4.2.** For the invariant part, the calculation goes exactly as the proof of Lemma 2.9 of [BC10], using that \(h^1(\Theta_s)^{inv} = h^1(\Theta_S)^{inv}\).

For the other character spaces, we use the same argument as in Lemma 2.12 of [BC10] to calculate \(\chi(\Omega_Y^1(\log D_i)(K_Y + L_i))\) (resp. \(\chi(\Omega_Y^1(\log \Delta_i)(K_Y + \Lambda_i))\)) for extended Burniat surfaces.

We first observe that

\[ \chi(\Omega_Y^1(\log D_i)(K_Y + L_i)) = \chi(\Omega_Y^1(K_Y + L_i)) + \chi(\mathcal{O}_{D_i}(K_Y + L_i)), \]

and analogously for \(\chi(\Omega_Y^1(\log \Delta_i)(K_Y + \Lambda_i))\) for extended Burniat surfaces. Moreover, note that with the same calculation as in Lemma 2.12 of [BC10], we see that \(\chi(\Omega_Y^1(K_Y + L_i)) = K_Y^2 - 12\).

Each \(D_i\) (resp. \(\Delta_i\)) consists of \(k_i\) irreducible connected components, each of them being a smooth rational curve. Write \(D_i = D_{i,1} + \ldots + D_{i,k_i}\) as disjoint union of smooth rational curves and let \(n_j := D_{i,j} \cdot (K_Y + L_i)\). Then

\[ \chi(\mathcal{O}_{D_i}(\log D_i)(K_Y + L_i)) = \sum_{j=1}^{k_i} \max(0, n_j + 1). \]

Therefore

\[ \chi(\Omega_Y^1(\log D_i)(K_Y + L_i)) = K_Y^2 - 12 + \sum_{j=1}^{k_i} \max(0, n_j + 1). \]

We summarize the calculations in the following table (note that we write \(\chi_i\) for \(\chi(\Omega_Y^1(\log D_i)(K_Y + L_i))\)). The values for \(h^2(\Theta_S)^i\) have been calculated in prop. 4.3. The notation: extended case (2), resp. (3), refers to Proposition 4.3.

Moreover, we use Lemma 9.22 of [Cat88] to compare \(h^1(\Theta_S)\) and \(h^1(\Theta_S)\); it asserts that for a single blow up of a point \(P\)

\[ \pi_* \Theta_S = m_P \Theta_S, \quad R^1 \pi_* \Theta_S = 0. \]
From the above calculations and from Propositions 3.1, 3.2 follow all the statements of our first main theorem, with the exception of the statement that $\mathcal{NEB}_4$ is a connected component. It follows that $\mathcal{NEB}_4$ is open, while the statement that $\mathcal{NEB}_4$ is closed will be shown in the forthcoming section.

**Theorem 0.1**

1) The subset $\mathcal{NEB}_4$ of the moduli space of canonical surfaces of general type $\mathcal{M}_{1,4}$ given by the union of the open set corresponding to extended Burniat surfaces with $K^2_S = 4$ with the irreducible closed set parametrizing nodal Burniat surfaces with $K^2_S = 4$ is an irreducible connected component, normal, unirational of dimension 3.

Moreover the base of the Kuranishi family of deformations of such a minimal model $S$ is smooth.

2) The subset $\mathcal{NEB}_3$ of the moduli space of canonical surfaces of general type $\mathcal{M}_{1,3}$ corresponding to extended and nodal Burniat surfaces with $K^2_S = 3$ is an irreducible open set, normal, unirational of dimension 4.

Moreover the base of the Kuranishi family of $S$ is smooth.
Proof. We show here the smoothness of the base $\mathcal{B}$ of the Kuranishi family. This follows, in the case $K^2 = 4$, from the fact that we have shown (Proposition 3.1) the existence of a family with smooth three dimensional base $B$ mapping to the moduli space, hence also to the base $\mathcal{B}$ of the Kuranishi family, in a quasi-finite way. Since the tangent dimension of $\mathcal{B}$ is also equal to three, it follows that the dimension of $\mathcal{B}$ is exactly three, and that $\mathcal{B}$ is smooth. The argument in the case $K^2 = 3$ is identical, using Proposition 5.2.

We are also almost done with the proof of our second main theorem

**Theorem 0.2** The deformations of nodal Burniat surfaces with $K^2_S = 4, 3$ to extended Burniat surfaces with $K^2_S = 4, 3$ yield examples where $\text{Def}(S, (\mathbb{Z}/2\mathbb{Z})^2) \to \text{Def}(X, (\mathbb{Z}/2\mathbb{Z})^2)$ is not surjective. Moreover, $\text{Def}(S, (\mathbb{Z}/2\mathbb{Z})^2) \subset \text{Def}(S)$, whereas for the canonical model we have: $\text{Def}(X, (\mathbb{Z}/2\mathbb{Z})^2) = \text{Def}(X)$.

The moduli space of pairs, of an extended (or nodal) Burniat surface with $K^2_S = 4, 3$ and a $(\mathbb{Z}/2\mathbb{Z})^2$-action, is disconnected; but its image in the moduli space is a connected open set.

**Proof.** By Propositions 3.1 and 3.2 we have two families with smooth connected rational base of dimension 3, resp. 4, parametrizing all the canonical models $X$ of the surfaces in $\mathcal{NEB}_4$, resp. $\mathcal{NEB}_3$.

In the previous Theorem 3.1 we showed that the base of the Kuranishi family of $S$ is smooth, hence base change of these families yield the Kuranishi family of $S$.

The above families of canonical models $X$ yield the Kuranishi family of $X$, e.g., by the Theorem of Burns and Wahl.

Propositions 3.1 and 3.2 exhibiting all the canonical models as bidouble covers of normal Del Pezzo surfaces, immediately show that $\text{Def}(X, (\mathbb{Z}/2\mathbb{Z})^2) = \text{Def}(X)$.

Let now $S$ be a nodal Burniat surface.

Since, by (7.1), page 23, of CatSS, $\text{Def}(S, (\mathbb{Z}/2\mathbb{Z})^2) \subset \text{Def}(S)$ is the intersection with $H^1(\Theta_S)^0$, which is the smooth subvariety corresponding to the nodal Burniat surfaces, we obtain that $\text{Def}(S, (\mathbb{Z}/2\mathbb{Z})^2) \subset \text{Def}(S)$.

On the other hand, for instance in the case $K^2_S = 4$, we explicitly see that $\mathcal{NEB}_4$ is the union of two families of bidouble covers, the family of nodal Burniat surfaces, respectively the family of extended Burniat surfaces; hence the moduli space of pairs $(S, (\mathbb{Z}/2\mathbb{Z})^2)$ has exactly two connected components.

5. One parameter limits of extended Burniat surfaces with $K^2_S = 4$

In this section we shall prove the following:
Theorem 5.1. The family of extended Burniat surfaces with $K_S^2 = 4$ yields, together with the family of nodal Burniat surfaces with $K_S^2 = 4$, a closed subset $NEB_4$ of the moduli space.

This will be accomplished through the study of limits of one parameter families of such extended Burniat surfaces: we shall indeed show that only nodal Burniat surfaces (or extended Burniat surfaces) occur.

Let $Y'$ be a normal $\mathbb{Q}$-Gorenstein surface and denote the dualizing sheaf of $Y'$ by $\omega_{Y'}$.

Then there is a minimal natural number $m$ such that $\omega_{Y'} \otimes m$ is an invertible sheaf and it makes sense to define $\omega_{Y'}$ to be ample, respectively anti-ample; $Y'$ is Gorenstein iff $m = 1$.

We recall the following result which was shown in [BC10].

Proposition 5.2. Let $T$ be a smooth affine curve, $t_0 \in T$, and let $f: X \to T$ be a flat family of canonical surfaces. Suppose that $X_t$ is the canonical model of a Burniat surface with $4 \leq K^{2}_{X_t}$ for $t \neq t_0 \in T$. Then there is a birational action of $G := (\mathbb{Z}/2\mathbb{Z})^2$ on $X$ yielding a one parameter family of finite $(\mathbb{Z}/2\mathbb{Z})^2$-covers

\[ X \twoheadrightarrow Y \rightarrow \rightarrow \rightarrow T, \]

(i.e., $X_t \to Y_t$ is a finite $(\mathbb{Z}/2\mathbb{Z})^2$-cover), such that $Y_t$ is a Gorenstein Del Pezzo surface for each $t \in T$.

Observe that the above result remains true if we replace “Burniat surface” by “extended Burniat surface”.

This implies immediately the following:

Corollary 5.3. Consider a one parameter family of bidouble covers $X \to Y$ as in prop. 5.2. Then the branch locus of $X_{t_0} \to Y_{t_0}$ is the limit of the branch locus of $X_t \to Y_t$, and it is reduced.

Note that the limit of a line on the del Pezzo surfaces $Y_t$ is a line on the del Pezzo surface $Y_{t_0}$, and, as a consequence of the above assertion, two lines in the branch locus in $Y_t$ cannot tend to the same line in $Y_{t_0}$.

Remark 5.4. Let $X$ be the canonical model of an extended or nodal Burniat surface with $K_X^2 = 4$. Recall that $X$ is smooth for a general member of the family of extended Burniat surfaces, whereas $X$ has one ordinary node if $X$ is the canonical model of a nodal Burniat surface with $K^2 = 4$.

In the extended case the branch locus consists of the union of 3 hyperplane sections, containing 8 lines, 2 conics and the node. In the nodal Burniat case one of the conics degenerates to two lines, hence the branch locus consists instead of 10 lines and one conic.

The first step towards proving Theorem 5.1 is the following:
Proposition 5.5. Consider a one parameter family of bidouble covers $X \to Y$ as in prop. 5.2 except that $X_t$ is an extended Burniat surface with $K_{X_t}^2 = 4$ for $t \neq 0$. Then $Y_0$ is a normal Del Pezzo surface with exactly one node as singularity.

Lemma 5.6. A normal singular Del Pezzo surface with $K_{Y_0}^2 = 4$ containing at least 8 lines has as singularities either 

1. one node, and then it contains 12 lines, or 
2. two nodes, and then it contains 9 lines, or 
3. an $A_2$ singularity, and then it contains 8 lines, 4 of which pass through the singular point.

Proof. The assertion is a generalization of Proposition 3.6 of [BC10], page 581, see especially the proof in the appendix ibidem, pages 585-587.

We blow up $r = 5$ points in the plane. By [BC10], p.586, we have that the loss of number of lines when one has a chain of $k$ infinitely near points is bounded by 

$$(k-1)(r-(k-1)) + \frac{1}{2}(k+1)(k-2) = (k-1)(5-(k-1)) + \frac{1}{2}(k+1)(k-2).$$

Since a smooth Del Pezzo surface of degree four has 16 lines, $k \geq 4$ implies that the number of lines is less than or equal to $16 - 11 = 5$.

If there is a chain with $k = 3$, the same estimate gives a loss of 8, and we cannot then have other (-2)-curves, else the number would be strictly smaller than $16 - 8 = 8$.

In this case we get an $A_2$ singularity and 8 lines.

In fact, in the chosen plane model we have 5 points lying on an irreducible conic $C$, of which $P_2$ infinitely near to $P_1$, and $P_3$ infinitely near to $P_2$. The lines are given by 

$E_3, E_4, E_5, |L - E_1 - E_4|, |L - E_1 - E_5|, |L - E_1 - E_2|, |L - E_4 - E_5|, C'$, 

where $C'$ is the strict transform of $C$.

In this case the 4 lines passing through the singular point are 

$E_3, |L - E_1 - E_4|, |L - E_1 - E_5|, |L - E_1 - E_2|.$

In the case where there is no chain of three infinitely near points by a standard Cremona transformation as in [BC10], ibidem, we may reduce to the case where there are no infinitely near points and then we have that the weak Del Pezzo surface is $\hat{Y}_0 := \hat{\mathbb{P}}^2(P_1, \ldots, P_3)$, where $P_1, P_2, P_3$ and $P_1, P_4, P_5$ are collinear.

Then $\hat{Y}_0$ contains nine lines. In fact, the set of lines of $\hat{Y}_0$ is: 

$L := \{E_1, \ldots, E_5, L - E_2 - E_4, L - E_2 - E_5, L - E_3 - E_4, L - E_4 - E_5\}.$

Proof of prop. 5.5. Since the branch locus of $X_t \to Y_t$ contains eight lines for $t \neq 0$, also the branch locus of $X_0 \to Y_0$ contains eight lines.

We want to show that cases (2) and (3) of the previous lemma cannot occur.
We start by eliminating case (3).
Here, the $A_2$ singularity must be a limit of the node of $Y_1$, hence the bidouble cover is branched at the singular point.

The bidouble cover is a RDP, hence, looking at table 2, page 90 of [Cat87], and table 3, page 93 ibidem, we see that the branch locus is analytically isomorphic to

- an ordinary cusp $\{y = 0 = z^2 + x^3 + t^4 = 0\}$ for $E_6 = \{z^2 + x^3 + y^2 = 0\}$,
- two lines $\{x = 0 = z^2 + y^2 = 0\}$ for $A_5 = \{z^2 + w^6 + y^2 = 0\} \to A_2 = \{z^2 + x^3 + y^2 = 0\}$,
- two lines $\{x = 0 = z^2 + y^2 = 0\}$ for the composition of $A_2 \to A_5$ (ramified only at the singular point) with the previous $A_5 = \{z^2 + w^6 + y^2 = 0\} \to A_2 = \{z^2 + x^3 + y^2 = 0\}$.

We observe however that by our previous arguments the branch locus contains the 8 lines, 4 of which pass through the $A_2$ singularity, contradicting the above local description of the branch locus.

Assume now by contradiction that we have case (2), i.e., $Y_0$ has two nodes. Then

**Claim 5.7.** $E_1$ is not a component of the total branch locus $\Delta$ of $\hat{X}_0 \to \hat{Y}_0$, i.e.,

$$E_2, \ldots, E_5, L - E_2 - E_4, L - E_2 - E_5, L - E_3 - E_4, L - E_3 - E_5$$

are exactly the 8 lines contained in $\Delta$.

**Proof of the claim.** Assume that $E_1$ is contained in the total branch locus $\Delta$ of the bidouble cover $\hat{X}_0 \to \hat{Y}_0$. Then $\Delta$ contains three lines intersecting one of the two $(-2)$ curves. But a bidouble cover of a node branched in at least three lines does not give a rational double point, as shown by the classification recalled in Section 2. A contradiction. $\Box$

Since for each node $\nu_1, \nu_2$ there are two lines in the total branch divisor passing through $\nu_i$, it follows by the classification given in Section 2 that $N_1, N_2 \leq \Delta$ and that $(\Delta - N_i)N_i = 2$.

Denote by $\pi: \hat{Y}_0 \to Y'$ be the desingularization map. Then $\pi_*(\Delta) \equiv -3K_{Y'}$, whence

$$\Delta \equiv -3K_{Y_0} + n_1N_1 + n_2N_2.$$ 

Then $2 = (\Delta - N_i)N_i = (n_i - 1)N_i^2 = 2(1 - n_i) \Leftrightarrow n_i = 0$.

We conclude that

$$\Delta \equiv -3K_{Y_0}.$$ 

Observe that

$$-3K_{Y_0} = \sum_{l \in L \setminus \{E_1\}} l - N_1 - N_2 \equiv 3L - E_1 - E_2 - \ldots - E_5.$$
Since no other component of $\Delta$ can intersect the $(-2)$-curves, we see immediately that the remaining two components of $\Delta$ are:

$$L - E_1, 2L - E_2 - E_3 - E_4 - E_5.$$ \[ \square \]

We write now

$$\Delta_1 = \lambda_1 L - E_1 - a_2 E_2 - a_3 E_3 - a_4 E_4 - a_5 E_5,$$

$$\Delta_2 = \lambda_2 L - E_1 - b_2 E_2 - b_3 E_3 - b_4 E_4 - b_5 E_5,$$

$$\Delta_3 = \lambda_3 L - E_1 - c_2 E_2 - c_3 E_3 - c_4 E_4 - c_5 E_5.$$ \[ \square \]

Here we have used that, since $E_1$ is not a component of $\Delta$ and since $\Delta_i + \Delta_j$ has to be divisible by two, the only possibility is

$$E_1 : (\Delta_1, \Delta_2, \Delta_3) = (1, 1, 1).$$ \[ \square \]

Note that, since $\lambda_1 + \lambda_2 + \lambda_3 = 9$ (and again since $\Delta_i + \Delta_j$ is divisible by two) we have:

$$(\lambda_1, \lambda_2, \lambda_3) \in \{(3, 3, 3), (1, 3, 5), (1, 1, 7)\}.$$ \[ \square \]

Moreover, since the branch divisor is reduced, for each $i$ it happens that, among the three numbers $a_i, b_i, c_i$, there cannot be two which are negative, and if one such a number is negative, then it is $=-1$; hence the only possibilities are:

$$\{a_i, b_i, c_i\} \in \{1, 1, 1\} \text{ or } \{-1, 1, 3\},$$

for $i \in \{2, \ldots, 5\}$. \[ \square \]

$$(\lambda_1, \lambda_2, \lambda_3) = (3, 3, 3); \text{ then we get for the character sheaves:}$$

$$\mathcal{L}_1 = \mathcal{O}(3L - E_1 - \frac{b_2 + c_2}{2} E_2 - \frac{b_3 + c_3}{2} E_3 - \frac{b_4 + c_4}{2} E_4 - \frac{b_5 + c_5}{2} E_5),$$

$$\mathcal{L}_2 = \mathcal{O}(3L - E_1 - \frac{a_2 + c_2}{2} E_2 - \frac{a_3 + c_3}{2} E_3 - \frac{a_4 + c_4}{2} E_4 - \frac{a_5 + c_5}{2} E_5),$$

$$\mathcal{L}_3 = \mathcal{O}(3L - E_1 - \frac{a_2 + b_2}{2} E_2 - \frac{a_3 + b_3}{2} E_3 - \frac{a_4 + b_4}{2} E_4 - \frac{a_5 + b_5}{2} E_5).$$ \[ \square \]

Note that $(a_i, b_i, c_i) = (1, 1, 1)$ for all $i \in \{2, \ldots, 5\}$ implies that $p_g(X_0) \neq 0$, whence w.l.o.g.

$$(a_2, b_2, c_2) = (-1, 1, 3).$$ \[ \square \]

Then $E_2 \leq \Delta_1$, and by the local calculations in Section 2 this implies that also $E_3 \leq \Delta_1$ (since the two lines of the branch locus intersecting a $(-2)$-curve belong to the same $\Delta_i$). Therefore

$$(a_3, b_3, c_3) \in \{(-1, *, *), (1, 1, 1)\}. \square$$

Again using $p_g(X_0) = 0$, we conclude (looking at $\mathcal{L}_3$) that (up to exchanging $P_4$ with $P_5$)

$$(a_4, b_4, c_4) \in \{(3, 1, -1), (1, 3, -1)\},$$

and again this implies that

$$(a_5, b_5, c_5) \in \{(*, *, -1), (1, 1, 1)\}. \square$$

But in all of these cases we have

$$\frac{a_i + c_i}{2} \in \{0, 1\} \ \forall \ i \in \{2, \ldots, 5\},$$

\[ \square \]
contradicting $p_9 = 0$.

$(\lambda_1, \lambda_2, \lambda_3) = (1, 3, 5)$ : here we have

\[
L_1 = O(4L - E_1 - \frac{b_2 + c_2}{2}E_2 - \frac{b_3 + c_3}{2}E_3 - \frac{b_4 + c_4}{2}E_4 - \frac{b_5 + c_5}{2}E_5),
\]

\[
L_2 = O(3L - E_1 - \frac{a_2 + c_2}{2}E_2 - \frac{a_3 + c_3}{2}E_3 - \frac{a_4 + c_4}{2}E_4 - \frac{a_5 + c_5}{2}E_5),
\]

\[
L_3 = O(2L - E_1 - \frac{a_2 + b_2}{2}E_2 - \frac{a_3 + b_3}{2}E_3 - \frac{a_4 + b_4}{2}E_4 - \frac{a_5 + b_5}{2}E_5).
\]

Again, $p_9 = 0$ implies that there is an $i \in \{2, \ldots, 5\}$ such that $\frac{a_i + c_i}{2} = 2$. W.l.o.g. we can assume that $\frac{a_2 + c_2}{2} = 2$. Therefore

\[
(a_2, b_2, c_2) \in \{(3, -1, 1), (1, -1, 3)\},
\]

whence

\[
(a_2, b_2, c_2) \in \{(3, -1, 1), (1, -1, 3), (1, 1, 1)\}.
\]

Then \(\frac{a_2 + c_2}{2}, \frac{a_3 + c_3}{2} \leq 1\) and \(\frac{a_2 + c_2}{2}, \frac{a_3 + c_3}{2} \leq 2\), which implies that $O(L - E_2 - E_4) \subset O(K_{Y_0} \otimes L_1)$, contradicting $p_9(X_0) = 0$.

$(\lambda_1, \lambda_2, \lambda_3) = (1, 1, 7)$ : this case can be excluded since

\[
4 = \Delta_3 \cdot (K_{Y_0}) = 3\lambda_3 - 1 - \sum_{i=2}^{5} c_i \geq \sum_{i=2}^{5} c_i = 16,
\]

a contradiction.

This proves the proposition. \(\square\)

Consider a one parameter family of bidouble covers $X \to Y$ as in prop. \[5.5\]

Then $Y' := Y_0$ is a normal Del Pezzo surface with exactly one node. Let $\hat{Y}$ be the blow up of $Y'$ in the node and denote the exceptional (-2)-curve of $\hat{Y}$ over the node by $A$.

The following result concludes the proof of Theorem \[5.4\].

**Proposition 5.8.** For the limit of a one parameter family of extended Burniat surfaces with $K_Y^2 = 4$ we have:

1) if $A$ does not intersect $\Delta - A$, then $X_0$ is an extended Burniat surface with $K_Y^2 = 4$ ;

2) if $A$ intersects $\Delta - A$, then $X_0$ is a nodal Burniat surface with $K_Y^2 = 4$ .

**Proof.** We can assume that $\hat{Y} = \mathbb{P}^2(P_1, \ldots, P_5)$, and w.l.o.g. $P_1, P_4, P_5$ collinear, i.e., $A \equiv L - E_1 - E_4 - E_5$.

Recall that we have shown that in both cases $A$ is contained in the branch locus, hence the two alternatives are that $A$ is a connected component of the branch locus, or not.

1) In the first case, arguing as in Proposition \[5.5\], we get that the total branch locus is $\Delta \equiv -3K_{\hat{Y}} + A$. 

\[ \text{Documenta Mathematica 18 (2013) 1089–1136} \]
It is easy to see that $\tilde{Y}$ contains exactly 8 lines $l_1, \ldots, l_8$ which do not intersect $A$. Then these 8 lines have to be contained in $\Delta$.

Then $\Delta - A - \sum_{i=1}^8 l_i \equiv 3L - \sum E_i$, which has to split into two Del Pezzo conics, which then have to be $L - E_1$ and $2L - \sum_{i=2}^5 E_i$. Hence we get an extended Burniat surface.

2) Here $L - E_1 - E_4 - E_5 \equiv A \leq \Delta \equiv -K_{\tilde{Y}}$.

Observe that $\tilde{Y}$ contains exactly 4 lines intersecting $A$: $E_1, E_4, E_5, L - E_2 - E_3$.

By our local calculations in Section 2 two of these four lines are components of the total branch divisor and the two other not.

W.l.o.g. we can assume $E_1, L - E_2 - E_3 \leq \Delta$. Since $E_4$ and $E_5$ are not contained in the branch divisor, we see (writing $\Delta_i$ as in the proof of Proposition 5.5) that $(a_4, b_4, c_4) = (a_5, b_5, c_5) = (1, 1, 1)$.

Now it is straightforward that $(\lambda_1, \lambda_2, \lambda_3) = (3, 3, 3)$ (use the same argument as in the proof of prop. 5.5 to exclude the cases $(1, 3, 5)$ and $(1, 1, 7)$).

Since $p_g = 0$, we have (up to a permutation of $\{1,2,3\}$)

$$\frac{b_1 + c_1}{2} = \frac{a_2 + c_2}{2} = \frac{a_3 + b_3}{2} = 2.$$

W.l.o.g. we can assume $(a_1, b_1, c_1) = (-1, 1, 3)$; then $E_1, L - E_2 - E_3 \leq \Delta_1$. Therefore

$$(a_2, b_2, c_2) \in \{(3, -1, 1), (1, -1, 3)\}$$

and

$$(a_3, b_3, c_3) \in \{(3, 1, -1), (1, 3, -1)\}.$$

But only $(a_2, b_2, c_2) = (3, -1, 1)$ and $(a_3, b_3, c_3) = (1, 3, -1)$ is possible (since a cubic cannot have two triple points, i.e., this would contradict the effectivity of $\Delta_i$ for some $i$).

Therefore we get a nodal Burniat surface. \hfill \Box

6. NODAL AND EXTENDED BURNIAT SURFACES DO NOT FORM A CLOSED SET FOR $K^2_S = 3$

We are going to exhibit surfaces which are in the closure of the family of nodal and extended Burniat surfaces, but for which the image of the bicanonical map is a normal cubic with other singularities than 3 nodes.

In our first example we exhibit a 3-dimensional family with a 4-nodal cubic as image. Consider a specialization of the 6 points $P_1, \ldots, P_6$ in $\mathbb{P}^2$ so that $P_1, P_2, P_3$ become collinear, and, more precisely, the point $P_2$ moves in the line joining $P_4$ and $P_6$ till it reaches the line joining $P_1$ and $P_3$.

Then $P_1, \ldots, P_6$ are the vertices of a complete quadrilateral with sides $N_1, N_2, N_3, N_4$: here we identify $N_4$ to the (-2) curve $N_4 \equiv L - E_1 - E_2 - E_3$ on the weak Del Pezzo $\tilde{Y}$ of degree 3 obtained blowing up the 6 points. Our notation for $N_1, N_2, N_3$ remains the same, and $\tilde{Y}$ is the minimal resolution of the 4-nodal cubic surface $Y'$ := $\Sigma$. 

Documenta Mathematica 18 (2013) 1089–1136
We consider exactly the same divisors as the strictly extended Burniat divisors in 4) of Definition 1.4. We obtain a three dimensional family of bidouble covers \(X\) of \(\Sigma\), with total branch locus consisting of 9 connected components, namely:

\[N_1, N_2, N_3; \Gamma_1, \Gamma_2, \Gamma_3; G_1, G_2, G_3.\]

\(G_1, G_2, G_3\) correspond to the three diagonals of the quadrilateral, and are the 3 lines of \(\Sigma\) not passing through the nodes, whereas \(\Gamma_1, \Gamma_2, \Gamma_3\) are conics as in Definition 1.4. The canonical models \(X\) have therefore 4 nodes lying over the node of \(\Sigma\) corresponding to \(N_4\).

We have therefore proven:

**Proposition 6.1.** The closure of the (4-dimensional) open set corresponding to nodal and extended Burniat surfaces with \(K^2_\Sigma = 3\) contains a 3-dimensional family of canonical models which are bidouble covers of a 4-nodal cubic surface \(\Sigma\).

Each such surface \(X\) has 4 nodes, lying over one fixed node of \(\Sigma\), and where the bicanonical map \(\Phi_2 : X \to \Sigma\) is unramified.

In our second example we obtain a 3-dimensional family of bidouble covers of a cubic surface \(Y'\) with a singularity of type \(D_4\).

We give this example using the different planar realization which was indeed the way we found our first description of the deformation of nodal Burniat surfaces with \(K^2_\Sigma = 3\) to extended Burniat surfaces.

To do this, we relabel the 6 points in the plane as follows:

\[P'_3 := P_4, \quad P'_2 := P_5, \quad P'_1 := P_6.\]

We have therefore irreducible rational curves

\[D_{i,1} := L - E_i - E_{i+1}, \quad D_{i,2} := N_i = L - E_i - E'_{i+1} - E''_{i+2},\]

\[D_{i,3} := G_i = L - E_i - E'_i\]

on the weak Del Pezzo \(\tilde{Y}\) of degree 3.

Blowing down the 3 (-1) curves \(D_{i,1}\) \((i = 1, 2, 3)\) first, and then the strict transform of the 3 (-2) curves \(D_{i,2}\) \((i = 1, 2, 3)\) we obtain another copy of the projective plane where one has blown up three points \(Q_i\) \((i = 1, 2, 3)\), where \(Q'_i\) is infinitely near to \(Q_i\).

We denote by slight abuse of notation by \(Q_i\) the full transform of the point \(Q_i\), namely, the divisor \(D_{i,1} + D_{i-1,2}\), and by \(Q'_i\) the full transform of the point \(Q'_i\), namely, the divisor \(D_{i,1}\).

The pull back of the system of lines in the new \(\mathbb{P}^2\) is, by the Hurwitz formula, the linear system

\[\mathcal{L} := 4L - 2 \sum E_i - \sum E'_i.\]

And the curve \(D_{i,3}\) is linearly equivalent to

\[D_{i,3} \equiv \mathcal{L} - 2D_{i+1,1} - D_{i,2} = \mathcal{L} - Q_{i+1} - Q'_{i+1}.\]
Hence $D_{i-1,2} = Q_1 - Q'_i$, and we can write the branch loci for the extended Burniat surfaces as:

$$\Delta_i \in D_{i,3} + D_{i+1,2} + |D_{i,3} + D_{i+1,3}| = D_{i,3} + |Q_{i-1} - Q'_{i-1}| + |D_{i,3} + D_{i+1,3}| =$$

$$= |\mathcal{L} - Q_{i+1} - Q'_{i+1}| + |Q_{i-1} - Q'_{i-1}| + |2\mathcal{L} - Q_{i+1} - Q'_{i+1} - Q_{i-1} - Q'_{i-1}| =$$

$$= |\mathcal{L} - Q_{i+1} - Q'_{i+1}| + N_{i-1} + |2\mathcal{L} - Q_{i+1} - Q'_{i+1} - Q_{i-1} - Q'_{i-1}|.$$

Now, we simply let the three points $Q_1, Q_2, Q_3$ become collinear, but we let the tangent directions $Q'_i$ remain general.

The blow up of the plane in the 6 points possesses now 4 (-2) curves, the three curves $N_1, N_2, N_3$ and the strict transform $N$ of the line through $Q_1, Q_2, Q_3$. Since $N$ intersects each $N_i$ and these are disjoint, the corresponding normal Del Pezzo surface $Y'$ has a singularity of type $D_4$.

Letting the branch divisor be as before (namely, take pull backs of general conics in $|2\mathcal{L} - Q_{i+1} - Q'_{i+1} - Q_{i-1} - Q'_{i-1}|$), we obtain

**Proposition 6.2.** The closure of the (4-dimensional) open set corresponding to nodal and extended Burniat surfaces with $K^2_Y = 3$ contains a 3-dimensional family of canonical models which are bidouble covers of a normal cubic surface $Y'$ with a singularity of type $D_4$.

The branch locus on $Y'$ has the singular point as an isolated point, and the local covering is determined by the epimorphism $D_4 \to (\mathbb{Z}/2\mathbb{Z})^2 = (D_4)^{ab}$ of the local fundamental group of the singularity to its abelianization.

**Proof.** The inverse image of the (-2) curves in the bidouble cover are: the inverse image $N'$ of $N$, which is a (-8) curve, and, for each $N_i$, there is a pair of (-1) curves meeting $N'$. After contracting the 6 (-1) curves we obtain a (-2) curves.

$$\square$$

7. **Appendix 1: a corrigendum to Burniat surfaces II**

Parts 1), 2) and 3) of the following lemma were contained in Lemma 2.10 of [BC10], while 4) corrects a wrongly stated assertion of 2) of loc. cit.

We also amend the proof for the correct assertions.

**Lemma 7.1.** Consider a finite set of distinct linear forms

$$l_\alpha := y - c_\alpha x, \alpha \in A$$

vanishing at the origin in $\mathbb{C}^2$.

Let $p: Z \to \mathbb{C}^2$ be the blow up of the origin, let $D_\alpha$ be the strict transform of the line $L_\alpha := \{l_\alpha = 0\}$, and let $E$ be the exceptional divisor.

Let $\Omega^1_{\mathbb{C}^2}((d\log l_\alpha)_{\alpha \in A})$ be the sheaf of rational 1-forms $\eta$ generated by $\Omega^1_{\mathbb{C}^2}$ and by the differential forms $d\log l_\alpha$ as an $\mathcal{O}_{\mathbb{C}^2}$-module and define similarly $\Omega^1_E((d\log D_\alpha)_{\alpha \in A})$. Then:

1) $p_*\Omega^1_{\mathbb{C}^2}(\log E)(-E) = \Omega^1_{\mathbb{C}^2}$,

2) $p_*\Omega^1_E(\log E, (\log D_\alpha)_{\alpha \in A}) = \Omega^1_{\mathbb{C}^2}((d\log l_\alpha)_{\alpha \in A})$.

---

Documenta Mathematica 18 (2013) 1089–1136
3) \( p_\ast \Omega^1_Z((\log D_\alpha)_{\alpha \in A}) = \{ \eta \in \Omega^1_{\mathcal{C}^2}((\log l_\alpha)_{\alpha \in A}) | \eta = \sum_\alpha g_\alpha \log l_\alpha + \omega, \omega \in \Omega^1_{\mathcal{C}^2}, \sum_\alpha g_\alpha(0) = 0 \} \).

4) \( p_\ast \Omega^1_Z((\log D_\alpha)_{\alpha \in A})(E) \supset \Omega^1_{\mathcal{C}^2}((\log l_\alpha)_{\alpha \in A}) \) and

\[
\dim_\mathbb{C}[p_\ast \Omega^1_Z((\log D_\alpha)_{\alpha \in A})(E)/\Omega^1_{\mathcal{C}^2}((\log l_\alpha)_{\alpha \in A})] = d - 2
\]

is supported at the origin, where \( d := |A| \). More precisely, we have an exact sequence

\[
0 \to \Omega^1_{\mathcal{C}^2} \to p_\ast \Omega^1_Z((\log D_\alpha)_{\alpha \in A})(E) \to \bigoplus_{\alpha = 1}^d \mathcal{O}_{\mathcal{D}_\alpha}(0) \to \mathbb{C}^2_0 \to 0.
\]

5) Assume w.l.o.g. \( c_1 = 0 \) in the following formulae: then

\[
p_\ast \Omega^1_Z(\log D_1)(-E) \subset \Omega^1_{\mathcal{C}^2}(\log l_1)
\]

is the subsheaf of forms

\[
\{ \omega = adx + \beta \frac{dy}{y} | \beta(0) = 0, \frac{\partial \beta}{\partial y}(0) = 0, \frac{\partial \beta}{\partial x}(0) + \alpha(0) = 0 \}.
\]

6) \( p_\ast \Omega^1_Z(-E) = m_0 \Omega^1_{\mathcal{C}^2} \), where \( m_0 \) is the maximal ideal of the origin 0 in the sheaf \( \mathcal{O}_{\mathcal{C}^2} \).

7) \( p_\ast \Omega^1_Z(\log D_1, \log D_2)(-E) \subset \Omega^1_{\mathcal{C}^2}(\log l_1, \log l_2) \) is the subsheaf of forms

\[
\{ \omega = \frac{dx}{x} + \beta \frac{dy}{y} | \alpha(0) = 0, \beta(0) = 0, \frac{\partial (\alpha + \beta)}{\partial x}(0) = 0, \frac{\partial (\alpha + \beta)}{\partial y}(0) = 0 \}.
\]

Proof. We show 2), 3), 4), 5) and 7).

Observe that

\[
p_\ast \Omega^1_Z((\log D_\alpha)_{\alpha \in A})(mE)
\]

consists of rational differential 1-forms \( \omega \) which, when restricted to \( \mathbb{C}^2 \setminus \{0\} \), yield sections of \( \Omega^1_{\mathcal{C}^2}((\log l_\alpha)_{\alpha \in A}) \).

Therefore, in particular, \( \omega \prod_{\alpha \in A} l_\alpha \) is a regular holomorphic 1-form on \( \mathbb{C}^2 \).

Hence \( \omega \), modulo holomorphic 1-forms, can be written as

\[
\omega = \frac{f}{\prod_{\alpha \in A} l_\alpha} \ dx + \frac{g}{\prod_{\alpha \in A} l_\alpha} \ dy,
\]

where \( f, g \) are Weierstrass pseudopolynomials of degree in \( y \) strictly less than \( d := \text{card}(A) \).

Since \( dy = dl_\alpha + c_\alpha dx \), the condition that \( \omega \) restricted to \( \mathbb{C}^2 \setminus \{0\} \) yields a section of \( \Omega^1_{\mathcal{C}^2}((\log l_\alpha)_{\alpha \in A}) \) implies that \( l_\alpha | (f + c_\alpha g) \).

Whence \( l_\alpha \) divides \( fx + yg \), and we conclude, since \( \prod_{\alpha \in A} l_\alpha \) is a pseudo polynomial of degree \( d \), that

\[
fx + yg = c(x) \prod_{\alpha \in A} l_\alpha.
\]

This allows us to write, modulo holomorphic 1-forms,

\[
\omega = \frac{g(dy - \frac{g}{f} dx)}{\prod_{\alpha \in A} l_\alpha} + \frac{c}{x} \ dx,
\]

\( \text{Documenta Mathematica 18 (2013) 1089–1136} \)
where now $c \in \mathbb{C}$.

Let us pull back $\omega$ to $Z$, using local coordinates $(x, t)$ such that $y = xt$, and where we make the assumption $c_{\alpha} \neq 0, \forall \alpha$.

$$p^* \omega = \frac{x^{-d}g(x, xt)(xdt)}{\prod_{\alpha \in A}(t - c_{\alpha})} + \frac{c}{x} dx.$$ 

The pull back form has logarithmic poles along $E = \{x = 0\}$ iff $g(x, y)$ has multiplicity at least $d - 1$ at the origin, and poles of order at most one along $E$ iff $g(x, y)$ has multiplicity at least $d - 2$ at the origin.

Observe that the $d$ polynomials $P_{\beta} := \prod_{\alpha \in A, \alpha \neq \beta} l_{\alpha}$ are linearly independent and homogeneous of degree $d - 1$, hence they generate the vector space of homogeneous polynomials of degree $d - 1$, hence they generate the ideal of holomorphic functions vanishing at the origin of multiplicity at least $d - 1$.

Hence $g(x, y)$ has multiplicity at least $d - 1$ iff we can write

$$g = \sum_{\alpha \in A} g_{\alpha} P_{\alpha}.$$ 

And since $g$ is a pseudo polynomial of degree $\leq d - 1$, the $g_{\alpha}$’s are just functions of $x$.

In this case we can write

$$\omega = \frac{c}{x} dx + \sum_{\alpha \in A} \frac{g_{\alpha}}{l_{\alpha}} (dy - \frac{y}{x} dx) = \frac{1}{x} \left( cdx + \sum_{\alpha \in A} \frac{g_{\alpha}}{l_{\alpha}} (xdy - ydx) \right).$$

$$= \frac{1}{x} \left( cdx + \sum_{\alpha \in A} \frac{g_{\alpha}}{l_{\alpha}} (xdl_{\alpha} + xc_{\alpha} dx - ydx) \right) = \sum_{\alpha \in A} \frac{g_{\alpha}}{l_{\alpha}} dl_{\alpha} + \frac{1}{x} dx \left( c - \sum_{\alpha \in A} g_{\alpha} \right).$$

The above form $\omega$ does not have poles on the line $x = 0$ if and only if $c = (\sum_{\alpha \in A} g_{\alpha}(0))$.

Observing that the strict transform of the line $x = 0$ is not among the divisors $D_{\alpha}$, we establish claim (2), while (3) follows since $c = 0$ iff there are no poles along $E$.

The very first assertion of (4) follows by (2), so let’s proceed to verify the other assertions.

Assume now that $p^* \omega$ has poles of order 1 along $E$; equivalently, assume that $g(x, y)$ has multiplicity at least $d - 2$ at the origin. Since we already dealt with the case where this multiplicity is at least $d - 1$, we may assume that $g(x, y)$ is homogeneous of degree $d - 2$, and that $c = 0$.

Arguing as before, the space of homogeneous polynomials of degree $d - 2$ has as basis the $d - 1$ polynomials $(\beta = 1, \ldots, d - 1)$

$$Q_{\beta} := \prod_{\alpha \in A, \alpha \neq \beta, \alpha \neq d} l_{\alpha}.$$ 

Whence $g = \sum_{\alpha \in A, \alpha \neq d} g_{\alpha} Q_{\alpha}$, where $g_{\alpha} \in \mathbb{C}$, and we may write:
\[ \omega = \sum_{\alpha=1}^{d-1} \frac{g_\alpha}{l_\alpha l_d} (dy - \frac{y}{x} dx). \]

Since we want no poles on the line \( x = 0 \), we must have
\[ \sum_{\alpha=1}^{d-1} \frac{g_\alpha y}{y^2} = 0 \iff \sum_{\alpha=1}^{d-1} g_\alpha = 0. \]

Under this condition we may then write
\[ \omega = \sum_{\alpha=1}^{d-1} \frac{g_\alpha}{l_\alpha l_d} (dl_\alpha), \]
which has logarithmic poles along \( l_\alpha = 0 \).

Logarithmic poles along \( l_d = 0 \) follow by writing
\[ \omega = \sum_{\alpha=1}^{d-1} \frac{g_\alpha}{l_\alpha l_d} (dl_\alpha) + \sum_{\alpha=1}^{d-1} \frac{g_\alpha (c_d - c_\alpha)}{l_\alpha l_d} dx, \]
and observing that \( \sum_{\alpha=1}^{d-1} \frac{g_\alpha (c_d - c_\alpha)}{y - c_\alpha} \) vanishes for \( l_d = 0 \) since on \( \{l_d = 0\} \) we have \( y = c_d x \).

Applying the residue sequence, we see that each such form \( \omega \) has as residue on \( D_\alpha \) a function with a single pole at most at the origin \( O \), and with coefficient of \( \frac{1}{x} \) respectively equal to \( r_\alpha = \sum_{\alpha=1}^{d-1} \frac{g_\alpha (c_d - c_\alpha)}{(c_d - c_\alpha)} \) in the case of \( D_d \), and \( r_\alpha := -\frac{g_\alpha}{c_d - c_\alpha} \) in the case of \( D_\alpha \).

In other words, the sum of the ‘double’ residues is 0, and the other condition \( \sum_{\alpha=1}^{d-1} g_\alpha = 0 \) can be also written down as \( \sum_{\alpha=1}^{d} c_\alpha r_\alpha = 0 \).

To show 5), observe that
\[ p_* \Omega^1_Z (\log D_1)(-E) \subset p_* \Omega^1_Z (\log D_1) \subset \Omega^1_Z (d\log l_1). \]

Take coordinates \( x, y \) such that \( l_1 = y \), and write \( \omega = \alpha dx + \beta dy \).

We just pull back \( \omega \) on the blow up \( Z \) in the chart where we have \( y = tx \), and impose that it lies in the span of
\[ \frac{dt}{t}, x dx. \]

We have
\[ \omega = \alpha(x, tx) dx + \beta(x, tx) \left( \frac{dt}{t} + \frac{dx}{x} \right) \]
and we must clearly have \( \beta(0) = 0 \).

Then \( \beta(x, tx) \frac{dt}{t} \) is a multiple of \( x \frac{dt}{t} \), and it suffices to require that \( \alpha(x, tx) + \frac{1}{t} \beta(x, tx) \) be divisible by \( x \).

Writing \( \beta(x, y) = \beta_1 x + \beta_2 y + \ldots \), our condition boils down to the divisibility by \( x \) of
\[ \alpha(0) + \beta_1 + \beta_2 t \iff \beta_2 = 0, \quad \alpha(0) + \beta_1 = 0. \]
Finally, let us show 7). Write
\[ \omega = \alpha \frac{dx}{x} + \beta \frac{dy}{y} \]
and pull back to the blow up in the chart where \( y = tx \): we get
\[ (\alpha + \beta) \frac{dx}{x} + \beta \frac{dt}{t}, \]
which must be divisible by \( x \), hence in particular \( \beta(0) = 0 \). Looking at the other chart we get symmetrically \( \alpha(0) = 0 \).
Now, \( \alpha + \beta \) must vanish of order two, in order that its pull back be divisible by \( x^2 \).

\[ \square \]

8. Appendix 2: Proof of Proposition 4.1

Proof of Proposition 4.1. We can prove 1) and 2) simultaneously for \( i = 1 \).
Observe that \( D_1 = \Delta_1 + N_1 \), that \( \Lambda_1 = L_1 + N_1 \), and apply Lemma 4.3 (observing that \( (K + 2N_1 + (E_1 - E_3))N_1 = -4 + 1 < 0 \)) in order to conclude that
\[ H^0(\Omega_X^1(\log(\Delta_1))(E_1 - E_3 + N_1)) \cong H^0(\Omega_Y^1(\log(D_1))(E_1 - E_3)). \]
Moreover we observe that, again by Lemma 4.3
\[ H^0(\Omega_Y^1(\log(D_1))(E_1 - E_3)) = H^0(\Omega_Y^1(\log(D_1 - (L - E_1)))(L - E_3)). \]

Let \( f: \tilde{Y} \to \mathbb{P}^2 \) be the blow down of \( E_1, \ldots, E_5 \). Then \( f_* (D_1 - (L - E_1)) \) splits as the sum of two lines \( l_1, l_2 \) in \( \mathbb{P}^2 \) intersecting in \( P_1 \).
W.l.o.g. we can assume that \( P_1 = (0 : 0 : 1), P_2 = (0 : 1 : 0), P_4 = (1 : 0 : 0) \) and \( P_5 = (1 : 0 : \lambda) \), with \( \lambda \neq 0 \).
Applying Proposition 4.1 several times for each blow down we get that
\[ H^0(\Omega_Y^1(\log(D_1 - (L - E_1)))(L - E_3)) = H^0(f_* \Omega_Y^1(\log(D_1 - (L - E_1)))(L - E_3)). \]
is the subspace \( V_1 \) of \( H^0(\Omega_{\mathbb{P}^2}^1(d\log l_1, d\log l_2)(1)) \) consisting of sections satisfying several linear conditions.
We write these conditions using the basis provided by Lemma 4.5 and its corollary, in order to show that \( V_1 = 0 \). By prop. 7.1, 3) we get for \( P_1 \):
\[ a_{13} + a_{23} = 0; \]
for \( P_2, P_4 \) and \( P_5 \) the three equations
\[ a_{12} = a_{21} = a_{21} + \lambda a_{23} = 0. \]
This shows that \( V_1 = 0 \).
We continue with the proof of 1).
For \( i = 2 \), again by Lemma 4.3 we have to calculate
\[ V_2 = H^0(\Omega_Y^1(\log(L - E_2 - E_3), log(L - E_2 - E_4))(L - E_3)), \]
which after blowing down \( E_1, \ldots, E_5 \) corresponds to a subspace of \( H^0(\Omega_{\mathbb{P}^2}^1(d\log l_1, d\log l_2)(1)) \).

\[ \square \]
I. Bauer and F. Catanese

W.l.o.g. we can assume that $P_2 = (0 : 0 : 1)$, $P_5 = (0 : 1 : 0)$, $P_4 = (1 : 0 : 0)$ and $P_3 = (1 : 1 : 1)$.

By prop. 7.1, 3), we get for $P_2, P_4, P_5$ the three linear equations:

\[ a_{13} + a_{23} = 0, \quad a_{21} = 0, \quad a_{12} = 0. \]

We evaluate $\omega$ in $P_3$, and get (using the above equalities)

\[ \omega(P_3) = a_{13}dx_1 + a_{23}dx_2, \]

whence by Proposition 7.1, 6) $a_{13} = a_{23} = 0$ and therefore we have verified that $V_2 = 0$.

For $i = 3$, using Lemma 4.3 we have to calculate:

\[ V_3 := H^0(\Omega^1_{Y}(\log(L - E_3 - E_4), \log(L - E_3 - E_5))(L - E_1)), \]

which, after blowing down $E_1, \ldots, E_5$, becomes a linear subspace of $H^0(\Omega^1_{\mathbb{P}^2}(\log(l_1, \log l_2)(1))$.

W.l.o.g. we can assume that $P_5 = (0 : 0 : 1)$, $P_4 = (0 : 1 : 0)$, $P_5 = (1 : 0 : 0)$, $P_3 = (1 : 1 : 0)$.

By prop. 7.1, 3), we get for $P_3, P_4, P_5$ the three linear equations:

\[ a_{13} + a_{23} = 0, \quad a_{12} = 0, \quad a_{21} = 0. \]

Setting the evaluation of $\omega$ in $P_3$ equal to zero is easily seen to give no new conditions, hence $V_3 \cong \mathbb{C}$.

Let’s proceed to prove 2) for $i = 2, 3$.

For $i = 2, 3$, by 4) of remark 1.5

\[ H^0(\Omega^1_{Y}(\log \Delta_i)(K_Y + \Lambda_i)) = H^0(\Omega^1_{Y}(\log \Delta_i)(E_i - E_{i+2})). \]

For $i = 2$, using again Lemma 4.3 observing that

\[ (K_Y + 2\Gamma_2 + (E_2 - E_1))\Gamma_2 < 0, \]

we see that we have to calculate

\[ V_2 := H^0(\Omega^1_{Y}(\log(L - E_2 - E_4), \log(L - E_2 - E_5))(2L - E_1 - E_3 - E_4 - E_5)), \]

which, after blowing down $E_1, \ldots, E_5$, becomes a linear subspace of $H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2)(2))$.

W.l.o.g. we can assume that $P_2 = (0 : 0 : 1)$, $P_4 = (0 : 1 : 0)$, $P_5 = (1 : 0 : 0)$, $P_3 = (1 : 1 : 1)$, and then $P_1 = (1 : \lambda : 0)$, where $\lambda \neq 0, 1$.

Using cor. 4.6 we get by prop. 7.1, 3) for $P_2$ the linear equation

\[ a_{213} + a_{323} = 0. \]

By prop. 7.1, 5) the conditions for $P_4$ are

\[ a_{212} = 0, \quad a_{12} = 0, \quad a_{23} = 0; \]

whereas the conditions for $P_5$ are

\[ a_{121} = 0, \quad a_{12} = 0, \quad a_{13} = 0. \]

Imposing that $\omega$ vanishes in $P_3$, we get

\[ \omega(P_3) = dx_1(a_{313} + 2a_{213} + a_{123}) + dx_2(a_{323} + 2a_{123} + a_{213}) = 0. \]
The above conditions yield:

\[ a_{123} = a_{313} = -a_{213} = -a_{323}. \]

Finally, imposing that \( \omega \) vanishes in \( P_1 \) we obtain:

\[ \omega(P_1) = -dx_3(\lambda a_{213} + a_{123}) = 0, \]

whence \( (\lambda - 1)a_{213} = 0 \). Since \( \lambda \neq 0, 1 \) this implies \( a_{213} = 0 \), and we have shown that \( V_2 = 0 \).

We are left with the case \( i = 3 \). Using repeatedly Lemma 4.3 and Proposition 7.1, we see that we have to calculate

\[ V_3 := H^0(\Omega^1_Y(\log \Delta_3)(E_3 - E_2)) = H^0(\Omega^1_Y(\log(L - E_1 - E_3), \log(L - E_1 - E_4 - E_5))(2L - E_3 - E_4 - E_5)). \]

After blowing down \( E_1, \ldots, E_5 \), we can assume w.l.o.g. that \( P_1 = (0 : 0 : 1) \), \( P_4 = (0 : 1 : 0) \), \( P_3 = (1 : 0 : 0) \), \( P_5 = (0 : 1 : 1) \), and \( V_3 \) becomes a linear subspace of \( H^0(\Omega^1_2(d\log l_1, d\log l_2)(2)) \).

Using cor. 4.6 we get by prop. 7.1, 3) for \( P_1 \) the linear equation

\[ a_{313} + a_{323} = 0. \]

By prop. 7.1, 5) the conditions for \( P_4 \) are

\[ a_{212} = 0, \quad a_{12} = 0, \quad a_{23} = 0; \]

whereas the conditions for \( P_3 \) are

\[ a_{121} = 0, \quad a_{12} = 0, \quad a_{13} = 0. \]

For \( P_5 \), we get instead (again by prop. 7.1, 5)), the two linear equations (the third is trivial):

\[ a_{213} + a_{313} = 0, \quad 2a_{313} = 0. \]

This implies that \( a_{313} = a_{213} = a_{323} = 0 \), but \( a_{123} \) is arbitrary. This shows that \( V_3 \cong C \).

Thus 2) is proven.

To prove 3), by symmetry, we may assume without loss of generality that \( i = 1 \).

We have to calculate \( V_1 := H^0(\Omega^1_Y(\log(D_1)(E_1 - E_3)) \), which by Lemma 4.3 is equal to

\[ H^0(\Omega^1_Y(\log(L - E_1 - E_2), \log(L - E_1 - E_4 - E_5))(L - E_6)), \]

which, after blowing down \( E_1, \ldots, E_5 \), becomes a linear subspace of \( H^0(\Omega^1_2(d\log l_1, d\log l_2)(1)) \).

W.l.o.g. we can assume that \( P_1 = (0 : 0 : 1) \), \( P_5 = (0 : 1 : 0) \), \( P_2 = (1 : 0 : 0) \), \( P_4 = (0 : 1 : 1) \). Since \( P_2, P_4, P_6 \) are collinear, \( P_5 = (1 : \mu : \mu) \), where \( \mu \neq 0 \).

By prop. 7.1, 3), we get for \( P_1, P_2, P_4 \) and \( P_5 \) the linear equations:

\[ a_{13} + a_{23} = 0, \quad a_{21} = 0, \quad a_{12} + a_{13} = 0, \quad a_{12} = 0. \]

This already shows that \( V_1 = 0 \).

Thus 3) is proven.
Let us treat the subcase of 4) where we have strictly extended Burniat divisors:
the situation is here symmetric in the indices $i$, hence it suffices to show the vanishing of

$$H^0(\Omega^1_Y(\log(\Delta_1))(K_Y + \Lambda_1)).$$

Recall that we have the decomposition in irreducible connected components

$$\Delta_1 = G_1 + \Gamma_1 + N_2,$$

where $G_1$ is the del Pezzo line $G_1 \equiv L - E_1 - E_6$.

By Lemma 4.3 we get:

$$H^0(\Omega^1_Y(\log(\Delta_1))(K_Y + \Lambda_1)) = H^0(\Omega^1_Y(\log(\Delta_1 + N_1))(E_1 - E_3)),$$

since $(K_Y + 2N_1 + (E_1 - E_3))N_1 < 0$. Using again Lemma 4.3 we see that

$$H^0(\Omega^1_Y(\log(\Delta_1 + N_1))(E_1 - E_3)) =$$

$$= H^0(\Omega^1_Y(\log(\Delta_1 + N_1 - \Gamma_1))((E_1 - E_3) + \Gamma_1)) =$$

$$= H^0(\Omega^1_Y(\log(G_1 + N_1 + N_2))(2L - E_2 - E_3 - E_5 - E_6)),$$

because $(K_Y + 2\Gamma_1 + (E_1 - E_3))\Gamma_1 < 0$.

Let $f: \tilde{Y} \to \mathbb{P}^2$ be the blow down of $E_1, \ldots, E_6$. Then $f_*(G_1 + N_1 + N_2)$ splits as the sum of three lines $l_1, l_2, l_3$ in $\mathbb{P}^2$ forming a triangle. W.l.o.g. we can assume that $P_0 = (1 : 0 : 0)$, $P_1 = (0 : 1 : 0)$, $P_4 = (0 : 0 : 1)$ and $P_3 = (1 : 1 : 1)$. Then $P_5 = (0 : 1 : 1)$, whereas $P_2$ is collinear with $P_1$, $P_4$, whence $P_2 = (1 : 0 : \lambda)$, with $\lambda \neq 0, 1$.

Then

$$H^0(\Omega^1_Y(\log(G_1 + N_1 + N_2))(2L - E_2 - E_3 - E_5 - E_6)) =$$

$$H^0(f_*(\Omega^1_Y(\log(G_1 + N_1 + N_2))(2L - E_2 - E_3 - E_5 - E_6))$$

is a subspace of

$$H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2, \log l_3)(2)),$$

where $l_i = x_i$, whence $P_1, P_3, P_5 \in \{l_1 = 0\}$, $P_0, P_4, P_2 \in \{l_2 = 0\}$, $P_1, P_6 \in \{l_3 = 0\}$, consisting of sections satisfying fourteen linear conditions described in Proposition 7.1.

We explicitly write these conditions using Lemma 4.3 and Lemma 4.6 in order to show that this subspace must be trivial.

Let $\omega \in H^0(\Omega^1_{\mathbb{P}^2}(\log l_1, \log l_2, \log l_3)(2))$ and we write $\omega$ in the basis of Lemma 4.3

$$\omega = a_{12}\eta_{12} + a_{13}\eta_{13} + a_{23}\eta_{23} + a_{212}x_2\omega_{12} + a_{313}x_3\omega_{13} + a_{323}x_3\omega_{23} +$$

$$+ a_{121}x_1\omega_{21} + a_{131}x_1\omega_{31} + a_{232}x_2\omega_{32} + a_{123}x_1\omega_{23} + a_{231}x_2\omega_{31} + a_{312}x_3\omega_{12}.$$

Then by prop. 7.1 3) the condition for $P_1 = (0 : 1 : 0)$ is

(1) $$a_{212} + a_{232} = 0.$$ The same argument shows that the linear condition for $P_3 = (0 : 0 : 1)$ is

(2) $$a_{313} + a_{323} = 0.$$
Next we work out the conditions for $P_5, P_2$ using prop. 7.1.5. For $P_5 := (0 : 1 : 1)$ we work in the chart $x_3 = 1$ and write $\omega$ locally around $(0,1)$ as $\alpha(x_1, x_2)dx_2 + \beta(x_1, x_2)\frac{dx}{x_1}$. Then we get (using Lemma 4.6):

\begin{align*}
\beta(0,1) &= a_{212} + a_{313} + a_{312} = 0; \\
\frac{\partial \beta}{\partial x_1}(0,1) &= -a_{12} - a_{13} - a_{231} = 0; \\
\frac{\partial \beta}{\partial x_2}(0,1) + \alpha(0,1) &= -a_{23} + 2a_{212} + a_{323} = 0.
\end{align*}

The same argument for $P_2 = (1 : 0 : \lambda)$ (working in the chart $x_1 = 1$ and writing $\omega$ locally around $(0,\lambda)$ as $\alpha(x_2, x_3)dx_3 + \beta(x_2, x_3)\frac{dx}{x_2}$) gives the following three linear equations ($\lambda \neq 0, 1$):

\begin{align*}
\beta(0, \lambda) &= a_{323}\lambda^2 + a_{121} + a_{123}\lambda = 0; \\
\frac{\partial \beta}{\partial x_2}(0, \lambda) &= a_{12} - \lambda a_{23} - \lambda a_{312} = 0; \\
\frac{\partial \beta}{\partial x_3}(0, \lambda) + \alpha(0, \lambda) &= a_{13} + \frac{1}{\lambda}a_{131} + \lambda a_{323} - \lambda a_{313} = a_{13} + \frac{1}{\lambda}a_{131} + 2\lambda a_{323} = 0.
\end{align*}

There are four linear conditions coming from $P_6 = (1 : 0 : 0)$, given in prop. 7.1.7. We work in the chart $x_1 = 1$ and write $\omega = \alpha(x_2, x_3)dx_3 + \beta(x_2, x_3)\frac{dx}{x_2}$. Then we get:

\begin{align*}
\alpha(0,0) &= a_{121} = 0; \\
\beta(0,0) &= a_{131} = 0; \\
\frac{\partial (\alpha + \beta)}{\partial x_2}(0,0) &= a_{12} + a_{231} = 0; \\
\frac{\partial (\alpha + \beta)}{\partial x_3}(0,0) &= a_{13} = 0.
\end{align*}

From equation (11) we get: $a_{12} = -a_{231}$.

Since $a_{13} = a_{131} = 0$, equation (8) implies $a_{323} = 0$, whence by (2) also $a_{113} = 0$. Moreover, by (6), we get $a_{123} = 0$.

We write finally the conditions coming from $P_3 = (1 : 1 : 1)$ (using again that certain coefficients are zero).

We evaluate $\omega$ in $P_3$ and work in the affine chart $x_2 = 1$ to obtain

\begin{align*}
\omega(P_3) &= (-a_{12} + a_{212} - a_{231} + a_{312})dx_1 + (a_{23} + a_{232} + a_{231})dx_3 = 0.
\end{align*}
I. Bauer and F. Catanese

Since:
\[ \omega(P_3) = (a_{212} + a_{312})dx_1 + ((a_{23} + a_{232} + a_{231})dx_3 \]

we get the last two linear equations:
(14) \[ a_{231} + a_{23} + a_{232} = 0; \]
(15) \[ a_{212} + a_{312} = 0. \]

These immediately imply that
\[ a_{312} = a_{232} = -\frac{a_{231}}{2}. \]

By (14) \[ a_{23} = -a_{232} - a_{231}, \] and using (5), we see that \[ a_{23} = 2a_{212}. \]

Again by (14) we get then that \[ a_{212} = -a_{231}. \] Therefore, we have:
\[ a_{212} = -a_{231} = a_{232} = -\frac{a_{231}}{2}. \]

By (7):
\[ 0 = a_{12} - \lambda a_{23} - \lambda a_{231} = a_{12} + \lambda a_{231}, \]
whence by (4) \( \lambda = 1 \), which gives a contradiction, or \( a_{12} = a_{231} = 0. \)

Hence the claim for strictly extended Burniat surfaces with \( K_S^2 = 3 \) is established.

Next we come to the case of (non strictly) extended Burniat surfaces. Here we have to consider two cases:

a) only one of the three conics \( \Gamma_i \) degenerates to two lines;
b) exactly two of the three conics \( \Gamma_i \) degenerate to two lines.

a) W.l.o.g. and by remark\(^\text{1.5, (5)}\) we may assume that \( \Gamma_1 \) splits as
\[ \Gamma_1 \equiv (L - E_1 - E_2) + N_3 + E_3. \]

Then we get the extended Burniat divisors:
\[ \{D'_1\} = |L - E_1 - E_6| + |L - E_1 - E_2| + E_3 + N_2, \]
\[ D'_2 \in |L - E_2 - E_5| + |2L - E_2 - E_3 - E_4 - E_6|, \]
\[ D'_3 \in |L - E_3 - E_4| + |2L - E_1 - E_3 - E_4 - E_6| + N_1 + N_3. \]

We make the assumption, for each \( D'_i, i = 2, 3 \) that the strict transform of the conic \( \Gamma_i \) is irreducible.

Then we have
\[ K_{\tilde{Y}} + L'_1 \equiv L - E_3 - E_4 - E_5 \equiv K_{\tilde{Y}} + \Lambda_1, \]
\[ K_{\tilde{Y}} + L'_2 \equiv L - E_1 - E_4 - E_6 \equiv K_{\tilde{Y}} + \Lambda_2, \]
\[ K_{\tilde{Y}} + L'_3 \equiv E_3 - E_2 \equiv K_{\tilde{Y}} + \Lambda_3 - N_3, \]

where the \( \Lambda_i \) are as for the strictly extended Burniat divisors.

Observe that \( D'_2 + N_3 = \Delta_2 \), whence
\[ H^0(\Omega_{\tilde{Y}}^1 \log D'_2)(K_{\tilde{Y}} + L'_2) \subset H^0(\Omega_{\tilde{Y}}^1 \log D'_2)(K_{\tilde{Y}} + L'_2 + N_3) = \]
\[ H^0(\Omega_{\tilde{Y}}^1 \log(D'_2 + N_3))(K_{\tilde{Y}} + L'_2) = H^0(\Omega_{\tilde{Y}}^1 \log(\Delta_2)(K_{\tilde{Y}} + \Lambda_2)) = 0, \]
where the first equality holds by Lemma 4.3 and the last holds by our previous computations for strictly extended Burniat surfaces.

Moreover, \( D'_3 = \Delta_3 + N_3 \), \( L'_3 = \Lambda_3 - N_3 \), whence the vanishing of \( H^0(\Omega^1_Y (\log D'_3)(K_Y + L'_3)) \) follows again using Lemma 4.3 from the analogous vanishing for strictly extended Burniat surfaces. It remains to prove the following

**Claim 8.1.** \( H^0(\Omega^1_Y (\log D'_1)(K_Y + L'_1)) = 0 \)

**Proof of the claim.** By Lemma 4.3 we see that

\[
H^0(\Omega^1_Y (\log D'_1)(K_Y + L'_1)) = H^0(\Omega^1_Y (\log(D'_1 - E_3))(L - E_4 - E_5)).
\]

Let \( f: \tilde{Y} \to \mathbb{P}^2 \) be the blow down of \( E_1, \ldots, E_6 \). Then \( f_*(D'_1 - E_3) \) splits as the sum of three lines \( l_1, l_2, l_3 \) in \( \mathbb{P}^2 \) forming a triangle. W.l.o.g. we can assume that \( P_1 = (1:0:0) \), \( P_2 = (0:1:0) \), \( P_3 = (0:0:1) \) and \( P_4 = (1:1:1) \). Then \( P_4 = (0:1:1) \).

We conclude that

\[
H^0(\Omega^1_Y (\log(D'_1 - E_3))(L - E_4 - E_5)) = H^0(f_*(\Omega^1_Y (\log(D'_1 - E_3))(L - E_4 - E_5))).
\]

We write these conditions using Lemma 4.3 in order to show that this subspace must be trivial.

By Lemma 4.3 we write \( \omega \in H^0(\Omega^1_{\mathbb{P}^2} (\log l_1, \log l_2, \log l_3)(1)) \) as

\[
\omega = \sum_{i \neq j} a_{ij} \omega_{ij}.
\]

Then the three equations for \( P_1, P_2, P_3 \) (cf. prop. 7.1, 3)) are

\[
a_{21} + a_{31} = 0, \quad a_{12} + a_{32} = 0, \quad a_{13} + a_{23} = 0.
\]

By prop. 7.1, 5), we get for \( P_4 \) the linear equations:

\[
a_{12} + a_{13} = 0, \quad -a_{21} - a_{31} = 0, \quad a_{23} - a_{32} = 0.
\]

The above conditions already imply:

\[
a_{13} = a_{21} = a_{23} = a_{32} = 0, \quad a_{21} = -a_{31}.
\]

We impose the vanishing of \( \omega \) in \( P_3 = (1:1:1) \) working in the affine chart \( x_3 = 1 \) and obtain

\[
\omega(1:1:1) = (-a_{21} - a_{31})dx_1 + (a_{21})dx_2 = 0,
\]

whence \( a_{21} = a_{31} = 0 \). \( \square \)

b) W.l.o.g. we can assume that each of the two conics \( \Gamma_1 \) and \( \Gamma_2 \) degenerate to two lines. Then we get the extended Burniat divisors:

\[
\begin{align*}
\{D'_1\} &= |L - E_1 - E_6| + |L - E_1 - E_2| + E_3 + N_1 + N_2, \\
\{D'_2\} &= |L - E_2 - E_5| + |L - E_2 - E_3| + E_4, \\
D'_3 &\in |L - E_3 - E_4| + |2L - E_1 - E_3 - E_4 - E_6| + N_3.
\end{align*}
\]
We make the assumption for $D''_a$ that the strict transform of the conic $\Gamma_3$ passing through $P_1, P_3, P_4, P_5$ is irreducible. Then we have
\begin{align*}
K_{\tilde{Y}} + \mathcal{L}'_1 &\equiv E_1 - E_3 = K_{\tilde{Y}} + \mathcal{L}'_1 - N_1, \ D''_1 = D'_1 + N_1; \\
K_{\tilde{Y}} + \mathcal{L}'_2 &\equiv L - E_1 - E_4 - E_5 \equiv K_{\tilde{Y}} + \mathcal{L}_2 + N_2, \ D''_2 = D_2 - N_2; \\
K_{\tilde{Y}} + \mathcal{L}'_3 &\equiv E_3 - E_2 \equiv K_{\tilde{Y}} + N_3', \ D''_3 = D_3' - N_3.
\end{align*}
Therefore for $i = 1, 2, 3$ the vanishing of $H^0(\Omega_1^1(K_{\tilde{Y}} + \mathcal{L}'_i))$ can be reduced via Lemma 4.3 to the analogous vanishing for extended Burniat surfaces of case a) ($i = 1, 3$) and to the analogous vanishing for Burniat divisors (for $i = 2$), which was already proved in part 3).

\[\square\]

9. Appendix 3: An alternative proof of statements 1), 2), 3) of Proposition 7.1

In this appendix we present other methods to calculate the space of sections of twisted logarithmic sheaves, in particular a fibration method. Assume that we have $d$ smooth rational curves $C_\alpha \subset Y$ contained in a smooth algebraic surface $Y$, meeting with distinct tangents in a point $O$, a divisor $B_\alpha$ on $C_\alpha$ of degree 0, 1 or 2, and disjoint from $O$, and let $Z$ be the blow up of $Y$ in the point $O$. Denote by $D_\alpha$ the strict transform of $C_\alpha$, and denote by $\Omega_Z^1((\log C_\alpha(-B_\alpha))_{\alpha \in A})$ the sheaf which is the inverse image, under the residue sequence, of $\oplus_{\alpha \in A} \mathcal{O}_{C_\alpha}(-B_\alpha)$. Then by 4) of Proposition 7.1 we have an exact sequence
\[0 \to \Omega_Y^1((\log C_\alpha(-B_\alpha))_{\alpha \in A}) \to \to p_* \Omega_Z^1((\log D_\alpha(-B_\alpha))_{\alpha \in A})(E) \to \mathbb{C}^{d^2-2} \to 0\]
which is exact on global sections if
\[h := \text{dim}_\mathbb{C} H^1(\Omega_Y^1((\log C_\alpha(-B_\alpha))_{\alpha \in A})) = 0.\]
Or, more generally, if $h = h'$, where
\[h' := \text{dim}_\mathbb{C} H^1(\Omega_Z^1((\log D_\alpha(-B_\alpha))_{\alpha \in A})).\]
Consider the exact sequence
\[0 \to \Omega_Y^1 \to \Omega_Y^1((\log C_\alpha(-B_\alpha))_{\alpha \in A}) \to \bigoplus_{\alpha = 1}^d \mathcal{O}_{C_\alpha}(-B_\alpha) \to 0,\]
and assume that $H^2(\Omega_Y^1) = 0$.
Then $h = a + b$, where $a$ is the number of $\alpha$’s such that $B_\alpha$ has degree 2, while $b$ is the difference of the dimensions between $H^1(\Omega_Y^1)$ and the subspace generated by the Chern classes of the $C_\alpha$’s such that $B_\alpha$ has degree 0. If we choose $Y = \mathbb{P}^2$ then $h = 0$ as soon as no $B_\alpha$ has degree 2, and some $B_\alpha$ has degree 0.
Otherwise, one can calculate $h'$ in a similar way. We assume for simplicity that $Y = \mathbb{P}^2$. We have a similar exact sequence

$$0 \to \Omega^1_Y(E) \to \Omega^1_Z(\log D_a(-B_a)) \to \bigoplus_{\alpha = 1}^d \mathcal{O}_{C_a}(O - B_a) \to 0,$$

and since $H^1(\mathcal{O}_{C_a}(O - B_a)) = 0$ by our assumption, we get that $h'$ is the dimension of the cokernel of

$$\bigoplus_{\alpha = 1}^d H^0(\mathcal{O}_{C_a}(O - B_a)) \to H^1(\Omega^1_Z(E)).$$

To calculate the last space, observe that

$$\Omega^1_Z \otimes \mathcal{O}_E = \mathcal{O}_E(-2) \oplus \mathcal{O}_E(1)$$

whence $h^1(\Omega^1_Z(E)) = h^1(\Omega^1_Z) + 1 = 3$.

These criteria can now be used in order to prove statements 1), 2), 3) of Proposition 4.1.

We can prove 1) and 2) simultaneously for $i = 1$.

Observe that $D_1 = \Delta_1 + N_1$, that $\Delta_1 = L_1 + N_1$, and apply Lemma 4.3 in order to conclude that

$$H^0(\Omega^1_Y(\log(\Delta_1))(E_1 - E_3 + N_1) = H^0(\Omega^1_Y(\log(D_1))(E_1 - E_3)).$$

By Lemma 2.4 we can blow down $E_3$ and obtain $H^0(\Omega^1_{\mathbb{P}^1}(\log(D'_1))(E_1))$. In this case the respective degrees of the divisors $B_a$ are 0, 1, 2 hence $h = 1$. We have to decide whether $h'$ is 0 or 1. We contract $E_3, E_4, E_5$, and we let $Z$ be the blow up of the plane in $P_1$. We must calculate $h^0(\Omega^1_Z(\log(C_a(-B_a)))(E_1))$. Here the curves $C_a$ are fibres of the ruling of $Z$, $f: Z \to \mathbb{P}^1$. Using the exact sequence

$$0 \to f^*\Omega^1_{\mathbb{P}^1} \to \Omega^1_Z \to \omega_{Z|\mathbb{P}^1} = \mathcal{O}_Z(-F - 2E_1) \to 0$$

we obtain the analogous sequence

$$0 \to f^*\mathcal{O}_{\mathbb{P}^1}(1)(E_1) \to \Omega^1_Z(\log(C_a))(E_1) \to \mathcal{O}_Z(-F - E_1) \to 0$$

to infer that

$$H^0(f^*\mathcal{O}_{\mathbb{P}^1}(1)(E_1)) = H^0(\Omega^1_Z(\log(C_a))(E_1)).$$

We are imposing some vanishing on three points lying in two fibres, hence we get the sections of $H^0(\mathcal{O}_Z(F + E_1)) = p^*H^0(\mathcal{O}_{\mathbb{P}^2}(1))$ vanishing in the three points $P_4, P_5, P_2$, whence we conclude that this space has dimension $= 0$.

This argument shows 1) also for $i = 2, 3$.

For a nodal Burniat with $m = 2$ the space $H^0(\Omega^1_Y(\log(D_i))(E_i - E_{i+2}))$ vanishes for $i = 2$, but it has dimension equal to 1 for $i = 3$, since then the three points $P_4, P_5, P_1$ are collinear.

Let’s proceed with 2).

For $i = 2, 3$

$$H^0(\Omega^1_Y(\log(\Delta_i))(K_Y + \Lambda_i)) = H^0(\Omega^1_Y(\log(\Delta_i))(E_i - E_{i+2})).$$
By applying again Lemma 7.3 for $i = 3$ we can blow down the curve $E_2$ and the curve $E_3$ and apply the residue sequence to the sheaf $\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)$. Since each component is smooth and rational, we find that $H^0(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)) = \ker(C^i \to H^1(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)))$, while $H^1(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)) = \text{Coker}(C^i \to H^1(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)))$.

The map is given by the Chern classes of $L - E_1, L - E_4, L - E_5, L - E_2 - E_4 - E_5$. These generate a rank 4 subspace of the 4-dimensional space $H^1(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i))$ (we are blowing up 3 points in the plane), whence $h^0(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)) = 0, h^1(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)) = 0$.

We conclude, by the exact cohomology sequence associated to

$$0 \to \Omega^n_{\mathbb{P}^i}(\log \Delta_4^i) \to f_*(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)(K + \Lambda_3)) \to \mathcal{C}_{P_3} \to 0,$$

that $h^0(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)(K + \Lambda_3)) = 1 + h^0(\Omega^n_{\mathbb{P}^i}(\log \Delta_4^i)) = 1$.

For the case $i = 2$ recall that $\Delta_2 \in [L - E_2 - E_4] + [L - E_2 - E_5] + [2L - E_2 - E_3 - E_4 - E_5]$ consists of three smooth connected components. Blow down $E_1, E_3, E_4, E_5$ and obtain the ruled surface $Z$ equal to the blow up of the plane in $P_2$. Denote by $f : Z \to \mathbb{P}^1$ the standard fibration. The direct image $\Delta'_2 := f_*\Delta_2$ decomposes as the union of two fibres $F_4$ and $F_5$ and a section $C$ with $C \cdot E_2 = 1$.

We have to calculate the space of global sections of

$$\mathcal{F} := \mathfrak{m}_{P_3} \Omega^n_{\mathbb{P}^2}(\log F_4, \log F_5, \log C(-P_3))(E_2)$$

satisfying two linear conditions imposed by the points $P_4, P_5$.

Using the exact sequence (***) we get the exact sequence

$$0 \to \mathfrak{m}_{P_3} \mathcal{O}_Z(E_2) \to \mathcal{F} \to \mathfrak{m}_{P_3} \mathfrak{m}_{P_3} \mathcal{O}_Z(-F - E_2 + C) \to 0.$$

Observe that $\mathcal{O}_Z(-F - E_2 + C)$ has degree 0 on each fibre and degree 1 on $E_2$.

If $D \equiv -F - E_2 + C \equiv L - E_2$ is effective, then $D$ is a fibre. Since no fibre contains both $P_1, P_3$, we obtain

$$H^0(\mathfrak{m}_{P_3} \mathfrak{m}_{P_3} \mathcal{O}_Z(-F - E_2 + C)) = 0.$$

Since $|E_2|$ consists of the curve $E_2$, which does not contain $P_1$, we conclude that $H^0(M_{P_1} \mathcal{O}_Z(E_2)) = H^0(\mathcal{F}) = 0$.

To prove 3), by symmetry, we may assume without loss of generality that $i = 1$.

Blow down all the curves $E_j$ except $E_1$, so that, as usual, we have the blow up $Z$ of the plane in a point ($P_1$) and the standard fibration $f : Z \to \mathbb{P}^1$.

By Lemma 7.1 and since $E_3$ is a connected component of $D_1$, the direct image $\mathcal{F}$ of $\Omega^n_{\mathbb{P}^2}(\log(D_1)(E_1 - E_3))$ is contained in $\Omega^n_{\mathbb{P}^3}(\log(F_2 + F_5 + F_{4,5}))(E_1)$ where $F_j$ denotes the unique fibre of $f$ passing through the point $P_j$.

More precisely, we have an exact sequence

$$0 \to \mathfrak{m}_{P_2} \mathfrak{m}_{P_1} \mathfrak{m}_{P_3} \mathfrak{m}_{P_3} \mathcal{O}_Z(F + E_1) \to \mathcal{F} \to \mathcal{O}_Z(-F - E_1) \to 0.$$

Clearly $H^0(\mathcal{O}_Z(-F - E_1)) = 0$ since $F \cdot (F + E_1) = 1$. On the other hand $H^0(\mathcal{O}_Z(F + E_1)) = H^0(\mathcal{O}_{\mathbb{P}^2}(1))$, hence the fact that the points $P_2, P_4, P_5, P_6$...
are not collinear implies the desired vanishing

$$H^0(m_{P_1}m_{P_2}m_{P_3}m_{P_4}O_Z(F + E_1)) = 0.$$  

Thus 3) is proven.

References


Documenta Mathematica 18 (2013) 1089–1136
Moduli, in honour of David Mumford, published by International Press.


[Ch12] Chen, Y., Two irreducible components of the moduli space $M_{1,3}^{can}$. Ph.D. Thesis (Bayreuth), 2012.


Random Walks in Compact Groups

Péter Pál Varjú

Received: December 17, 2012
Communicated by Ursula Hamenstädt

Abstract. Let $X_1, X_2, \ldots$ be independent identically distributed random elements of a compact group $G$. We discuss the speed of convergence of the law of the product $X_1 \cdots X_1$ to the Haar measure. We give poly-log estimates for certain finite groups and for compact semi-simple Lie groups. We improve earlier results of Solovay, Kitaev, Gamburd, Shahshahani and Dinai.

2010 Mathematics Subject Classification: 60B15, 22E30, 05E15
Keywords and Phrases: random walk, spectral gap, diameter, poly-log, Solovay-Kitaev, compact group, Cayley graph

1 Introduction

Let $G$ be a group and $S \subset G$ a finite set. We study the distribution of the product of $l$ random elements of $S$. In particular, we are interested in how fast this distribution becomes uniform as $l$ grows.

We discuss the problem in two different but very related settings: profinite groups, and compact Lie groups.

1.1 Two ways to measure uniformity

We begin by describing the details in the first setting. Let $G$ be a finite group and $S \subset G$ be a finite generating set. For simplicity, we assume that $1 \in S$.

The unit element of any group is denoted by $1$ in this paper. Write

$$S^l = \{g_1 \cdots g_l : g_1, \ldots, g_l \in S\}$$

for the $l$-fold product set of $S$.

The diameter of $G$ with respect to $S$ is defined by

$$\text{diam}(G, S) = \min\{l : S^l = G\}.$$
The diameter is the minimal length of a product that can express any element of the group. Hence it is a (very weak) quantity to measure uniformity. We quantify uniformity in a stronger sense, too. To this end, we introduce the notion of random walks. Denote by $\mu_S$ the normalized counting measure on the set $S$, and let $X_1, X_2, \ldots$ be a sequence of independent random elements of $S$ with law $\mu_S$. The (simple) random walk is the sequence of random elements $Y_0, Y_1, \ldots$ of $G$ such that $Y_0 = 1$ almost surely, and $Y_{t+1} = X_{t+1}Y_t$ for all $t \geq 0$.

Denote by $\mu^{*}_S(l)$ the $l$-fold convolution of the measure $\mu_S$ with itself. Convolution of two measures (or functions) $\mu, \nu$ is defined by the usual formula

$$\mu * \nu(g) = \sum_{h \in G} \mu(gh^{-1})\nu(h).$$

Observe that $\mu^{*}_S(l)$ is the law of $Y_l$.

We want to understand, how large $l$ is needed to be taken such that $\mu^{*}_S(l)$ is “very close” to the uniform distribution. We make this precise with the following construction. Consider the space $L^2(G)$ which is simply the vector space of complex valued functions on $G$ endowed with the standard scalar product. The group $G$ acts on this by

$$\text{Reg}_G(g)f(h) = f(g^{-1}h) \quad \text{for} \quad f \in L^2(G).$$

This is a unitary representation called the regular representation.

To a measure $\mu$, we associate an operator (linear transformation) on $L^2(G)$:

$$\text{Reg}_G(\mu) = \sum_{g \in G} \mu(g) \cdot \text{Reg}_G(g).$$

This is an analogue of the Fourier transform of classical harmonic analysis. In particular, it has the following property:

$$\text{Reg}_G(\mu * \nu) = \text{Reg}_G(\mu) \cdot \text{Reg}_G(\nu),$$

which is easy to verify from the definitions. We can recover $\mu$ by the formula

$$\mu = \text{Reg}_G(\mu)\delta_1, \quad (1)$$

where $\delta_1$ is the Dirac measure supported at 1. (Recall that $G$ is finite, hence we can embed the space of probability measures into $L^2$.)

The operator $\text{Reg}_G(\mu_S)$ is of norm 1 and it acts trivially on the one dimensional space of constant functions. We denote by $\text{Reg}_G^0$ the restriction of $\text{Reg}_G$ to the 1 codimensional space orthogonal to constants. We define by

$$\text{gap}(G, S) := 1 - \|\text{Reg}^0_G(\mu_S)\|$$

the spectral gap of the random walk on $G$ generated by $S$. Later, we will consider a slightly more general situation and replace $\mu_S$ by an arbitrary probability measure $\mu$. Then we write

$$\text{gap}(G, \mu) := 1 - \|\text{Reg}^0_G(\mu)\|.$$
It is clear that
\[ \| \text{Reg}_G (\mu_S^{*l}) - \text{Reg}_G (\mu_G) \| < e^{-l \text{gap}(G,S)}. \] (2)

This shows that if we take say \( l = 10 \log |G|/\text{gap}(G,S) \), then \( \mu_S^{*l} \) is very close to the uniform distribution (see (1)). In particular, the support of \( \mu_S^{*l} \) is the whole group. Thus spectral gap is a stronger quantity to measure uniformity than diameter. Somewhat surprisingly we can obtain a bound in the other direction, as well.

**Lemma 1.** Let \( G \) be an arbitrary finite group, and \( 1 \in S \subset G \). Suppose that \( S \) is symmetric, i.e. \( g \in S \) if and only if \( g^{-1} \in S \). We have:
\[ (\text{diam}(G,S) - 1)/\log |G| \leq \text{gap}(G,S)^{-1} \leq |S| \text{diam}(G,S)^2. \]

This lemma is well-known. The second (and more difficult) inequality of it can be found for example in [13, Corollary 1 in Section 3]. The other estimate follows directly from (2). The assumption on symmetricity is not an essential one. Later, at the beginning of Section 2.1 we show how to reduce the problem to the symmetric case.

We stop for a moment to connect our terminology to the computer science and combinatorics literature. If \( S \) is symmetric, then the matrix of the operator \( \text{Reg}_G (\mu_S) \) is proportional to the adjacency matrix of the Cayley graph of \( G \) with respect to the generating set \( S \). In that case \( \text{gap}(G,S) \) is proportional to the spectral gap of the Cayley graph. If \( \text{gap}(G,S) \geq c > 0 \) for a family of groups and generators than the corresponding family of Cayley graphs are called expanders.

However, in this paper we are looking for weaker bounds of the form \( \text{gap}(G,S) \geq \log^{-A} |G| \) which, in light of the above Lemma, is equivalent to \( \text{diam}(G,S) \leq \log^{A'} |G| \) as long as say \( |S| < 10 \) and one does not care about the value of \( A \) and \( A' \). We call such bounds poly-logarithmic.

1.2 Prior results

It was conjectured by Babai and Seress [1, Conjecture 1.7] that the family of non-Abelian finite simple groups have poly-logarithmic diameter, i.e. there is a constant \( A \) such that \( \text{diam}(G,S) \leq \log^{A} |G| \) holds for any non-Abelian finite simple group \( G \) and generating set \( S \subset G \). This has been verified for the family \( SL_2(F_p) \) by Helfgott [22, Main Theorem] and for finite simple groups of Lie type of bounded rank by Breuillard, Green and Tao [11, Theorem 7.1] and Pyber and Szabó [30, Theorem 2] independently. The conjecture is still open for other families of finite simple groups. The best known bound to date on the diameter of alternating groups is due to Helfgott and Seress [24] and it is slightly weaker than poly-logarithmic.

However, the first results on poly-logarithmic diameter were obtained for non-simple groups. Fix a prime \( p \) and a symmetric set \( S \subset SL_2(Z) \) such that
P. P. Varjú

Gamburd and Shahshahani [20, Theorem 2.1] proved the poly-log diameter estimate
\[ \text{diam}(SL_2(\mathbb{Z}/p^n\mathbb{Z}), S) \leq C \log^4 |SL_2(\mathbb{Z}/p^n\mathbb{Z})|, \]
where \( C \) depends only on \( S \) and \( A \) is an absolute constant. Dinai [15, Theorem 1.2] observed that the result holds with \( C \) depending only on \( p \) and he also improved the parameter \( A \). Thus the family \( SL_2(\mathbb{Z}/p^n\mathbb{Z}) \) enjoys a uniform poly-log diameter bound with respect to arbitrary generators. Using the result of Helfgott [22], the constant \( C \) can be made absolute. In [16, Theorem 1.1], Dinai extended the result to the quotients of other Chevalley groups over local rings.

The result of this paper is also about non-simple finite groups. In fact, it is part of our assumptions that all simple quotients of the groups we study has poly-logarithmic diameter. Our result has a huge overlap with [20], [15] and [16]. We will remark on this later.

1.3 The role of quasirandomness

Our approach is based on representation theory, but we will use only very basic facts of the theory. We explain the key idea of the paper, which allows us to estimate the spectral gap based on lower bounds for the dimension of nontrivial representations of the group. Let \( \mu \) be a probability measure on \( G \). Write \( \chi_G \) and \( \chi_G^{\circ} \) respectively for the characters of \( \text{Reg}_G \) and \( \text{Reg}_G^{\circ} \). Furthermore, we write
\[ \chi(\mu) = \sum_{g \in G} \mu(g) \cdot \chi(g), \]
where \( \chi \) is the character of a representation of \( G \). Then \( \chi_G(\mu) \) is the trace of the operator \( \text{Reg}_G(\mu) \).

For the moment, assume that \( \mu \) is symmetric that is \( \mu(g^{-1}) = \mu(g) \) for all \( g \in G \). Then \( \text{Reg}_G(\mu) \) is selfadjoint. We can decompose \( L^2(G) \) as the orthogonal sum of irreducible subrepresentations of \( \text{Reg}_G \). As is well known, if \( \pi \) is an irreducible representation of \( G \), then exactly \( \dim \pi \) isomorphic copies of \( \pi \) appear in the decomposition of \( \text{Reg}_G \). If \( \lambda \) is an eigenvalue of \( \text{Reg}_G(\mu) \) then there is a corresponding eigenvector in one of the irreducible representations of \( G \). Moreover, there is one in each isomorphic copy. This leads us to the following inequality which is fundamental to us:
\[ \dim(\pi) \cdot \lambda^2 \leq \chi_G(\mu * \mu), \]
where \( \pi \) is an irreducible representation of \( G \) such that there is an eigenvector corresponding to \( \lambda \) in \( \pi \). (Note that all eigenvalues of \( \text{Reg}_G(\mu * \mu) \) are non-negative.)

This idea goes back to Sarnak and Xue [32] and it is also one of the major steps in the work of Bourgain and Gamburd [2] on estimating spectral gaps and also in several papers [4], [5], [7], [34], [9], [31] which follow it. Gowers [21]...
also exploited the idea, and introduced the term quasirandom for groups that
does not have low dimensional non-trivial representations. He proved several
properties of such groups, in particular that they do not have large product-
free subsets. Nikolov and Pyber [29, Corollary 1] pointed out that Gowers’s
result implies that any element of a quasirandom group can be expressed as
the product of three elements of a sufficiently large subset.

In what follows, \( G \) is a profinite group and \( \Omega \) denotes the family of finite index
normal subgroups of it. An interesting example to have in mind is \( G = \text{SL}_d(\hat{\mathbb{Z}}) \),
where \( \hat{\mathbb{Z}} \) is the profinite (congruence) completion of the integers. (The reader
unfamiliar with the notion of profinite groups may assume that \( G \) is finite
without loss of generality.) Inspired by Gowers’s terminology, we make the
following definition.

**Definition 2.** We say that a profinite group \( G \) is \((c,\alpha)\)-quasirandom if for
every irreducible unitary representation \( \pi \) of \( G \), we have
\[
\dim \pi \geq c[G: \text{Ker}(\pi)]^\alpha.
\]

### 1.4 Results about profinite groups

Let \( G \) be a profinite group and \( \Gamma \) a finite index normal subgroup. Our plan is
to prove the estimate
\[
\text{gap}(G/\Gamma, S) \geq c \log^{-A}[G : \Gamma]
\]
with constants \( c, A \) independent of \( \Gamma \) and \( S \). We prove this statement by
induction as follows: We find a larger subgroup \( \Gamma' \supseteq \Gamma \in \Omega \) and assume that
the above spectral gap estimate holds for \( \Gamma' \). We use this to bound the trace
of the operator \( \text{Reg}_{G/\Gamma'}(\mu_S^{(l)}) \) for a suitable integer \( l \). This in turn gives an
estimate for the trace of \( \text{Reg}_{G/\Gamma}(\mu_S^{(l)}) \), and by (3) we can estimate \( \text{gap}(G/\Gamma, S) \)
and continue by induction.

In order that the induction step works, we need to ensure that \([\Gamma' : \Gamma]\) is “not
very large” compared to \([G : \Gamma]\). Of course, we cannot always find a suitable
\( \Gamma' \), in particular when \( G/\Gamma \) is simple.

The statement of the following theorem is very technical, so we first explain it
informally: We suppose that \( G \) is a quasirandom profinite group and partition
its finite index normal subgroups into two sets: \( \Omega_1 \cup \Omega_2 \). We assume that for
\( \Gamma \in \Omega_1 \) we can find a larger subgroup \( \Gamma' \in \Omega \) so that \([\Gamma' : \Gamma]\) is “not very large”. We
assume further that for \( \Gamma \in \Omega_2 \), the quotient \( G/\Gamma \) has poly-log spectral
gap (i.e. (4) is satisfied). Then we can conclude a poly-log estimate (4) for all
\( \Gamma \in \Omega \).

**Theorem 3.** Let \( c_1, \alpha, \beta \) and \( A \) be positive numbers satisfying \( \beta < \alpha \) and
\[
A \geq (\alpha - \beta)^{-1} - 1.
\]
Then there is a positive number \( C_1 \) depending only on \( c_1, \alpha, \beta \) and \( A \) such that
the following holds.
Let $G$ be a $(c_1, \alpha)$-quasirandom profinite group. Let $\Omega_1 \cup \Omega_2 = \Omega$ be a partition of the family of finite index normal subgroups of $G$. Suppose that for all $\Gamma \in \Omega_1$, we have $[G: \Gamma] > C_1$ and there is $\Gamma' \in \Omega$ with $\Gamma \lhd \Gamma'$ such that
\[ [\Gamma' : \Gamma] \leq [G : \Gamma']^\beta. \] (5)

Let $\mu$ be a Borel probability measure on $G$ and suppose that
\[ \text{gap}(G/\Gamma, \mu) \geq B \cdot \log^{-A}[G : \Gamma] \] (6)
for all $\Gamma \in \Omega_2$. Then for all $\Gamma \in \Omega$
\[ \text{gap}(G/\Gamma, \mu) \geq C^{-1}_1 B \cdot \log^{-A}[G : \Gamma]. \]

To verify assumption (6), we need to use known cases of the Babai–Seress conjecture mentioned above. Fortunately, this is available for many cases in the papers [22, Main Theorem], [23, Corollary 1.1], [11, Theorem 7.1], [30, Theorem 2]. In particular we can deduce the following.

**Corollary 4.** Let $S \subset \text{SL}_d(\hat{\mathbb{Z}})$ be a set generating a dense subgroup. Then for all integers $q > 1$, we have
\[ \text{gap}(\text{SL}_d(\mathbb{Z}/q\mathbb{Z}), S) \geq c \log^{-A}(q), \]
where $A > 0$ is a number depending on $d$ and $c$ is a number depending on $d$ and $|S|$.

We stress here that $c$ depends only on the cardinality of $S$. Note that this dependence is necessary, owing to the following example: It may happen that all but one element of $S$ projects into a proper subgroup of $\text{SL}_d(\mathbb{Z}/q\mathbb{Z})$ in which case the characteristic function of the subgroup is an almost invariant function if $|S|$ is large.

In the case $S \subset \text{SL}_d(\mathbb{Z})$ one can improve the poly-log bound to a uniform one. More precisely, Bourgain and Varjú [9, Theorem 1] proved $\text{gap}(\text{SL}_d(\mathbb{Z}/q\mathbb{Z}), S) \geq c$ with a constant $c > 0$ independent of $q$ but dependent on $S$. This strongly resonates with the state of affairs for compact semi-simple Lie groups to be discussed below.

Finally, we compare our result to the papers of Gamburd and Shahshahani [20, Theorem 2.1] and Dinai [15, Theorem 1.2], [16, Theorem 1.1]. These papers give results similar to Corollary 4, and even more, the proofs have some common features with our approach. However, they obtain the diameter bound directly and they use properties of commutators instead of representation theory. A flaw of our approach that it is not constructive, i.e. we can not give an efficient algorithm to express an element of $G/\Gamma$ as a product of elements of $S$. On the contrary, [20, Theorem 2.1], [15, Theorem 1.2], [16, Theorem 1.1] give such algorithms. The advantage of our paper that it seems to apply in more general cases.
situations, e.g. [20, Theorem 2.1], [15, Theorem 1.2], [16, Theorem 1.1] are restricted to the case when \( q \) is the power of a prime.

A comment on the exponents: If one considers the group \( G = \text{SL}_2(\mathbb{Z}_p) \), where \( p \) is a fixed prime, then the conditions of Theorem 3 can be satisfied for any \( A > 2 \). This via Lemma 1 exactly recovers the diameter bound for \( \text{SL}_2(\mathbb{Z}/p^k\mathbb{Z}) \) in the paper [15, Theorem 1.2], but our bound for the spectral gap is better than what could be obtained from [15]. If one considers \( \text{SL}_d \) with \( d \) large than our bounds deteriorate compared to [15]. However, it seems possible that a more careful version of our argument could give better bounds, but this requires a more precise understanding of the representations. In Section 2.2 we include some remarks about what this would require. These ideas are worked out in the setting of compact Lie groups.

1.5 Results about compact Lie groups

We turn to the second setting of our paper. Let \( G \) be a semi-simple compact Lie group endowed with the bi-invariant Riemannian metric. We denote by \( \text{dist}(g, h) \) the distance of two elements \( g, h \in G \). Let \( \varepsilon > 0 \) be a number and \( S \subset G \) be a finite subset which generates a dense subgroup. Again, for simplicity, we assume \( 1 \in S \). We define the diameter of \( G \) at scale \( \varepsilon \) with respect to  \( S \) by

\[
\text{diam}_\varepsilon(G, S) = \min \{ l : \text{ for every } g \in G \text{ there is } h \in S^l \text{ such that } \text{dist}(g, h) < \varepsilon \}.
\]

We also introduce the relevant spectral gap notion. This requires some basic facts about the representation theory of compact Lie groups. We follow the notation in [12], but the results we need can be found in many of the textbooks on the subject, as well.

Let \( T \) be a maximal torus in \( G \) and denote by \( LT \) its tangent space at 1. Then \( T \) can be identified with \( LT/I \) via the exponential map, where \( I \) is a lattice in \( LT \). Denote by \( LT^* \) the dual of \( LT \) and by \( I^* \subset LT^* \) the lattice dual to \( I \). We denote by \( R \subset I^* \) the set of roots, and by \( R_+ \) (a choice of) the positive roots. We fix an inner product \( \langle \cdot, \cdot \rangle \) on \( LT \) which is invariant under the Weyl group. Denote by

\[
K = \{ u \in LT : \langle u, v \rangle > 0 \text{ for every } v \in R_+ \}
\]

the positive Weyl chamber and by \( \overline{K} \) its closure.

It is well known (see [12, Chapter IV (1.7)]) that the irreducible representations of \( G \) can be parametrized by the elements of \( I^* \cap \overline{K} \). For \( v \in I^* \cap \overline{K} \), we denote by \( \pi_v \) the unitary representation of \( G \) with highest weight \( v \).

For a finite Borel measure \( \mu \) on \( G \), we write

\[
\pi_v(\mu) = \int \pi_v(g) \, d\mu(g),
\]

which is an operator (linear transformation) on the representation space of \( \pi_v \).

Let \( r > 1 \) be a number, and \( S \subset G \) a finite set which contains 1 and generates
P. P. Varjú

a dense subgroup. Define the spectral gap at scale $1/r$ with respect to $S$ by

$$\text{gap}_r(G,S) := 1 - \max_{0 < |v| \leq r} \| \pi_v(\mu_S) \|.$$  

As in the finite case, the notions of spectral gap and diameter are closely related:

**Lemma 5.** Let $G$ be a compact connected semi-simple Lie group, and let $1 \in S \subset G$ be finite. Then there is a constant $C > 0$ depending only on $G$ such that for any $\varepsilon > 0$

$$\text{diam}_\varepsilon(G,S) \leq \frac{C \log(\varepsilon^{-1})}{\text{gap}_{C\varepsilon}(G,S)}$$

and for any $r \geq 1$

$$\text{gap}_r(G,S) \geq \frac{1}{|S| \text{diam}_{(C \varepsilon)^{-1}}(G,S)^2}.$$  

This lemma is also well-known. We give a proof in Section 4 for completeness. In Section 3, we develop an analogue for compact Lie groups of the ideas explained in the previous section. A replacement for (3) was given by Gamburd, Jacobson and Sarnak [19] and it appeared in various forms in [3] and [6], as well. However, these are based on direct calculation with characters rather than on multiplicities of eigenvalues. A more direct analogue of (3) and also of the results of Gowers [21] and Nikolov and Pyber [29] was developed very recently by Saxcélé [33].

Our main result in the setting of compact Lie groups is the following:

**Theorem 6.** For every semi-simple compact connected Lie group $G$, there are numbers $c, r_0$ and $A$ such that the following holds. Let $\mu$ be an arbitrary probability measure on $G$. Then

$$\text{gap}_r(G,\mu) \geq c \cdot \text{gap}_{r_0}(G,\mu) \cdot \log^{-A} r.$$  

For simple groups, the value of $A$ can be found in Table 1. For semi-simple groups $A$ is the maximum of the corresponding values over all simple quotients of $G$. In particular, $A \leq 2$ for all groups.

<table>
<thead>
<tr>
<th>$A_n$</th>
<th>$B_n$</th>
<th>$C_n$</th>
<th>$D_n$</th>
<th>$E_6$</th>
<th>$E_7$</th>
<th>$E_8$</th>
<th>$F_4$</th>
<th>$G_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 + \frac{2}{n+1}$</td>
<td>$1 + \frac{1}{n}$</td>
<td>$1 + \frac{1}{n}$</td>
<td>$1 + \frac{1}{n}$</td>
<td>$\frac{7}{5}$</td>
<td>$\frac{16}{5}$</td>
<td>$\frac{24}{7}$</td>
<td>$\frac{12}{5}$</td>
<td>$\frac{2}{3}$</td>
</tr>
</tbody>
</table>

Table 1: The value of $A(G)$ in terms of the Dynkin diagram

A poly-logarithmic estimate for $\text{diam}_\varepsilon(\text{SU}(2),S)$ was found by Solovay and Kitaev independently. Nice expositions are in the paper of Dawson and Nielsen [14, Theorem 1] and in the book [26, Chapter 8.3], where they obtain the bound $\text{diam}_\varepsilon(\text{SU}(2),S) < \log^3 \varepsilon^{-1}$. Our theorem provides the same bound for the
diameter of SU(2). On the other hand, our spectral gap bound in Theorem 6 beats anything that could be obtained from a diameter bound via Lemma 5. Dolgopyat [17, Theorems A.2 and A.3] gave an estimate for the spectral gap which is weaker than poly-logarithmic but his argument would give a poly-log estimate without significant changes. His proof consists of a version of the Solovay-Kitaev argument (that he discovered independently) and a variant of Lemma 5. The connection between the present paper and [17] was pointed out to me by Breuillard, see also his survey [10]. We note that Bourgain and Gamburd [3, Corollary 1.1], [6, Theorem 1] showed that when \( \mu = \mu_S \) for some finite set \( 1 \in S \subset \text{SU}(d) \) and the entries of the elements of \( S \) are algebraic numbers, then 

\[
gap_r(G,S) \geq c
\]

for some constant \( c \) depending on \( G \) and \( S \). Their argument is likely to carry over to arbitrary semi-simple compact Lie groups, however the assumption on algebraicity is essential for the proof. It is a very interesting open problem whether this assumption can be removed. Moreover, we raise the following question: Is it true that there are numbers \( c, r_0 \) depending only on \( G \) such that 

\[
gap_r(G,\mu) \geq c \cdot \gap_{r_0}(G,\mu)
\]

for all probability measures \( \mu \) on \( G \)?

Finally, we state a technical result which almost immediately follow from Theorem 6. Its purpose is that this is the version used in the paper [35] to study random walks in the group of Euclidean isometries.

For a measure \( \nu \) on \( G \), we define the measure \( \tilde{\nu} \) by

\[
\int f(x) \, d\tilde{\nu}(x) = \int f(x^{-1}) \, d\nu(x)
\]

for all continuous functions \( f \). We write \( m_G \) for the Haar measure on \( G \).

**Corollary 7.** Let \( G \) be a compact Lie group with semi-simple connected component. Let \( \mu \) be a probability measure on \( G \) such that \( \text{supp}(\tilde{\mu} \ast \mu) \) generates a dense subgroup in \( G \). Then there is a constant \( c > 0 \) depending only on \( \mu \) such that the following holds. Let \( \varphi \in \text{Lip}(G) \) be a function such that \( \|\varphi\|_2 = 1 \) and \( \int \varphi \, dm_G = 0 \). Then

\[
\left\| \int \varphi(h^{-1}g) \, d\mu(h) \right\|_2 < 1 - c \log^{-A}(\|\varphi\|_{\text{Lip}} + 2),
\]

where \( A \) depends on \( G \) and is the same as in Theorem 6.

The rest of the paper is organized as follows. In Section 2 we give the proof of Theorem 3 and Corollary 4. The proof of Theorem 6 is given in Section 3. Sections 2 and 3 are independent, but there is a strong analogy between the two arguments. Finally we prove Corollary 7 and Lemma 5 in Section 4.
Throughout the paper we use the letters $c$ and $C$ to denote numbers which may depend on several other quantities and their value may change at each occurrence. We follow the convention that $c$ tends to denote numbers that we consider “small” and $C$ denotes those that we consider “large”.

1.6 Motivation

In recent years there was a lot of progress on the Babai-Seress conjecture mentioned above, although it has not been settled yet. In addition, poly-log spectral gap and diameter is known to hold for some families of non-simple finite groups as well. What the scope of this phenomenon is, is an interesting question. Our result on finite groups is a (very modest) step towards understanding this. In Section 2.2 we include some remarks on how to exploit our approach for non-quasirandom groups.

Our main motivation for Theorem 6 is the application in the paper [35]. Although it seems easy to extend the Solovay-Kitaev approach to prove similar poly-log type bounds, we believe that our method gives better exponents, at least for spectral gaps.

There are many recent applications of spectral gaps. Many of these require stronger bounds than what we obtain in this paper, e.g. the results in [2], [4], [5], [7], [34], [9], [31] mentioned above. However, for some applications, the poly-log type bounds are enough, at least to obtain the same qualitative result. Prominent examples are the work of Ellenberg, Hall and Kowalski [18], the Group Large Sieve developed by Lubotzky and Meiri [28], the study of curvatures in Apollonian Circle Packings by Bourgain and Kontorovich [8] and the study of random walks on Euclidean isometries by Varjú [35]. However, our results are relevant only for the last two of the above papers, because [18] and [28] requires spectral gaps only for products of two simple groups. In addition, in the case of [8] a better uniform spectral gap is available.

ACKNOWLEDGMENTS. I thank Jean Bourgain for discussions related to this project, in particular, for explaining to me the relation between the diameter and the spectral gap. I thank Nicolas de Saxcé for careful reading of my manuscript and for his suggestions which greatly improved the presentation of the paper. I thank Emmanuel Breuillard for calling my attention on the work of Dolgopyat [17].

2 Profinite groups

2.1 Proof of Theorem 3

Recall that $G$ is a profinite group and $\Omega$ is the family of finite index normal subgroups of it. Assume that the hypothesis of the theorem is satisfied. We note that $\alpha \leq 1/2$, since $(\dim \pi)^2 \leq |G/\Gamma|$ for any irreducible representation $\pi$ of the group $G/\Gamma$. Consequently $A \geq 1$. 
We first explain how to reduce to the case when $\mu$ is symmetric. Define the probability measure $\tilde{\mu}$ by
\[ \int f(g) \, d\tilde{\mu}(g) = \int f(g^{-1}) \, d\mu(g). \]
(If $G$ is finite, this is can be expressed as $\tilde{\mu}(g) = \mu(g^{-1})$.) Clearly $\tilde{\mu} * \mu$ is symmetric and
\[ \text{Reg}_{G/\Gamma}(\tilde{\mu} * \mu) = \text{Reg}_{G/\Gamma}(\mu)^* \cdot \text{Reg}_{G/\Gamma}(\mu). \]
Hence
\[ \|\text{Reg}_{G/\Gamma}(\tilde{\mu} * \mu)\| = \|\text{Reg}_{G/\Gamma}(\mu)\|^2. \]
This yields
\[ \text{gap}(G/\Gamma, \tilde{\mu} * \mu) \geq \text{gap}(G/\Gamma, \mu) \geq \text{gap}(G/\Gamma, \tilde{\mu} * \mu)/2. \]

This shows that the spectral gap of $\tilde{\mu} * \mu$ is roughly proportional to that of $\mu$, hence it suffices to prove the theorem for the first one. From now on, we assume that $\mu$ is symmetric, hence the operator $\text{Reg}_{G/\Gamma}(\mu)$ is selfadjoint and has an eigenbasis with real eigenvalues.

We define
\[ l_\Gamma = 2[C_\Gamma \log^{A+1}[G : \Gamma]], \]
where
\[ C_\Gamma := C_0\left(10 - \log^{-1/10}([G : \Gamma])\right), \]
and
\[ C_0 = \max_{\Gamma \in \Omega_2} \frac{1}{\text{gap}(G/\Gamma, \mu) \cdot \log^A [G : \Gamma]} \]
(The role of the subtracted term in the definition of $C_\Gamma$ is simply to cancel lower order terms later.)

Recall that we denote by $\chi_{G/\Gamma}$ the character of $\text{Reg}_{G/\Gamma}$. Our goal is to prove
\[ \chi_{G/\Gamma}(\mu^{*(l_\Gamma)}) \leq M, \quad (7) \]
where $M \geq 2$ is a suitably large number depending on $\alpha$ and $\beta$. Once we proved this, the claim of the theorem will be concluded easily.

The proof is by induction with respect to the partial order $\prec$ on $\Gamma \in \Omega$. Suppose that (7) holds for all $\Gamma' \in \Omega$ with $\Gamma \not\prec \Gamma'$.

We distinguish two cases. First we suppose that $\Gamma \in \Omega_2$. Note that $\chi_{G/\Gamma}(\mu^{*(l_\Gamma)})$ is the trace of the operator $\text{Reg}_{G/\Gamma}(\mu^{*(l_\Gamma)})$, hence it is the sum of its eigenvalues. The non-trivial eigenvalues are bounded by $e^{-l_\Gamma \cdot \text{gap}(G/\Gamma, \mu)}$, hence
\[ \chi_{G/\Gamma}(\mu^{*(l_\Gamma)}) \leq 1 + [G : \Gamma] \cdot e^{-l_\Gamma \cdot \text{gap}(G/\Gamma, \mu)} \leq 1 + [G : \Gamma] \cdot e^{-C_0 \log^{A+1}[G : \Gamma] \cdot \text{gap}(G/\Gamma, \mu)} \leq 1 + 1 \]
using the definitions of $l_{\Gamma}$, $C_{\Gamma}$ and $C_{0}$. Hence (7) follows.

Now we suppose that $\Gamma \notin \Omega_{2}$, hence $\Gamma \in \Omega_{1}$, in particular $[G : \Gamma] \geq C_{1}$, where $C_{1}$ can be taken as large as we need depending on $\alpha, \beta, c_{1}, A$. Let $\Gamma'$ be such that $\Gamma' \in \Omega$ and $[\Gamma' : \Gamma]$ is minimal but larger than 1. Then by assumption, $[\Gamma' : \Gamma] \leq [G : \Gamma']^{\alpha}$.

Since $\chi_{G/\Gamma}$ is pointwise majorized by the function $[\Gamma' : \Gamma] \cdot \chi_{G/\Gamma'}$, and $\mu^{*}(l_{\Gamma})$ is a positive measure, we have

$$\chi_{G/\Gamma}(\mu^{*}(l_{\Gamma})) \leq [\Gamma' : \Gamma] \cdot \chi_{G/\Gamma}(\mu^{*}(l_{\Gamma})) \leq M[\Gamma' : \Gamma].$$  

(8)

We applied (7) for $\Gamma'$ in the second inequality.

Denote by $\lambda_{0} = 1, \lambda_{1}, \ldots, \lambda_{k}$ the eigenvalues of the operator $\text{Reg}_{G/\Gamma}(\mu)$ each listed as many times as its multiplicity. Then

$$\chi_{G/\Gamma}(\mu^{*}(l_{\Gamma})) = 1 + \sum_{i=1}^{k} \lambda_{i}^{l_{\Gamma}} \leq 1 + \left( \sum_{i=1}^{k} \lambda_{i}^{l_{\Gamma}} \right) \cdot \max \{ \lambda_{i} \}^{l_{\Gamma}-l_{\Gamma'}} \leq 1 + M[\Gamma' : \Gamma] \max \{ \lambda_{i} \}^{l_{\Gamma}-l_{\Gamma'}}.$$  

(9)

We applied (8) in the last line. Also note, that all terms are positive because $l_{\Gamma}$ is even by construction.

Our next goal is to obtain a sufficient bound on the $\lambda_{i}$. This can be deduced from (8) and the assumption about the dimension of faithful representations.

The representation $\text{Reg}(G/\Gamma)$ can be decomposed as the orthogonal sum of irreducible subrepresentations. Each irreducible representation occur with multiplicity equal to its dimension.

Consider now an eigenvalue $\lambda_{i}$. There is a corresponding eigenvector which is contained in an irreducible subrepresentation of $\text{Reg}(G/\Gamma)$. Denote this representation by $\pi$. Write $\Gamma'' = \text{Ker}(\pi)$.

First we consider the case that $\Gamma'' = \Gamma$. It follows that $\lambda_{i}^{l_{\Gamma'}}$ is an eigenvalue of $\text{Reg}_{G/\Gamma}(\mu^{*}(l_{\Gamma}))$ with multiplicity at least

$$\dim \pi \geq c_{1}[G : \Gamma]^{|\alpha|},$$

(by the assumption on quasirandomness). Hence by (8) we can conclude that

$$\lambda_{i}^{l_{\Gamma'}} \leq \frac{M[\Gamma' : \Gamma]}{c_{1}[G : \Gamma]^{|\alpha|}}.$$  

Next, we consider the case when $\Gamma'' \neq \Gamma$. By the assumption on the minimality of $[\Gamma' : \Gamma]$, we have $[G : \Gamma''] \leq [G : \Gamma']$. We apply (7) for $\Gamma''$ along with the bound for the multiplicity of eigenvalues and get

$$\lambda_{i}^{l_{\Gamma''}} \leq \frac{M}{c_{1}[G : \Gamma'']^{|\alpha|}}.$$  

An easy calculation shows that the bound we obtain for $|\lambda_{i}|$ is worsening when $[G : \Gamma'']$ grows.
Thus in both cases we obtain:
\[
\lambda'_i \leq \frac{M[\Gamma' : \Gamma]}{c_1[G : \Gamma]^\alpha}.
\] (10)

We plug this into (9) and we want to conclude \(\chi_{G/\Gamma}(\mu^{s(l)}) \leq M\). To this end, we need
\[
M[\Gamma' : \Gamma] \left( \frac{M[\Gamma' : \Gamma]}{c_1[G : \Gamma]^\alpha} \right)^{(l-1)/l_{\Gamma'}} \leq M - 1
\]
that is
\[
\frac{M}{M - 1}[\Gamma' : \Gamma] \leq \left( \frac{c_1[G : \Gamma]^\alpha}{M[\Gamma' : \Gamma]} \right)^{(l-1)/l_{\Gamma'}}.
\] (11)

For simplicity, we write \(X = \log[G : \Gamma']\) and \(Y = \log[\Gamma' : \Gamma]\). Then by the definition of \(l_{\Gamma'}\):
\[
\frac{l_{\Gamma'} - l_{\Gamma'}}{l_{\Gamma'}} = \frac{2[C_{\Gamma'}(X + Y)^{A+1}] - 2[C_{\Gamma'}X^{A+1}]}{2[C_{\Gamma'}X^{A+1}]}
\geq \frac{C_{\Gamma'}X^{A+1} - C_{\Gamma'}X^{A+1} - 1}{C_{\Gamma'}X^{A+1}}
\geq \frac{C_{\Gamma'}(X + Y)^{A+1} - C_{\Gamma'}X^{A+1}}{C_{\Gamma'}X^{A+1}} + \frac{C_{\Gamma'} - C_{\Gamma'}}{C_{\Gamma'}X^{A+1}}
\geq (A + 1)\frac{Y}{X} + c_2\frac{Y}{X^{1/10}},
\] (12)

where \(c_2\) is an absolute constant if \(X\) is larger than an absolute constant. (Using the definition of \(C_{\Gamma'}\), we evaluate \((C_{\Gamma'} - C_{\Gamma'})/C_{\Gamma'}\) and get the second term of (12). Notice that this is of larger order of magnitude than \(1/X^{A+1}\).)

Then the logarithm of the right hand side of (11) is at least:
\[
\left( (A + 1)\frac{Y}{X} + c_2\frac{Y}{X^{1/10}} \right) (\alpha X - \log(M/c_1) - Y)
\geq \alpha(A + 1)Y - \log(M/c_1)(A + 1)\frac{Y}{X} - (A + 1)\frac{Y^2}{X} + c_3\frac{Y}{X^{1/10}}
\geq \alpha(A + 1)Y - (A + 1)\frac{Y^2}{X} + c_3\frac{Y}{X^{1/10}}
\] if \(X\) is sufficiently large depending on \(\alpha, \beta, c_1, A\). Here \(c_3\) is a sufficiently small constant depending on \(\alpha\) and \(\beta\) satisfying \(c_3X \leq c_2(\alpha X - \log(M/c_1) - Y)\). (Recall that \(Y \leq \beta X\) by assumption.)

To get (11), we need
\[
\log(M/(M - 1)) + Y \leq \alpha(A + 1)Y - (A + 1)\frac{Y^2}{X} + c_3\frac{Y}{X^{1/10}}
\]
that is
\[
A + 1 \geq \frac{1 + \log(M/(M - 1))/Y - c_4X^{-1/10}}{\alpha - Y/X + c_4X^{-1/10}},
\] (13)
where $c_4$ is yet another number depending on $\alpha, \beta, c_1, A$.

We consider two cases. If $Y \leq \sqrt{X}$ (and $X$ is sufficiently large) then $Y/X \leq c_4 X^{-1/10}$. Clearly $Y \geq \log 2$, hence (13) holds if

$$A + 1 \geq \frac{1 + \log(M/(M - 1))/\log 2}{\alpha}.$$  

On the other hand if $Y \geq \sqrt{X}$ (and $X$ is sufficiently large) then $\log(M/(M - 1))/Y \leq c_4 X^{-1/10}$. Recall that $Y/X \leq \beta$ by assumption, hence (13) holds if $A + 1 \geq (\alpha - \beta)^{-1}$.

By choosing $M$ sufficiently large depending on $\alpha, \beta$, we can ensure that

$$1 + \frac{\log(M/(M - 1))/\log 2}{\alpha} \leq \frac{1}{\alpha - \beta}.$$  

Thus (13) holds in either case if $A + 1 \geq (\alpha - \beta)^{-1}$, which was assumed in the theorem. This completes the induction to prove (7).

Let $\lambda$ be an eigenvalue of $\text{Reg}_G^\mu$. We want to show that $|\lambda| < 1 - C_1^{-1} B \cdot \log^{\alpha - A} [G : \Gamma]$.

Denote by $\pi$ an irreducible representation of $G/\Gamma$ that contains an eigenvector corresponding to $\lambda$. Without loss of generality, we can assume that $\Gamma = \text{Ker}(\pi)$.

From quasirandomness and (7) we get

$$\lambda^{lr} \leq \frac{M}{c_1 [G : \Gamma]^{\alpha}}.$$  

Thus

$$\log |\lambda| \leq \frac{\log M/c_1 - \alpha \log [G : \Gamma]}{l^{lr}}.$$  

If we compare this with the definition of $l^{lr}$ we can conclude the theorem.

2.2 Some remarks about weakening the hypothesis on quasirandomness

In this section we present some ideas that lead to a refined version of (9) and (10). Using this, one could obtain a version of Theorem 3 with a weaker hypothesis instead of quasirandomness. Namely one would require that the quotient groups does not have “many” irreducible representations with “small” dimension. This weaker form of quasirandomness would be very technical hence we do not state a theorem. However, (based on analogy with compact Lie groups) it seems possible that these ideas lead to better bounds for the groups $\text{SL}_d(\mathbb{Z}_p)$ than Theorem 3.

Proposition 8. Let $G$ be a finite group and $N$ a normal subgroup. Let $\rho_1, \ldots, \rho_n$ be all irreducible representations of $N$ up to isomorphism. Denote by
Random Walks in Compact Groups

\( a(\rho_j) \) the number of \( G \)-conjugates of \( \rho_j \). Denote by \( d(\rho_j) \) the smallest possible dimension of a representation of \( G \) whose restriction to \( N \) contains \( \rho_j \). Let \( \mu \) be a symmetric probability measure on \( G \) and suppose that

\[
\chi_{G/N} \left( \mu^{*2l} \right) \leq M
\]

for some numbers \( l \) and \( M > 1 \).

Then for all integers \( l' > l \) we have:

\[
\chi_{G} \left( \mu^{*2l'} \right) \leq \sum_{j=1}^{n} \dim(\rho_j)^2 M \cdot \left( \frac{a(\rho_j) \dim(\rho_j)^2 M}{d(\rho_j)} \right)^{(l'-l)/l}.
\]

(15)

Observe that \( \sum \dim(\rho_j)^2 = |N| \) and \( a(\rho_j) \dim(\rho_j)^2 \leq |N| \), hence we obtain (9) combined with (10), if we estimate \( d(\rho_j) \) using quasirandomness. If \( d(\rho_j) \) is small only for a very few \( j \), Proposition 8 significantly improves the argument given in the previous section.

Proof. We recall some facts from representation theory. Let \( \pi \) be an irreducible representation of \( G \), and denote by \( \chi_{\pi} \) its character. Denote by \( \pi|_{N} \) the restriction of \( \pi \) to \( N \). By Clifford’s theorem, the irreducible components of \( \pi|_{N} \) is a \( G \)-conjugacy class of representations and each appears with the same multiplicity. We denote by \( a(\pi) \) the number of different irreducible components of \( \pi|_{N} \), by \( b(\pi) \) their common multiplicity and by \( c(\pi) \) their common dimension. Thus \( \dim(\pi) = a(\pi) \cdot b(\pi) \cdot c(\pi) \).

Denote one of the irreducible components of \( \pi|_{N} \) by \( \rho \) and its character by \( \chi_{\rho} \). Let \( \rho^{g_1}, \cdots, \rho^{g_{a(\pi)}} \) be all \( G \)-conjugates of \( \rho \). In what follows, the function

\[
\varphi = \sum_{i=1}^{a(\pi)} \chi_{\rho^{g_i}}
\]

will play an important role. We also extend it to \( G \) by setting it 0 in the complement of \( N \), i.e. we write \( \tilde{\varphi}(g) = \varphi(g) \) for \( g \in N \) and \( \tilde{\varphi}(g) = 0 \) otherwise. We use two inner products, one on \( L^2(G) \) and one on \( L^2(N) \) defined by:

\[
\langle f_1, f_2 \rangle_G = \frac{1}{|G|} \sum_{g \in G} f_1(g) f_2(g), \quad \langle f_1, f_2 \rangle_N = \frac{1}{|N|} \sum_{g \in N} f_1(g) f_2(g).
\]

These are the inner products with respect to which the irreducible characters of the corresponding groups are orthonormal.

We write:

\[
\langle \chi_{\pi}, \varphi \rangle_G = \frac{|N|}{|G|} \langle \chi_{\pi}|_{N}, \varphi \rangle_N = \frac{|N|a(\pi)b(\pi)}{|G|}.
\]

A similar calculation shows that the inner product of \( \tilde{\varphi} \) with an irreducible character of \( G \) is always non-negative.
Since \( \tilde{\phi} \) is a class function on \( G \), it can be decomposed as a linear combination of irreducible characters. According to the above calculation, the coefficient of \( \chi_\pi \) is \( \frac{|N|a(\pi)b(\pi)}{|G|} \). Thus
\[
\chi_\pi \left( \mu^{\ast(2l)} \right) \leq \frac{|G|}{|N|a(\pi)b(\pi)} \tilde{\phi} \left( \mu^{\ast(2l)} \right),
\]
(16)
(Note that \( \chi_\pi(\mu^{\ast(2l)}) \geq 0 \) being the trace of a positive operator.)

On the other hand,
\[
\|\tilde{\phi}\|_\infty \leq \sum_{i=1}^{a(\pi)} \|\chi_{\rho_i}\|_\infty = a(\pi)c(\pi).
\]
Thus for every \( g \in G \), we have
\[
|\tilde{\phi}(g)| \leq \frac{|N|a(\pi)c(\pi)}{|G|} \chi_{G/N}(g).
\]
(Note that for \( g \notin N \) both sides are 0.) Therefore
\[
\tilde{\phi} \left( \mu^{\ast(2l)} \right) \leq \frac{|N|a(\pi)c(\pi)}{|G|} \chi_{G/N} \left( \mu^{\ast(2l)} \right).
\]
(17)

We combine (16) and (17) and get
\[
\chi_\pi \left( \mu^{\ast(2l)} \right) \leq \frac{c(\pi)}{b(\pi)} \chi_{G/N} \left( \mu^{\ast(2l)} \right).
\]
This implies that for all eigenvalues \( \lambda \) of \( \pi(\mu^{\ast(2l)}) \), we have
\[
|\lambda| < \frac{c(\pi)}{b(\pi)} M = \frac{a(\pi)c(\pi)^2}{\text{dim}(\pi)} M.
\]
(Here we also used the hypothesis (14).)

Now let \( \pi_1, \pi_2, \ldots, \pi_k \) denote all the irreducible representations (up to isomorphism) of \( G \) whose restriction to \( N \) contain \( \rho \). By a calculation very similar to the one leading to (16) we get
\[
\sum_{i=1}^{k} \frac{|N|a(\pi_i)b(\pi_i)}{|G|} \chi_{\pi_i} \left( \mu^{\ast(2l)} \right) \leq \tilde{\phi} \left( \mu^{\ast(2l)} \right).
\]
Combining with (17), we get
\[
\sum_{i=1}^{k} \frac{b(\pi_i)}{c(\pi_i)} \chi_{\pi_i} \left( \mu^{\ast(2l)} \right) \leq \chi_{G/N} \left( \mu^{\ast(2l)} \right) \leq M.
\]
Random Walks in Compact Groups

We used again the hypothesis \((14)\). Multiplying by \(a(\pi)c(\pi)^2 = a(\pi_i)c(\pi_i)^2\) (which is independent of \(i\)) we get

\[
\sum_{i=1}^{k} \dim(\pi_i) \chi_{\pi_i}(\mu^{*(2l)}) \leq a(\pi)c(\pi)^2 M. \tag{19}
\]

We use \((18)\) and write for an integer \(l' \geq l\):

\[
\sum_{i=1}^{k} \dim(\pi_i) \chi_{\pi_i}(\mu^{*(2l')}) \leq a(\pi)c(\pi)^2 M \left(\frac{a(\pi)c(\pi)^2 M}{\dim(\pi)}\right)^{(l'-1)/l}. \tag{20}
\]

We sum \((20)\) for \(\rho = \rho_1, \ldots, \rho_n\) and get \((15)\).

2.3 Proof of Corollary 4

We first discuss quasirandomness. This was already proved for \(\text{SL}_d(\hat{\mathbb{Z}})\) by Bourgain and Varjú [9]. In fact, it is easy to deduce it from the corresponding result about \(\text{SL}_d(\mathbb{Z}_p)\) which was obtained by Bourgain and Gamburd [4, Lemma 7.1] for \(d = 2\) and by Saxcé [33, Lemme 5.1] for \(d \geq 2\). For completeness, we explain this deduction.

For an integer \(q > 0\) denote by

\[
\Gamma_q = \{ g \in \text{SL}_d(\hat{\mathbb{Z}}) : g \equiv 1 \pmod{q}\}
\]

the mod \(q\) congruence subgroup of \(\text{SL}_d(\hat{\mathbb{Z}})\).

Let \(p\) be a prime, \(k \geq 1\) be an integer, and let \(\pi\) be a representation of \(\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_{p^k}\) which is not a representation of \(\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_{p^{k-1}}\). By [33, Lemme 5.1], \(\pi\) is of dimension at least

\[
c \cdot \dim(\pi) \leq c' \cdot \dim(\pi) \leq c \cdot [\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma_{p^k}]^{1/(d+1)},
\]

where \(c > 0\) is a number depending on \(d\). For any \(\epsilon > 0\), we can replace this bound by \([\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma_{p^k}]^{1/(d+1) - \epsilon}\) if \(p\) is sufficiently large depending on \(\epsilon\). Let \(\pi\) be an irreducible unitary representation of \(\text{SL}_d(\hat{\mathbb{Z}})\). Since \(\Gamma_q, q > 1\) form a system of neighborhoods of 1 in \(\text{SL}_d(\hat{\mathbb{Z}})\), there is \(q\) such that \(\text{Ker}(\pi) \supset \Gamma_q\). Let \(q\) be minimal with this property. Let \(q = p_1^{k_1} \cdots p_n^{k_n}\) such that \(p_i\) are primes. Then

\[
\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_q = \left(\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_{p_1^{k_1}}\right) \times \cdots \times \left(\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_{p_n^{k_n}}\right).
\]

Any representation of this group is a tensor product of representations of \(\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_{p_i^{k_i}}\). Hence

\[
\dim \pi \geq c^m [\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma_{p^k}]^{1/(d+1) - \epsilon}, \tag{21}
\]

where \(m\) is the number of not large enough primes in the sense of the previous paragraph. Thus \(\text{SL}_d(\hat{\mathbb{Z}})\) is \((c^m, 1/(d + 1) - \epsilon)\)-quasirandom.

Documenta Mathematica 18 (2013) 1137–1175
We refer the interested reader to the paper of Kelmer and Silberman \cite[Section 4]{K1}, where quasirandomness is proved with optimal parameter $\alpha$ for some other arithmetic groups.

We define $\Omega_1$ and $\Omega_2$. We fix an integer $M$ that we will set later depending on $d$. Let $\Gamma \triangleleft \text{SL}_d(\hat{\mathbb{Z}})$ be a finite index normal subgroup. Denote by $q$ the smallest integer such that $\Gamma_q \triangleleft \Gamma$. We put $\Gamma$ in $\Omega_1$, if $q$ has at least $M + 1$ prime factors (taking multiplicities into account) and $q$ is sufficiently large. We put $\Gamma$ in $\Omega_2$ otherwise.

Let $\Gamma \in \Omega_1$ and let $q$ be the same as above. Let $p|q$ be the smallest prime divisor of $q$. By simple calculation:

$$[\Gamma_{q/p} : \Gamma_q] \leq [\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma_{q/p}]^{1/M}.$$ 

Now we define $\Gamma' := \Gamma_{q/p} \Gamma$. Clearly

$$[\Gamma' : \Gamma] \leq [\Gamma_{q/p} : \Gamma_q],$$

so we only need to estimate $[\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma']$ in terms of $[\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma_{q/p}]$. For our purposes the very crude bound

$$[\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma'] \geq c^m [\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma_{q/p}]^{1/(d+1)-\varepsilon}$$

is sufficient which follows from (21). This implies

$$[\Gamma' : \Gamma] \leq [\text{SL}_d(\hat{\mathbb{Z}}) : \Gamma^{(d+2)/M}].$$

if $q$ is sufficiently large (and $\varepsilon$ is sufficiently small).

We choose $\alpha$ and $\beta$ in such a way that $\beta < \alpha < 1/(d + 1)$. Then the quasirandomness is satisfied, and also (5) if we set $M \geq (d + 2)/\beta$.

It is left to verify (6) for $\mu = \mu_S$. First we note that by the same argument as at the beginning of the proof of Theorem 3, we can assume that $S$ is symmetric. We show that the groups $\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_q$ for $\Gamma_q \in \Omega_2$ have poly-logarithmic diameter with respect to any generating set $S$. In light of Lemma 1 this implies (6).

Let now $\Gamma_q \in \Omega_2$. There are two possibilities. If $q$ is small (e.g. $q \leq C_1$ or as in the definition of $\Omega_1$), then we have the trivial bound

$$\text{diam}(\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_q, S) \leq |\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_q| \leq C \log^A(\text{SL}_d(\hat{\mathbb{Z}})/\Gamma_q)$$

for some suitably large constant $C$.

The other situation that may happen is that $q$ contains at most $M$ prime factors counting multiplicities. In this case we can easily deduce the poly-log diameter bound from \cite[Theorem 7.1]{K2} and \cite[Theorem 2]{K3} which contain this result in the case when $q$ is prime. This deduction is very similar to \cite[Proof of Proposition 3]{K4}.

Let $q_0 = 1, q_1, q_2, \ldots, q_n = q$ be a sequence of integers such that $q_{i+1}/q_i$ is a prime number for all $i$. We will apply the following lemma repeatedly to prove the diameter bound we are looking for.
LEMMA 9. Fix $i$ and write $p = q_i/q_{i-1}$. Then

$$\text{diam}(\text{SL}_d(\mathbb{Z}/q_i\mathbb{Z}), S) \leq C(\text{diam}(\text{SL}_d(\mathbb{Z}/q_{i-1}\mathbb{Z}), S) + \text{diam}(\text{SL}_d(\mathbb{Z}/p\mathbb{Z}), S)),$$

where $C$ is a number depending on $d$.

Proof. Let

$$D = \max\{\text{diam}(\text{SL}_d(\mathbb{Z}/q_{i-1}\mathbb{Z}), S), \text{diam}(\text{SL}_d(\mathbb{Z}/p\mathbb{Z}), S)\}.$$

Then

$$S^D \cdot \Gamma_{q_{i-1}} = \text{SL}_d(\mathbb{Z}/q_i\mathbb{Z}).$$

Since $S$ is generating, $S^{D+1}$ must intersect some $\Gamma_{q_{i-1}}$-coset in at least two points. Thus there is

$$1 \neq g_0 \in S^{2D+1} \cap \Gamma_{q_{i-1}}.$$

If $p \nmid q_{i-1}$, and hence $\Gamma_{q_{i-1}}/\Gamma_{q_i} = \text{SL}_d(\mathbb{Z}/p\mathbb{Z})$, we also want to show that $g_0$ can be taken non-central in $\text{SL}_d(\mathbb{Z}/p\mathbb{Z})$. With the same argument as above, we can show that $S^{(j+1)D+j}$ intersects all $\Gamma_{q_{i-1}}$-cosets in at least $j + 1$ points.

Taking $j = |Z(\text{SL}_d(\mathbb{Z}/p\mathbb{Z}))|$, we can find a suitable $g_0$ in $S^{(j+1)D+j}$.

We put

$$X = \{g^{-1}g_0g, g^{-1}g_0^{-1}g : g \in S^D\}.$$

We show that

$$X^C = \Gamma_{q_{i-1}}/\Gamma_{q_i}$$

for some constant $C$ depending on $d$.

We have two cases. First, we suppose that $p \nmid q_{i-1}$. Then $\Gamma_{q_{i-1}}/\Gamma_{q_i} = \text{SL}_d(\mathbb{Z}/p\mathbb{Z})$, and $X$ is a non-trivial conjugacy class. In this case (22) is a result of Lev [27, Theorem 2].

Now suppose that $p|q_{i-1}$. In this case $\Gamma_{q_{i-1}}/\Gamma_{q_i}$ is isomorphic to $\text{sl}_d(\mathbb{Z}/p\mathbb{Z})$, and the conjugation action

$$h \mapsto g^{-1}hg, \quad g \in \text{SL}_d(\mathbb{Z}/q_i\mathbb{Z}), \quad h \in \Gamma_{q_{i-1}}/\Gamma_{q_i}$$

factors through $G/\Gamma_p = \text{SL}_d(\mathbb{Z}/p\mathbb{Z})$. Now the claim (22) follows from [9, Lemma 5].

Therefore, we have $S^C \cdot D \supset \Gamma_{q_{i-1}}/\Gamma_{q_i}$ for some other constant $C$ depending on $d$. This implies $S^{(C+1)D} = \text{SL}_d(\mathbb{Z}/q_i\mathbb{Z})$ which was to be proved.

Now Corollary 4 is immediate. By [11, Theorem 7.1] and [30, Theorem 2] we have

$$\text{diam}(\text{SL}_d(\mathbb{Z}/p\mathbb{Z}), S) \leq C \log^A (p)$$

for all primes $p$. We can use Lemma 9 repeatedly to show

$$\text{diam}(\text{SL}_d(\mathbb{Z})/\Gamma_q, S) \leq C \log^A (p)$$

for $\Gamma_q \in \Omega_2$, where $p$ is the largest prime factor of $q$ and $C$ is a different constant depending on $M$. This is precisely the poly-log diameter estimate we were looking for.
3 Compact Lie groups

The purpose of this section is to prove Theorem 6. Recall the definitions of $T, LT, LT^*, I, I^*, R, R_+, \mathcal{K}$ and $\pi_v$ from Section 1. Let $\mu$ be a probability measure on $G$. Denote by $\chi_v$ the character of $\pi_v$. For a continuous function $f$ and a measure $\nu$ on $G$, we write

$$f(\nu) = \int f(x) \, d\nu(x).$$

If $f$ is a continuous class function on $G$, then by the Peter-Weyl theorem, we can decompose it as a linear combination of irreducible characters. Denote by $m_v(f)$ the coefficient of $\chi_v$ in this decomposition.

We introduce two partial orders on the space of continuous class functions on $G$. We write $f_1 \leq f_2$, if $f_1(g) \leq f_2(g)$ for every $g \in G$. We write $f_1 \sqsubseteq f_2$ if $m_v(f_1) \leq m_v(f_2)$ for all $v \in \mathcal{K} \cap I^*$. Denote by $\preceq$ the transitive closure of the union of $\leq$ and $\sqsubseteq$, i.e. we write $f_1 \preceq f_2$ if there is a sequence of class functions $\varphi_i$ such that

$$f_1 \preceq \varphi_1 \sqsubseteq \varphi_2 \sqsubseteq \ldots \sqsubseteq \varphi_n \sqsubseteq f_2.$$

These relations have a crucial property contained in the following Lemma. Recall that for a measure $\nu$ on $G$, we define the measure $\tilde{\nu}$ by

$$\int f(x) \, d\nu(x) = \int f(x^{-1}) \, d\nu(x)$$

for all continuous functions $f$. We say that $\nu$ is symmetric if $\nu = \tilde{\nu}$.

**Lemma 10.** Let $\nu$ be a symmetric probability measure and let $f_1 \preceq f_2$ be two continuous class functions on $G$. Then

$$f_1(\nu \ast \nu) \leq f_2(\nu \ast \nu).$$

**Proof.** Clearly, it is enough to prove the statements for $\leq$ and $\sqsubseteq$ in place of $\preceq$. For $\leq$ it easily follows from the definitions and from the fact that $\nu \ast \nu$ is a positive measure.

Suppose $f_1 \sqsubseteq f_2$. Observe that

$$f_i(\nu \ast \nu) = \sum_{v \in \mathcal{K} \cap I^*} m_v(f_i) \chi_v(\nu \ast \nu).$$

Hence the claim follows from $m_v(f_1) \leq m_v(f_2)$ once we prove that $\chi_v(\nu \ast \nu) \geq 0$ for all $v$. This follows from $\chi_v(\nu \ast \nu) = \text{Tr}(\pi_v(\nu \ast \nu))$ and from the fact that $\pi_v(\nu \ast \nu) = \pi_v(\nu) \cdot \pi_v(\nu)^*$ is a positive selfadjoint operator.

Now we explain the strategy of the proof. First of all, we note that by the argument at the beginning of Section 2.1 we can assume that $\mu$ is symmetric. Hence Lemma 10 applies for $\nu = \mu^{*2l}$ for all positive integers $l$. 

**Documenta Mathematica 18 (2013) 1137–1175**
We write for $r \geq 1$
\[
\chi_r = \left( \sum_{|\nu| \leq r} \chi_{\nu} \right)^2 \cdot r^{-\dim LT}
\]
which plays the role of $\chi_{G/T}$ used in the previous section. We also write
\[
l_r = 2[C_r \log^{A+1} r],
\]
where
\[
C_r = C_0 \left(10 - \log^{-1/10}(r)\right)
\]
and $C_0$ is a suitably large constant to be set later.

Our goal is to prove the inequality
\[
\chi_r \left(\mu^*(l_{r_2})\right) \leq E
\]
for some constant $E$ independent of $r$. This will easily imply the theorem.

We assume that (23) holds for some range $1 \leq r \leq r_1$. (This can be verified easily for $r_1 = r_0$ if $C_0$ is suitably large in terms of $\text{gap}_{r_0}(\mu)$.) And then we show that (23) also holds for a suitable $r = r_2$. Iterating this argument, we can prove the claim for all $r$.

We prove the “induction step” in the following way. We find suitable functions $\varphi_1, \ldots, \varphi_n$ such that
\[
\chi_{r_2} \subseteq \varphi_1 + \cdots + \varphi_n.
\]
Then it will be enough to estimate $\varphi_i(\mu^*(l_{r_2}))$. We will show that $\varphi_i \leq B_i \chi_{r_1}$, where $B_i$ is a number depending on $i, r_1$ and $r_2$. This allows us to estimate $\varphi_i(\mu^*(l_{r_1}))$. We will also show that $m_v(\varphi_i)$ is either “large” or 0, and this will yield an estimate on the eigenvalues of $\pi_v(\mu^*(l_{r_1}))$ for all $v$ which contributes to $\varphi_i$. Finally this allows us to get a refined estimate on $\varphi_i(\mu^*(l_{r_2}))$.

To implement the above plan we need methods to estimate $m_v(f)$. In our examples $f$ will always be the character of a representation which is obtained from other representations using tensor products. Explicit formulas are available for $m_v(f)$ in such cases, however, they do not seem very practical for our purposes.

Instead, we will use elementary methods to estimate these coefficients based on double-counting dimensions. However, we still need some very basic facts about the representations $\pi_v$.

The first fact is Weyl’s dimension formula [12, Chapter VI. (1.7) (iv)]:
\[
\dim \pi_v = \prod_{u \in \mathfrak{h}^+} \frac{(u, v + \rho)}{(u, \rho)}
\]
where
\[
\rho = \frac{1}{2} \sum_{u \in \mathfrak{h}_+} u
\]
is the half sum of the positive roots.
The second fact is the content of the following lemma which bounds the highest weight of possible irreducible constituents of $\pi_v \otimes \pi_u$. Recall that we denote the Haar measure on $G$ by $m_G$.

**Lemma 11.** There is a constant $D$ depending only on $G$ such that if

$$\int \chi_v \chi_u \overline{\chi_w} \, dm_G \neq 0$$

for some $v, u, w \in \overline{K} \cap I^*$, then $|v - w| < D|u|$.

**Proof.** If (24) holds then $\pi_w$ is contained in $\pi_v \otimes \pi_u$. By [12, Chapter VI, Lemma (2.8)], $v + u$ dominates $w$ that is

$$\langle v + u, t \rangle \geq \langle w, t \rangle$$

for every $t \in \overline{K}$. Similarly, if (24) holds then $\pi_v$ is contained in $\pi_w \otimes \overline{\chi_u}$, hence

$$\langle w + \overline{\chi_u}, t \rangle \geq \langle v, t \rangle.$$ 

Here $\overline{\chi_u}$ is the highest weight of $\chi_u$.

Now let $t_1, \ldots, t_{\dim T}$ be a basis of $LT^*$ consisting of unit vectors in $\overline{K}$. By the above inequalities, we have

$$|\langle w - v, t_j \rangle| \leq |u|$$

for all $1 \leq j \leq \dim T$. The claim follows from this with the constant $D$ equal to the length of the longest vector in the set

$$\{x \in LT^* : |\langle x, t_j \rangle| \leq 1\}.$$

\[\square\]

We proceed by some Lemmata which bound the multiplicities of some irreducible constituents in certain tensor products.

**Lemma 12.** We have

$$\int \left( \sum_{|v| \leq r} \chi_v \right)^2 \, dm_G = |\{v \in \overline{K} \cap I^* : |v| \leq r\}|.$$

In particular

$$c \leq m_{\chi_0}(\chi_r) \leq C,$$

where $c, C > 0$ are constants depending on $G$. 

Documenta Mathematica 18 (2013) 1137–1175
Proof. Since characters form an orthonormal basis, we have
\[
\int \left( \sum_{|v| \leq r} \chi_v \right)^2 \chi_0 \, dm_G = \int \sum_{|v| \leq r} \chi_v \chi_0 \, dm_G = \{ v \in \mathcal{K} \cap I^* : |v| \leq r \}.
\]
For the second statement notice that
\[
m_{\chi_0}(\chi_r) = \frac{1}{r \dim LT} \int \left( \sum_{|v| \leq r} \chi_v \right)^2 \chi_0 \, dm_G = \frac{\{ v \in \mathcal{K} \cap I^* : |v| \leq r \}}{r \dim LT}.
\]

**Lemma 13.** There is a constant $C > 0$ depending on $G$ such that the following holds. Let $u \in \mathcal{K} \cap I^*$, and $r \geq 1$.

\[
m_{\chi_u}(\chi_r) \leq C \dim \chi_u.
\]
Moreover, $m_{\chi_u}(\chi_r) = 0$ if $|u| > Cr$.

Proof. For the first part of the lemma we write
\[
r \dim LT \, m_{\chi_u}(\chi_r) = \int \left( \sum_{|v| \leq r} \chi_v \right)^2 \chi_u \, dm_G \leq 2 \sum_{|v|=|w| \leq r, \dim \chi_v \leq \dim \chi_w} \chi_v \chi_w \chi_u \, dm_G.
\]
The sum of the multiplicities of some irreducible components of $\pi_v \otimes \pi_u$ can not be bigger than the dimension of $\pi_v \otimes \pi_u$ divided by the minimal dimension of the irreducible components we consider. Thus
\[
r \dim LT \, m_{\chi_u}(\chi_r) \leq 2 \sum_{|v| \leq r} \frac{\dim(\chi_v \chi_u)}{\dim(\chi_v)} \leq C r \dim LT \, \dim \chi_u.
\]
The second part follows immediately from Lemma 11. □

For $u \in \mathcal{K} \cap I^*$ and $r \geq 1$, we write
\[
\psi_{u,r} = \chi_r \cdot \sum_{w : |u-w| \leq 3Dr} \chi_w.
\]
We show that $\chi_z$ is contained in $\psi_{u,r}$ with high multiplicity if $|u - z| \leq Dr$. Documenta Mathematica 18 (2013) 1137–1175
Lemma 14. Let $r \geq 1$ and $z, u \in \overline{K} \cap I^*$ with $|z - u| \leq Dr$. Then
\[ m_{\chi_z}(\psi_{u,r}) \geq c \cdot \frac{\dim(\chi_z)}{\max_w |u - w| < D r \dim G} \cdot r^\dim G, \]
where $c > 0$ is a constant depending only on $G$.

Proof. We can write
\[
 r^\dim LT m_{\chi_z}(\psi_{u,r}) = \sum_{w: |u - w| < 3 Dr} \left( \sum_{|v_1|,|v_2| \leq r} \int \chi_{v_1} \chi_{v_2} \chi_w \overline{\chi_z} \ dm_G \right) 
\]
\[
 = \sum_{w: |u - w| < 3 Dr} m_{\chi_w} \left( \sum_{|v_1|,|v_2| \leq r} \chi_{v_1} \chi_{v_2} \chi_z \right) 
\]
\[
 \geq \sum_{w: |u - w| < 3 Dr} \dim \chi_{v_1} \cdot \dim \chi_{v_2} \cdot \dim \chi_w. \tag{25}
\]
In the last line we used the fact that all irreducible components of $\chi_{v_1} \chi_{v_2} \chi_z$ has highest weight $w$ satisfying $|w - z| \leq 2Dr$ and hence $|u - w| < 3Dr$ which follows from two applications of Lemma 11. Hence all possible irreducible component appears in the range of summation, and the inequality follows by comparing dimensions.

Let $K'$ be a closed convex cone strictly contained in the positive Weyl chamber $K$. For $v \in K' \cap I^*$ and $|v| \geq r/2$ it follows from Weyl’s dimension formula that $\dim(\chi_v) \geq c r^{1|R_+|}$ for some constant $c > 0$ depending only on $G, K'$. Thus the numerator in (25) is bounded below by
\[
 c \dim(\chi_z) \cdot r^2 \dim LT + 2 |R_+|.
\]
The proof is finished by noting that $\dim G = \dim LT + 2 |R_+|$. \hfill \qed

We continue implementing our plan described above. Recall that $r_2 > r_1 \geq 1$, are numbers that we will choose later and that we assume
\[
 \chi_r \left( \mu^{(l_r)} \right) \leq E \tag{26}
\]
for $r \leq r_1$ and for some number $E$ to be chosen later, as well. Our goal is to prove (26) with $r = r_2$.

Let $u_0 = 0, u_1, \ldots, u_m$ be a maximal $Dr_1$ separated subset of $\{ v \in \overline{K} \cap I^* : |v| \leq Cr_2 \}$, where $C$ is the constant from Lemma 13. For $i = 0, \ldots, m$, let
\[
 M_i = \frac{C \max_w |u_i - w| < 3 D r_1 \{ \dim \chi_w \}}{c r_1^\dim G},
\]
where $c$ is the constant from Lemma 14 and $C$ is as above. Write
\[
 \varphi_i = M_i \sum_{w: |u_i - w| < D r_1} m_{\chi_w} (\psi_{u_i,r_1}) \chi_w.
\]
i.e. we removed from $\psi_{u_i, r_1}$ those irreducible components whose multiplicities we cannot bound below. Hence $m_{\chi_w}(\varphi_i) \geq C \cdot \dim(\chi_w)$ if $|w - u_i| < D r_1$ by Lemma 14 (applied with $r = r_1$).

It follows from Lemma 13 (applied with $r = r_2$) that

$$\chi_{r_2} \equiv \sum_{i=0}^{m} \varphi_i.$$  

Moreover, we have

$$\varphi_i \leq M_i \psi_{u_i, r_1} \leq \left[ M_i \sum_{w:|u_i - w| < 3D r_1} \dim(\chi_w) \right] \chi_{r_1},$$

because $\chi_{r_1}$ is non-negative, and

$$\sum_{w:|u_i - w| < 3D r_1} \chi_w(g) \leq \sum_{w:|u_i - w| < 3D r_1} \dim \chi_w$$

for every $g \in G$. Clearly

$$\sum_{w:|u_i - w| < 3D r_1} \dim \chi_w \leq C r_1^{\dim L T} \max_{w:|u_i - w| < 3D r_1} \{ \dim \chi_w \},$$

hence

$$\varphi_i \leq C \frac{(\max_{w:|u_i - w| < 3D r_1} (\dim \chi_w))^2}{r_1^{2|R_+|}} \cdot \chi_{r_1}. \quad (27)$$

Denote by $N_i$ the number of positive roots $v \in R_+$ such that $\langle u_i, v \rangle \leq 4 D r_1 |v|$. Then it follows from Weyl’s dimension formula that

$$\max_{w:|u_i - w| < 3D r_1} \dim \chi_w \leq C r_1^{N_i} \min_{w:|u_i - w| < 3D r_1} \dim \chi_w. \quad (28)$$

After these preparations, we can give an estimate on $\varphi_i (\mu^{*(l_r)})$. This is done in the next two Lemmata.

**Lemma 15.** Fix $0 \leq i \leq m$ such that $N_i < |R_+|$. If $r_1$ is sufficiently large and

$$\log^{1/3} r_1 \leq \log r_2 - \log r_1 \leq \log^{1/2} r_1,$$

then

$$\varphi_i \left( \mu^{*(l_r)} \right) \leq \left( \frac{r_1}{r_2} \right)^{(A-1)(|R_+|-N_i)}.$$
Proof. For notational simplicity, write
\[ X = \max_{w : |u_i - w| < 3Dr_1} \dim \chi_w, \]
and note that by Weyl's dimension formula, we have
\[ X \leq Cr_1N_i^{\frac{|R_+| - N_i}}{r_2^{2|R_+| - 2N_i}}. \tag{29} \]
By (27), the induction hypothesis and Lemma 10, we have
\[ \varphi_i \left( \mu^{* \ell_{r_1}} \right) \leq CE \frac{X^2}{r_1^{2|R_+|}}. \tag{30} \]
Denote by \( \lambda_{\max} \) the maximum of the absolute values of the eigenvalues of the operators \( \pi_v(\mu) \) for \( |v - u_i| \leq Dr_1 \), i.e. for the irreducible characters contained in \( \varphi_i \). Clearly
\[ \lambda_{\max}^{\ell_{r_1}} \leq CE \frac{X^2}{r_1^{2|R_+|}} \frac{1}{m_{\chi_v}(\varphi_i)}, \tag{31} \]
where \( \chi_v \) is the character of the representation, which contains \( \lambda_{\max} \). Recall that by Lemma 14 and the definition of \( \varphi_i \), we have
\[ m_{\chi_v}(\varphi_i) \geq c \dim \chi_v. \]
If we combine this with (28) and (31), we get
\[ \lambda_{\max}^{\ell_{r_1}} \leq CE \frac{X^2}{r_1^{2|R_+| - 2N_i}} \leq CE \frac{r_2^{2|R_+| - 2N_i}}{r_1^{2|R_+| - 2N_i}}. \tag{32} \]
(For the second inequality we used (29).)
By (30), we clearly have
\[ \varphi_i \left( \mu^{* \ell_{r_2}} \right) \leq CE \frac{X^2}{r_1^{2|R_+|}} \lambda_{\max}^{\ell_{r_2} - \ell_{r_1}} \leq CE \frac{r_2^{2|R_+| - 2N_i}}{r_1^{2|R_+| - 2N_i}} \left( CE \frac{r_2^{2|R_+| - 2N_i}}{r_1^{2|R_+| - 2N_i}} \right)^{\ell_{r_2} - \ell_{r_1}}. \tag{33} \]
By a computation very similar to (12), we get
\[ \frac{l_{r_2} - l_{r_1}}{l_{r_1}} \geq \left( A + 1 + \frac{c}{\log^{1/10} r_1} \right) \frac{\log r_2 - \log r_1}{\log r_1}, \tag{34} \]
where \( c \) is an absolute constant.
Now we write
\[ (\log r_2 - 2 \log r_1) \frac{\log r_2 - \log r_1}{\log r_1} = (\log r_1 - \log r_2) \left( 1 - \frac{\log r_2 - \log r_1}{\log r_1} \right). \tag{35} \]
\begin{equation}
\left( A + 1 + \frac{c}{\log^{1/10} r_1} \right) \left( 1 - \frac{\log r_2 - \log r_1}{\log r_1} \right) \geq A + 1 + \frac{c}{2 \log^{1/10} r_1}
\end{equation}

which follows from our assumption \( \log r_2 - \log r_1 \leq \log^{1/2} r_1 \).

Combining (34), (35) and (36) we get
\[
\left( \frac{r_2}{r_1} \right)^{\frac{1}{r_1} - \frac{1}{r_2}} \leq \left( \frac{r_1}{r_2} \right)^{(A+1)(|R_+| - N_i) + c \log^{-1/10} r_1}.
\]

If we plug this into (33), we get
\[
\phi_i \left( \mu^\star \left( \frac{l_{r_2}}{l_{r_1}} \right) \right) \leq \left( \frac{r_1}{r_2} \right)^{(A-1)(|R_+| - N_i) + c \log^{-1/10} r_1} (CE)^{l_{r_2}/l_{r_1}}.
\]

Finally we note that by \( \log r_2 - \log r_1 \geq \log^{1/3} r_1 \), we have
\[
\left( \frac{r_2}{r_1} \right)^{c / \log^{1/10} r_1} \geq e^{c \log^{7/30} r_1} \geq (CE)^2
\]
if \( r_1 \) is sufficiently large depending on \( G \) and \( E \). On the other hand, by the computation leading to (34), we have \( l_{r_2}/l_{r_1} < 2 \). This finishes the proof.

**Lemma 16.** Fix \( 0 \leq i \leq m \) such that \( N_i = |R_+| \). If \( \log r_2 - \log r_1 > \log^{1/3} r_1 \) and \( r_1 \) is sufficiently large then
\[
\phi_i (\mu^\star \left( \frac{l_{r_2}}{l_{r_1}} \right)) < C,
\]
where \( C \) is a constant depending only on \( G \) (and not on \( E \)).

**Proof.** Let \( v \in K \cap I^* \), such that \( m_{\chi_v} (\phi_i) > 0 \). By Weyl’s formula, we have \( \dim(\chi_v) \leq Cr_1^{\frac{|R_+|}{|N_i|}} \), since \( N_i = |R_+| \). Let \( \lambda \) be an eigenvalue of \( \pi_v(\mu) \). Then by (31) we get
\[
\lambda^{l_{r_1}} \leq \frac{CE}{\dim \chi_v}.
\]
In fact, we can get a better bound if \( |v| \leq r_1 \). By a similar argument and using the induction hypothesis for \( r = |v| \leq r_1 \), we get
\[
\lambda^{l_{|v|}} \leq \frac{CE}{\dim \chi_v}.
\]
Weyl’s dimension formula gives \( \dim \chi_v \geq c|v| \), and an easy calculation shows that
\[
\lambda^{l_{r_1}} \leq \left( \frac{CE}{|v|} \right)^{l_{r_1}/l_{|v|}} \leq \frac{CE}{r_1}.
\]

Documenta Mathematica 18 (2013) 1137–1175
if $|v|$ is sufficiently large (depending on $G$ and $E$).
We suppose that $r_0$ is so large that $|v| < r_0$ for those $v$ which are too small for the above argument. We set $C_0 > \text{gap}_{r_0}^{-1}(G, \mu)$, hence for $r_0 \geq |v| \neq 0$ we have
$$\lambda_{r_1} \leq e^{-\text{gap}_{r_0}(G, \mu)} \leq e^{-\log^{A+1}(r_1)},$$
which is stronger than (37).
By (27), the induction hypothesis and Lemma 10, we have
$$\varphi_i \left( \mu^* (l_{r_1}) \right) \leq CE,$$
hence
$$\varphi_i \left( \mu^* (l_{r_1}) \right) \leq m_{\chi_0}(\varphi_i) + CE \left( \frac{CE}{r_1} \right)^{l_{r_2}/l_{r_1}}.$$ 
Since $\log r_2 - \log r_1 > \log^{1/3} r_1$ and
$$\frac{l_{r_2} - l_{r_1}}{l_{r_1}} > \frac{\log r_2 - \log r_1 \log r_1}{\log r_1},$$
we easily get
$$\varphi_i \left( \mu^* (l_{r_1}) \right) \leq m_{\chi_0}(\varphi_i) + 1$$
if $r_1$ is sufficiently large (depending on $G$ and $E$) which was to be proved.

It is left to estimate the number of $u_i$ for which $N_i$ takes a particular value. This is done with the help of the next lemma.

Denote by $S$ the set of simple roots. (This is not to be confused with the generating set of the random walk, which is denoted by $S$ in other sections.)

**Lemma 17.** Let $S' \subseteq S$ be a set of simple roots. Denote by $R' \subset R_+$ the set of all positive roots which can be expressed as a combination of elements of $S'$.

We have
$$\frac{|S| - |S'|}{|R_+| - |R'|} \leq A(G) - 1,$$
where the value of $A(G)$ is given in Table 1, for simple Lie groups and for non-simple ones it is defined to be $A(G) = \max \{ A(H) \}$ where $H$ runs through all simple quotients.

**Proof.** If the Dynkin diagram of $G$ is not connected, we can write $R_+ = R_1 \cup \ldots \cup R_n$, where $R_i$ is a system of positive roots in a root system with connected diagram. Clearly
$$\frac{|S| - |S'|}{|R_+| - |R'|} \leq \max_{1 \leq i \leq n} \frac{|S \cap R_i| - |S' \cap R_i|}{|R_i| - |R' \cap R_i|}.$$ 
Hence we can assume without loss of generality that the diagram of $G$ is connected.
By a simple calculation, one can verify that $|S|/|R_+| = A(G) - 1$ as given in Table 1. Now notice that $R'$ is itself (the set of positive roots in) a root system and its diagram is the subgraph spanned by $S'$ in the diagram of $G$. Examining Table 1 it is easy to check that

$$
\frac{|S'|}{|R'|} \geq A(G) - 1 = \frac{|S|}{|R_+|}
$$

(In fact, it is enough to check that the value in the table is never smaller for connected subdiagrams.) Then

$$
\frac{|S| - |S'|}{|R_+| - |R'|} \leq \frac{|S| - |S'||R'|/|R_+|}{|R_+| - |R'|} = \frac{|S|}{|R_+|} = A(G) - 1.
$$

$\Box$

Let $S' \subset S$ be a subset of the simple roots. We write $\Omega(S')$ for the set of indices $i$ such that $\langle u_i, v \rangle \leq 4Dr_1$ for $v \in S$ if and only if $v \in S'$. We estimate $|\Omega(S')|$. Since $S'$ consist of linearly independent vectors, the elements of $\Omega(S')$ are in a $Cr_1$ neighborhood of a subspace of $LT$ of dimension $|S| - |S'|$. Since they are $Dr_1$-separated, we have

$$
|\Omega(S')| \leq C \left( \frac{r_2}{r_1} \right)^{|S| - |S'|}.
$$

(38)

All positive roots are positive linear combinations of simple roots, hence $N_i \leq |R'|$ for $i \in \Omega(S')$, where $R'$ is the set of positive roots which are combinations of the elements of $S'$. If $S' \neq S$, then Lemmata 15 and 17 together with (38) gives

$$
\sum_{i \in \Omega(S')} \varphi_i \left( \mu^{(i_r)} \right) \leq C
$$

with a constant $C$ depending on $G$. If $S' = S$, the same follows from Lemma 16. Summing this up for all $S' \subset S$, we get

$$
\sum_i \varphi_i \left( \mu^{(i_r)} \right) \leq C.
$$

(39)

This completes the proof of (26) for $r = r_2$ with $E = C$, where $C$ is the constant in (39).

We explain how to set the various parameters and how to complete the induction. We set $E = C$ with the constant $C$ from (39) in the previous paragraph. Then we pick $r_0$ to be sufficiently large (depending on $E$ and $G$) so that all of the above arguments are valid with $r_1 \geq r_0$. 
For $r_0 \geq r \geq 1$, we have
\[
\chi_r \left( \mu^*(l_r) \right) \leq m_{\chi_0} (\chi_r) + e^{-l_r \cdot \text{gap}_{C_r}(G,\mu)} \cdot \sum_{0 < |v| \leq C_r} m_{\chi_v} (\chi_r) \\
\leq C + C r^{-C_0 \cdot \text{gap}_{C_r}(G,\mu)} \cdot \sum_{0 < |v| \leq C_r} \dim(\chi_v) \\
\leq C + C r^{C(\dim L + |R|) - C_0 \cdot \text{gap}_{C_r}(G,\mu)} \cdot \sum_{0 < |v| \leq C_r} \dim(\chi_v).
\] (40)

Here we first used the definition of $\text{gap}_{C_r}(G,\mu)$, then the definition of $l_r$ and Lemmata 12 and 13, finally Weyl’s dimension formula.

We put $C_0 = C \cdot \text{gap}_{C_r}(G,\mu)$, where $C$ is a suitable constant depending on the constant in (40) such that
\[
\chi_r \left( \mu^*(l_r) \right) \leq E
\] for $1 \leq r \leq r_0$. (Recall that the only constraint we had for $C_0$ above is in the proof of Lemma 16 and it is satisfied with this choice.)

Thus we see that (26) hold for $1 \leq r \leq r_0$. Once we know that (26) holds on an interval $r \in [1, a]$, we can extend it to $r \in [1, e^{\log(a)} + \log^{1/2}(a)]$. This follows from the above argument with the choice $r_2 = r$ and any $r_1 \leq a$ which satisfies
\[
\log^{1/3} r_1 \leq \log r_2 - \log r_1 \leq \log^{1/2} r_1.
\]

If we apply this repeatedly, we can conclude that (26) holds for all $r \geq 1$.

Finally, we conclude the proof of the theorem. Fix $r \geq 1$ and suppose that $\pi_{v_0}$ is the representation for which the maximum in the definition of $\text{gap}_r(G,\mu)$ is attained. We use Lemma 14 with $z = u = v_0$ and (26) to get
\[
\chi_{v_0} \left( \mu^*(l_r) \right) \leq C \cdot \frac{\max_{w : \|v_0 - w\| < 3 Dr} \{ \dim(\chi_w) \}}{\dim(\chi_{v_0}) \cdot r \dim G} \cdot \psi_{v_0, r} \left( \mu^*(l_r) \right) \\
\leq CE \cdot \frac{\max_{w : \|v_0 - w\| < 3 Dr} \{ \dim(\chi_w) \}}{\dim(\chi_{v_0}) \cdot r \dim G} \cdot \sum_{w : \|v_0 - w\| < 3 Dr} \dim(\chi_w).
\]

We evaluate dimensions using Weyl’s formula and get
\[
\| \pi_{v_0} \left( \mu^*(l_r) \right) \| \leq \chi_{v_0} \left( \mu^*(l_r) \right) \leq CE / \dim(\chi_{v_0}) \leq CE / r \leq r^{-1/2},
\]
if $r \geq r_0$ and $r_0$ is sufficiently large, as we may assume. This implies
\[
\text{gap}_r(G,\mu) \geq (1/10) \frac{\log r}{l_r}.
\]

Inspecting the definition of $l_r$ and the above choice of $C_0$, we see that this is exactly what was to be proved.
4 Some technicalities

We begin this section by proving Corollary 7. First we give a lemma which will be used for reducing the problem to the connected case:

**Lemma 18.** Let $G$ be a Lie group and let $\mu$ be a symmetric probability measure on it such that $\supp \mu$ generates a dense subgroup and $1 \in \supp \mu$. Write $G^o$ for the connected component of $G$ and let $n = [G : G^o]$. Then $\supp (\mu^{(2n-1)}) \cap G^o$ generates a dense subgroup of $G^o$.

**Proof.** Suppose to the contrary that for some $h \in G^o$ and $\varepsilon > 0$ there is no $h'$ in the group generated by $\supp (\mu^{(2n-1)}) \cap G^o$ such that $\dist(h, h') < \varepsilon$. On the other hand, by assumption, there is $g = g_1 \cdots g_l$ with $g_i \in \supp (\mu)$ and $\dist(h, g) < \varepsilon$. We show that $g$ is in the group generated by $\supp (\mu^{(2n-1)}) \cap G^o$, a contradiction. If $l \leq 2n - 1$, then by definition, $g \in \supp (\mu^{(2n-1)}) \cap G^o$. Suppose $l > 2n - 1$. By the pigeon hole principle, there are $i \leq j \leq n$ such that $g_i \cdots g_j \in G^o$. We can write $g = g' \cdot g''$, where
\[
    g' = g_1 \cdots g_{i-1} g_i \cdot g_j \cdots g_l^{-1} \quad \text{and} \quad g'' = g_1 \cdots g_{i-1} g_{j+1} \cdots g_l.
\]
Now clearly $g', g'' \in G^o$, and $g' \not\in \supp (\mu^{(2n-1)})$, since it is a product of length at most $2n - 1$. Since the length of $g''$ is strictly less than that of $g$ the proof can be completed by induction.

If $G$ is a compact connected semi-simple Lie group, the space $L^2(G)$ can be decomposed as an orthogonal sum of finite dimensional irreducible representations. We write $\mathcal{H}_r \subset L^2(G)$ for the sum of those constituents which have highest weight $\nu$ with $|\nu| \leq r$. In the next lemma we construct an approximate identity in $\mathcal{H}_r$.

**Lemma 19.** Let $G$ be a compact connected semi-simple Lie group. Then for each $r$, there is a non-negative function $f_r \in \mathcal{H}_r$ such that
\[
    \int f_r \, dm_G = 1, \quad \|f_r\|_2 \leq C r^{\dim G/4} \quad \text{and} \quad \int f_r(g) \dist(g, 1) \, dm_G(g) \leq C/\sqrt{r},
\]
where $C$ is a constant depending on $G$.

**Proof.** We fix a maximal torus $T \subset G$. Let $\pi$ be a faithful (not necessarily irreducible) finite dimensional unitary representation of $G$ with real character $\chi$. We can decompose the representation space as the sum of weight spaces, i.e. there is an orthonormal basis $\varphi_1, \ldots, \varphi_m$ (where $m = \dim \pi$) such that the following holds. For each $\varphi_i$, there is a weight $u_j \in \mathfrak{t}^*$ such that for the elements $g \in T$ we have $\pi(g) \varphi_j = e^{2\pi i (\log g, u_j)} \varphi_j$. (Here $\log : T \to LT$ is a branch of the inverse of the exponential map.) Since $\chi$ is real, we have
\[
    \chi(g) = \sum_{j=1}^m e^{2\pi i (\log g, u_j)} = \sum_{j=1}^m \cos(2\pi (\log g, u_j)).
\]

Documenta Mathematica 18 (2013) 1137–1175
Since $\pi$ is faithful, $\chi(g) < m$ for $g \neq 1$ and we can deduce from the above formula that
\[ m - c_1 \text{dist}(g, 1)^2 \leq \chi(g) \leq m - c_2 \text{dist}(g, 1)^2, \]
where $c_1, c_2 > 0$ are constants depending on $G$.

Denote by $r_0$ the length of the highest weight of the irreducible components of $\pi$. We define
\[ f_r(g) = c_r(\chi(g) + m)^{\lfloor r/r_0 \rfloor}, \]
where $c_r$ is a normalizing constant so that $\int f_r \, dm_G = 1$. Observe that $f_r \in \mathcal{H}_r$.

By simple calculation based on (41), we get
\[ c \frac{(r/r_0)^{\dim G/2}}{(2m)^{\lfloor r/r_0 \rfloor}} \leq c_r \leq C \frac{(r/r_0)^{\dim G/2}}{(2m)^{\lfloor r/r_0 \rfloor}}, \]
where $c, C > 0$ are numbers depending on $G$.

We have
\[ \|f_r\|_\infty = c_r(2m)^{\lfloor r/r_0 \rfloor} \leq C(r)^{\dim G/2}. \]

The $L^2$ bound now follows from $\|f_r\|_2^2 \leq \|f_r\|_1 \|f_r\|_\infty$.

Now using again (41), we get
\[
\int f_r(g) \text{dist}(1, g) \, dm_G(g)
\leq c_r \int (2m - c_2 \text{dist}(g, 1)^2)^{\lfloor r/r_0 \rfloor} \text{dist}(1, g) \, dm_G(g)
\leq C \frac{(r/r_0)^{\dim G/2}}{(2m)^{\lfloor r/r_0 \rfloor}} \int e^{-c_3 \text{dist}(g, 1)^2} \text{dist}(1, g) \, dm_G(g)
\leq C \frac{(r/r_0)^{\dim G/2}}{(2m)^{\lfloor r/r_0 \rfloor}} \int e^{-c_3 |g|^2} |x|^{\lfloor r/r_0 \rfloor} |x| \, dx
= C \frac{(r/r_0)^{-1/2}}{(2m)^{\lfloor r/r_0 \rfloor}} \int e^{-c_3 |y|^2} |y| \, dy \leq C/\sqrt{r},
\]
which was to be proved. \[\square\]

**Proof of Corollary 7.** Write $\text{Reg}(g)f(h) = f(g^{-1}h)$ for $f \in L^2(G)$, which is the left regular representation of $G$.

Assume to the contrary that $f \in \text{Lip}(G)$, $\int f = 0$, $\|f\|_2 = 1$ and yet
\[
\left\| \int \text{Reg}(g)f \, d\mu(g) \right\|_2 = \|\text{Reg}(\mu)f\|_2 \geq 1 - c_0 \log^{-A}(\|f\|_{\text{Lip}} + 2)
\]
with a constant $c_0$ which will be chosen to be sufficiently small depending on $\mu$.

By the same argument as in the beginning of Section 2, we have $\|\text{Reg}(\mu)f\|_2^2 = \|\text{Reg}(\hat{\mu} \ast \mu)f\|_2$. Thus (42) holds (with a different $c_0$) for $\mu$ replaced by $\hat{\mu} \ast \mu$, hence we can assume that $\mu$ is symmetric and $1 \in \text{supp} \mu$. 

Documenta Mathematica 18 (2013) 1137–1175
Write $n = [G : G^o]$ as in Lemma 18. Furthermore, we define $\mu_1$ to be the probability measure on $G^o$ that we obtain from $\mu^{(2n-1)}$ by restricting it to $G^o$ and then normalizing it. By Lemma 18, we can apply Theorem 6 for the measure $\mu_1$. Now we need to find a suitable test function related to $f$.

First, we want to rule out the possibility that $f$ is “almost constant” on cosets of $G^o$. It follows from (42) that there is a set $X \subset G$ with $\mu(X) > 1 - \varepsilon$ such that

$$\|f - \text{Reg}(g)f\|_2 < 1/(10n)$$

(43)

where $\varepsilon > 0$ is as small as we wish, if we choose $c_0$ in (42) sufficiently small. (In fact, we could obtain a much stronger estimate from (42)). In particular, we can ensure that $X \cdot G^o$ generates $G/G^o$. Since $\int f \, dm_G = 0,$

$$\int \langle \text{Reg}(g)f, f \rangle \, dm_G(g) = 0.$$ 

Hence there is $g \in G$ such that $\langle \text{Reg}(g)f, f \rangle \leq 0$, in particular $\|\text{Reg}(g)f - f\|_2 \geq \sqrt{2}$. We can write $g = h_1 \cdots h_n g_0$, where $h_i \in X$ and $g_0 \in G^o$. By the triangle inequality, either

$$\|\text{Reg}(h_1 \cdots h_n g_0)f - \text{Reg}(h_1 \cdots h_n)f\|_2 > 1$$

(44)

or

$$\|\text{Reg}(h_1 \cdots h_j)f - \text{Reg}(h_1 \cdots h_{j-1})f\|_2 > (\sqrt{2} - 1)/n$$

(45)

for some $1 \leq j \leq n$. Since Reg is unitary, the second case yields $\|\text{Reg}(h_j)f - f\|_2 > (\sqrt{2} - 1)/n$ which is a contradiction to (43). Thus only the first case is possible, from which we conclude $\|\text{Reg}(g_0)f - f\|_2 > 1$ again by unitarity.

This shows that $f$ can not be “almost constant” on all cosets of $G^o$. Let $x_1, \ldots, x_n \in G$ be a system of representatives for $G^o$-cosets. Write $f_i$ for the restriction of $f$ to the coset $G^o \cdot x_i$ considered a function on $G^o$. More formally:

$$f_i(g) = f(g x_i) \in L^2(G^o).$$

We have

$$\|\text{Reg}(g_0)f - f\|_2^2 = \sum_{i=1}^n \|\text{Reg}(g_0)f_i - f_i\|_2^2,$$

hence there is $1 \leq i_0 \leq n$ such that $\|\text{Reg}(g_0)f_{i_0} - f_{i_0}\|_2 \geq 1/\sqrt{n}$.

We define

$$\varphi = \frac{f_{i_0} - \int f_{i_0} \, dm_G}{\|f_{i_0} - \int f_{i_0} \, dm_G\|_2}$$

aiming to estimate $\|\text{Reg}(\mu_1)\varphi\|_2$ using Theorem 6.

It follows from the above considerations that $\|f_{i_0} - \int f_{i_0} \, dm_G\|_2 \geq 1/(2\sqrt{n})$. Thus $\|\varphi\|_{\text{Lip}} < 2\sqrt{n}/\|f\|_{\text{Lip}}$.

From Lemma 18, we know that supp $\mu_1$ is not contained in a proper closed subgroup of $G^o$. Thus we have

$$\text{gap}_r(G^o, \mu_1) > 0$$
for all $r$. Then Theorem 6 implies that
\[ \text{gap}_r(G^\circ, \mu_1) > c \log^{-A(G)} r \]
with a constant $c > 0$ depending on $\mu$.
To apply this spectral gap estimate, we need to approximate $\varphi$ by a function in $\mathcal{H}_r$ with small $r$. We use Lemma 19 with $r = D(\|\varphi\|_{\text{Lip}} + 2)^4$; we will chose the sufficiently large number $D$ later. Then the Lemma gives:
\[ \|f_r \ast \varphi - \varphi\|_{\infty} \leq \int f_r(g) \text{dist}(g, 1) \|\varphi\|_{\text{Lip}} dm_G(g) \leq \frac{C}{D^{1/2}(\|\varphi\|_{\text{Lip}} + 2)}. \]
Clearly $f_r \ast \varphi \in \mathcal{H}_r$, and moreover
\[ \int f_r \ast \varphi dm_G = 0. \]
Thus
\[ \|\text{Reg}(\mu_1) \varphi\|_2 \leq \|\text{Reg}(\mu_1)(f_r \ast \varphi)\|_2 + \frac{C}{D^{1/2}(\|\varphi\|_{\text{Lip}} + 2)} \]
\[ \leq 1 - \text{gap}_r(G^\circ, \mu_1) + \frac{C}{D^{1/2}(\|\varphi\|_{\text{Lip}} + 2)} \]
\[ \leq 1 - c \log^{-A(G)}(D^{1/2}(\|\varphi\|_{\text{Lip}} + 2)) + \frac{C}{D^{1/2}(\|\varphi\|_{\text{Lip}} + 2)}. \]
Now we choose $D$ sufficiently large depending on $\mu$ so that for the quantities in the last line we have
\[ c \log^{-A(G)}(D^{1/2}(\|\varphi\|_{\text{Lip}} + 2)) \geq \frac{2C}{D^{1/2}(\|\varphi\|_{\text{Lip}} + 2)}. \]
This in turn implies with a different constant $c'$:
\[ \|\text{Reg}(\mu_1) \varphi\|_2 \leq 1 - c \log^{-A(G)}(\|\varphi\|_{\text{Lip}} + 2). \]
Furthermore, by the definition of $\varphi$, we have
\[ \|\text{Reg}(\mu_1) f_\alpha\|_2 \leq \|f_\alpha\|_2 - c \log^{-A(G)}(\|f\|_{\text{Lip}} + 2). \]
This yields
\[ \|\text{Reg}(\mu_1) f\|_2 \leq 1 - c \log^{-A(G)}(\|f\|_{\text{Lip}} + 2). \]
By selfadjointness
\[ \|\text{Reg}(\mu) f\|^{2n-1}_2 \leq \|\text{Reg}(\mu^{(2n-1)}) f\|_2 \leq 1 - c'(1 - \|\text{Reg}(\mu_1) f\|_2), \]
where $c' = \mu^{(2n-1)}(G^n)$ is the normalizing constant in the definition of $\mu_1$.
This is a contradiction to (42), if we choose there $c_0$ to be sufficiently small. \[ \Box \]
The rest of the section is devoted to the proof of Lemma 5. We start with a Lemma which provides an estimate on the Lipschitz norm of a function in \( H_r \).

**Lemma 20.** Let \( G \) be a connected compact semi-simple Lie group and let \( f \in H_r \) with \( \|f\|_2 = 1 \). Then \( \|f\|_{\text{Lip}} \leq C r^{\dim G/2 + 1} \), where \( C \) is a constant depending on \( G \).

**Proof.** We fix a maximal torus \( T \) and write

\[
\|f\|_{\text{Lip}(T)} = \max_{t \in T, g \in G} \left\{ \frac{|f(tg) - f(g)|}{\text{dist}(1, t)} \right\}
\]

for \( f \in \text{Lip}(G) \). Since every element of \( G \) is contained in a maximal torus, it is enough to bound this new semi-norm which a priori could be smaller.

As in the proof of Lemma 19, we decompose \( H_r \) as the sum of weight spaces for \( T \). There is an orthonormal basis \( \varphi_1, \ldots, \varphi_m \in H_r \) (where \( m = \dim H_r \)) such that for \( g \in T \) we have

\[
\pi(g) \varphi_j = e^{2\pi i \langle \log g, u_j \rangle} \varphi_j,
\]

where \( u_j \in I^\ast \) is a weight and \( |u_j| \leq r \). Thus \( \|\varphi_j\|_{\text{Lip}(T)} \leq C r \). We can write

\[
f = \sum \alpha_j \varphi_j
\]

such that \( \sum \alpha_j^2 = 1 \). Then

\[
\|f\|_{\text{Lip}(T)} \leq C r \sum \alpha_j \leq C r \sqrt{m}.
\]

By Weyl’s dimension formula it follows that each irreducible representation in \( H_r \) is of dimension at most \( C r^{\dim T} \) and each appears with multiplicity equal to its dimension. The number of irreducible components is at most \( C r^{\dim T} \).

Putting these together we get \( m \leq C r^{\dim G} \) which proves the Lemma.

**Proof of Lemma 5.** First we bound the diameter in terms of the spectral gap. Let \( \epsilon > 0 \) be a number, and \( g_0 \in G \). Suppose that for some integer \( l \), there is no \( g \in S_l \) such that \( \text{dist}(g, g_0) \leq \epsilon \).

We fix a number \( D \) that we will specify later, and write \( r = D \epsilon^{-2} \dim G^{-2} \) and let \( f_r \) be the approximate identity constructed in Lemma 19. Then

\[
\int_{\text{dist}(g, g_0) < \frac{\epsilon}{2}} \text{Reg}(\mu_S)^l f_r(g) \, dm_G(g) = \sum_{h \in S^l} \int_{\text{dist}(g, g_0) < \frac{\epsilon}{2}} \mu_S^{\ast(l)}(h) f_r(h^{-1} g) \, dm_G(g) \leq \int_{\text{dist}(g_0, 1) > \frac{\epsilon}{2}} f_r(g) \, dm_G(g) \leq \frac{C}{\epsilon \sqrt{r}}. \quad (46)
\]

For the inequality between the first and second lines, we used that \( \text{dist}(h, g) \geq \epsilon/2 \), hence \( \text{dist}(1, h^{-1} g) \geq \epsilon/2 \).

On the other hand, we have

\[
\|1 - \text{Reg}(\mu_S)^l f_r\|_2 \leq C r^{\dim G/4} e^{-l \text{gap}_p(G, S)}.
\]
(Recall $\|f_r\|_2 \leq C r^{\dim G/4}$ from Lemma 19.) Hence

$$\int_{\text{dist}(g,g_0) < \frac{r}{\sqrt{2}}} \operatorname{Reg}(\mu_S)^f f_r(g) \, dm_G(g) \geq \int_{\text{dist}(g,g_0) < \frac{r}{\sqrt{2}}} 1 \, dm_G(g) - \frac{C r^{\dim G/4}}{\varepsilon \operatorname{gap}_r(G,S)} \geq c \varepsilon^{\dim G},$$

provided

$$l > \frac{\log \left( 2 C r^{\dim G/4} / \left( \varepsilon \dim G \right) \right)}{\operatorname{gap}_r(G,S)} \geq C' \cdot \frac{\log(\varepsilon^{-1})}{\operatorname{gap}_r(G,S)},$$

where $C'$ depends on $G$ and $D$. Now we choose $D$ such that

$$\frac{C}{\varepsilon \sqrt{r}} \leq c \varepsilon^{\dim G},$$

where $C$ and $c$ are the constants form (46) and (47), respectively. This is impossible. We can conclude

$$\operatorname{diam}(G,S) \leq \frac{C \log(\varepsilon^{-1})}{\operatorname{gap}_D^{-2 \dim G/2} \cdot (G,S)}.$$ 

Now we estimate the spectral gap in terms of the diameter. This argument was communicated to me by Jean Bourgain. Let $r > 0$ be a number, and set $\varepsilon = D r^{- \dim G/2}$. Let $f \in \mathcal{H}_r$ and assume that $\|f\|_2 = 1$ and $\int f = 0$. Then $\int (\operatorname{Reg}(g) f, f) \, dm_G(g) = 0$, hence there is $g \in G$ such that $(\operatorname{Reg}(g) f, f) \leq 0$. Thus $\|\operatorname{Reg}(g) f - f\|_2 \geq \sqrt{2}$. Let $l = \operatorname{diam}_r(G,S)$ and $g_0 = g_1 \cdots g_l \in S^l$ such that $\operatorname{dist}(g,g_0) \leq \varepsilon$. By Lemma 20 we have

$$\|\operatorname{Reg}(g_j) f - f\|_2 \geq \sqrt{2} - \varepsilon C r^{\dim G/2+1} \geq 1$$

if we choose $D$ to be sufficiently small in the definition of $\varepsilon$. By the triangle inequality, there is $1 \leq j \leq l$ such that

$$\|\operatorname{Reg}(g_j) f - f\|_2 = \|\operatorname{Reg}(g_1 \cdots g_j) f - \operatorname{Reg}(g_1 \cdots g_j-1) f\|_2 \geq 1/l.$$ 

This implies

$$\|\operatorname{Reg}(g_j) f + f\|_2 \leq 2 - 1/l^2.$$ 

Finally, we can conclude

$$\|\operatorname{Reg}(\mu_S) f\|_2 \leq \frac{1}{|S|} \left( \|\operatorname{Reg}(g_j) f + f\|_2 + \left\| \sum_{g \in S \setminus \{1,g_j\}} \operatorname{Reg}(g) f \right\|_2 \right) \leq 1 - \frac{1}{|S| \operatorname{diam}_G(G,S)^2}$$

which was to be proved. (Recall the assumption $1 \in S.$)
Random Walks in Compact Groups

References


Péter P. Varjú  
Centre for Mathematical Sciences  
Wilberforce Road  
Cambridge CB3 0WA  
England  
pv270@dpmms.cam.ac.uk  
and  
The Einstein Institute of Mathematics  
Edmond J. Safra Campus  
Givat Ram  
The Hebrew University of Jerusalem  
Jerusalem, 91904  
Israel
Torelli Theorem for the Deligne–Hitchin Moduli Space, II

Indranil Biswas, Tomás L. Gómez, and Norbert Hoffmann

Received: April 9, 2012
Communicated by Ulf Rehmann

Abstract. Let $X$ and $X'$ be compact Riemann surfaces of genus at least three. Let $G$ and $G'$ be nontrivial connected semisimple linear algebraic groups over $\mathbb{C}$. If some components $\mathcal{M}_{\text{DH}}^d(X, G)$ and $\mathcal{M}_{\text{DH}}^{d'}(X', G')$ of the associated Deligne–Hitchin moduli spaces are biholomorphic, then $X'$ is isomorphic to $X$ or to the conjugate Riemann surface $\overline{X}$.

2010 Mathematics Subject Classification: 14D20, 14C34
Keywords and Phrases: Principal bundle, Deligne–Hitchin moduli space, Higgs bundle, vector field

1. Introduction

Let $X$ be a compact connected Riemann surface of genus $g \geq 3$. Let $\overline{X}$ denote the conjugate Riemann surface; by definition, it consists of the real manifold underlying $X$ and the almost complex structure $J_{\overline{X}} := -J_X$. Let $G$ be a nontrivial connected semisimple linear algebraic group over $\mathbb{C}$. The topological types of holomorphic principal $G$–bundles $E$ over $X$ correspond to elements of $\pi_1(G)$. Let $\mathcal{M}_{\text{Higgs}}^d(X, G)$ denote the moduli space of semistable Higgs $G$–bundles $(E, \theta)$ over $X$ with $E$ of topological type $d \in \pi_1(G)$.

The Deligne–Hitchin moduli space $[Si3]$ is a complex analytic space $\mathcal{M}_{\text{DH}}^d(X, G)$ associated to $X$, $G$ and $d$. It is the twistor space for the hyper-Kähler structure on $\mathcal{M}_{\text{Higgs}}^d(X, G)$; see $[Hi2]$ §9. Deligne $[De]$ has constructed it together with a surjective holomorphic map

$$
\mathcal{M}_{\text{DH}}^d(X, G) \longrightarrow \mathbb{C}P^1 = \mathbb{C} \cup \{\infty\}.
$$

This work was supported by the grant MTM2010-17389, the ICMAT Severo Ochoa project SEV-2011-0087 of the Spanish Ministerio de Economía y Competitividad, and by the German SFB 647: Raum - Zeit - Materie. The first-named author acknowledges the support of the J. C. Bose Fellowship.
The inverse image of $C \subseteq \mathbb{CP}^1$ is the moduli space $\mathcal{M}_\text{Hod}^d(X, G)$ of holomorphic principal $G$–bundles over $X$ endowed with a $\lambda$–connection. In particular, every fiber over $C^* \subseteq \mathbb{CP}^1$ is isomorphic to the moduli space of holomorphic $G$–connections over $X$. The fiber over $0 \in \mathbb{CP}^1$ is $\mathcal{M}_\text{Higgs}^d(X, G)$, and the fiber over $\infty \in \mathbb{CP}^1$ is $\mathcal{M}_\text{Higgs}^{-d}(X, G)$.

In this paper, we study the dependence of these moduli spaces on $X$. Our main result, Theorem 5.3, states that the complex analytic space $\mathcal{M}_\text{DH}^d(X, G)$ determines the unordered pair $\{X, \bar{X}\}$ up to isomorphism. We also prove that $\mathcal{M}_\text{Higgs}^d(X, G)$ and $\mathcal{M}_\text{Hod}^d(X, G)$ each determine $X$ up to isomorphism; see Theorem 5.1 and Theorem 5.2.

The key technical result is Proposition 3.1, which says the following: Let $Z$ be an irreducible component of the fixed point locus for the natural $C^*$–action on a moduli space $\mathcal{M}_\text{Higgs}^d(X, G)$ of Higgs $G$–bundles. Then,

$$\dim Z \leq (g - 1) \cdot \dim \mathfrak{C} G,$$

with equality holding only if $Z$ is the moduli space $\mathcal{M}_d^d(X, G)$ of holomorphic principal $G$-bundles of topological type $d$ over $X$.

In [BGHL], the case of $G = \text{SL}(r, \mathbb{C})$ was considered.

If $g = 2$ and $G = \text{SL}(2, \mathbb{C})$, then the above theorems are no longer valid. So we have assumed that $g \geq 3$.

2. Some moduli spaces associated to a compact Riemann surface

Let $X$ be a compact connected Riemann surface of genus $g \geq 3$. Let $G$ be a nontrivial connected semisimple linear algebraic group defined over $\mathbb{C}$, with Lie algebra $\mathfrak{g}$.

2.1. Principal $G$–bundles. We consider holomorphic principal $G$–bundles $E$ over $X$. Recall that the topological type of $E$ is given by an element $d \in \pi_1(G)$ [R]; this is a finite abelian group. The adjoint vector bundle of $E$ is the holomorphic vector bundle

$$\text{ad}(E) := E \times^G \mathfrak{g}$$

over $X$, using the adjoint action of $G$ on $\mathfrak{g}$. $E$ is called stable (respectively, semistable) if

$$\text{degree}(\text{ad}(E_P)) < 0 \quad \text{(respectively, } \leq 0)$$

for every maximal parabolic subgroup $P \varsubsetneq G$ and every holomorphic reduction of structure group $E_P$ of $E$ to $P$; here $\text{ad}(E_P) \subset \text{ad}(E)$ is the adjoint vector bundle of $E_P$.

Let $\mathcal{M}_d^d(X, G)$ denote the moduli space of semistable holomorphic principal $G$–bundles $E$ over $X$ of topological type $d \in \pi_1(G)$. It is known that $\mathcal{M}_d^d(X, G)$ is an irreducible normal projective variety of dimension $(g - 1) \cdot \dim \mathfrak{C} G$ over $\mathbb{C}$.
2.2. Higgs G-bundles. The holomorphic cotangent bundle of $X$ will be denoted by $K_X$.

A Higgs $G$-bundle over $X$ is a pair $(E, \theta)$ consisting of a holomorphic principal $G$-bundle $E$ over $X$ and a holomorphic section

$$\theta \in H^0(X, \text{ad}(E) \otimes K_X),$$

the so-called Higgs field [Hi1, Si1]. The pair $(E, \theta)$ is called stable (respectively, semistable) if the inequality (1) holds for every holomorphic reduction of structure group $E_P$ of $E$ to a maximal parabolic subgroup $P \subseteq G$ such that $\theta \in H^0(X, \text{ad}(E_P) \otimes K_X)$.

Let $\mathcal{M}^d_{\text{Higgs}}(X, G)$ denote the moduli space of semistable Higgs $G$-bundles $(E, \theta)$ over $X$ such that $E$ is of topological type $d \in \pi_1(G)$. It is known that $\mathcal{M}^d_{\text{Higgs}}(X, G)$ is an irreducible normal quasiprojective variety of dimension $2(g-1) \cdot \dim_C G$ over $\mathbb{C}$ [Si2]. We regard $\mathcal{M}^d_{\text{Higgs}}(X, G)$ as a closed subvariety of $\mathcal{M}^d_{\text{Higgs}}(X, G)$ by means of the embedding

$$\mathcal{M}^d_{\text{Rep}}(X, G) \hookrightarrow \mathcal{M}^d_{\text{Higgs}}(X, G), \quad E \mapsto (E, 0).$$

There is a natural algebraic symplectic structure on $\mathcal{M}^d_{\text{Higgs}}(X, G)$; see [Hi1, BR].

2.3. Representations of the surface group in $G$. Fix a base point $x_0 \in X$. The fundamental group of $X$ admits a standard presentation

$$\pi_1(X, x_0) \cong \langle a_1, \ldots, a_g, b_1, \ldots, b_g \mid \prod_{i=1}^g a_i b_i a_i^{-1} b_i^{-1} = 1 \rangle$$

which we choose in such a way that it is compatible with the orientation of $X$. We identify the fundamental group of $G$ with the kernel of the universal covering $\tilde{G} \to G$. The type $d \in \pi_1(G)$ of a homomorphism $\rho : \pi_1(X, x_0) \to G$ is defined by

$$d := \prod_{i=1}^g \alpha_i \beta_i \alpha_i^{-1} \beta_i^{-1} \in \pi_1(G) \subset \tilde{G}$$

for any choice of lifts $\alpha_i, \beta_i \in \tilde{G}$ of $\rho(a_i), \rho(b_i) \in G$. This is also the topological type of the principal $G$-bundle $E_\rho$ over $X$ given by $\rho$. The space Hom$^d(\pi_1(X, x_0), G)$ of all homomorphisms $\rho : \pi_1(X, x_0) \to G$ of type $d \in \pi_1(G)$ is an irreducible affine variety over $\mathbb{C}$, and $G$ acts on it by conjugation. The GIT quotient

$$\mathcal{M}^d_{\text{Rep}}(X, G) := \text{Hom}^d(\pi_1(X, x_0), G) \sslash G$$

doesn’t depend on $x_0$. It is an affine variety of dimension $2(g-1) \cdot \dim_C G$ over $\mathbb{C}$, which carries a natural symplectic form [AB, Go]. Its points represent equivalence classes of completely reducible homomorphisms $\rho$. There is a natural bijective map

$$\mathcal{M}^d_{\text{Rep}}(X, G) \to \mathcal{M}^d_{\text{Higgs}}(X, G)$$

Documenta Mathematica 18 (2013) 1177–1189
given by a variant of the Kobayashi–Hitchin correspondence [Si1]. This bijective map is not holomorphic.

2.4. Holomorphic $G$–connections. Let $p : E \rightarrow X$ be a holomorphic principal $G$–bundle. Because the vertical tangent space at every point of the total space $E$ is canonically isomorphic to $g$, there is a natural exact sequence

$$0 \rightarrow E \times g \rightarrow TE \xrightarrow{dp} p^*TX \rightarrow 0$$

of $G$-equivariant holomorphic vector bundles over $E$. Taking the $G$-invariant direct image under $p$, it follows that the Atiyah bundle

$$At(E) := p_*(TE)$$

sits in a natural exact sequence of holomorphic vector bundles

$$0 \rightarrow \text{ad}(E) \rightarrow At(E) \xrightarrow{dp} TX$$

over $X$. This exact sequence is called the Atiyah sequence. A holomorphic connection on $E$ is a splitting of the Atiyah sequence, or in other words a holomorphic homomorphism

$$D : TX \rightarrow At(E)$$

such that $dp \circ D = \text{id}_{TX}$. It always exists if $E$ is semistable [AzBi, p. 342, Theorem 4.1], [BG, p. 20, Theorem 1.1]. The curvature of $D$ is a holomorphic 2–form with values in $\text{ad}(E)$, so $D$ is automatically flat.

A holomorphic $G$–connection is a pair $(E, D)$ where $E$ is a holomorphic principal $G$–bundle over $X$, and $D$ is a holomorphic connection on $E$. Such a pair is automatically semistable, because the degree of a flat vector bundle is zero.

Let $\mathcal{M}^d_{\text{conn}}(X, G)$ denote the moduli space of holomorphic $G$–connections $(E, D)$ over $X$ such that $E$ is of topological type $d \in \pi_1(G)$. It is known that $\mathcal{M}^d_{\text{conn}}(X, G)$ is an irreducible quasiprojective variety of dimension $2(g - 1) \cdot \text{dim}_C G$ over $\mathbb{C}$.

Sending each holomorphic $G$–connection to its monodromy defines a map

$$\mathcal{M}^d_{\text{conn}}(X, G) \rightarrow \mathcal{M}^d_{\text{Rep}}(X, G)$$

which is biholomorphic, but not algebraic; it is called Riemann–Hilbert correspondence. The inverse map sends a homomorphism $\rho : \pi_1(X, x_0) \rightarrow G$ to the associated principal $G$–bundle $E_\rho$, endowed with the induced holomorphic connection $D_\rho$.

2.5. $\lambda$–connections. Let $p : E \rightarrow X$ be a holomorphic principal $G$–bundle. For any $\lambda \in \mathbb{C}$, a $\lambda$–connection on $E$ is a holomorphic homomorphism of vector bundles

$$D : TX \rightarrow At(E)$$

such that $dp \circ D = \lambda \cdot \text{id}_{TX}$ for the epimorphism $dp$ in the Atiyah sequence (2). Therefore, a 0–connection is a Higgs field, and a 1–connection is a holomorphic connection.

If $D$ is a $\lambda$–connection on $E$ with $\lambda \neq 0$, then $\lambda^{-1}D$ is a holomorphic connection on $E$. In particular, the pair $(E, D)$ is automatically semistable in this case.
Let $\mathcal{M}_{\text{Hod}}^d(X, G)$ denote the moduli space of triples $(\lambda, E, D)$, where $\lambda \in \mathbb{C}$, $E$ is a holomorphic principal $G$–bundle over $X$ of topological type $d \in \pi_1(G)$, and $D$ is a semistable $\lambda$–connection on $E$; see [Si2]. There is a canonical algebraic map

$$\text{pr} = \text{pr}_X : \mathcal{M}_{\text{Hod}}^d(X, G) \longrightarrow \mathbb{C}, \quad (\lambda, E, D) \longmapsto \lambda.$$ 

Its fibers over $\lambda = 0$ and $\lambda = 1$ are $\mathcal{M}_{\text{Higgs}}^d(X, G)$ and $\mathcal{M}_{\text{Conn}}^d(X, G)$, respectively. The Riemann–Hilbert correspondence (3) allows to define a holomorphic open embedding

$$j = j_X : \mathbb{C}^* \times \mathcal{M}_{\text{Rep}}^d(X, G) \hookrightarrow \mathcal{M}_{\text{Hod}}^d(X, G), \quad (\lambda, \rho) \longmapsto (\lambda, E_{\rho}, \lambda D_{\rho})$$

with image $\text{pr}^{-1}(\mathbb{C}^*)$. This map commutes with the projections onto $\mathbb{C}^*$.

2.6. **The Deligne–Hitchin moduli space.** The compact Riemann surface $X$ provides an underlying real $\mathcal{C}^\infty$ manifold $X_\mathbb{R}$, and an almost complex structure $J_X : T X_\mathbb{R} \longrightarrow T X_\mathbb{R}$. Since any almost complex structure in real dimension two is integrable, $X := (X_\mathbb{R}, -J_X)$ is a compact Riemann surface as well. It has the opposite orientation, so

$$\mathcal{M}_{\text{Rep}}^d(X, G) = \mathcal{M}_{\text{Rep}}^d(X, G).$$

The Deligne–Hitchin moduli space $\mathcal{M}_{\text{DH}}^d(X, G)$ is the complex analytic space obtained by gluing $\mathcal{M}_{\text{Hod}}^d(X, G)$ and $\mathcal{M}_{\text{Hod}}^d(X, G)$ along their common open subspace

$$\mathcal{M}_{\text{Hod}}^d(X, G) \overset{j_X}{\leftarrow} \mathbb{C}^* \times \mathcal{M}_{\text{Rep}}^d(X, G) \cong \mathbb{C}^* \times \mathcal{M}_{\text{Rep}}^d(X, G) \overset{j_X}{\longrightarrow} \mathcal{M}_{\text{Hod}}^d(X, G)$$

where the isomorphism in the middle sends $(\lambda, \rho)$ to $(1/\lambda, \rho)$; see [Si3, De]. The projections $\text{pr}_X$ on $\mathcal{M}_{\text{Hod}}^d(X, G)$ and $1/\text{pr}_X$ on $\mathcal{M}_{\text{Hod}}^d(X, G)$ patch together to a holomorphic map

$$\mathcal{M}_{\text{DH}}^d(X, G) \longrightarrow \mathbb{C} \mathbb{P}^1 = \mathbb{C} \cup \{\infty\}.$$ 

Its fiber over any $\lambda \in \mathbb{C}^*$ is biholomorphic to the representation space [4], whereas its fibers over $\lambda = 0$ and $\lambda = \infty$ are $\mathcal{M}_{\text{Higgs}}^d(X, G)$ and $\mathcal{M}_{\text{Higgs}}^d(X, G)$, respectively.

3. **Fixed points of the natural $\mathbb{C}^*$–action**

The group $\mathbb{C}^*$ acts algebraically on the moduli space $\mathcal{M}_{\text{Higgs}}^d(X, G)$, via the formula

$$t \cdot (E, \theta) := (E, t\theta).$$

The fixed point locus $\mathcal{M}_{\text{Higgs}}^d(X, G)^{\mathbb{C}^*}$ contains the closed subvariety $\mathcal{M}^d(X, G)$. 

\textbf{Documenta Mathematica 18 (2013) 1177–1189}
Proposition 3.1. Let $Z$ be an irreducible component of $\mathcal{M}_d^{dHiggs}(X,G)^{\mathbb{C}^*}$. Then one has
\[ \dim Z \leq (g-1) \cdot \dim C, \]
with equality holding only for $Z = \mathcal{M}_d(X,G)$.

Proof. Let $(E,\theta)$ be a stable Higgs $G$–bundle over $X$. Its infinitesimal deformations are, according to [BR, Theorem 2.3], governed by the complex of vector bundles
\[ C^0 := \text{ad}(E) \xrightarrow{\text{ad}(\theta)} \text{ad}(E) \otimes K_X =: C^1 \]
over $X$. Since $(E,\theta)$ is stable, it has no infinitesimal automorphisms, so
\[ H^0(X,C^*) = 0. \]
The Killing form on $g$ induces isomorphisms $g^* \cong g$ and $\text{ad}(E)^* \cong \text{ad}(E)$. Hence the vector bundle $\text{ad}(E)$ has degree 0. Serre duality allows us to conclude
\[ H^2(X,C^*) = 0. \]
Using all this, the Riemann–Roch formula yields
\[ \dim H^1(X,C^*) = 2(g-1) \cdot \dim C. \]
From now on, we assume that the point $(E,\theta)$ is fixed by $\mathbb{C}^*$, and we also assume $\theta \neq 0$. Then $(E,\theta) \cong (E,t\theta)$ for all $t \in \mathbb{C}^*$, so the sequence of complex algebraic groups
\[ 1 \longrightarrow \text{Aut}(E,\theta) \longrightarrow \text{Aut}(E,\mathbb{C}\theta) \longrightarrow \text{Aut}(\mathbb{C}\theta) = \mathbb{C}^* \longrightarrow 1 \]
is exact. Because $(E,\theta)$ is stable, $\text{Aut}(E,\theta)$ is finite. Consequently, the identity component of $\text{Aut}(E,\mathbb{C}\theta)$ is isomorphic to $\mathbb{C}^*$. This provides an embedding
\[ \mathbb{C}^* \hookrightarrow \text{Aut}(E), \quad t \mapsto \varphi_t, \]
and an integer $w \neq 0$ with $\varphi_t(\theta) = t^w \cdot \theta$ for all $t \in \mathbb{C}^*$. We may assume that $w \geq 1$.
Choose a point $e_0 \in E$. Then there is a unique group homomorphism
\[ \iota : \mathbb{C}^* \longrightarrow G \]
such that $\varphi_t(e_0) = e_0 \cdot \iota(t)$ for all $t \in \mathbb{C}^*$. The conjugacy class of $\iota$ doesn’t depend on $e_0$, since the space of conjugacy classes $\text{Hom}(\mathbb{C}^*,G)/G$ is discrete. The subset
\[ E_H := \{ e \in E : \varphi_t(e) = e \cdot \iota(t) \text{ for all } t \in \mathbb{C}^* \} \]
of $E$ is a holomorphic reduction of structure group to the centralizer $H$ of $\iota(\mathbb{C}^*)$ in $G$. Let
\[ \mathfrak{g} = \bigoplus_{n \in \mathbb{Z}} \mathfrak{g}_n \]
denote the eigenspace decomposition given by the adjoint action of $\mathbb{C}^*$ on $\mathfrak{g}$ via $\iota$. 

Documenta Mathematica 18 (2013) 1177–1189
Let $N \in \mathbb{Z}$ be maximal with $g_N \neq 0$. Let $P \subset G$ be the parabolic subgroup with

$$\text{Lie}(P) = \bigoplus_{n \geq 0} g_n \subset \mathfrak{g}.$$ 

Since $H \subset G$ has Lie algebra $\mathfrak{g}_0$, it is a Levi subgroup in $P$. Choose subgroups

$$\iota(\mathbb{C}^*) \subseteq T \subseteq B \subseteq P \subset G$$

such that $T$ is a maximal torus in $H \subset G$ and $B$ is a Borel subgroup in $G$. Let

$$\alpha_j : T \rightarrow \mathbb{C}^*$$

be the resulting simple roots and coroots of $G$. We denote by $\langle \cdot, \cdot \rangle$ the natural pairing between characters and cocharacters of $T$. Let $\alpha_j$ be a simple root of $G$ with $\langle \alpha_j, \iota \rangle > 0$, and let $\beta$ be a root of $G$ with $\langle \beta, \iota \rangle = N$. Then the elementary reflection $s_j(\beta) = \beta - \langle \beta, \alpha_j^\vee \rangle \alpha_j$ is a root of $G$, so $\langle s_j(\beta), \iota \rangle \leq N$; this implies that $\langle \beta, \alpha_j^\vee \rangle \geq 0$. The sum of all such roots $\beta$ with $\langle \beta, \iota \rangle = N$ is the restriction $\chi|_T$ of the determinant $\chi : P \rightarrow \text{Aut}(g_N) \xrightarrow{\text{det}} \mathbb{C}^*$ of the adjoint action of $P$ on $g_N$. Hence we conclude $\langle \chi|_T, \alpha_j^\vee \rangle \geq 0$ for all simple roots $\alpha_j$ with $\langle \alpha_j, \iota \rangle > 0$. This means that the character $\chi$ of $P$ is dominant.

The decomposition (9) of $g$ induces a vector bundle decomposition

$$\text{ad}(E) = \bigoplus_{n \in \mathbb{Z}} E_{H \times H g_n}.$$ 

Since $\mathbb{C}^*$ acts with weight $w$ on the Higgs field $\theta$ by construction, we have

$$\theta \in H^0(X, (E_{H \times H g_w}) \otimes K_X).$$ 

In particular, $\theta \in H^0(X, \text{ad}(E_P) \otimes K_X)$ for the reduction $E_P := E_{H \times H P} \subset E$ of the structure group to $P$. The Higgs version of the stability criterion [Ra, Lemma 2.1] yields

$$\text{degree}(E_{H \times H g_N}) \leq 0$$

since $P$ acts on $\text{det}(g_N)$ via the dominant character $\chi$ in (10). Now Riemann–Roch implies the following:

$$\dim H^1(X, E_{H \times H g_N}) \geq (g - 1) \cdot \dim \mathbb{C} \mathfrak{g}_N > 0.$$ 

The complex $C^\bullet$ in (7) is, due to (11), the direct sum of its subcomplexes $C_n^\bullet$ given by

$$C_n^0 := E_{H \times H \mathfrak{g}_n} \xrightarrow{\text{ad}(\theta)} (E_{H \times H \mathfrak{g}_n+w}) \otimes K_X =: C_n^1.$$ 

Thus the hypercohomology of $C^\bullet$ decomposes as well; in particular, we have

$$H^1(X, C^\bullet) = \bigoplus_{n \in \mathbb{Z}} H^1(X, C_n^\bullet).$$

In the last nonzero summand $C_N^\bullet$, we have $C_N^1 = 0$ and hence

$$\dim H^1(X, C_N^\bullet) = \dim H^1(X, E_{H \times H \mathfrak{g}_N}) > 0.$$
due to (12). Since $g_{\mathfrak{n}}^* \cong g_{-n}$ via the Killing form on $g$, Serre duality yields in particular
\[ \dim H^1(X, C^*_0) = \dim H^1(X, C^*_w). \]
Taken together, the last three formulas and the equation (8) imply that
\[ \dim H^1(X, C^*_0) < \frac{1}{2} \dim H^1(X, C^*) = (g-1) \cdot \dim_{\mathbb{C}} G. \]
But $H^1(X, C^*_0)$ parameterizes infinitesimal deformations of pairs $(E_H, \theta)$ consisting of a principal $H$–bundle $E_H$ and a section $\theta$ as in (11); see [BR, Theorem 2.3]. This proves that
\[ \dim Z < (g-1) \cdot \dim_{\mathbb{C}} G \]
for every irreducible component $Z$ of the fixed point locus $M_{\text{Higgs}}^d(X, G)_{\mathbb{C}^*}$ such that $Z$ contains stable Higgs $G$–bundles $(E, \theta)$ with $\theta \neq 0$.

The non-stable points in $M_{\text{Higgs}}^d(X, G)$ correspond to polystable Higgs $G$–bundles $(E, \theta)$. Polystability means that $E$ admits a reduction of structure group $E_L$ to a Levi subgroup $L \subset G$ of a parabolic subgroup in $G$ such that $\theta$ is a section of the subbundle
\[ \text{ad}(E_L) \otimes K_X \subset \text{ad}(E) \otimes K_X \]
and the pair $(E_L, \theta)$ is stable. Let $C \subseteq L$ be the identity component of the center, and let $c \subseteq \mathfrak{l}$ be their Lie algebras. Then $E_{L/C} := E_L/C$ is a principal $(L/C)$–bundle over $X$, and
\[ \text{ad}(E_L) \cong (c \otimes \mathcal{O}_X) \oplus \text{ad}(E_{L/C}) \]
since $\mathfrak{l} = c \oplus [\mathfrak{l}, \mathfrak{l}]$, where the subalgebra $[\mathfrak{l}, \mathfrak{l}] \subseteq \mathfrak{l}$ is also the Lie algebra of $L/C$.

We have
\[ \dim_{\mathbb{C}} G - \dim_{\mathbb{C}} L \geq 2 \dim_{\mathbb{C}} C \]
because maximal Levi subgroups in $G$ have 1-dimensional center and at least one pair of opposite roots less than $G$; the other Levi subgroups can be reached by iterating this.

Now suppose that $\mathbb{C}^*$ fixes the point $(E, \theta)$. Then $(E_L, \theta) \cong (E_L, t\theta)$ for all $t \in \mathbb{C}^*$. But the action of $\text{Aut}(E_L)$ on the direct summand $c \otimes \mathcal{O}_X$ of $\text{ad}(E_L)$ is trivial, since the adjoint action of $L$ on $c$ is trivial. So $\theta$ lives in the other summand of $\text{ad}(E_L)$, meaning
\[ \theta \in H^0(X, \text{ad}(E_{L/C}) \otimes K_X). \]
The Higgs $(L/C)$–bundle $(E_{L/C}, \theta)$ is still stable and fixed by $\mathbb{C}^*$; we have already proved that the locus of such has dimension $\leq (g-1) \cdot \dim_{\mathbb{C}} (L/C)$.

The abelian variety $M^d(X, C)$ acts simply transitively on lifts of $E_{L/C}$ to a principal $L$–bundle $E_L$; so these lifts form a family of dimension $g \cdot \dim_{\mathbb{C}} C$. Hence the pairs $(E_L, \theta)$ in question have at most
\[ (g-1) \cdot \dim_{\mathbb{C}} (L/C) + g \cdot \dim_{\mathbb{C}} C < (g-1) \cdot \dim_{\mathbb{C}} G \]
moduli. This implies that $\dim Z < (g-1) \cdot \dim_{\mathbb{C}} G$ for each non-stable component $Z$ of the fixed point locus, since there are only finitely many possibilities for $L$ up to conjugation. \qed
The algebraic \( C^* \)-action (13) on \( \mathcal{M}_{dHiggs}^d(X,G) \) extends naturally to an algebraic \( C^* \)-action on \( \mathcal{M}_{dHod}^d(X,G) \), which is given by the formula
\[
\lambda \cdot (\lambda, E, D) := (\lambda, tE, t^2D).
\]
A point \((\lambda, E, D)\) can only be fixed by this action if \( \lambda = 0 \), so Proposition 3.1 yields the following corollary:

**Corollary 3.2.** Let \( Z \) be an irreducible component of \( \mathcal{M}_{dHod}^d(X,G) \). Then one has
\[
\dim Z \leq (g - 1) \cdot \dim C_G,
\]
with equality only for \( Z = \mathcal{M}_{dHod}^d(X,G) \).

The algebraic \( C^* \)-action (13) on \( \mathcal{M}_{dHod}^d(X,G) \) extends naturally to a holomorphic \( C^* \)-action on \( \mathcal{M}_{dDH}^d(X,G) \), which is on the other open patch \( \mathcal{M}_{dHod}^{-d}(X,G) \) given by the formula
\[
\lambda \cdot (\lambda, E, D) := (\lambda, E, t^{-1}D).
\]
Applying Corollary 3.2 to both \( \mathcal{M}_{dHod}^d(X,G) \) and \( \mathcal{M}_{dHod}^{-d}(X,G) \), one immediately gets

**Corollary 3.3.** Let \( Z \) be an irreducible component of \( \mathcal{M}_{dDH}^d(X,G) \). Then one has
\[
\dim Z \leq (g - 1) \cdot \dim C_G,
\]
with equality only for \( Z = \mathcal{M}^d(X,G) \) and for \( Z = \mathcal{M}^{-d}(X,G) \).

4. **Vector fields on the moduli spaces**

A stable principal \( G \)-bundle \( E \) over \( X \) is called regularly stable if the automorphism group \( \text{Aut}(E) \) is just the center of \( G \). The regularly stable locus
\[
\mathcal{M}_{d,rs}^d(X,G) \subseteq \mathcal{M}^d(X,G)
\]
is open, and coincides with the smooth locus of \( \mathcal{M}^d(X,G) \); see [BH, Corollary 3.4].

**Proposition 4.1.** There are no nonzero holomorphic vector fields on \( \mathcal{M}_{d,rs}^d(X,G) \).

**Proof.** This statement is contained in [Fa, p. 549, Corollary III.3].

**Proposition 4.2.** There are no nonzero holomorphic 1-forms on \( \mathcal{M}_{d,rs}^d(X,G) \).

**Proof.** The moduli space of Higgs \( G \)-bundles is equipped with the Hitchin map
\[
\mathcal{M}_{Higgs}^d(X,G) \longrightarrow \bigoplus_{i=1}^{\text{rank}(G)} H^0(X, K_X^{n_i})
\]
where the \( n_i \) are the degrees of generators for the algebra \( \text{Sym}(g^*)^G \); see [HH § 4], [La]. Any sufficiently general fiber of this Hitchin map is a complex abelian variety \( A \) (see [Da], [Fa], [DP] for the details), and
\[
\varphi : A \longrightarrow \mathcal{M}_{d,rs}^d(X,G), \quad (E, \theta) \longmapsto E,
\]

Documenta Mathematica 18 (2013) 1177–1189
is a dominant rational map. This rational map $\varphi$ is defined outside a closed subscheme of codimension at least two; see [Fa] p. 534, Theorem II.6.

Let $\omega$ be a holomorphic 1–form on $\mathcal{M}^{d,\text{rs}}(X,G)$. Then $\varphi^*\omega$ extends to a holomorphic 1–form on $A$ by Hartog’s theorem. As any holomorphic 1–form on $A$ is closed, it follows that $\omega$ is closed. Since $H^1(\mathcal{M}^{d,\text{rs}}(X,G), \mathbb{C}) = 0$ by [AB] Ch. 10, we conclude $\omega = df$ for a holomorphic function $f$ on $\mathcal{M}^{d,\text{rs}}(X,G)$. But any such function $f$ is constant, so $\omega = 0$. □

We denote by $\mathcal{M}^{d,\text{rs}}_{\text{Higgs}}(X,G) \subseteq \mathcal{M}^d_{\text{Higgs}}(X,G)$ the open locus of Higgs $G$–bundles $(E,\theta)$ for which $E$ is regularly stable. The forgetful map

$$\mathcal{M}^{d,\text{rs}}_{\text{Higgs}}(X,G) \longrightarrow \mathcal{M}^{d,\text{rs}}(X,G), \quad (E,\theta) \longmapsto E,$$

is an algebraic vector bundle with fibers $H^0(X, \text{ad}(E) \otimes K_X) \cong H^1(X, \text{ad}(E))^*$, so it is the cotangent bundle of $\mathcal{M}^{d,\text{rs}}(X,G)$.

**Corollary 4.3.** The restriction of the algebraic tangent bundle $T\mathcal{M}^{d,\text{rs}}_{\text{Higgs}}(X,G) \longrightarrow \mathcal{M}^{d,\text{rs}}_{\text{Higgs}}(X,G)$ to the subvariety $\mathcal{M}^{d,\text{rs}}(X,G) \subseteq \mathcal{M}^{d,\text{rs}}_{\text{Higgs}}(X,G)$ has no nonzero holomorphic sections.

**Proof.** The subvariety in question is the zero section of the vector bundle (14). Given a vector bundle $V \longrightarrow M$ with zero section $M \subseteq V$, there is a natural isomorphism

$$T(V)|_M \cong TM \oplus V$$

of vector bundles over $M$. In our situation, both summands have no nonzero holomorphic sections, according to Proposition 4.1 and Proposition 4.2. □

Let $\mathcal{M}^{d,\text{rs}}_{\text{conn}}(X,G) \subseteq \mathcal{M}^d_{\text{conn}}(X,G)$ denote the open locus of holomorphic $G$–connections $(E,D)$ for which $E$ is regularly stable.

**Proposition 4.4.** There are no holomorphic sections for the forgetful map

$$\mathcal{M}^{d,\text{rs}}_{\text{conn}}(X,G) \longrightarrow \mathcal{M}^{d,\text{rs}}(X,G), \quad (E,D) \longmapsto E.$$

**Proof.** The map (16) is a holomorphic torsor under the cotangent bundle of $\mathcal{M}^{d,\text{rs}}(X,G)$. As such, it is isomorphic to the torsor of holomorphic connections on the line bundle $\mathcal{L} \longrightarrow \mathcal{M}^{d,\text{rs}}(X,G)$ with fibers $\det H^1(X, \text{ad}(E))$; see [Fa] p. 554, Lemma IV.4. Since $\mathcal{L}$ is ample [KNR], its first Chern class is nonzero, so $\mathcal{L}$ admits no global holomorphic connections. □

Let $\mathcal{M}^{d,\text{rs}}_{\text{Hod}}(X,G) \subseteq \mathcal{M}^d_{\text{Hod}}(X,G)$ denote the open locus of triples $(\lambda, E, D)$ for which $E$ is regularly stable. The forgetful maps in (14) and (16) extend to the forgetful map

$$\mathcal{M}^{d,\text{rs}}_{\text{Hod}}(X,G) \longrightarrow \mathcal{M}^{d,\text{rs}}(X,G), \quad (\lambda, E, D) \longmapsto E,$$
which is an algebraic vector bundle. It contains the cotangent bundle (14) as a subbundle; the quotient is a line bundle, which is trivialized by the projection \( \text{pr} \) in (11).

**Corollary 4.5.** The vector bundle (17) has no nonzero holomorphic sections.

**Proof.** Let \( s \) be a holomorphic section of the vector bundle (17). Then \( \text{pr} \circ s \) is a holomorphic function on \( \mathcal{M}_{\text{Higgs}}^{d,rs}(X,G) \), and hence constant. This constant vanishes because of Proposition 4.4. So \( \text{pr} \circ s = 0 \), which implies that \( s = 0 \) using Proposition 4.2. \( \square \)

**Corollary 4.6.** The restriction of the algebraic tangent bundle

\[
T\mathcal{M}_{\text{Hod}}^{d,rs}(X,G) \to \mathcal{M}_{\text{Hod}}^{d,rs}(X,G)
\]

to the subvariety \( \mathcal{M}_{\text{Hod}}^{d,rs}(X,G) \subseteq \mathcal{M}_{\text{Hod}}^{d,rs}(X,G) \) has no nonzero holomorphic sections.

**Proof.** Use the decomposition (15), Proposition 4.1, and Corollary 4.5. \( \square \)

5. **Torelli theorems**

Let \( X, X' \) be compact connected Riemann surfaces of genus \( \geq 3 \). Let \( G, G' \) be nontrivial connected semisimple linear algebraic groups over \( \mathbb{C} \). Fix \( d \in \pi_1(G) \) and \( d' \in \pi_1(G') \).

**Theorem 5.1.** If \( \mathcal{M}_{\text{Higgs}}^{d}(X',G') \) is biholomorphic to \( \mathcal{M}_{\text{Higgs}}^{d}(X,G) \), then \( X' \cong X \).

**Proof.** Corollary 4.3 implies that the subvariety \( \mathcal{M}^{d}(X,G) \) is fixed pointwise by every holomorphic \( \mathbb{C}^* \)–action on \( \mathcal{M}_{\text{Higgs}}^{d}(X,G) \). All other complex analytic subvarieties with that property have smaller dimension, due to Proposition 3.1. Thus we get a biholomorphic map from \( \mathcal{M}^{d'}(X',G') \) to \( \mathcal{M}^{d}(X,G) \) by restriction. Using \( [BH] \), this implies that \( X' \cong X \). \( \square \)

**Theorem 5.2.** If \( \mathcal{M}_{\text{Hod}}^{d'}(X',G') \) is biholomorphic to \( \mathcal{M}_{\text{Hod}}^{d'}(X,G) \), then \( X' \cong X \).

**Proof.** The argument is exactly the same as in the previous proof. It suffices to replace Corollary 4.3 by Corollary 4.0 and Proposition 3.1 by Corollary 3.2. \( \square \)

**Theorem 5.3.** If \( \mathcal{M}_{\text{DH}}^{d'}(X',G') \) is biholomorphic to \( \mathcal{M}_{\text{DH}}^{d'}(X,G) \), then \( X' \cong X \text{ or } X' \cong \overline{X} \).

**Proof.** The argument is similar. Corollary 4.0 implies that the two subvarieties \( \mathcal{M}^{d}(X,G) \text{ and } \mathcal{M}^{-d}(X,G) \) are fixed pointwise by every holomorphic \( \mathbb{C}^* \)–action on \( \mathcal{M}_{\text{DH}}^{d}(X,G) \). All other complex analytic subvarieties with that property have smaller dimension, due to Corollary 3.3. Thus we get a biholomorphic map from \( \mathcal{M}^{d'}(X',G') \) to either \( \mathcal{M}^{d}(X,G) \text{ or } \mathcal{M}^{-d}(X,G) \) by restriction. Using \( [BH] \), this implies that either \( X' \cong X \text{ or } X' \cong \overline{X} \). \( \square \)
References


[Si3] C. T. Simpson, A weight two phenomenon for the moduli of rank one local systems on open varieties, *From Hodge theory to integrability and

Documenta Mathematica 18 (2013) 1177–1189
Deligne–Hitchin Moduli Space, II


Indranil Biswas
School of Mathematics
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 400005
India
indranil@math.tifr.res.in

Tomás L. Gómez
Instituto de Ciencias Matemáticas
(CSIC-UAM-UC3M-UCM)
Nicolás Cabrera 15
Campus Cantoblanco UAM
28049 Madrid
Spain
tomas.gomez@icmat.es

Norbert Hoffmann
Department of Mathematics and Computer Studies
Mary Immaculate College
South Circular Road
Limerick
Ireland
norbert.hoffmann@mic.ul.ie
**Abstract.** We give a constructive proof of the existence of the essential Whittaker function of a generic representation of $GL(n,F)$, for $F$ a non-archimedean local field, using mirabolic restriction techniques.

2010 Mathematics Subject Classification: 22E50

**Introduction**

Let $F$ be nonarchimedean local field, we denote by $\mathcal{O}$ its ring of integers, and by $\mathfrak{p} = \varpi \mathcal{O}$ the maximal ideal of this ring, where $\varpi$ is a uniformiser of $F$. We denote by $q$ the cardinality of $\mathcal{O}/\mathfrak{p}$ and by $|.|$ the absolute value on $F$ normalised such that $|\varpi|$ is equal to $q^{-1}$.

For $n \geq 1$, we denote the group $GL(n,F)$ by $G_n$, the group $GL(n,\mathcal{O})$ by $G_n(\mathcal{O})$, and we set $G_0 = \{1\}$. We denote by $A_n$ the torus of diagonal matrices in $G_n$, and by $N_n$ the unipotent radical of the Borel subgroup of $G_n$ given by upper triangular matrices. For $m \geq 1$, we denote by $K_n(m)$ the subgroup of $G_n$, given by matrices $\begin{pmatrix} g & v \\ l & t \end{pmatrix}$, for $g$ in $G_{n-1}(\mathcal{O})$, $v$ in $\mathcal{O}^{n-1}$, $l$ with every coefficient in $\mathfrak{p}^m$, and $t$ in $1 + \mathfrak{p}^m$. We set $K_n(0) = G_n(\mathcal{O})$.

If $\pi$ is a generic representation of $G_2$, the essential vector of $\pi$ was first considered in [C], for $G_n$ with $n \geq 2$, it was studied in [J-P-S]. Here is one of its main properties: if one calls $d$ the conductor (the power of $q^{-\varepsilon}$ in the $\varepsilon$ factor with respect to an unramified additive character of $F$) of the representation $\pi$, the complex vector space $\pi^{K_n(d)}$ of vectors in $\pi$ fixed under $K_n(d)$, is generated by the essential vector of $\pi$, and $\pi^{K_n(k)}$ becomes the null space for $k < d$. However, to prove its existence, one has to study properties of the Rankin-Selberg integrals associated to the pairs $(\pi,\pi')$, where $\pi'$ varies through the...
set of unramified generic representations of $G_{n-1}$.

We set a few more notations before explaining this.

We denote by $\nu$ the positive character $|.| \circ \det$ of $G_n$. We use the product notation for normalised parabolic induction (see Section 1.2). For any sequence of complex numbers $s_1, \ldots, s_n$, the representation $|.|^{s_1} \times \cdots \times |.|^{s_n}$ of $G_n$ is unramified, and its subspace of $G_n(\mathcal{O})$-invariant vectors is of dimension 1 (see Section 1.3).

We choose a character $\theta$ of $(F,+)$ trivial on $\mathcal{O}$ but not on $\mathfrak{P}^{-1}$, and use it to define a non degenerate character, still denoted $\theta$, of the standard unipotent subgroup $N_n$ of $G_n$, by $\theta(n) = \theta(\sum_{i=1}^{n-1} n_{i,i+1})$.

For $n \geq 2$, let $\pi$ and $\pi'$ be representations of Whittaker type (see Section 1.4) of $G_n$ and $G_{n-1}$ respectively, and denote by $W(\pi,\theta)$ and $W(\pi',\theta^{-1})$ their respective Whittaker models (which are quotients of $\pi$ and $\pi'$) with respect to $\theta$ and $\theta^{-1}$.

If $W$ and $W'$ belong respectively to $W(\pi,\theta)$ and $W(\pi',\theta^{-1})$, we denote $I(W,W',s)$ the associated Rankin-Selberg integral (see Section 1.4).

For example, for a sequence of complex numbers $a_1, \ldots, a_m$, the induced representation $|.|^{a_1} \times \cdots \times |.|^{a_m}$ of $G_m$, is of Whittaker type. If moreover $\Re(a_1) \geq \cdots \geq \Re(a_m)$, the representation $|.|^{a_1} \times \cdots \times |.|^{a_m}$ is of Langlands' type, and its Whittaker model contains a unique normalised spherical Whittaker function $W(q^{-a_1}, \ldots, q^{-a_m})$. It is the unique Whittaker function on $G_m$, fixed by $G_m(\mathcal{O})$, which equals 1 on $G_m(\mathcal{O})$, and associated to the Satake parameter $\{q^{-a_1}, \ldots, q^{-a_m}\}$ (see [S]). For fixed $g$ in $G_m$, the function $W(q^{-s_1}, \ldots, q^{-s_m})(g)$ is an element of the ring $\mathbb{C}[q^{\pm 1}, \ldots, q^{\pm m}]^{S_m}$ of invariant Laurent polynomials. To define the essential vector of $\pi$, one needs to show as in [1-P-S], the following theorem (see [G-J] for the definition of the $L$ function of an irreducible representation of $G_n$):

**Theorem.** Let $\pi$ be a generic representation of $G_n$ with Whittaker model $W(\pi,\theta)$, then there exists in $W(\pi,\theta)$ a unique $G_{n-1}(\mathcal{O})$-invariant function $W_{\pi}^{\text{ess}}$, such that for every sequence of complex numbers $s_1, \ldots, s_{n-1}$, one has the equality $I(W_{\pi}^{\text{ess}}, W(q^{-s_1}, \ldots, q^{-s_{n-1}}), s) = \prod_{i=1}^{n-1} L(\pi, s + s_i)$.

Hence, the statement of the theorem is equivalent to say that for any unramified representation $\pi'$ of Langlands' type of $G_{n-1}$ with normalised spherical Whittaker function $W_{\pi'}^0$, in $W(\pi',\theta)$, one has the equality $I(W_{\pi'}^{\text{ess}}, W_{\pi'}^0, s) = L(\pi,\pi',s)$ (see Section 1.4 for the definition of $L(\pi,\pi',s)$ and the equality $L(\pi,\pi',s) = \prod_{i=1}^{n-1} L(\pi, s + s_i)$ when $\pi' = |.|^{s_1} \times \cdots \times |.|^{s_m}$).

Using this theorem, it is then shown in [1-P-S], using the functional equation of $L(\pi,\pi',s)$, that the space $W(\pi,\theta)^{K_n(d)}$ is a complex line spanned by $W_{\pi}^{\text{ess}}$, and that $W(\pi,\theta)^{K_n(\mathcal{O})}$ is zero for $k < d$.

In this paper, we will show the following result, using the interpretation in terms of restriction of Whittaker functions of the Bernstein-Zelevinsky derivatives.
Let $\pi$ be a ramified generic representation of $G_n$, and $\pi_u$ be the unramified component of the first nonzero spherical Bernstein-Zelevinsky derivative $\pi^{(n-r)}$ of $\pi$ (see Definition 1.3 for the precise definition). The representation $\pi_u$ is an unramified representation of Langlands’ type of $G_r$ when $r \geq 1$. In this situation, we show in Corollary 3.2, that there is a unique Whittaker function $W_{ess}\pi$ in $W(\pi, \theta)$, which is right $G_{n-1}$-invariant, and which satisfies, for $a = diag(a_1, \ldots, a_{n-1}) \in A_{n-1}$ and $a' = diag(a_1, \ldots, a_r) \in A_r$, the equality:

$$W(diag(a, 1)) = W_{0\pi_u}(a') \nu(a')^{(n-r)/2} \prod_{r < i < n} 1_{O^*}(a_i),$$  \hspace{1cm} (1)$$

when $r \geq 1$, and

$$W(diag(a, 1)) = \prod_{0 < i < n} 1_{O^*}(a_i)$$

when $r = 0$.

Computing the integral $I(W_{\pi}, W_{\pi}', s)$ for an unramified representation $\pi'$ of Langlands’ type of $G_{n-1}$, we will obtain in Corollary 3.3 the statement (more precisely a slightly more general statement) of the theorem stated above.

For $GL(2, F)$, a detailed account about newforms can be found in [Sc], the author obtains Formula (1) (see Section 2.4 of [loc. cit.]) up to normalisation by an $\epsilon$-factor. For $GL(n, F)$, Miyauchi ([Mi]) recently obtained Formula (1), assuming the existence of the essential vector, by using Hecke algebras, i.e. generalising Shintani’s method for spherical representations.

REMARK. The reason why we got interested in reproving the existence of such a vector is the following. In [J-P-S], the uniqueness of such a vector is proved. The proof of the existence is valid only for generic representations $\pi$ appearing as subquotients of representations parabolically induced by ramified characters of $GL(1, F)$ and cuspidal representations of $GL(r, F)$ for $r \geq 2$, i.e. generic representations with $L$-function equal to one.

Before we explain this, let us mention that Jacquet (see [J]) found a simple fix for the proof of [J-P-S], so that the motivation of writing our note is really to give a constructive proof of the existence of this vector, which provides a nice application of the techniques developed in [C-P].

In [J-P-S], the following is shown: for fixed $W$ in $W(\pi, \theta)$, the function

$$P(W, q^{-s_1}, \ldots, q^{-s_{n-1}}) = I(W, W(q^{-s_1}, \ldots, q^{-s_{n-1}}), 0)/\prod_{i=1}^{n-1} L(\pi, s_i)$$

belongs to the ring $\mathbb{C}[q^{\pm s_1}, \ldots, q^{\pm s_{n-1}}]^{S_{n-1}}$ of symmetric Laurent polynomials in the variables $q^{-s_i}$. It is also shown that the existence of the essential vector is equivalent to the fact that the vector space

$$I(\pi) = \{P(W, q^{-s_1}, \ldots, q^{-s_{n-1}}), W \in W(\pi, \theta)\},$$

is valid only for generic representations $\pi$ appearing as subquotients of representations parabolically induced by ramified characters of $GL(1, F)$ and cuspidal representations of $GL(r, F)$ for $r \geq 2$, i.e. generic representations with $L$-function equal to one.

Before we explain this, let us mention that Jacquet (see [J]) found a simple fix for the proof of [J-P-S], so that the motivation of writing our note is really to give a constructive proof of the existence of this vector, which provides a nice application of the techniques developed in [C-P].

In [J-P-S], the following is shown: for fixed $W$ in $W(\pi, \theta)$, the function

$$P(W, q^{-s_1}, \ldots, q^{-s_{n-1}}) = I(W, W(q^{-s_1}, \ldots, q^{-s_{n-1}}), 0)/\prod_{i=1}^{n-1} L(\pi, s_i)$$

belongs to the ring $\mathbb{C}[q^{\pm s_1}, \ldots, q^{\pm s_{n-1}}]^{S_{n-1}}$ of symmetric Laurent polynomials in the variables $q^{-s_i}$. It is also shown that the existence of the essential vector is equivalent to the fact that the vector space

$$I(\pi) = \{P(W, q^{-s_1}, \ldots, q^{-s_{n-1}}), W \in W(\pi, \theta)\},$$

is valid only for generic representations $\pi$ appearing as subquotients of representations parabolically induced by ramified characters of $GL(1, F)$ and cuspidal representations of $GL(r, F)$ for $r \geq 2$, i.e. generic representations with $L$-function equal to one.

Before we explain this, let us mention that Jacquet (see [J]) found a simple fix for the proof of [J-P-S], so that the motivation of writing our note is really to give a constructive proof of the existence of this vector, which provides a nice application of the techniques developed in [C-P].

In [J-P-S], the following is shown: for fixed $W$ in $W(\pi, \theta)$, the function

$$P(W, q^{-s_1}, \ldots, q^{-s_{n-1}}) = I(W, W(q^{-s_1}, \ldots, q^{-s_{n-1}}), 0)/\prod_{i=1}^{n-1} L(\pi, s_i)$$

belongs to the ring $\mathbb{C}[q^{\pm s_1}, \ldots, q^{\pm s_{n-1}}]^{S_{n-1}}$ of symmetric Laurent polynomials in the variables $q^{-s_i}$. It is also shown that the existence of the essential vector is equivalent to the fact that the vector space

$$I(\pi) = \{P(W, q^{-s_1}, \ldots, q^{-s_{n-1}}), W \in W(\pi, \theta)\},$$

is valid only for generic representations $\pi$ appearing as subquotients of representations parabolically induced by ramified characters of $GL(1, F)$ and cuspidal representations of $GL(r, F)$ for $r \geq 2$, i.e. generic representations with $L$-function equal to one.
which is actually an ideal, is equal to the the full ring
\[ \mathbb{C}[q^{\pm s_1}, \ldots, q^{\pm s_{n-1}}] S_{n-1}. \]

The argument used to prove it goes like this:
For \( W \) well chosen, \( P(W, q^{s_1}, \ldots, q^{-s_{n-1}}) \) is equal to
\[ \prod_{i=1}^{n-1} 1/L(\pi, s_i). \]
We denote by \( Q \) the element \( 1/L(\pi, s) \) of \( \mathbb{C}[q^{-s}] \), so that
\[ P(W, q^{-s_1}, \ldots, q^{-s_{n-1}}) = \prod_{i=1}^{n-1} Q(q^{-s_i}). \]
Because of the functional equation of the \( L \)-function \( L(\pi,|.|^{s_1} \times \cdots \times |.|^{s_{n-1}}, s) \), denoting \( \pi^\vee \) the smooth contragredient of \( \pi \), one shows that \( I(\pi) \) also contains the product \( \prod_{i=1}^{n-1} Q'(q^{-1}q^{s_i}) \), where
\[ Q'(q^{-s}) = 1/L(\pi^\vee, s) \]
Proposition 2.1. of the paper then shows that \( Q'(q^{-1}q^s) \) and \( Q(q^s) \) are prime to one another in \( \mathbb{C}[q^{\pm s}] \), and they deduce from this that no maximal ideal
\[ I_{q^{-s_1}, \ldots, q^{-s_{n-1}}} = \{ R \in \mathbb{C}[q^{\pm a_1}, \ldots, q^{\pm a_{n-1}}], R(q^{\pm a_1}, \ldots, q^{\pm a_{n-1}}) = 0 \} \]
for \( (a_1, \ldots, a_{n-1}) \) in \( \mathbb{C}^{n-1} \), contains \( \prod_{i=1}^{n-1} Q'(q^{-1}q^{s_i}) \) and \( \prod_{i=1}^{n-1} Q(q^{-s_i}) \) together, which implies the result.
This last step is false as soon as \( n \geq 3 \), and there are \( a \) and \( b \) in \( \mathbb{C}^* \) such that \( Q(a) = Q'(q^{-1}b^{-1}) = 0 \), because then both products belong to any ideal \( I_{a, b, \ldots, z_{n-1}} \). This is the case as soon as the degree \( d(Q) \) of \( Q \) satisfies \( d(Q) \geq 1 \). However, using the functional equation of \( L(\pi,|.|^{s_1} \times \cdots \times |.|^{s_{n-1}}, s) \), and the cyclicity of \( W(q^{-z_1}, \ldots, q^{-z_{n-1}}) \) in \( W(|.|^{s_1} \times \cdots \times |.|^{s_{n-1}}, \theta) \) when \( Re(z_i) \geq Re(z_{i+1}) \). Jacquet noticed (see [3]) that one can find for every \( (a_1, \ldots, a_{n-1}) \) in \( \mathbb{C}^{n-1} \), a polynomial in \( I(\pi) \), taking the value 1 when evaluated at \( (q^{\pm a_1}, \ldots, q^{\pm a_{n-1}}) \), so \( I(\pi) \) is indeed equal to \( \mathbb{C}[q^{\pm s_1}, \ldots, q^{\pm s_{n-1}}] S_{n-1} \).

1 Preliminaries

In this section, we first recall basic facts about smooth representations of locally profinite groups. We then focus on \( G_n \), recall results from [B-Z] about derivatives, then introduce the \( L \)-function of a pair of representations of Whittaker type, we discuss espacially the unramified case.
1.1 Smooth representations, restriction and induction

When $G$ is an $l$-group (locally compact totally disconnected group), we denote by $\text{Alg}(G)$ the category of smooth complex $G$-modules. We denote by $\tilde{G}$ the group of smooth characters (smooth representations of dimension 1) of $G$. If $(\pi, V)$ belongs to $\text{Alg}(G)$, $H$ is a closed subgroup of $G$, and $\chi$ is a character of $H$, we denote by $V(H, \chi)$ the subspace of $V$ generated by vectors of the form $\pi(h)v - \chi(h)v$ for $h$ in $H$ and $v$ in $V$. This space is stable under the action of the subgroup $N_G(\chi)$ of the normalizer $N_G(H)$ of $H$ in $G$, which fixes $\chi$.

We denote by $\delta_G$ the positive character of $G$ such that if $\mu$ is a right Haar measure on $G$, and $\text{int}$ is the action of $G$ on smooth functions $f$ with compact support in $G$, given by $(\text{int}(g)f)(x) = f(g^{-1}xg)$, then $\mu \circ \text{int}(g) = \delta_G(g)\mu$ for $g$ in $G$.

The space $V(H, \chi)$ is $N_G(\chi)$-stable. Thus, if $L$ is a closed subgroup of $N_G(\chi)$, and $\delta'$ is a (smooth) character of $L$ (which will be a normalising character dual to that of normalised induction later), the quotient $V_{H,\chi} = V/V(H, \chi)$ (that we simply denote by $V_H$ when $\chi$ is trivial) becomes a smooth $L$-module for the (normalised) action $L(x + V(H, \chi)) = \delta'(l)\pi(l)v + V(H, \chi)$ of $L$ on $V_{H,\chi}$.

We denote by $V^H$ the subspace of vectors of $V$ fixed by $H$: for $H$ compact and open, the functor $V \mapsto V^H$ from $\text{Alg}(G)$ to $\text{Alg}(G_0)$ is exact ([B-H], 2.3., Corollary 1).

We say that $(\pi, V)$ in $\text{Alg}(G)$ is admissible if for any compact open subgroup $H$ of $G$, the vector space $V^H$ is finite dimensional.

If $H$ is a closed subgroup of an $l$-group $G$, and $(\rho, W)$ belongs to $\text{Alg}(H)$, we define the objects $(\text{ind}_H^G(\rho), V_c = \text{ind}_H^G(W))$ and $(\text{Ind}_H^G(\rho), V = \text{Ind}_H^G(W))$ of $\text{Alg}(G)$ as follows. The space $V$ is the space of smooth functions from $G$ to $W$, fixed under right translation by the elements of a compact open subgroup $U_H$ of $G$, and satisfying $f(hg) = \rho(h)f(g)$ for all $h$ in $H$ and $g$ in $G$. The space $V_c$ is the subspace of $V$, consisting of functions with support compact mod $H$, in both cases, the action of $G$ is by right translation on the functions.

We recall that by Frobenius reciprocity law ([B-H], 2.4.), the spaces $\text{Hom}_G(\pi, \text{Ind}_H^G(\rho))$ and $\text{Hom}_H(\pi|_H, \rho)$ are isomorphic when $\pi$ (resp. $\rho$) belongs to $\text{Alg}(G)$ (resp. $\text{Alg}(H)$).

If the group $G$ is exhausted by compact subsets (which is the case of closed subgroups of $G_n$), and $(\pi, V)$ is irreducible, it is known ([B-H], 2.6., Corollary 1) that the center $Z$ of $G$ acts on $V$ by the so-called central character of $\pi$ which we will denote $c_\pi$. When $G = G_n$, then $Z$ identifies with $F^*$. By definition, the real part $Re(\chi)$ of a character $\chi$ of $F^*$ is the real number $r$ such that $|\chi(t)|_C = |t|^r$, where $|z|_C = \sqrt{z\overline{z}}$ for $z$ in $C$.

1.2 Parabolic induction and segments for $\text{GL}(n)$

Now we focus on the case $G = G_n$, we will only consider smooth representations of its closed subgroups. It is known that irreducible representations of $G_n$ are admissible (see [C2]).
If \( n \geq 1 \), let \( \bar{n} = (n_1, \ldots, n_t) \) be a partition of \( n \) of length \( t \) (i.e. an ordered set of \( t \) positive integers whose sum is \( n \)), we denote by \( M_{\bar{n}} \) to be the Levi subgroup of \( G_n \), of matrices \( \text{diag}(g_1, \ldots, g_t) \), with each \( g_i \) in \( G_{n_i} \), by \( N_{\bar{n}} \) the unipotent subgroup of matrices \( \begin{pmatrix} I_{n_1} & * & * \\ \vdots & \ddots & \vdots \\ I_{n_t} \end{pmatrix} \), and by \( P_{\bar{n}} \) the standard parabolic subgroup \( M_{\bar{n}}N_{\bar{n}} \) (where \( M_{\bar{n}} \) normalises \( N_{\bar{n}} \)). Note that \( M_{\{1, \ldots, 1\}} \) is equal to \( A_n \), and \( N_{\{1, \ldots, 1\}} = N_n \). For each \( i \), let \( \pi_i \) be a smooth representation of \( G_{n_i} \), then the tensor product \( \pi_1 \otimes \cdots \otimes \pi_t \) is a representation of \( M_{\bar{n}} \), which can be considered as a representation of \( P_{\bar{n}} \) trivial on \( N_{\bar{n}} \). We will use the product notation

\[
\pi_1 \times \cdots \times \pi_t = \text{Ind}_{P_{\bar{n}}}^{G_n}(\otimes_{i=1}^{t} \pi_i)
\]

for the normalised parabolic induction. Parabolic induction preserves finite length and admissibility (see [13Z] or [22]).

We say that an irreducible representation \((\rho, V)\) of \( G_n \) is cuspidal, if the Jacquet module \( V_{N_n} \) is zero whenever \( \bar{n} \) is a proper partition of \( n \) (i.e. we exclude \( \bar{n} = (n) \)).

Suppose that \( \bar{n} = (m, \ldots, m) \) is a partition of \( n \) of length \( l \), and that \( \rho \) is a cuspidal representation of \( G_m \). Then Theorem 9.3. of [Z] implies that the \( G_n \)-module \( \nu^{-l(l-1)}\rho \times \nu^{-l(l-2)}\rho \times \cdots \times \nu^{-1}\rho \times \rho \) has a unique irreducible quotient which we denote \( [\nu^{-l(l-1)}\rho, \nu^{-l(l-2)}\rho, \ldots, \nu^{-1}\rho, \rho] \). We will call such a representation a segment, it is known that segments are the quasi square integrable representations of \( G_n \), but we won’t need this result.

We end this paragraph with a word about induced representations of Langlands’ type:

**Definition 1.1.** Let \( \Delta_1, \ldots, \Delta_t \) be segments of respectively \( G_{n_1}, \ldots, G_{n_t} \), and suppose that \( \text{Re}(c_{\Delta_i}) \geq \text{Re}(c_{\Delta_{i+1}}) \). Let \( n = n_1 + \cdots + n_t \), then the representation \( \Delta_1 \times \cdots \times \Delta_t \) of \( G_n \) is said to be induced of Langlands’ type.

These representations enjoy many remarkable properties, some of which we will recall later, here is a first one (which is the main result of [33]).

**Proposition 1.1.** Let \( \pi \) be induced of Langlands’ type, then \( \pi \) has a unique irreducible quotient \( Q(\pi) \). Moreover, considering that isomorphic representations are equal, the map \( \pi \mapsto Q(\pi) \) gives a bijection between the set of induced representations of Langlands’ type of \( G_n \), and the set of irreducible representations of \( G_n \).

### 1.3 Bernstein-Zelevinsky derivatives

For \( n \geq 2 \) we denote by \( U_n \) the group of matrices of the form \( \begin{pmatrix} I_{n-1} & \nu \\ 1 & 1 \end{pmatrix} \).
For $n > k \geq 1$, the group $G_k$ embeds naturally in $G_n$, and is given by matrices of the form $\text{diag}(g, I_{n-k})$. We denote by $P_n$ the mirabolic subgroup $G_{n-1}U_n$ of $G_n$ for $n \geq 2$, and $P_1 = \{ 1_{G_1} \}$. If one sees $P_{n-1}$ as a subgroup of $G_{n-1}$ itself embedded in $G_n$, then $P_{n-1}$ is the normaliser of $\theta_1$ in $G_{n-1}$ (i.e. if $g \in G_{n-1}$, then $\theta(g^{-1}ug) = \theta(u)$ for all $u \in U_n$ if and only if $g \in P_{n-1}$). We define the following functors:

- The functor $\Phi^-$ from $\text{Alg}(P_k)$ to $\text{Alg}(P_{k-1})$ such that, if $(\pi, V)$ is a smooth $P_k$-module, $\Phi^- V = V_{U_k, \theta}$, and $P_{k-1}$ acts on $\Phi^-(V)$ by $\Phi^- \pi(p)(v + V(U_k, \theta)) = \delta_{P_k}(p)^{-1/2} \pi(p)(v + V(U_k, \theta))$.
- The functor $\Phi^+$ from $\text{Alg}(P_{k-1})$ to $\text{Alg}(P_k)$ such that, for $\pi$ in $\text{Alg}(P_{k-1})$, one has $\Phi^+ \pi = \text{ind}_{P_{k-1}}^{P_k} (\delta_{P_k}^{1/2} \pi \otimes \theta)$.
- The functor $\hat{\Phi}^+$ from $\text{Alg}(P_{k-1})$ to $\text{Alg}(P_k)$ such that, for $\pi$ in $\text{Alg}(P_{k-1})$, one has $\hat{\Phi}^+ \pi = \text{Ind}_{P_{k-1}}^{P_k} (\delta_{P_k}^{1/2} \pi \otimes \theta)$.
- The functor $\Psi^-$ from $\text{Alg}(P_k)$ to $\text{Alg}(G_{k-1})$, such that if $(\pi, V)$ is a smooth $P_k$-module, $\Psi^- V = V_{U_k, 1}$, and $G_{k-1}$ acts on $\Psi^-(V)$ by $\Psi^- \pi(g)(v + V(U_k, 1)) = \delta_{P_k}(g)^{-1/2} \pi(g)(v + V(U_k, 1))$.
- The functor $\Psi^+$ from $\text{Alg}(G_{k-1})$ to $\text{Alg}(P_k)$, such that for $\pi$ in $\text{Alg}(G_{k-1})$, one has $\Psi^+ \pi = \text{ind}_{G_{k-1}}^{P_k} (\delta_{P_k}^{1/2} \pi \otimes 1) = \delta_{P_k}^{1/2} \pi \otimes 1$.

These functors have the following properties which can be found in [B-Z]:

**Proposition 1.2.** a) The functors $\Phi^-, \Phi^+, \Psi^-$, and $\Psi^+$ are exact.
b) $\Psi^-$ is left adjoint to $\Psi^+$.
c) $\Phi^- \Psi^+ = 0$ and $\Psi^- \Phi^+ = 0$.
d) $\Psi^- \Psi^+ \simeq \text{Id}$ and $\Phi^- \Phi^+ \simeq \text{Id}$.
e) One has the exact sequence $0 \to \Phi^+ \Phi^- \to \text{Id} \to \Psi^- \Psi^+ \to 0$.

Following [C-P], if $\tau$ belongs to $\text{Alg}(P_n)$, we will denote $(\Phi^-)^{k} \tau$ by $\tau_{(k)}$, and as usual, $\tau^{(k)}$ will be defined as $\Psi^{-\tau_{(k-1)}}$.

Because of e), $\tau$ has a natural filtration of $P_n$-modules $0 \subset \tau_n \subset \cdots \subset \tau_1 = \tau$, where $\tau_k = \Phi^{+k-1} \Phi^{-k-1} \tau$. We will use the notation $\tau_{(k),i}$ for $(\tau_{(k)})_i$. The following observation is just a restatement of the definitions:

**Lemma 1.1.** If $\tau$ belongs to $\text{Alg}(P_n)$, then $\tau_k = \Phi^+(\tau_{(1),k-1})$ for $k \geq 1$. 

*Documenta Mathematica 18 (2013) 1191–1214*
1.4 Representations of Whittaker Type and their L-functions

We recall that we fixed a character θ of conductor δ in the introduction.

**Definition 1.2.** Let π be an admissible representation of \( G_n \), we say that π is of Whittaker type if \( Hom(\pi, Ind_{N_n}^{G_n}(\theta)) \) is of dimension 1, or equivalently, according to Frobenius reciprocity law, if the space \( Hom_N(\pi, \theta) \) is of dimension 1. We denote by \( W(\pi, \theta) \) the image of \( \pi \) in \( Ind_{N_n}^{G_n}(\theta) \), it is called the Whittaker model of \( \pi \) with respect to \( \theta \), it is a quotient of \( \pi \).

Being of Whittaker type does not depend on the character \( \theta \) of \( (F, +) \), as another non trivial character \( \theta' \) of \( (F, +) \) will give birth to a character \( \theta' \) of \( N_n \), conjugate to \( \theta \) by \( A_n \).

In terms of derivatives, as the representation \( Ind_{N_n}^{G_n}(\theta) \) is isomorphic to \((\h^+)^n-1\psi^+(1)\), where 1 is the trivial representation of \( G_0 \), applying \( b \) and \( b' \) of Proposition [1.2], we obtain that \( Hom_N(\pi, \theta) \simeq \mathbb{C} \) if and only if \( \pi^{(n)} = 1 \).

Applying this to product of segments, and using the rules of “derivation” given in Lemma 3.5 of [B-Z] and Proposition 9.6. of [Z], we obtain that if \( \Delta_1, \ldots, \Delta_t \) are segments of \( G_n \), respectively, the representation \( \pi = \Delta_1 \times \cdots \times \Delta_t \) of \( G_n \) (for \( n = n_1 + \cdots + n_t \)) is of Whittaker type. If the segments \( \Delta_i \) are ordered so that \( \pi \) is of Langlands’ type, we can say more according to the main result of [LS3].

**Proposition 1.3.** For \( n \geq 1 \), let \( \pi \) be a representation of \( G_n \), which is induced of Langlands’ type, then it has an injective Whittaker model, i.e. \( \pi \simeq W(\pi, \theta) \) (equivalently \( \pi \) embeds in \( Ind_{N_n}^{G_n}(\theta) \)).

If \( \pi \) is irreducible and embeds in \( Ind_{N_n}^{G_n}(\theta) \), it is a well-known theorem of Gelfand and Kazhdan ([G-K]) that the multiplicity of \( \pi \) in \( Ind_{N_n}^{G_n}(\theta) \) is 1, we then say that \( \pi \) is generic. We recall (Theorem 9.7 of [Z]), that every generic representation \( \pi \) of \( GL(n, F) \) can be written uniquely, up to permutation of the terms in the product, as a commutative product of unlinked (see 4.1. of [Z]) segments

\[
[\nu^{-(k_1(\pi)-1)}\rho_1(\pi), \ldots, \rho_1(\pi)] \times \cdots \times [\nu^{-(k_t(\pi)-1)}\rho_t(\pi), \ldots, \rho_t(\pi)].
\]

In particular, generic representations are the representations of Langlands’ type which are irreducible.

We now recall from [LFPS2], some results about the \( L \)-function of a pair of representations of Whittaker type. Let \( \pi \) be a representation of \( G_n \) of Whittaker type, and \( \pi' \) be a representation of Whittaker type of \( G_m \), with respective Whittaker models \( W(\pi, \theta) \) and \( W(\pi', \theta') \), for \( n \geq m \geq 1 \).

When \( n > m \), and \( W \) and \( W' \) are respectively in \( W(\pi, \theta) \) and \( W(\pi', \theta') \), we write

\[
I(W, W', s) = \int_{N_n \setminus G_m} W \left( \begin{pmatrix} g \\ I_{n-m} \end{pmatrix} \right) W'(g) \nu(g)^{s-(n-m)/2} \, dg.
\]
When \( n = m \), and \( W \) and \( W' \) are respectively in \( W(\pi, \theta) \) and \( W(\pi', \theta^{-1}) \), \( \phi \) is in \( \mathcal{C}_c^\infty(F^n) \), and \( \eta \) is the row vector \((0, \ldots, 0, 1)\) in the space \( \mathcal{M}(1, n, F) \) of row matrices \( 1 \) by \( n \) with entries in \( F \), we write
\[
I(W, W', \phi, s) = \int_{N_n \backslash G_n} W(g)W'(g)\phi(\eta g)\nu(g)^sdg.
\]
It is shown in [J-P-S 2] that these integrals converge absolutely for \( Re(s) \) large, and define elements of \( \mathbb{C}(q^{-s}) \). If \( n > m \), the integrals \( I(W, W', s) \) (which we shall also write \( I(W, W, s) \) when convenient) span, when \( (W, W') \) varies in \( W(\pi, \theta) \times W(\pi', \theta^{-1}) \), a fractional ideal of \( \mathbb{C}[q^s, q^{-s}] \), which is generated by a unique Euler factor \( L(\pi, \pi', s) \). If \( n = m \), the integrals \( I(W, W', \phi, s) \) span, when \( (W, W', \phi) \) varies in \( W(\pi, \theta) \times W(\pi', \theta^{-1}) \times \mathbb{C}_c^\infty(F^n) \), a fractional ideal of \( \mathbb{C}[q^s, q^{-s}] \), which is generated by a unique Euler factor \( L(\pi, \pi', s) \). If \( n < m \), we define \( L(\pi, \pi', s) \) to be \( L(\pi', \pi, s) \). We recall Proposition 9.4 of [J-P-S 2].

**Proposition 1.4.** For \( n \geq m \geq 1 \), if \( \pi = \Delta_1 \times \cdots \times \Delta_i \) is a representation of \( G_n \) induced of Langlands’ type, and \( \pi' = \Delta'_1 \times \cdots \times \Delta'_j \) is a representation of \( G_m \) induced of Langlands’ type, then \( L(\pi, \pi', s) = \prod_{i,j} L(\Delta_i, \Delta'_j, s) \).

Finally, we recall that it is proved in Section 5 of [J-P-S 2], that if \( \pi \) is a generic representation of \( G_n \), and \( \chi \) is a character of \( G_1 \), one has the equality \( L(\pi, \chi, s) = L(\chi \otimes \pi, s) \) between the Rankin-Selberg \( L \)-function on the left, and the Godement-Jacquet \( L \)-function on the right.

### 1.5 Unramified representations

We say that a representation \((\pi, V)\) of \( G_n \) is **unramified** (or **spherical**), if it admits a nonzero \( G_n(\mathbb{O}) \)-fixed vector in its space. If it is the case, we recall that the Hecke convolution algebra \( \mathcal{H}_n \) (whose elements are the functions with compact support on \( G_n \), which are left and right invariant under \( G_n(\mathbb{O}) \)), acts on \( V^{G_n(\mathbb{O})} \) (when \( V^{G_n(\mathbb{O})} \) is of dimension 1, the action is necessarily by a character). The Hecke algebra \( \mathcal{H}_n \) is commutative and isomorphic by the Satake isomorphism to the algebra \( \mathbb{C}[X_1^{\pm 1}, \ldots, X_n^{\pm 1}]^{S_n} \) according to [Sat]. Hence, a character of \( \mathcal{H}_n \) is associated to a unique set of nonzero complex numbers \( \{z_1, \ldots, z_n\} \), corresponding to the evaluation \( P \mapsto P(z_1^{\pm 1}, \ldots, z_n^{\pm 1}) \) from \( \mathbb{C}[X_1^{\pm 1}, \ldots, X_n^{\pm 1}]^{S_n} \) to \( \mathbb{C} \). It is known that when \( \pi \) is unramified and irreducible (see Section 4.6 in [BG] for example), then \( V^{G_n(\mathbb{O})} \) is one dimensional, and that the corresponding character of \( \mathcal{H}_n \) determines \( \pi \), in which case the associated set of nonzero complex numbers is called the **Satake parameter** of \( \pi \).

We recall with proofs, some classical facts about parabolically induced spherical representations.

**Proposition 1.5.** Let \( \pi_i \) be a representation of \( G_n_i \) for \( i \) between 1 and \( t \), \( n = n_1 + \cdots + n_t \), \( M_n \) be the standard Levi subgroup of \( n \) corresponding to the partition \((n_1, \ldots, n_t)\), and \( \pi = \pi_1 \times \cdots \times \pi_t \). The vector space \( \text{Hom}_{G_n(\mathbb{O})}(1, \pi) \) is isomorphic to \( \text{Hom}_{M_n \cap G_n(\mathbb{O})}(1, \pi_1 \otimes \cdots \otimes \pi_t) \).
Proposition 1.5, as the representations by Theorem 4.2 of \([Z]\), hence it is its irreducible quotient. It is spherical from and submodule of \(\nu\). If \(\chi\) is spherical, then it is the irreducible quotient of \(\pi\) is spanned by \(\nu\). We have the following corollaries to this.

Corollary 1.1. Let \(\chi_1, \ldots, \chi_n\) be unramified characters of \(F^*\), then the representation \(\pi = \chi_1 \times \cdots \times \chi_n\) of \(G_n\) is unramified, and \(\pi G_n(\mathcal{O})\) is one dimensional. In particular, by the exactness of the functor \(V \mapsto V_{G_n(\mathcal{O})}\) from \(\text{Alg}(G_n)\) to \(\text{Alg}(G_0)\), the representation \(\pi\) has only one irreducible spherical subquotient.

When \(\pi = |^{s_1} \times \cdots \times |^{s_n}\) is an unramified representation of \(G_n\) induced from the Borel subgroup, we just saw that \(\pi G_n(\mathcal{O})\) is of dimension 1. The character of \(\mathcal{H}_n\), given by its action on \(\pi G_n(\mathcal{O})\) corresponds to the set \(\{q^{-s_1}, \ldots, q^{-s_n}\}\). If \(\pi\) is moreover of Langlands’ type, it is the Satake parameter of \(Q(\pi)\) thanks to the next corollary.

Corollary 1.2. If \(\pi = \chi_1 \times \cdots \times \chi_n\) is as above, but moreover of Langlands’ type, then it is the irreducible quotient of \(\pi\) which is spherical. In particular, \(\pi\) is spanned by \(\pi G_n(\mathcal{O})\).

Proof. One checks by induction on \(n\) the following assertion: the representation \(\pi\) can be written \(\pi_1 \times \cdots \times \pi_t\), where for each \(i\), the representation \(\pi_i\) is equal to \(\nu^{(l_i-1)/2} \mu_i \times \nu^{(l_i-3)/2} \mu_i \times \cdots \times \nu^{(1-l_i)/2} \mu_i\) for a positive integer \(l_i\) and an unramified character \(\mu_i\) of \(F^*\), and such that the segments \([\pi_i]\) are unlinked, the character \(\nu^{(1-l_i)/2} \mu_i\) is equal to \(\chi_i\), and the segment \([\pi_i]\) contains any other \([\pi_j]\) in which \(\chi_j\) occurs (i.e. if \(\chi_n = \nu^r \mu_r\) for \(r \in \mathbb{Q}\) between \((1-l_i)/2\) and \((l_i-1)/2\)), then \(r = (1-l_i)/2\) and \(l_i \leq l_j\). From \([Z]\), Section 2, the irreducible quotient of \(\pi_i\) is the irreducible submodule of \(\nu^{(1-l_i)/2} \mu_i \times \cdots \times \nu^{(l_i-1)/2} \mu_i\), i.e. the character \(\mu_i = \mu_i \circ \det\) of \(G_i\). The representation \(\mu_i \times \cdots \times \mu_i\) is thus a quotient of \(\pi\), which is irreducible by Theorem 4.2 of \([Z]\), hence it is its irreducible quotient. It is spherical from Proposition 1.1 as the representations \(\mu_i\) are spherical.

Corollary 1.3. A segment \(\Delta\) of \(G_n\), for \(n \geq 2\), is always ramified.

Proof. If \(\Delta\) was unramified, as it is irreducible, its Satake parameter would be equal to a set \(\{q^{-s_1}, \ldots, q^{-s_n}\}\), hence the same as that of \(Q(\pi)\), for \(\pi = |^{s_1} \times \cdots \times |^{s_n}\) (if we order the \(s_i\)’s in a correct way). In particular, \(\Delta\) would be equal to \(Q(\pi)\), which is absurd according to Proposition 1.1.

For unramified representations of Langlands’ type, normalised spherical Whittaker functions are test functions for \(L\) factors. If \(\pi\) is an unramified representation of Langlands’ type, we denote by \(W_\pi^0\) the spherical function in \(W(\pi, \theta)\) which is equal to 1 on \(G_n(\mathcal{O})\), and call it the normalised spherical Whittaker function of \(\pi\). In \([S]\), Shintani gave an explicit formula for \(W_\pi^0\) in terms of the Satake parameter of \(\pi\).
Using this formula, Jacquet and Shalika found (Proposition 2.3 of [J-S, and equality (3) in Section 1 of [J-S 2]), for \( \pi \) and \( \pi' \) two unramified representations of Langlands' type (see the discussion after Equations (3) and (4) below) of \( G_n \) and \( G_m \) respectively, and for correct normalisations of Haar measures, the equalities:

\[
L(\pi,\pi',s) = I(W^{0}_\pi W^{0}_{\pi'}, 1_{\Omega^n}, s) = \int_{A_n} W^{0}_\pi(a) W^{0}_{\pi'}(a) 1_{\Omega}(a) d_B(a)^{-1} \nu(a)^s d^* a,
\]

(3)

when \( n = m \), and

\[
L(\pi,\pi',s) = I(W^{0}_\pi W^{0}_{\pi'}, s) = \int_{A_n} W^{0}_\pi \left( a I_{n-m}\right) W^{0}_{\pi'}(a) d_B(a)^{-1} \nu(a)^{s-(n-m)/2} d^* a
\]

(4)

when \( n > m \).

In the aforementioned papers, Jacquet and Shalika work with generic representations, however, their proofs extend verbatim to unramified representations of Langlands' type. Indeed, the formula for \( W^0_\pi \) in terms of Satake parameters is still valid, and thanks to Proposition (4), the factor \( L(\pi,\pi',s) \) is still equal to (using notations of [J-S] and [J-S 2]) the Artin factor \( 1/det(1-q^{-s}A \otimes A') \) for \( A \) and \( A' \) diagonal matrices corresponding to the Satake parameters of \( \pi \) and \( \pi' \) respectively. In particular, when \( \pi \) is generic, \( W^0_\pi \) is the essential vector of \( \pi \).

Now let \( \pi = \Delta_1 \times \cdots \times \Delta_t \) be a generic representation of \( G_n \), written as a unique product of the unlinked segments \( \Delta_i = [\nu^{-(k_i)}(\pi_i)^{\frac{1}{2}}] \), and \( \pi' = \mu_1 \times \cdots \times \mu_m \) be an unramified representation of \( G_m \) of Langlands' type, for \( 1 \leq m \leq n \). One has, according to Proposition (4) and Theorem 8.2. of [J-P-S 2] (whose proof is independent of [J-P-S]), the equality of Rankin-Selberg \( L \)-functions \( L(\pi,\pi',s) = \prod_{i,j} L(\rho_i(\pi), \mu_j, s) \).

We notice that \( L(\rho_i(\pi), \mu_j, s) \) is equal to 1 unless \( \rho_i(\pi) \) is an unramified character of \( G_1 \). Hence, one has the equality

\[
L(\pi,\pi',s) = \prod_{(i,\rho_i(\pi) \in F^*/D^*)} L(\rho_i(\pi), \mu_j, s).
\]

This incites us to introduce the following representation.

**Definition 1.3.** Let \( \pi = \Delta_1 \times \cdots \times \Delta_t \) be a generic representation of \( G_n \), with \( \Delta_i = [\nu^{-(k_i)}(\pi_i)^{\frac{1}{2}}] \). Let \( r \) be the cardinality of the set \( \{\rho_j(\pi), \rho_j(\pi) \in F^*/D^*\} \). When this set is non empty, denote by \( \chi_1, \ldots, \chi_r \) its elements ordered such that \( \text{Re}(\chi_i) \geq \text{Re}(\chi_{i+1}) \) for \( 1 \leq i \leq r - 1 \). We define \( \pi_u \) as the trivial representation of \( G_0 \) when \( r = 0 \), and as the unramified representation of Langlands type \( \chi_1 \times \cdots \times \chi_r \) of \( G_r \) when \( r \geq 1 \).
Let $\pi$ a generic representation of $G_n$, and $\pi'$ be an unramified representation of $G_m$ of Langlands' type, with $1 \leq m \leq n$. If we set $L(\pi_u, \pi', s) = 1$ when $\pi_u$ is the trivial representation of $G_0$, we have, according to Proposition [L4]

$$L(\pi, \pi', s) = L(\pi_u, \pi', s).$$ (5)

From now on, we will order the segments $\Delta_i$ in the generic representation $\pi$, such that $\rho_i(\pi)$ is an unramified character $\chi_i$ of $G_i$ for $1 \leq i \leq r$, is not such a character for $i > r + 1$, and $\text{Re}(\chi_i) \geq \text{Re}(\chi_{i+1})$ for $1 \leq i \leq r - 1$.

2 Mirabolic restriction, sphericity, and restriction of Whittaker functions

In this section, we first give results on the derivative functors and how they act on subspaces fixed by compact subgroups, then we recall some results from [C-P] about their interpretation in terms of restriction of Whittaker functions.

We introduce a few more notations, in order to get a handy parametrisation of the diagonal torus of $G_n$, in terms of simple roots. For $k \leq n$, let $Z_k$ be the center of $G_k$ naturally embedded in $G_n$; we parametrise it by $F^*$ using the morphism $\beta_k : z_k \mapsto \text{diag}(z_k I_1, I_{n-k})$. Hence the maximal torus $A_n$ of $G_n$ is the direct product $Z_1 Z_2 \ldots Z_{n-1} Z_n$. We will sometimes (but not always) omit the $\beta_k$'s in this parametrisation and write $(z_1, \ldots, z_n)$ for the element $\beta_1(z_1) \ldots \beta_n(z_n)$ of $A_n$. Notice that the $i$-th simple root $\alpha_i$ has the property that $\alpha_i(z_1, \ldots, z_n) = z_i$.

2.1 Mirabolic restriction and sphericity

We first give a corollary of Proposition [L2] about a concrete interpretation of $\Phi^-$, when restricted to $\Phi^+(\text{Alg}(P_n))$. Property d) of the aforementioned proposition says that $\Phi^-$ sends $\Phi^+\tau$ surjectively onto a $P_n$-module isomorphic to $\tau$. Writing $\Phi^+\tau$ as $\text{Ind}_{P_{n+1}}^{P_n} (\delta_{P_{n+1}}^{1/2} \tau \otimes \theta)$, we want to make the map $\Phi^-$ explicit between $\text{Ind}_{P_{n+1}}^{P_n} (\delta_{P_{n+1}}^{1/2} \tau \otimes \theta)$ and $\tau$.

**Proposition 2.1.** For $n \geq 1$, if $\tau$ belongs to $\text{Alg}(P_n)$, then $\Phi^-$ identifies with the map $f \mapsto f(I_{n+1})$ from $\Phi^+\tau$ to $\tau$.

**Proof.** Call $E_I$ the map $f \mapsto f(I_{n+1})$ from $\Phi^+\tau$ to $\tau$. Let’s first show that $E_I$ is surjective. If $v$ belongs to the space $V$ of $\tau$, let $U = I_n + M(n, \mathbb{F}^2)$ be a congruence subgroup of $G_n$, with $l$ large enough for $v$ to be fixed by $U' = U \cap P_n$. Call $f$ the function from $P_n$ to $V$, defined by

$$f \begin{pmatrix} pu \\ x \\ 1 \end{pmatrix} = \delta_{P_{n+1}}^{1/2} (p) \theta(x_n) \tau(p)v,$$

for $p \in P_n$, $u \in U$, $x \in F^n$ with bottom coordinate $x_n$, and $f(p') = 0$ when $p'$ is not in $P_n U_{n+1} U$. The map $f$ is well defined, because $v$ is fixed by $U'$. It
is smooth as it is right invariant under $U_{n+1}(\mathfrak{O})U$ (as $\theta$ is trivial on $\mathfrak{O}$). One checks (see the proof of next Proposition for a detailed similar computation), that $f$ satisfies the requested left invariance under $P_nU_{n+1}$, so that $f \in \Phi^+(V)$. Finally $f(I_{n+1}) = v$, thus $f$ is a preimage of $v$ via $E_I$.

An easy adaptation of the Proposition 1.1. of [C-P] then shows that the $P_n$-submodule $\Phi^+\tau(U_{n+1}, \theta)$ of $\Phi^+(\tau)$ is equal to $\ker(E_I)$. As a consequence, the map $E_I$ induces an isomorphism $\overline{E_I}$ between $\Phi^{-}\Phi^+(\tau)$ and $\tau$, which is $P_n$-equivariant. Hence, the following diagram, with the right isomorphism equal to $\overline{E_I}$, commutes:

\[
\begin{array}{ccc}
\Phi^+ (\tau) & \xrightarrow{\Phi^{-}} & \Phi^+ (\tau) \\
\| & \| & \|
\Phi^+ (\tau) & \xrightarrow{E_I} & \tau
\end{array}
\]

We recall that for $n \geq 1$, as a consequence of the Iwasawa decomposition, any element $g$ of $G_n$ can be written in the form $zpk$ with $z \in F^*$, $p \in P_n$, and $k$ in $G_n(\mathfrak{O})$. We now notice that the restriction of $\Phi^-$ to $(\Phi^+\tau)_{P_n(\mathfrak{O})}$ is surjective onto $\tau_{P_n(\mathfrak{O})}$.

**Proposition 2.2.** For $n \geq 1$, the map $f \mapsto f(I_{n+1})$ from $(\Phi^+\tau)_{P_n(\mathfrak{O})}$ to $\tau_{P_n(\mathfrak{O})}$ is surjective.

**Proof.** Let $v_0$ be a vector in the space of $\tau$ which is $P_n(\mathfrak{O})$-invariant, we claim that the function $f$ defined by

\[
f \left( \begin{array}{c}
zpk \\
x \\
1
\end{array} \right) = \delta^{1/2}_{P_n(\mathfrak{O})} (\tau(p)(x_0)1_{\Omega} \cdot (z)\tau(p)\varepsilon_0),
\]

for $z \in F^*$, $p \in P_n$, $k$ in $G_n(\mathfrak{O})$, and $x \in F^n$ (with $x$ the transpose of $(x_1, \ldots, x_n)$), is a preimage of $v_0$ in $(\Phi^+\tau)_{P_n(\mathfrak{O})}$.

First we check that $f$ is well-defined: if $zpk = z'p'k'$, this implies that $z'$ is equal to $z$ mod $\mathcal{O}^*$, and $p'$ is equal to $p$ mod $P_n(\mathfrak{O})$. Hence $\delta^{1/2}_{P_n(\mathfrak{O})}(\tau=p') = \delta^{1/2}_{P_n(\mathfrak{O})}(\tau(p'))$, $1_\Omega (z') = 1_\Omega (z)$, and $\tau(p')\varepsilon_0 = \tau(p)\varepsilon_0$ as $v_0$ is $P_n(\mathfrak{O})$-invariant.

Then we check that $f$ indeed belongs to $\Phi^+(\tau)$. Let $p_0$ belong to $P_n$ embedded in $P_{n+1}$ by $p \mapsto \text{diag}(p,1)$, let $u_0$ belong to $U_{n+1}$, and $p_1$ belong to $P_{n+1}$, we need to check the relations

\[
f(p_0 p_1) = \delta^{1/2}_{P_{n+1}} (\tau(p_0)\tau(p_1))f(p_1)
\]

and

\[
f(u_0 p_1) = \theta(u_0)f(p_1).
\]

Write $p_1 = \left( \begin{array}{c}
zpk \\
x \\
1
\end{array} \right)$, we have $f(p_0 p_1) = f \left( \begin{array}{c}
\text{zpopk} \\
p_0 x \\
1
\end{array} \right)$. As $\theta((p_0 x)_n) = \theta(x_0)$, we have

\[
f(p_0 p_1) = \delta^{1/2}_{P_{n+1}} (\text{zpopk})\theta(x_0)1_{\Omega} \cdot (z)\tau(p_0)\varepsilon_0 = \delta^{1/2}_{P_{n+1}} (\tau(p_0))f(p_1)
\]

**Documenta Mathematica 18 (2013) 1191–1214**
Proposition 2.3. For \( n \geq 1 \), if \( \tau \) belongs to \( \text{Alg}(P_n) \), then \( \Phi^- \) maps \( \tau_{P_n(D)} \) surjectively onto \( \tau_{P_{n-1}(D)}^{(1)} \), and \( \Psi^- \) maps \( \tau_{G_{n-1}(D)} \) surjectively onto \( \tau_{(1)G_{n-1}(D)} \).

Proof. For the first part, we use the filtrations \( 0 \subset \tau_n \subset \ldots \subset \tau_1 = \tau \) of \( \tau \), and \( 0 \subset \tau_{(1),n-1} \subset \ldots \subset \tau_{(1),1} = \tau_{(1)} \) of \( \tau_{(1)} \). But \( \tau_{(1)} \) equals \( \Phi^+(\tau_{(1),i-1}) \) because of Lemma 1.1, so that \( \Phi^- \) maps \( \tau_{P_n(D)} \) onto \( \tau_{P_{n-1}(D)}^{(1)} \) surjectively according to Proposition 2.2. In particular, \( \Phi^- \) maps \( \tau_{P_n(D)} \), hence \( \tau_{P_n(D)} \) (as \( \tau_{P_n(D)} \subset \tau_{P_n(D)} \)), onto \( \tau_{P_{n-1}(D)}^{(1)} = \tau_{P_{n-1}(D)}^{(1)} \) surjectively.

\( \Psi^- \) maps \( \tau_{G_{n-1}(D)} \) surjectively onto \( \tau_{(1)G_{n-1}(D)} \), because \( \Psi^- \) is surjective from \( \tau \) to \( \tau_{(1)} \), and the functor \( V \mapsto V^{G_{n-1}(D)} \) is exact from \( \text{Alg}(G_{n-1}) \) to \( \text{Alg}(G_0) \) as \( G_{n-1}(D) \) is compact open in \( G_{n-1}(\mathbb{A}) \) (\( \tau \) is a \( G_{n-1}(\mathbb{A}) \)-module by restriction). \( \square \)

2.2 Mirabolic restriction for Whittaker functions

We start by recalling Proposition 1.1 of [C-P], which gives an interpretation of \( \Phi^- \) in terms of restriction of Whittaker functions.

Proposition 2.4. For \( k \geq 2 \), and any submodule \( \tau \) of \( (\rho, C^\infty(N_k \backslash K, \theta)) \) (where \( \rho \) denotes the action of \( P_k \) by right translation), the map \( R : W \mapsto \delta_{P_{k-1}}^{-1/2}W |_{P_{k-1}} \) is \( P_{k-1} \)-equivariant from \( (\rho, C^\infty(N_k \backslash K, \theta)) \) to \( (\rho, C^\infty(N_{k-1} \backslash K_{k-1}, \theta)) \), with kernel \( \tau(U_k, \theta) \). Hence it induces a \( P_{k-1} \)-modules isomorphism between \( \Phi^- \tau \) and \( \text{Im}(R) \subset C^\infty(N_{k-1} \backslash K_{k-1}, \theta) \), so that \( (\rho, \text{Im}(R)) \) is a model for \( \Phi^- \tau \).
Notice that for \( k \geq 2 \), if \( g \in G_{k-1} \) equals \( zpk \) with \( z \in F^*, p \in P_{k-1}, \) and \( k \in G_{n-1}(\mathbb{D}) \), then the absolute value of \( z \) depends only on \( g \), so we can write \( |z(g)| \). We now state a proposition that follows from the proofs of Proposition 1.6. of [C-P], about the interpretation of \( \Psi^- \) in terms of Whittaker functions.

**Proposition 2.5.** Let \( \tau \) be a \( P_k \)-submodule of \( C^\infty(N_k \backslash P_k, \theta) \), and suppose that \( \tau^{(1)} \) is a \( G_{k-1} \)-module with central character \( c \). Then, for any \( W \) in \( \tau \), for any \( g \) in \( G_{k-1} \), the quantity \( c^{-1}(z)|z|^{1-(k-1)/2}W(diag(zg,1)) \) is constant whenever \( z \) is in a punctured neighbourhood of zero (maybe depending on \( g \) in \( F^* \)).

**Remark 2.1.** Notice that in the proof of the Proposition 1.6. of [C-P], \( \tau \) is of a particular form, and \( \tau^{(1)} \) is supposed to be irreducible. The only fact that is actually needed is that \( \tau^{(1)} \) has a central character.

This has the following consequence.

**Corollary 2.1.** For \( k \geq 2 \), let \( \tau \) be a \( P_k \)-submodule of \( (\rho, C^\infty(N_k \backslash P_k, \theta)) \), and suppose that \( \tau^{(1)} \) is a \( G_{k-1} \)-module with central character \( c \), then \( \Psi^- \) identifies with the map

\[
S : W \mapsto [g \mapsto \lim_{z \to 0} c^{-1}(z)|z|^{1-(k-1)/2}W(diag(zg,1))\delta_{P_k}^{-1/2}(g)]
\]

from \( \tau \) to \( C^\infty(N_{k-1} \backslash G_{k-1}, \theta) \). To be more precise, \( S \) has kernel \( \tau(U_k,1) \), and it induces a \( G_{k-1} \)-modules isomorphism \( \overline{S} \) between \( \tau^{(1)} \) and \( S(\tau) \subset C^\infty(N_{k-1} \backslash G_{k-1}, \theta) \).

**Proof.** For \( W \) in \( \tau \), call \( \overline{W} \) its image in \( \tau^{(1)} \), and call \( S(W) \) the function

\[
[g \mapsto \lim_{z \to 0} c^{-1}(z)|z|^{1-(k-1)/2}W(diag(zg,1))\delta_{P_k}^{-1/2}(g)]
\]

in \( C^\infty(N_{k-1} \backslash G_{k-1}, \theta) \), which is well defined according to Proposition 2.5. If \( u(x) = \left( \begin{array}{c} \rho(x) \\ 1 \end{array} \right) \) belongs to \( U_k \), then \( \rho(u(x))W(diag(zg,1)) = \theta(z(gx)_{k-1})W(diag(zg,1)) \). As \( \theta(z(gx)_{k-1}) = 1 \) for \( z \) small enough, we deduce that \( S(\rho(u(x))W) = S(W) \), hence the kernel of \( S \) contains \( \tau(U_k,1) \). Conversely, if \( S(W) = 0 \), the smoothness of \( W \) and the Iwasawa decomposition imply that \( W(g) \) is null for \( |z(g)| \) in a punctured neighbourhood of zero depending only on \( W \). According to Proposition 2.3. of [M] (which is a restatement of Proposition 1.3. of [C-P]), this means that \( W \) belongs to \( \tau(U_k,1) \).

The \( \mathbb{C} \)-linear map \( S : W \mapsto S(W) \) induces a \( \mathbb{C} \)-linear isomorphism \( \overline{S} : \overline{W} \mapsto S(W) \) between \( \tau^{(1)} \) and its image in \( (\rho, C^\infty(N_{k-1} \backslash G_{k-1}, \theta)) \). Moreover, it is a \( G_{k-1} \)-equivariant because for \( g_0 \in G_{k-1} \), one has that \( \overline{S}(\tau^{(1)}(g_0)\overline{W}) = \delta_{P_k}^{-1/2}(g_0)S(\rho(g_0)W) = \rho(g_0)S(W) = \rho(g_0)\overline{S}(W) \).
We end this section by stating two technical lemmas about Whittaker functions fixed under a maximal compact subgroup, the first is inspired from Lemma 9.2 of [L-P-S 2].

**Lemma 2.1.** For $n \geq 2$, let $\tau$ be a $P_n$-submodule of $C^\infty(N_n \setminus P_n, \theta)$, and let $W$ belong to $\tau^{P_n}(\mathcal{O})$, then there exists $W'$ in $\tau^{P_n}(\mathcal{O})$, such that $W'(p^\beta_{n-1}(z)) = W(p^\lambda_1^\beta_0)1_{\mathcal{O}^\ast}(z)$ for $p$ in $P_{n-1}$ and $z$ in $F^\ast$.

**Proof.** For $l$ in $\mathbb{Z}$, we denote by $\phi_l$ the characteristic function $1_{\mathfrak{p}^l}$. We fix a Haar measure $dt$ on $F$. The Fourier transform $\hat{\phi}_l$ with respect to $\theta$ and $dt$ is equal to $\lambda_l \phi_{-l}$ for $\lambda_l = dt(\mathfrak{p}^l) > 0$. We denote by $\Phi_l$ the function $\otimes_{l=1}^{n-1} \phi_l$, which is the characteristic of the lattice $\mathfrak{z}^l\mathcal{O}^{n-1}$ in $F^{n-1}$. We denote by $u$ the natural isomorphism between $F^{n-1}$ and $U_n$. We also recall that any element of $\tau$ is determined by its restriction to $G_{n-1}$.

We set $W^l(p) = \int_{x \in F_{n-1}} W(p^l(x)) \Phi_l(x) dx$ for $p$ in $P_n$ and $dx = dt_1 \otimes \cdots \otimes dt_{n-1}$, hence $W^l$ belongs to $\tau_l$. Moreover if $k$ belongs to $G_{n-1}$, and $g$ belongs to $G_{n-1}$, then $W^l(gk)$ is equal to $\int_{x \in F_{n-1}} W(g^l(kx)) \Phi_l(x) dx$ because $W$ is $P_{l}(\mathcal{O})$-invariant, and this last integral is equal to $\int_{x \in F_{n-1}} W(g^l(x)) \Phi_l(k^{-1}x) dx = \int_{x \in F_{n-1}} W(g^l(x)) \Phi_l(x) dx = W^l(g)$ because of the invariance of $dx$ and $\Phi_l$ under $G_{n-1}(\mathcal{O})$. It is also clear that $W^l$ is invariant $U_n(\mathcal{O})$ because $W$ is, hence $W^l$ belongs to $\tau^{P_n}(\mathcal{O})$.

Now, for $p$ in $P_{n-1} \subset P_n$, and $z$ in $F^\ast$, we obtain:

$$W^l(p^\beta_{n-1}(z)) = \int_{x \in F_{n-1}} W(p^\beta_{n-1}(z)u(x)) \Phi_l(x) dx$$

$$= \int_{x \in F_{n-1}} W(u(zpx)p^\beta_{n-1}(z)) \Phi_l(x) dx$$

$$= \int_{x \in F_{n-1}} \theta((zpx)_{n-1}) W(p^\beta_{n-1}(z)) \Phi_l(x) dx$$

$$= W(p^\beta_{n-1}(z)) \int_{x \in F_{n-1}} \Phi_l(x) \theta(x_{n-1}) dx$$

$$= W(p^\beta_{n-1}(z)) \int_{g \in \mathfrak{z}^n \mathcal{O}^{n-2}} dg \int_{t \in F} \Phi_l(t) \theta(t) dt$$

$$= \lambda_{n-2}^l W(p^\beta_{n-1}(z)) \phi_{-l}(z) = \lambda_{n-1}^l W(p^\beta_{n-1}(z)) \phi_{-l}(z)$$

The function $W' = W^l/\lambda_{n-1}^l - W^l/\lambda_{n-1}^l$ thus satisfies

$$W'(p^\beta_{n-1}(z)) = W(p^\beta_{n-1}(z))(\phi_0 - \phi_1)(z) = W(p^\beta_{n-1}(z))1_{\mathcal{O}^\ast}(z) = W(p^\lambda_1^\beta_0)1_{\mathcal{O}^\ast}(z)$$

because $W$ is invariant under $\beta_{n-1}(\mathcal{O}^\ast) \subset P_n(\mathcal{O})$. \hfill $\square$

**Lemma 2.2.** For $n \geq 2$, let $\tau$ be a $P_n$-submodule of $C^\infty(N_n \setminus P_n, \theta)$, and let $W$ belong to $\tau^{G_{n-1}(\mathcal{O})}$, then there exists $W'$ in $\tau^{P_n}(\mathcal{O})$, such that $W'(z_1, \ldots, z_{n-1}, 1) = W(z_1, \ldots, z_{n-1}, 1)1_{\mathcal{O}^\ast}(z_{n-1})$ for $z_i$ in $F^\ast$.

**References:**

1. L-P-S 2. Documenta Mathematica 18 (2013) 1191–1214
Proof. Let \( du \) be the Haar measure on \( U_n \), corresponding to the Haar measure \( dx = dt_1 \otimes \cdots \otimes dt_{n-1} \) on \( F^{n-1} \), normalised by \( dt_i(\mathfrak{O}) = 1 \) for \( i \) between 1 and \( n-1 \). Now set \( W'(g) = \int_{u \in U_n(\mathfrak{O})} W(gu) du \). The vector \( W' \) is a linear combination of right translates of \( W \) by elements of \( U_n(\mathfrak{O}) \), so it belongs to \( \pi \). It is clearly invariant under \( U_n(\mathfrak{O}) \), and still invariant under \( G_{n-1}(\mathfrak{O}) \), as \( G_{n-1}(\mathfrak{O}) \) normalises \( U_n(\mathfrak{O}) \). The following computation then gives the result:

\[
W'(z_1, \ldots, z_{n-1}, 1) = \int_{x \in \mathfrak{O}^{n-1}} W(\beta_1(z_1) \cdots \beta_{n-1}(z_{n-1})u(x)) dx
= \int_{x \in \mathfrak{O}^{n-1}} W(u(\beta_1(z_1) \cdots \beta_{n-1}(z_{n-1})x)\beta_1(z_1) \cdots \beta_{n-1}(z_{n-1})) dx
= \int_{x \in \mathfrak{O}^{n-1}} \theta(z_{n-1}x_{n-1}) W(\beta_1(z_1) \cdots \beta_{n-1}(z_{n-1})) dx
= \int_{\mathfrak{O}^{n-1}} \theta(z_{n-1}) W(z_1, \ldots, z_{n-1}, 1) = 1_{\mathfrak{O}}(z_{n-1}) W(z_1, \ldots, z_{n-1}, 1),
\]
the last equality because of the normalisation of the Haar measure on \( F \).

3 Construction of the essential Whittaker function

We are now going to produce the essential vector of a generic representation \( \pi \) of \( G_r \), which will now be fixed until the end. We recall that we associated to \( \pi \), an integer \( 0 \leq r \leq n \), and an unramified representation of Langlands’ type \( \pi_u \) of \( G_r \) in Section 1.5.

We first notice that the subspace of \( \pi^{(n-r)} \) fixed by \( G_r(\mathfrak{O}) \) is a complex line.

Proposition 3.1. Let \( \pi \) be generic representation of \( G_n \). Then \( (\pi^{(n-r)})^{G_r(\mathfrak{O})} \) is of dimension 1. If \( v^0 \) is a generator of \( (\pi^{(n-r)})^{G_r(\mathfrak{O})} \), then the submodule \( \langle v^0 \rangle \) of \( \pi^{(n-r)} \) spanned by \( v^0 \) surjects onto \( \pi_u \).

Proof. Write \( \pi = \Delta_1 \times \cdots \times \Delta_r \) for the ordering of the \( \Delta_i \)’s fixed after Definition 1.3. According to Lemma 3.5. of [3Z] the representation \( \pi^{(n-r)} \) has a filtration with subquotients \( \Delta_1^{(a_1)} \times \cdots \times \Delta_r^{(a_r)} \), with \( \sum_i a_i = n - r \). According to Proposition 9.6 of [3Z], \( \pi_u \) appears as one of these subquotients, and by the choice of \( r \), the other nonzero subquotients amongst them all contain either a segment as a factor (in the product notation) of some \( G_k \) for \( k \geq 2 \), or a ramified character of \( G_1 \). According to Proposition 1.8 and Proposition 1.9 of [3Z], these other subquotients contain no nonzero \( G_n(\mathfrak{O}) \)-invariant vector. The result then follows from the exactness of the functor \( V \mapsto V^{G_n(\mathfrak{O})} \) from \( Alg(G_n) \) to \( Alg(G_0) \).

We also notice the following facts. First, from the theory of Kirillov models (see [3Z], Theorem 4.9), for \( n \geq 2 \), the map \( W \in W(\pi, \theta) \mapsto W|_{P_n} \) is injective, we denote by \( W(\pi_u, \theta) \) its image. We choose this notation because \( P_n \)-module \( \pi_u(0) = \pi|_{P_n} \) is isomorphic to the submodule \( W(\pi_u(0), \theta) \) of
(p, C^\infty(\mathfrak{n}/\mathfrak{p}_n, \theta)). Now if one applies Proposition 2.4 repeatedly to \(\pi(0)\), then for \(r \leq n - 1\), the \(P_{r+1}\)-module \(\pi_n\) is isomorphic to the submodule of \((p, C^\infty(\mathfrak{n}/\mathfrak{p}_{r+1}, \theta))\), whose vectors are the functions \(\prod_{k=r+2}^{\infty} \delta^{1/2}_{\mathfrak{F}_k} W|_{P_{r+1}}\) for \(W \in W(\pi, \theta)\) (where \(P_{r+1}\) is embedded in \(P_n\) via \(p \mapsto diag(p, I_{\mathfrak{n}/\mathfrak{p}_{r-1}})\)), we denote by \(W(\pi_n, \theta)\) this \(P_{r+1}\)-module.

Proposition 2.4 has the following corollary.

**Corollary 3.1.** Under the condition 1 \(\leq r \leq n - 1\), there exists \(W_0\) in \(W(\pi_n, \theta)\) such that

\[
W_0^0(g) = \lim_{z \to 0} c^{-1}(z)|z|^{-r/2}W_0(diag(zg, 1))\delta^{1/2}_{\mathfrak{F}_{r+1}}(g)
\]

for all \(g \in G_r\). This implies that the representation \(\pi_n\) occurs as a submodule of \(\pi^{(n-r)}\).

**Proof.** We take \(\Psi^{-}(W(\pi_n, \theta))\) as a model for \(\pi^{(n-r)}\), i.e. \(\pi^{(n-r)} = \Psi^{-}(W(\pi_n, \theta))\). Let \(W_0\) be a preimage (which we shall normalise later), of \(v^0\) under \(\Psi^{-}\), which we take in \(W(\pi_n, \theta)\) thanks to Proposition 2.4.

We denote by \(< P_{r+1}, W_0 >\) the \(P_{r+1}\)-submodule of \(W(\pi_n, \theta)\) spanned by \(W_0\). By definition of \(\Psi^{-}\), we have

\[
\Psi^{-}(< P_{r+1}, W_0 >) = \Psi^{-}(< G_r, W_0 >) = < G_r, v^0 > .
\]

Now, \(Z_n\) acts by a character \(c\) on \(< G_r, v^0 >\) (which is the central character of \(\pi_n\) as well according to Proposition 2.4). Let \(S\) be the map defined in Corollary 2.4 from \(< P_{r+1}, W_0 >\) to \(C^\infty(\mathfrak{n}/\mathfrak{g}_r, \theta)\). We know from this corollary, that \(S\) factors to give an isomorphism \(\mathcal{S}\) between \(< G_r, v^0 >\) and \(S(< P_{r+1}, W_0 >)\). Define \(W^0 = S(W_0)\) in \(C^\infty(\mathfrak{n}/\mathfrak{g}_r, \theta)\), so that \(S(< P_{r+1}, W_0 >)\) is equal to \(G_rW^0\). As the \(G_r\)-module \(< G_r, W^0 >\) is isomorphic to \(< G_r, v^0 >\), there is a surjective \(G_r\)-module morphism from \(< G_r, W^0 >\) onto \(W(\pi_n, \theta)\) according to Proposition 2.4. It sends \(W^0\) to a nonzero multiple of \(W_{\pi_n}^0\). We normalise \(W_0\), such that the Whittaker function \(W^0 = S(W_0)\) is equal to 1 on \(G_n(\mathfrak{D})\). The Hecke algebra \(H_r\) thus multiplies \(W^0\) and \(W_{\pi_n}^0\) by the same character, as \(W_{\pi_n}^0\) is the image of \(W^0\) via a \(G_r\)-intertwining operator. Both are normalised spherical Whittaker functions, they are thus equal according to [S]. In particular, we have \(S(W_0) = W_{\pi_n}^0\), which is the first statement of the corollary. Next, this implies the equalities \(< G_r, W^0 > = < G_r, W_{\pi_n}^0 > = W(\pi_n, \theta)\), so the surjection from \(< G_r, v^0 >\) onto \(W(\pi_n, \theta)\) is actually equal to the isomorphism \(\mathcal{S}\), hence \(\pi_n\) occurs as a submodule of \(\pi^{(n-r)}\).

The following proposition then holds.

**Proposition 3.2.** Under the condition 1 \(\leq r \leq n - 1\), there exists in \(W(\pi_n, \theta)\) an element \(W_0\), such that \(W_0(z_1, \ldots, z_r, 1) = \delta^{1/2}_{\mathfrak{F}_{r+1}}(z_1, \ldots, z_r)W_{\pi_n}^0(z_1, \ldots, z_r)1(\mathfrak{D})\) for \(z_r\) in \(F^*\).
They imply the equality

$$W_0(z_1, \ldots, z_{r}a, 1) = c_{\pi_n}(a)|a|^{r/2}W_0(z_1, \ldots, z_{r}, 1)$$

(parametrizing $A_{r+1}$ with the $\beta_i$'s) for $|z_r| \leq q^{-N}$ and $|a| \leq 1$. For $b$ in $F^*$, call $\tilde{W}_{0,b}$ the function $p \mapsto \tilde{W}_0(p\beta_r(b))/(c_{\pi_n}(b)|b|^{r/2})$ defined on $P_{r+1}$, then $\tilde{W}_{0,b}$ still belongs to $W(\pi_{(n-r-1)}, \theta^{P_{r+1}|O|})$, and $\tilde{W}_{0,b}(z_1, \ldots, z_{r}, 1)/(c_{\pi_n}(z_r)|z_r|^{r/2})$ is constant with respect to $z_r$ whenever $|z_r| \leq q^{-N}/|b|$. We choose $b$ in $F^*$ satisfying $|b| = q^{-N}$, so that the function $\tilde{W}_{0,b}(z_1, \ldots, z_{r}, 1)/(c_{\pi_n}(z_r)|z_r|^{r/2})$ is constant with respect to $z_r$ for $|z_r| \leq 1$. Hence, according to Corollary 5.1 for $|z_r| \leq 1$, we have the equalities

$$\tilde{W}_{0,b}(z_1, \ldots, z_{r}, 1)/(c_{\pi_n}(z_r)|z_r|^{r/2}) = \tilde{W}_0(z_1, \ldots, z_{r}b, 1)/(c_{\pi_n}(z_r)|z_r|^{r/2})$$

$$= \delta_{P_{r+1}}^{1/2}(z_1, \ldots, z_{r-1}, b)W_{\pi_n}^0(z_1, \ldots, z_{r-1}, b)/(c_{\pi_n}(b)|b|^{r/2})$$

$$= \delta_{P_{r+1}}^{1/2}(z_1, \ldots, z_{r-1}, 1)W_{\pi_n}^0(z_1, \ldots, z_{r-1}, 1).$$

They imply the equality

$$\tilde{W}_{0,b}(z_1, \ldots, z_{r}, 1) = \delta_{P_{r+1}}^{1/2}(z_1, \ldots, z_{r-1}, z_{r})W_{\pi_n}^0(z_1, \ldots, z_{r-1}, z_{r})$$

for $|z_r| \leq 1$. On the other hand, applying Lemma 2.2 there is $W_0$ is in $W(\pi_{(n-r-1)}, \theta^{P_{r+1}|O|})$ such that $W_0(z_1, \ldots, z_{r}, 1)$ is equal to $\tilde{W}_{0,b}(z_1, \ldots, z_{r}, 1)1_{O}(z_{r})$, it is then clear that $W_0$ has the desired property.

We now prove the main result of this paper.

**Theorem 3.1.** For $n \geq 2$, let $\pi$ be a ramified generic representation of $G_n$ (i.e. $r \leq n-1$). Then one can produce a $G_{n-1}(O)$-invariant function $W_{\pi}^{\text{ess}}$ in $W(\pi, \theta)$, whose restriction to $A_{n-1}$ (when $A_{n-1}$ is parametrised by its simple roots), is given by formula

$$W_{\pi}^{\text{ess}}(z_1, \ldots, z_{n-1}, 1) = W_{\pi_n}^0(z_1, \ldots, z_r)\nu(z_1, \ldots, z_r)^{(n-r)/2}1_{O}(z_{r}) \prod_{j=r+1}^{n-1} 1_{O\cdot}(z_{j})$$

when $r \geq 1$, and by

$$W_{\pi}^{\text{ess}}(z_1, \ldots, z_{n-1}, 1) = \prod_{j=1}^{n-1} 1_{O\cdot}(z_{j})$$

when $r = 0$. A function $W_{\pi}^{\text{ess}}$ with such properties is unique, and has image $W_{\pi_n}^0$ in $\pi^{(n-r)}$.
Proof. Suppose first that we have $r \geq 1$. We already constructed in the previous proposition a vector $W_0$ in $W(\pi_{(n-r-1)}, \theta)^{P'_{r+1}(\Delta)}$ such that

$$W_0(z_1, \ldots, z_r, 1) = \delta_{P'_{r+1}}^{1/2}(z_1, \ldots, z_r)W^0_{\pi_n}(z_1, \ldots, z_r)1_{\Delta}(z_r).$$

Then, applying Proposition 2.3 and then Lemma 2.1 we obtain $W_1$ in $W(\pi_{(n-r-2)}, \theta)^{P'_{r+2}(\Delta)}$, that satisfies

$$W_1(z_1, \ldots, z_{r+1}, 1) = \delta_{P'_{r+2}}^{1/2}(z_1, \ldots, z_{r+1})W_0(z_1, \ldots, z_r, 1)1_{\Delta}(z_{r+1})$$

$$= \delta_{P'_{r+2}}^{1/2}(z_1, \ldots, z_r, 1)\delta_{P'_{r+1}}^{1/2}(z_1, \ldots, z_r)W^0_{\pi_n}(z_1, \ldots, z_r)1_{\Delta}(z_r)1_{\Delta}(z_{r+1}).$$

Repeating this last step (Proposition 2.3 and then Lemma 2.1), we obtain by induction for all $k$ between 1 and $n - r - 1$, an element $W_k$ in $W(\pi_{(n-r-1-k)}, \theta)^{P'_{r+k+1}(\Delta)}$, that satisfies

$$W_k(z_1, \ldots, z_{r+k}, 1) = \delta_{P'_{r+k+1}}^{1/2}(z_1, \ldots, z_{r+k})W_{k-1}(z_1, \ldots, z_{r+k-1}, 1)1_{\Delta}(z_{r+k})$$

$$= W^0_{\pi_n}(z_1, \ldots, z_r)1_{\Delta}(z_r)\prod_{j=r+1}^{r+k} 1_{\Delta}(z_j) \prod_{i=r+1}^{r+k+1} \delta_{P'_{i}}^{1/2}(z_1, \ldots, z_r, 1, \ldots, 1).$$

We define $W^\text{ess}_\pi$ to be the element of $W(\pi, \theta)$ which restricts to $P_n$ as $W_{n-r-1}$, it is thus $G_{n-1}(\Delta)$-invariant and satisfies Equation (6) of the statement of the theorem, because

$$\prod_{i=r+1}^{n} \delta_{P_i}^{1/2}(z_1, \ldots, z_r, 1, \ldots, 1) = \text{det}(z_1, \ldots, z_r)$$

as $\delta_{P_i}(z_1, \ldots, z_r, 1, \ldots, 1) = |\text{det}(z_1, \ldots, z_r)|$ for $i > r$.

If $r = 0$, we take for $W_0$ the constant function on the trivial group $P_1$ equal to 1 in $W(\pi_{(n-1)}, \theta) = W(\pi_{(n-1)}, \theta)^{P_1(\Delta)}$. Again, thanks to Proposition 2.3 and Lemma 2.1, there is $W_1$ in $W(\pi_{(n-2)}, \theta)^{P_2(\Delta)}$ such that $W_1(z_1, 1) = 1_{\Delta}(z_1)$ for $z_1$ in $F^*$, and we end as in the case $r \geq 1$.

The function $W^\text{ess}_\pi$ is unique by the theory of Kirillov models, and its image in $\pi^{(n-r)}$ is $W^0_{\pi_n}$ by construction. \qed

The expression of the restriction of $W^\text{ess}_\pi$ to $A_{n-1}$ in the usual coordinates is the same.

Corollary 3.2. Let $\pi$ be a ramified generic representation of $G_n$, then if $a = \text{diag}(a_1, \ldots, a_{n-1})$ belongs to $A_{n-1}$, and $a' = \text{diag}(a_1, \ldots, a_r) \in A_r$, we obtain Formulas (7) and (8) of the introduction:

$$W^\text{ess}_\pi(\text{diag}(a, 1)) = W^0_{\pi_n}(a')^{(n-r)/2}1_{\Delta}(a_r)\prod_{j=r+1}^{n-1} 1_{\Delta}(a_j)$$

Documenta Mathematica 18 (2013) 1191–1214
when $r \geq 1$, and
\[
W_{\pi}^{\text{ess}}(\text{diag}(a,1)) = \prod_{j=1}^{n-1} 1_{\Omega^*}(a_j)
\]
when $r = 0$.

**Proof.** We do the case $r \geq 1$, the case $r = 0$ being simpler. If 
\[
\text{diag}(a_1, \ldots, a_{n-1}) = (z_1, \ldots, z_{n-1}) \text{ belongs to } A_{n-1},
\]
we have $a_i = z_i \ldots z_{n-1}$, hence if $a' = \text{diag}(a_1, \ldots, a_r)$, we have $(z_1, \ldots, z_r) = (\prod_{i=r+1}^{n-1} z_i)^{-1} a'$ in $A_r$.

Equation (6) can thus be read:
\[
W_{\pi}(\text{diag}(a,1)) = W_{\pi}(a') \nu(a')^{(n-r)/2} \prod_{i=r+1}^{n-1} 1_{\Omega^*}(a_i),
\]
the last equality because if it is not $0 = 0$, this means that $z_{r+1}, \ldots, z_{n-1}$ all belong to $\Omega^*$, hence the inverse of their product as well, and $W_{\pi,0}, \nu, 1_{\Omega^*}$ and $1_{\Omega}$ are all invariant under $\Omega^*$. \qed

We then have the following corollary.

**Corollary 3.3.** Let $\pi$ be a generic representation of $G_n$ with Whittaker model $W(\pi, \theta)$. There exists in $W(\pi, \theta)$ a unique $G_{n-1}(\Omega)$-invariant function $W_{\pi}^{\text{ess}}$ equal to 1 on $G_{n-1}(\Omega)$, such that for every $1 \leq m \leq n-1$, and every unramified representation $\pi'$ of Langlands' type of $G_m$, with normalised spherical function $W_{\pi'}^0$ in $W(\pi', \theta^{-1})$, the equality $I(W_{\pi}^{\text{ess}}, W_{\pi'}^0, s) = L(\pi, \pi', s)$ holds for an appropriate normalisation of the invariant measure on $N_m\backslash G_m$.

**Proof.** The unicity of a function $W_{\pi}^{\text{ess}}$ with such properties follows from [J-P-S].

If $\pi$ is unramified (i.e. $r = n$), we set $W_{\pi}^{\text{ess}} = W_{\pi}^0$ and Equations (3) and (4) show that it is the correct choice.

When $r \leq n-1$, we again only treat the case $r \geq 1$, the case $r = 0$ being similar, but simpler (using Equation (3) instead of Equation (1)). We show that the function $W_{\pi}^{\text{ess}}$ from the previous corollary satisfies the wanted equalities.

Thanks to Iwasawa decomposition, we have
\[
I(W_{\pi}^{\text{ess}}, W_{\pi'}^0, s) = \int_{A_m} W_{\pi}^{\text{ess}}(\text{diag}(a, I_{n-m})) W_{\pi'}^0(a) \delta_{B_m}^{-1}(a) \nu(a)^s d^* a'.
\]

If $m > r$, using Equations (1) and $\delta_{B_m}(a I_{m-r}) = \delta_{B_m}(a) \nu(a)^{m-r}$, we find
\[
I(W_{\pi}^{\text{ess}}, W_{\pi'}^0, s) = \int_{A_r} W_{\pi}(a') W_{\pi'}^0(a') \delta_{B_r}^{-1}(a') 1_{\Omega^*}(a) \nu(a')^{s-\frac{m-r}{2}} d^* a'
\]
Nadir Matringe

\[ I(W^{\text{cis}}, W^{0}_{\pi'}, s) = \int_{A_0} W^{0}_{\pi}(\alpha') W^{0}_{\pi'}(\alpha') \delta_{\mathcal{B}_0}(\alpha') \nu(\alpha') s^{-(\frac{r-m}{2})} d^* a' = I(W^{0}_{\pi}, W^{0}_{\pi'}, s), \]

as \( W^{0}_{\pi'}(\alpha') \delta_{\mathcal{B}_0}(\alpha') \nu(\alpha') s^{-(\frac{r-m}{2})} d^* a' \) vanishes for \( |\alpha_r| > 1 \). Hence, by Equations (4) and (5), we obtain

\[ I(W^{\text{cis}}, W^{0}_{\pi'}, s) = L(\pi_u, \pi', s) = L(\pi, \pi', s). \]

If \( m = r \), using Equation (1), we find

\[ I(W^{\text{cis}}, W^{0}_{\pi'}, s) = \int_{A_m} W^{0}_{\pi}(\alpha') W^{0}_{\pi'}(\alpha') \delta_{\mathcal{B}_0}(\alpha') \nu(\alpha') s^{-(\frac{r-m}{2})} d^* a, \]

and this integral is equal to

\[ I(W^{0}_{\pi}, W^{0}_{\pi'}, 1_{\mathcal{D}_m}, s) = L(\pi_u, \pi', s) = L(\pi, \pi', s) \]

by Equations (3) and (5).

If \( m < r \), Equation (1) gives

\[ I(W^{\text{cis}}, W^{0}_{\pi'}, s) = \int_{A_m} W^{0}_{\pi}(\alpha') W^{0}_{\pi'}(\alpha') \delta_{\mathcal{B}_0}(\alpha') \nu(\alpha') s^{-(\frac{r-m}{2})} d^* a, \]

and this integral is equal to

\[ I(W^{0}_{\pi}, W^{0}_{\pi'}) = L(\pi_u, \pi', s) = L(\pi, \pi', s) \]

by Equations (4) and (5).

In all cases, we have

\[ I(W^{\text{cis}}, W^{0}_{\pi'}, s) = L(\pi, \pi', s). \]

Acknowledgements. I thank the p-adic workgroup in Poitiers for allowing me to give talks about [J-P-S]. I thank Guy Henniart, Hervé Jacquet, and Michitaka Miyauchi, for useful comments and corrections about this note. Finally and most importantly, I thank the referee for his very careful reading of the paper, which allowed to correct some mistakes, and improve many arguments.

References


Evidence for a Generalization of Gieseker’s Conjecture on Stratified Bundles in Positive Characteristic

Lars Kindler

Received: March 25, 2013
Communicated by Takeshi Saito

Abstract. Let $X$ be a smooth, connected, projective variety over an algebraically closed field of positive characteristic. In [Gie75], Gieseker conjectured that every stratified bundle (i.e. every $\mathcal{O}_X$-coherent $\mathcal{D}_{X/k}$-module) on $X$ is trivial, if and only if $\pi_1^\et(X) = 0$. This was proven by Esnault-Mehta, [EM10]. Building on the classical situation over the complex numbers, we present and motivate a generalization of Gieseker’s conjecture, using the notion of regular singular stratified bundles developed in the author’s thesis and [Kin12a]. In the main part of this article we establish some important special cases of this generalization; most notably we prove that for not necessarily proper $X$, $\pi_1^{\text{tame}}(X) = 0$ implies that there are no nontrivial regular singular stratified bundles with abelian monodromy.

2010 Mathematics Subject Classification: 14E20, 14E22, 14F10
Keywords and Phrases: Fundamental group, coverings, stratified bundles, D-modules, tame ramification

1 Introduction and statement of the conjecture

Let $X$ be a smooth, separated, connected scheme of finite type over an algebraically closed field $k$, and fix a base point $x \in X(k)$. For readability, we write $\Pi^\et_X := \pi_1^\et(X,x)$, and we denote by $\text{Rep}^\text{cont}_k \Pi^\et_X$ the category of continuous representations of the profinite group $\Pi^\et_X$ on finite dimensional $k$-vector spaces equipped with the discrete topology.
If $k = \mathbb{C}$, then $\Pi^\et_X$ is the profinite completion of the abstract group $\Pi_X^{\text{top}} := \pi_1^{\text{top}}(X(\mathbb{C}),x)$, and if we write $\text{Rep}_{\mathbb{C}} \Pi_X^{\text{top}}$ for the category of representations
of $\Pi^\text{top}_X$ on finite dimensional $\mathbb{C}$-vector spaces, then $\text{Rep}_{\mathbb{C}}^\text{cont} \Pi^\text{\acute{e}t}_X$ can be considered as the full subcategory of $\text{Rep}_{\mathbb{C}} \Pi^\text{top}_X$ having as objects precisely those representations which factor through a finite group. Since $\Pi^\text{top}_X$ is finitely generated, the category $\text{Rep}_{\mathbb{C}} \Pi^\text{top}_X$ is controlled by $\text{Rep}_{\mathbb{C}}^\text{cont} \Pi^\text{\acute{e}t}_X$, according to the following theorem:

**Theorem 1.1 (Grothendieck [Gro70], Malcev [Mal40]).** If $\phi : G \to H$ is a morphism of finitely generated groups, then the following statements are equivalent:

(a) The induced morphism $\hat{G} \to \hat{H}$ is an isomorphism, where $\hat{(-)}$ denotes profinite completion.

(b) The induced $\otimes$-functor $$\text{Rep}_{\mathbb{C}} H \to \text{Rep}_{\mathbb{C}} G$$ is a $\otimes$-equivalence.

(c) The induced $\otimes$-functor $$\text{Rep}_{\mathbb{C}}^\text{cont} \hat{H} \to \text{Rep}_{\mathbb{C}}^\text{cont} \hat{G}$$ is a $\otimes$-equivalence.

Accordingly, if $f : Y \to X$ is a morphism of smooth, connected, complex varieties, then the induced morphism $\Pi^\text{\acute{e}t}_Y \to \Pi^\text{\acute{e}t}_X$ (with respect to compatible base points) is an isomorphism if and only if $\text{Rep}_{\mathbb{C}} \Pi^\text{top}_X \to \text{Rep}_{\mathbb{C}} \Pi^\text{top}_Y$ is a $\otimes$-equivalence. This consequence was already noted in [Gro70].

To study the category $\text{Rep}_{\mathbb{C}} \Pi^\text{top}_X$, we invoke the Riemann-Hilbert correspondence as developed in [Del70]: It states that the choice of a base point $x \in X(\mathbb{C})$ gives a $\otimes$-equivalence $\|_{x}$ of Tannakian categories between the category of regular singular flat connections on $X$ and the category $\text{Rep}_{\mathbb{C}} \Pi^\text{top}_X$. Theorem 1.1 then translates into the following completely algebraic statement, in which we suppress the choice of base points from the notation:

**Corollary 1.2.** If $f : Y \to X$ is a morphism of smooth, connected, separated, finite type $\mathbb{C}$-schemes, then the following are equivalent:

(a) The morphism $\Pi^\text{\acute{e}t}_Y \to \Pi^\text{\acute{e}t}_X$ induced by $f$ is an isomorphism.

(b) Pull-back along $f$ induces an equivalence on the categories of regular singular flat connections.

(c) Pull-back along $f$ induces an equivalence on the categories of regular singular flat connections with finite monodromy.

In particular: $\Pi^\text{\acute{e}t}_Y = 0$ if and only if every regular singular flat connection on $Y$ is trivial.
This article is devoted to the study of analogous statements in positive characteristics. Let \( k \) be an algebraically closed field of characteristic \( p > 0 \), and \( X \) a smooth, connected, separated, finite type \( k \)-scheme. In general, neither the category of vector bundles with flat connection, nor the category of coherent \( \mathcal{O}_X \)-modules with flat connection are Tannakian categories over \( k \). Instead, we consider left-\( \mathcal{D}_{X/k} \)-modules which are coherent (and then automatically locally free) as \( \mathcal{O}_X \)-modules, where \( \mathcal{D}_{X/k} \) is the ring of differential operators relative to \( k \), as developed in [EGA4, §16]. Following [Gro68] and [SR72], we call such objects stratified bundles, and we write \( \text{Strat}(X) \) for the category of stratified bundles on \( X \). Recall that in characteristic 0, giving a stratified bundle is equivalent to giving a vector bundle with flat connection. In positive characteristic, these notions are not equivalent. A stratified bundle is called trivial if it is isomorphic to \( \mathcal{O}_X^n \) with the canonical diagonal left-\( \mathcal{D}_{X/k} \)-action. In [Kin12a] (see also [Kin12b]), a notion of regular singularity for stratified bundles is defined and studied, generalizing work of Gieseker ([Gie75]); for a summary see Section 2. We write \( \text{Strat}^{rs}(X) \) for the full subcategory of \( \text{Strat}(X) \) with objects regular singular stratified bundles; after choosing a base point, it is a neutral Tannakian category over \( k \).

Using the theory of Tannakian categories, there still is a procedure to attach to a stratified bundle a monodromy group and a monodromy representation at the base point \( x \in X(k) \). The main result of [Kin12a] states that this procedure induces a \( \otimes \)-equivalence between the category of regular singular stratified bundles with finite monodromy and \( \text{Rep}_{\text{cont}}^{\text{tame}} \Pi_X \), where \( \Pi_X^{\text{tame}} := \pi_1^{\text{tame}}(X, x) \) is the tame fundamental group as defined in [KS10]. This result suggests the following conjecture, completely analogous to Corollary 1.2:

**Conjecture 1.3.** Let \( k \) be an algebraically closed field of characteristic \( p > 0 \). If \( f : Y \to X \) is a morphism of smooth, connected, separated, finite type \( k \)-schemes, then the following are equivalent:

(a) The morphism \( \Pi_Y^{\text{tame}} \to \Pi_X^{\text{tame}} \) induced by \( f \) is an isomorphism.

(b) Pull-back along \( f \) induces an equivalence on the categories of regular singular stratified bundles.

(c) Pull-back along \( f \) induces an equivalence on the categories of regular singular stratified bundles with finite monodromy.

In particular: \( \Pi_Y^{\text{tame}} = 0 \) if and only if every regular singular stratified bundle on \( Y \) is trivial.

For a slightly different exposition of this conjecture, see [Esn12]. Note that by the main result of [Kin12a] we always have \((b) \Rightarrow (c) \Rightarrow (a)\).

The main part of this article is concerned with establishing special cases of and evidence for the validity of Conjecture 1.3, i.e. for the direction \((a) \Rightarrow (b)\). We give a brief summary: Gieseker’s original conjecture from [Gie75] corresponds to Conjecture 1.3 for a projective morphism \( f : Y \to \text{Spec}(k) \). His conjecture was proven by Esnault-Mehlt, [EM10]. Thus Conjecture 1.3 generalizes...
Gieseker’s conjecture in two “directions”: It allows non-projective varieties by using the notion of regular singularity, and it gives a relative formulation. The main result of this article, proven in Section 4, is the following:

**Theorem 1.4** (see Theorem 4.2). In the situation of Conjecture 1.3, assume that \( X = \text{Spec}(k) \), and that \( Y \) admits a good compactification. If \( \Pi_Y^{\text{ab},(p')} = 0 \) and if \( \Pi_Y^{\text{tame}} \) does not have a quotient isomorphic to \( \mathbb{Z}/p\mathbb{Z} \), then every regular singular stratified bundle with abelian monodromy is trivial.

In other words: The abelianization of the pro-algebraic group associated with \( \text{Strat}^0(Y) \) (and the choice of a base point) is trivial.

Here a good compactification of \( Y \) is a smooth, proper \( k \)-scheme \( \overline{Y} \) containing \( Y \) as a dense open subscheme, such that \( \overline{Y} \setminus Y \) is the support of a strict normal crossings divisor, and \( (-)^{\text{ab},(p')} \) denotes the maximal abelian pro-prime-to-\( p \)-quotient. This quotient is independent of the choice of the base point.

Theorem 1.4 is a consequence of the fact that under the given assumptions there are no nontrivial rank 1 stratified bundles, and no nontrivial regular singular extensions of two rank 1 stratified bundles. Hence we establish these facts first: in Section 3 we study stratified bundles of rank 1, and we prove a relative version of Conjecture 1.3 for stratified line bundles:

**Theorem 1.5** (see Theorem 3.10). Let \( X \) and \( Y \) be smooth, connected, separated, finite type \( k \)-schemes, which admit good compactifications \( \overline{X} \) and \( \overline{Y} \) over \( k \). If \( f : Y \to X \) is a map extending to a morphism \( f : \overline{Y} \to \overline{X} \) such that \( f \) induces an isomorphism \( \Pi_Y^{\text{ab},(p')} \cong \Pi_X^{\text{ab},(p')} \) and such that the cokernel of the induced map \( H^0(X, O_X^*) \to H^0(Y, O_Y^*) \) is a \( p \)-group, then pull-back along \( f \) induces an isomorphism

\[
\text{Pic}^{\text{Strat}}(X) \cong \text{Pic}^{\text{Strat}}(Y)
\]

where \( \text{Pic}^{\text{Strat}} \) denotes the group of isomorphism classes of stratified line bundles.

Note that contrary to the situation over the complex numbers, a stratified line bundle in our context is always regular singular (Proposition 2.7). The assumption on the cokernel of \( H^0(X, O_X^*) \to H^0(Y, O_Y^*) \) is trivially fulfilled if \( Y \) and \( X \) are proper. For further comments see Remark 3.11.

In particular we obtain:

**Corollary 1.6.** Let \( Y, X \) be proper, smooth, connected \( k \)-schemes. If \( f : Y \to X \) is a morphism such that \( f \) induces an isomorphism \( \Pi_Y^{\text{ab},(p')} \cong \Pi_X^{\text{ab},(p')} \), then \( f \) induces an isomorphism \( \text{Pic}^{\text{Strat}}(X) \cong \text{Pic}^{\text{Strat}}(Y) \).

In the case that \( X = \text{Spec} k \), the assumption on the existence of a good compactification of \( Y \) is not necessary:

**Proposition 1.7** (see Proposition 3.16). Let \( Y \) be a smooth, connected, separated, finite type \( k \)-scheme, such that \( \Pi_Y^{\text{ab},(p')} = 0 \). Then \( \text{Pic}^{\text{Strat}}(Y) = 0 \).
In Section 4 we study regular singular extensions of rank 1 stratified bundles, and prove Theorem 1.4.

For regular singular stratified bundles with not necessarily abelian monodromy, we establish the following two special cases of Conjecture 1.3:

It is well-known that $\pi^\text{tame}_1(\mathbb{A}_k^n) = 0$ and in Section 5 we give a short proof of:

**Theorem 1.8** (see Theorem 5.1). *Every regular singular stratified bundle on $\mathbb{A}_k^n$ is trivial.*

This was already sketched in [Esn12] using a slightly different approach.

If $f : Y \to X$ is a universal homeomorphism, then $f$ induces an isomorphism $\Pi^\text{tame}_Y \cong \Pi^\text{tame}_X$ ([Vid01]), and in Section 6 we show:

**Theorem 1.9** (see Theorem 6.6). *If $f : Y \to X$ is a universal homeomorphism, then pull-back along $f$ is an equivalence $\text{Strat}^\text{tr}(X) \to \text{Strat}^\text{tr}(Y).$

We write Strat$(X)$ for the category of stratified bundles which will be needed in the course of the text.

**Acknowledgements:** Large parts of the contents of this note were developed while the author wrote his doctoral dissertation under the supervision of Hélène Esnault. Her influence is undeniable, and the author is immensely grateful for her support. The author also wishes to thank Oliver Bräunling and Jakob Stix for helpful discussions, the referee for carefully reading the manuscript, and Brian Conrad for pointing out a helpful result by Lang; see the discussion after Proposition 3.16.

This work was supported by the Sonderforschungsbereich/Transregio 45 “Periods, moduli spaces and the arithmetic of algebraic varieties” of the DFG and the European Research Council Advanced Grant 0419744101.

**2 Generalities on stratified bundles**

Let $k$ be an algebraically closed field of characteristic $p > 0$, and fix a smooth, separated, connected, finite type $k$-scheme $X$.

**Definition 2.1.** A stratified bundle on $X$ is a $\mathcal{D}_{X/k}$-module which is $\mathcal{O}_X$-coherent. A morphism of stratified bundles is a morphism of $\mathcal{D}_{X/k}$-modules. A stratified bundle $E$ is called trivial if it is isomorphic to $\mathcal{O}_X^n$ with its canonical diagonal $\mathcal{D}_{X/k}$-action. Here $\mathcal{D}_{X/k}$ is the ring of differential operators on $X$, see [EGA4, §16].

We write Strat$(X)$ for the category of stratified bundles.

**Remark 2.2.** The name “stratified module” goes back to Grothendieck: In [Gro68] he defines the notion of a stratification relative to $k$ on an $\mathcal{O}_X$-module $E$ as an “infinitesimal descent datum”, and since $X$ is smooth over $k$, such a
datum is equivalent to the datum of a $\mathcal{D}_{X/k}$-action on $E$, compatible with its $\mathcal{O}_X$-structure, see e.g. [BO78, Ch. 2]. As in characteristic 0, one shows that a $\mathcal{O}_X$-coherent $\mathcal{D}_{X/k}$-module is automatically locally free, [BO78, 2.17]. Hence the name “stratified bundle”.

The following result of Katz gives a different perspective on the notion of a stratified bundle:

**Theorem 2.3 (Katz, [Gie75, Thm. 1.3]).** Let $F : X \to X$ denote the absolute Frobenius. The category $\text{Strat}(X)$ is equivalent to the following category:

- **Objects:** Sequences of pairs $E := (E_n, \sigma_n)_{n \geq 0}$ with $E_n$ a vector bundle on $X$ and $\sigma_n : E_n \xrightarrow{\sim} F^* E_{n+1}$ an $\mathcal{O}_X$-linear isomorphism.

- **Morphisms:** A morphism $\phi : (E_n, \sigma_n) \to (E'_n, \sigma'_n)$ is a sequence of morphisms of vector bundles $\phi_n : E_n \to E'_n$, such that $(F^* \phi_{n+1}) \sigma_n = \sigma'_n \phi_n$.

The functor giving the equivalence assigns to a stratified bundle $E$ the sequence $(E_n, \sigma_n)_{n \geq 0}$, with

$$E_n(U) = \{ e \in E(U) | D(e) = 0 \text{ for all } D \in \mathcal{D}_{X/k}^n(U) \text{ with } D(1) = 0 \}$$

and $\sigma_n : E_n \to F^* E_{n+1}$ the isomorphism given by Cartier’s theorem [Kat70, §5]. This functor is compatible with tensor products.

In the sequel we will freely switch between the two perspectives on stratified bundles. The description by the above theorem is especially nice when $X$ is proper over $k$:

**Proposition 2.4 ([Gie75, Prop. 1.7]).** If $X$ is proper over $k$, then the isomorphism class of a stratified bundle $E = (E_n, \sigma_n)_{n \geq 0}$ only depends on the isomorphism classes of the vector bundles $E_1, E_2, \ldots$.

## 2.1 Regular singular stratified bundles

We recall from [Kin12a] the definition of regular singular stratified bundles.

**Definition 2.5.** Let $X$ be a smooth, separated, finite type $k$-scheme.

(a) If $\overline{X}$ is a smooth, separated, finite type $k$-scheme containing $X$ as an open dense subscheme such that $\overline{X} \setminus X$ is a strict normal crossings divisor, then the pair $(X, \overline{X})$ is called good partial compactification of $X$. If $\overline{X}$ is also proper, then $(X, \overline{X})$ is called good compactification of $X$.

(b) If $(X, \overline{X})$ is a good partial compactification, then $\mathcal{D}_{\overline{X}/k}(\log \overline{X} \setminus X)$ is by definition the sheaf of subalgebras of $\mathcal{D}_{\overline{X}/k}$ spanned over an open $U \subset \overline{X}$ by those differential operators in $\mathcal{D}_{\overline{X}/k}(U)$, fixing all powers of the ideal of the divisor $(\overline{X} \setminus X) \cap U$. Let’s make this explicit: If $x_1, \ldots, x_n \in H^0(U, \mathcal{O}_U)$
Evidence for a Generalization of Gieseker’s Conjecture

are coordinates on $\overline{U}$, i.e. if they define an étale morphism $\overline{U} \to \mathbb{A}_k^n$, then $D_{\mathbb{A}_k^n}$ is spanned by operators $\partial_{x_i}^{(m)}$, $i = 1, \ldots, r$, $m \geq 0$, which “behave” like $\frac{1}{m!}\partial^{m}/\partial x_i^m$. If $U \cap (X \setminus X)$ is defined by $x_1 \cdot \ldots \cdot x_r$, then the subring $D_{\mathbb{A}_k^n}(\log X \setminus X)_{\overline{U}}$ is spanned by $\delta_{x_i}^{(m)} := x_i^m \partial_{x_i}^{(m)}$, $i = 1, \ldots, r$, $\delta_{x_i}^{(m)}$, $i > r$, $m \geq 0$.

This ring can also be defined intrinsically as the ring of differential operators associated with a morphism of log-schemes, which, for example, is studied in [Mon02].

(c) If $(X, \overline{X})$ is a good partial compactification, then a stratified bundle $E \in \text{Strat}(X)$ is called $(X, \overline{X})$-regular singular if $E$ extends to an $\mathcal{O}_{\overline{X}}$-torsion free, $\mathcal{O}_{\overline{X}}$-coherent $\mathcal{D}_{\mathbb{A}_k^n}(\log X \setminus X)$-module.

We write $\text{Strat}^{\text{rs}}((X, \overline{X}))$ for the full subcategory of $\text{Strat}(X)$ with objects the $(X, \overline{X})$-regular singular stratified bundles.

(d) A stratified bundle $E$ is called regular singular if $E$ is $(X, \overline{X})$-regular singular for every good partial compactification $(X, \overline{X})$ of $X$.

We write $\text{Strat}^{\text{rs}}(X)$ for the full subcategory of $\text{Strat}(X)$ with objects the regular singular stratified bundles.

Remark 2.6. The notion of $(X, \overline{X})$-regular singularity for stratified bundles first appeared (to the author’s knowledge) in [Gie75] for good compactifications $(X, \overline{X})$. Gieseker attributes some of the ideas used in [loc. cit.] to Katz. In [Kin12a] and [Kin12b] this notion of regular singularity is extended to varieties for which resolution of singularities is not known to hold, and its connection with tame ramification is studied.

For the purpose of this article, the following facts are of importance:

Proposition 2.7. Let $X$ be a smooth, separated, finite type $k$-scheme.

(a) If $X$ admits a good compactification $(X, \overline{X})$, then a stratified bundle $E$ on $X$ is regular singular if and only if it is $(X, \overline{X})$-regular singular.

(b) If $E$ is a stratified bundle of rank 1, then $E$ is regular singular.

(c) If $\iota : \text{Strat}^{\text{rs}}(X) \to \text{Strat}(X)$ denotes the inclusion functor, then for every object $E \in \text{Strat}^{\text{rs}}(X)$, $\iota$ induces an equivalence $(E)_{\overline{X}} \cong (\iota(E))_{\overline{X}}$.

Proof. Statement (a) is [Kin12a, Prop. 7.5], (c) is [Kin12a, Prop. 4.5], and (b) is [Gie75, Lemma 3.12]. We recall some arguments for (b) from [loc. cit.] for convenience: Let $A$ be a finite type $k$-algebra with coordinates $x_1, \ldots, x_n$, and $M$ a free rank 1 module over $A[x_1^{-1}, \ldots, x_r^{-1}]$, $1 \leq r \leq n$. Assume that $M$ carries an action of $D_{A[x_1^{-1}, \ldots, x_r^{-1}]/k}$, and that $e$ is a basis of $M$. With the notation from Definition 2.5, (b), it suffices to show that $\delta_{x_i}^{(m)}(e) \in eA$ for every $m > 0$. 

Documenta Mathematica 18 (2013) 1215–1242
and every $1 \leq i \leq r$. By Theorem 2.3, for $N \in \mathbb{N}$ there exists a nonzero section $s \in A$ such that $\delta^{(m)}_{2_i}(se) = 0$ for all $m \leq p^N$. But then in $\text{Frac}A$ we have
\[
\delta^{(m)}_{2_i}(e) = \delta^{(m)}_{2_i}(s^{-1}se) = \delta^{(m)}_{2_i}(s^{-1})se
\]
and $\delta^{(m)}_{2_i}(s^{-1})s \in A$.

**Remark 2.8.** Recall that the analogous statement to (b) in characteristic 0 is false.

We recall the definition of the monodromy group of a stratified bundle:

**Definition 2.9.** If $E \in \text{Strat}(X)$, then $\langle E \rangle_\otimes$ is the Tannakian subcategory of $\text{Strat}(X)$ generated by $E$. If $\omega: \langle E \rangle_\otimes \to \text{Vect}_k$ is a fiber functor, then the associated $k$-group scheme is called the monodromy group of $E$ (with respect to $\omega$). By [dS07], the monodromy group of a stratified bundle is always smooth.

Note that if $E$ is regular singular, then by Proposition 2.7, (c), we get the same result whether we compute the monodromy group of $E$ as an object of $\text{Strat}^{rs}(X)$ or $\text{Strat}(X)$.

Moreover, note that since $k$ is algebraically closed, the isomorphism class of the monodromy group does not depend on the choice of the fiber functor.

To finish this subsection, we briefly recall the notion of exponents of regular singular stratified bundles. For more details see [Kin12a] or [Kin12b].

**Proposition 2.10.** Let $(\overline{X}, X)$ be a good partial compactification of $X$ with $D := \overline{X} \smallsetminus X$ is irreducible. Let $E$ be a $\mathcal{O}_{\overline{X}/k}(\log \overline{X} \smallsetminus X)$-module, which is $\mathcal{O}_{\overline{X}}$-locally free of finite rank. Then the following are true:

(a) ([Gie75, Lemma 3.8]) There exists a decomposition
\[
\overline{E}|_D = \bigoplus_{\alpha \in \mathbb{Z}_p} F_{\alpha},
\]
with the property that if $x_1, \ldots, x_n$ are local coordinates around the generic point of $D$, such that $D = (x_1)$, then $e + x_1\overline{E} \in F_{\alpha}$ if and only if
\[
\delta^{(m)}_{2_1}(e) = \left(\frac{\alpha}{m}\right)e + x_1\overline{E}.
\]
For the definition of $\delta^{(m)}_{2_1}$, see Definition 2.5, (b). Write $\text{Exp}(\overline{E}) \subset \mathbb{Z}_p$ for the set of those $\alpha \in \mathbb{Z}_p$ for which $F_{\alpha} \neq 0$.

(b) ([Kin12a, Prop. 4.12]) If $\overline{E}$ is a second $\mathcal{O}_{\overline{X}}$-locally free, finite rank, $\mathcal{D}_{\overline{X}/k}(\log \overline{X} \smallsetminus X)$-module, such that the stratified bundles $\overline{E}|_X$ and $\overline{E}|_X$ are isomorphic, then the sets $\text{Exp}(\overline{E})$ and $\text{Exp}(\overline{E})$ have the same image in $\mathbb{Z}_p/\mathbb{Z}$.
Definition 2.11 ([Kin12a, Def. 4.13]). If \((X, \overline{X})\) is a good partial compactification, \(E\) an \((X, \overline{X})\)-regular singular stratified bundle and \(D\) an irreducible component of \(\overline{X} \setminus X\), then we define the exponents of \(E\) along \(D\) as follows: Let \(U\) be a sufficiently small open neighborhood of the generic point of \(D\), such that there exists an \(O_U\)-locally free \(\mathcal{O}_{\overline{U}/U}((\log U \setminus X))\)-module extending \(E|_{U \cap X}\). Then the set of exponents of \(E\) along \(D\) is the image of the set \(\text{Exp}(E)\) in \(\mathbb{Z}_{/p}\) from Proposition 2.10. This construction is well-defined by Proposition 2.10.

Having all exponents equal to 0 mod \(\mathbb{Z}\) is a stronger condition in characteristic \(p > 0\) than in characteristic 0. In particular there are no regular singular stratified bundles with nontrivial "nilpotent residues":

Proposition 2.12 ([Kin12a, Cor. 5.4]). Let \((X, \overline{X})\) be a good partial compactification and \(E\) an \((X, \overline{X})\)-regular singular stratified bundle. If the exponents of \(E\) along all components of \(\overline{X} \setminus X\) are 0 in \(\mathbb{Z}_{/p}\), then there exists a stratified bundle \(E\) on \(\overline{X}\) extending \(E\).

3 Special case I: Stratified line bundles

We continue to denote by \(k\) an algebraically closed field of characteristic \(p > 0\), and by \(X\) a smooth, separated, connected, finite type \(k\)-scheme. We start with a group theoretic definition.

Definition 3.1. If \(G\) is an abelian group, write \((G)_p\) for the projective system

\[
G \leftarrow G_p \leftarrow \ldots.
\]

This defines an exact functor from the category of abelian groups into the category of pro-systems of abelian groups. We write \(\lim \leftarrow_p G := \lim \leftarrow_p ((G)_p)\), and \(R^1 \lim \leftarrow_p G := R^1 \lim \leftarrow_p ((G)_p)\). This construction defines a left exact functor \(\lim \leftarrow_p\) from the category of abelian groups to itself, and for a short exact sequence

\[
0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0
\]

of abelian groups, we have the long exact sequence

\[
0 \rightarrow \lim \leftarrow_p A \rightarrow \lim \leftarrow_p B \rightarrow \lim \leftarrow_p C \rightarrow R^1 \lim \leftarrow_p A \rightarrow R^1 \lim \leftarrow_p B \rightarrow R^1 \lim \leftarrow_p C \rightarrow 0.
\]

Definition 3.2. We write \(\text{Pic}^{\text{Strat}}(X)\) for the set of isomorphism classes of stratified bundles of rank 1. The tensor product of stratified bundles gives this set an abelian group structure.

By Theorem 2.3 we can associate with every stratified line bundle a sequence of elements \(L_n \in \text{Pic}^X\), such that \(L_n^{p+1} = L_n\). This gives a group homomorphism \(\text{Pic}^{\text{Strat}}(X) \rightarrow \lim \leftarrow_p \text{Pic}^X\).
**Definition 3.3.** Denote by $\mathbb{I}(X)$ the subgroup of $\text{Pic}^{\text{Strat}}(X)$ of isomorphism classes of stratified line bundles $(L_n, \sigma_n)$ with $L_n \cong O_X$ for all $n$.

We clearly have the following:

**Proposition 3.4.** The morphisms described above fit in a functorial short exact sequence

$$0 \to \mathbb{I}(X) \to \text{Pic}^{\text{Strat}}(X) \to \lim_{\rightarrow p} \text{Pic}(X) \to 0.$$ 

Moreover, it is not difficult to describe the group $\mathbb{I}(X)$ concretely:

**Proposition 3.5.** There is a functorial exact sequence

$$0 \longrightarrow k^\times \longrightarrow H^0(X, O_X^\times) \xrightarrow{\Delta} \lim_{\longrightarrow n} H^0(X, O_X^\times)/H^0(X, O_X^\times)^{p^n} \longrightarrow \mathbb{I}(X) \longrightarrow 0$$

where the morphism $k^\times \to H^0(X, O_X^\times)$ is the canonical inclusion, and $\Delta$ the diagonal map.

**Proof.** Since the base field $k$ is algebraically closed of characteristic $p > 0$, the sequence is exact at $H^0(X, O_X^\times)$. Indeed, $H^0(X, O_X^\times)/k^\times$ is a finitely generated free abelian group, because of the exact sequence

$$0 \longrightarrow H^0(X, O_X^\times) \longrightarrow H^0(X, O_X^\times) \longrightarrow \Phi, D, \mathbb{Z} \longrightarrow \text{Cl}(X) \longrightarrow \text{Cl}(X) \longrightarrow 0,$$

where $\overline{X}$ is any normal compactification of $X$, and $D_i$ the irreducible components of $\overline{X} \setminus X$. This shows that the kernel of $\Delta$ is precisely $k^\times$.

Thus we only need to construct the morphism

$$\phi : \text{coker}(\Delta) \to \mathbb{I}(X),$$

and show that it is a natural isomorphism.

Given an element $\alpha = (a_1, a_2, \ldots) \in \lim_{\longrightarrow n} H^0(X, O_X^\times)/H^0(X, O_X^\times)^{p^n}$, take any sequence of lifts $a := (a_1, a_2, \ldots) \in H^0(X, O_X^\times)^{\mathbb{N}}$, and define the stratified bundle $\Psi(a) := (\Psi_n(a), \sigma_n^a)_{n \geq 0}$ by setting $\Psi_n(a) = O_X$, and by defining the transition isomorphisms $\sigma_n^a : \Psi_n(a) \to \Phi_n^a \Psi_n(a+n+1)$ as follows: $\sigma_0^a$ is multiplication by $a_1$, and $\sigma_n^a = (a_{n+1}^{-1}a_n^{-1})^{1/p^n}$, $n > 0$. This works, since by definition $a_{n+1} \equiv a_n \mod p^n$. We have to check that this construction gives a well-defined map $\Psi$: If we pick a different sequence of lifts $a' := (a'_1, a'_2, \ldots)$ of $\alpha = (a_1, a_2, \ldots)$, then the resulting stratified line bundle $\Psi(a')$ is isomorphic to $\Psi(a)$. Indeed, if $b_n^p = a'_n a_n^{-1}$, then the sequence of isomorphisms $\phi_n = \text{id}, \phi_n : \Psi_n(a) \xrightarrow{b_n} \Psi_n(a'), n > 1$, defines an isomorphism of stratified bundles. We write $\Psi(a)$ for the isomorphism class of the stratified bundle $\Psi(a)$.

It is readily checked that this map is surjective: If the stratified bundle $L = (L_n, \sigma_n)_{n \geq 0}$ is given by $L_n = O_X$ and the transition morphisms $\sigma_n$, then $\sigma_n \in H^0(X, O_X^\times)$, and $L = \Psi((\sigma_0, \sigma_0, \sigma_1, \sigma_0, \sigma_0^2, \sigma_2^2, \ldots))$. 

*Documenta Mathematica 18 (2013) 1215–1242*
To compute the kernel, note that a stratification \( L \) on \( O_X \) which is given by transition morphisms \( \sigma_n \), is trivial if and only if there exists a sequence \( \phi_0, \phi_1, \ldots \in H^0(X, O_X^\cdot) \) such that \( \sigma_n \phi_{n+1}^{p^n} \equiv \phi_n, n > 0 \). By recursion, this means that
\[
\phi_0 = \sigma_0 \phi_1^{p^1} \cdots \sigma_n^{p^n} \phi_{n+1}^{p^{n+1}}
\]
for every \( n \). In other words: \( L = \Psi(\Delta(\phi_0)) \), so the kernel of \( \Psi \) is precisely \( H^0(X, O_X^\cdot) \). This completes the proof. \( \square \)

**Remark 3.6.** (a) The description of \( \|I(X) \) in Proposition 3.5 was inspired by a similar description for stratified bundles line bundles on \( k((t)) \) in [MvdP03].

(b) Note that the abelian group \( \lim_{\rightarrow} H^0(X, O_X^\cdot)/H^0(X, O_X^\cdot)^{p^n} \) is just the \( p \)-adic completion of the free finitely generated abelian group \( H^0(X, O_X^\cdot)/k^\times \), and the map \( \Delta \) induces the canonical map from \( H^0(X, O_X^\cdot)/k^\times \) into its \( p \)-adic completion.

**Corollary 3.7.** The group \( \|I(X) \) is trivial if and only if \( H^0(X, O_X^\cdot) = k^\times \).

**Proof.** This follows from Proposition 3.5 and Remark 3.6, (b), because the map from \( \mathbb{Z}^r \) into its \( p \)-adic completion is surjective if and only if \( r = 0 \). \( \square \)

Corollary 3.7 allows us to exhibit a class of examples of varieties \( X \) such that \( \|I(X) \) is trivial:

**Proposition 3.8.** If \( k \) is an algebraically closed field, and if \( X \) is a connected normal \( k \)-scheme of finite type, such that the maximal abelian pro-\( \ell \)-quotient \( \pi_1(X)_{\text{ab},(\ell)} \) is trivial for some \( \ell \not\equiv \text{char}(k) \), then \( H^0(X, O_X^\cdot) = k^\times \). In particular, if \( k \) has positive characteristic, then \( \|I(X) \) = 0. \( \square \)

**Proof.** This argument is due to Hélène Esnault. Assume \( f \in H^0(X, O_X^\cdot) \setminus k^\times \). Then \( f \) induces a dominant morphism \( f' : X \to \mathbb{G}_{m,k} \cong \mathbb{A}_k^1 \setminus \{0\} \), as \( f' \) is given by the map \( k[x^\pm] \to H^0(X, O_X^\cdot), x \mapsto f \), which is injective if and only if \( f \) is transcendental over \( k \). Thus \( f' \) induces an open morphism \( \pi_1(X) \to \pi_1(\mathbb{G}_{m,k}) \), see e.g. [Sti02, Lemma 4.2.10]. But under our assumption, the maximal abelian pro-\( \ell \)-quotient of the image of this morphism is trivial, so the image of \( \pi_1(X) \) cannot have finite index in the group \( \pi_1(\mathbb{G}_{m,k}) \), as in fact \( \pi_1(\mathbb{G}_{m,k})^{(\ell)} \cong \mathbb{Z}_\ell \).

**Remark 3.9.** We can modify the argument of Proposition 3.8 slightly to obtain: If \( k \) has positive characteristic \( p \), and \( \pi_1(X)_{\text{ab},(p)} = 0 \), then \( H^0(X, O_X^\cdot) = k \).

A consequence is a proof of the folklore fact that over a field \( k \) of positive characteristic, unlike in characteristic 0, no normal, connected, affine, finite type \( k \)-scheme \( X \) of positive dimension is simply-connected, or even has \( \pi_1(X)_{\text{ab},(p)} = 0 \).
Indeed, if \( f \in H^0(X, \mathcal{O}_X) \) is nonconstant, then \( f \) induces a dominant morphism \( X \to \mathbb{A}^1_k \) and hence an open morphism \( \pi_1^{\text{ab}}(p)(X) \to \pi_1^{\text{ab}}(p)(\mathbb{A}^1_k) \). But by [Kat86, 1.4.3, 1.4.4] we have \( H^2(\pi_1(\mathbb{A}^1_k, \mathbb{Z}_p)) = 0 \), so \( \pi_1(\mathbb{A}^1_k) \) is free pro-\( p \) of rank \( \dim \pi_2 H^1(\mathbb{A}^1_k, \mathbb{Z}_p) = \# k \). Thus the image of \( \pi_1^{\text{ab}}(p)(X) \) in this group can only have finite index if \( \pi_1(X)^{\text{ab}}(p) \neq 0 \).

Recall that a good compactification of a smooth \( k \)-scheme \( X \) is a proper, smooth, finite type \( k \)-scheme \( \overline{X} \), with a dominant open immersion \( X \hookrightarrow \overline{X} \), such that \( \overline{X} \setminus X \) is the support of a strict normal crossings divisor.

The main goal of this section is to prove the following theorem:

**Theorem 3.10.** Let \( X, Y \) be smooth, separated, finite type \( k \)-schemes, and let \( f : Y \to X \) be a morphism such that the following conditions are satisfied:

(a) There exist good compactifications \( \overline{X} \) and \( \overline{Y} \) of \( X \) and \( Y \), such that \( f \) extends to a morphism \( \overline{f} : \overline{Y} \to \overline{X} \).

(b) \( f \) induces an isomorphism

\[
\pi_1^{\text{ab}}(p')(Y) \to \pi_1^{\text{ab}}(p') \pi_1(X),
\]

(2)

(c) The cokernel of the morphism

\[
H^0(X, \mathcal{O}_X^\wedge) \to H^0(Y, \mathcal{O}_Y^\wedge).
\]

induced by \( f \) is a p-group.

Then pull-back along \( f \) induces an isomorphism

\[
\text{Pic}^{\text{Strat}}(X) \xrightarrow{\sim} \text{Pic}^{\text{Strat}}(Y).
\]

**Remark 3.11.** (a) The morphism (3) is an isomorphism in the following two situations:

(i) \( X, Y \) proper over \( k \).

(ii) \( \pi_1^{\text{ab}}(p')(Y) = \pi_1^{\text{ab}}(p')(X) = 1 \) (e.g. (2) is an isomorphism and \( X = \text{Spec}(k) \)). See Proposition 3.8.

Thus in these two cases, the theorem establishes that \( f \) induces an isomorphism of abelian groups \( \text{Pic}^{\text{Strat}}(X) \to \text{Pic}^{\text{Strat}}(Y) \), if (2) is an isomorphism.

(b) If (2) is an isomorphism, then (3) is always injective. Indeed, if \( \alpha \in H^0(X, \mathcal{O}_X^\wedge) \) is a global unit, then \( \alpha \) induces a morphism \( X \to \mathbb{G}_m \), which is dominant if and only if \( \alpha \) is non-constant. Hence the induced continuous morphism \( \pi_1^{\text{ab}}(p')(X) \to \mathbb{Z}(p') \) is open if and only if \( \alpha \) is non-constant. If (2) is an isomorphism, then \( f^* \alpha \) is non-constant whenever \( \alpha \) is. In particular, it follows that (3) is injective.
Evidence for a Generalization of Gieseker’s Conjecture

(c) Clearly (3) is not necessarily an isomorphism, even if (2) is: Just take the purely inseparable morphism $\mathbb{G}_m \to \mathbb{G}_m$ defined by taking the $p$-th root of a coordinate on $\mathbb{G}_m$.

(d) The question whether (3) has a $p$-group as cokernel whenever (2) is an isomorphism, belongs to the area of Grothendieck’s anabelian geometry: The cokernel of (3) is finitely generated by [Kah06, Lemme 1], so we can split off the $p$-power torsion. If $\alpha \in H^1(Y, \mathcal{O}_Y^\times)$ is a global unit of order prime to $p$ in the cokernel of (3), and if (2) is an isomorphism, then $\alpha$ comes from a global unit on $X$ if and only if the induced morphism on fundamental groups

$$\pi^\text{ab}(p')_1(X) \xrightarrow{\sim} \pi^\text{ab}(p')_1(Y) \xrightarrow{\alpha_*} \pi^\text{ab}(p')_1(\mathbb{G}_m) = \mathbb{Z}(p')$$

is induced by a morphism of $k$-schemes $X \to \mathbb{G}_m$. The author does not know whether the condition that (2) is an isomorphism always implies that the cokernel of (3) is a $p$-group.

We need a few lemmas to prepare the proof of Theorem 3.10.

**Lemma 3.12.** If $G$ is a finitely generated abelian group, then the functor $\lim_{\leftarrow}^{-p}$ is naturally isomorphic to the functor which assigns to $G$ its subgroup $G[p']$ of torsion elements of order prime to $p$. If $G$ is a finite abelian group or an abelian group (not necessarily finitely generated) on which multiplication by $p$ is surjective, then $R^1\lim_{\leftarrow}^{-p} G = 0$.

**Proof.** In a finitely generated abelian group, an element is infinitely $p$-divisible if and only if it is torsion of order prime to $p$. Moreover, an element $x \in G[p']$ admits a unique $p$-th root in $G[p']$. It follows that the map $\lim_{\leftarrow}^{-p} G \to G[p']$, $(x_1, x_2, \ldots) \mapsto x_1$ is an isomorphism. For the second claim, if $G$ is finite or if multiplication by $p$ on $G$ is surjective, then the projective system (1) satisfies the Mittag-Leffler condition, so $R^1\lim_{\leftarrow}^{-p} G = 0$. \hfill \Box

**Lemma 3.13.** Consider the following morphism of short exact sequences of abelian groups

\[
\begin{array}{cccccc}
0 & \to & D_1 & \to & G_1 & \to & F_1 & \to & 0 \\
& f & \downarrow & g & \downarrow & h & & & \\
0 & \to & D_2 & \to & G_2 & \to & F_2 & \to & 0
\end{array}
\]

with $F_1, F_2$ finitely generated, and such that multiplication by $p$ on $D_1, D_2$ is surjective (e.g. $D_1, D_2$ divisible).

Assume that the following conditions are satisfied:

(a) $f$ is surjective with finite kernel.
(b) \( \ker(g) \) contains no torsion elements of order prime to \( p \).

(c) The restriction \( h|_{F_1[p']} : F_1[p'] \to F_2[p'] \) is surjective.

Then the induced morphism \( \lim_{\leftarrow p} G_1 \to \lim_{\leftarrow p} G_2 \) is an isomorphism.

Proof. Since \( R^1 \lim_{\leftarrow p} D_i = 0 \) for \( i = 1,2 \) by Lemma 3.12, we get the following morphism of short exact sequences:

\[
\begin{array}{ccccccccc}
0 & \to & \lim_{\leftarrow p} D_1 & \to & \lim_{\leftarrow p} G_1 & \to & \lim_{\leftarrow p} F_1 & \to & 0 \\
& & \downarrow f & & \downarrow g & & \downarrow h & & \\
0 & \to & \lim_{\leftarrow p} D_2 & \to & \lim_{\leftarrow p} G_2 & \to & \lim_{\leftarrow p} F_2 & \to & 0
\end{array}
\]

By the left-exactness of \( \lim_{\leftarrow p} \), we have \( \lim_{\leftarrow p} (\ker g) = \ker(\lim_{\leftarrow p} g) \). But by assumption (a), \( \ker g \) is an extension of two finitely generated groups, and hence finitely generated itself. It has no prime-to-\( p \) torsion by (b), so by Lemma 3.12, \( \lim_{\leftarrow p} (\ker g) = 0 \). Thus \( \lim_{\leftarrow p} g \) is injective.

Next, by (a), \( f \) is surjective with finite kernel, which implies that \( \lim_{\leftarrow p} f \) is surjective, since \( R^1 \lim_{\leftarrow p} (\ker(f)) = 0 \) by Lemma 3.12.

Finally, by Lemma 3.12 we know that \( \lim_{\leftarrow p} h = h|_{F_1[p']} : F_1[p'] \to F_2[p'] \). This morphism is surjective by (c). It follows that \( \lim_{\leftarrow p} g \) is also surjective, which completes the proof.

**Lemma 3.14.** If \( f : Y \to X \) is a morphism of connected, smooth, separated, finite type \( k \)-schemes, such that \( f \) induces an isomorphism \( \pi_1(Y)^{\text{ab.}(p')} \xrightarrow{\sim} \pi_1(X)^{\text{ab.}(p')} \), and if \( X,\overline{X},Y \) are smooth compactifications of \( X,Y \), such that \( f \) extends to \( f : \overline{Y} \to \overline{X} \), then the map

\[
\pi_1(\overline{Y})^{\text{ab.}(p')} \to \pi_1(\overline{X})^{\text{ab.}(p')}
\]

induced by \( \bar{f} \) is surjective with finite kernel.

Proof. The surjectivity is clear, as \( \pi_1(X)^{\text{ab.}(p')} \) and \( \pi_1(Y)^{\text{ab.}(p')} \) surject onto \( \pi_1(\overline{X})^{\text{ab.}(p')} \) and \( \pi_1(\overline{Y})^{\text{ab.}(p')} \), respectively. We use the theory of the Albanese variety: After choosing a base point \( x \in X(k) \), there exists a unique semi-abelian variety \( \text{Alb}_X \), together with a map \( \text{alb}_{X,x} : X \to \text{Alb}_X \), such that \( \text{alb}_{X,x}(x) = 0 \), and such that any map \( g : X \to A \) from \( X \) to a semi-abelian variety \( A \) with \( g(x) = 0 \) factors uniquely through \( \text{alb}_{X,x} \). For more details, see e.g. [SS03]. The Albanese variety classifies abelian coverings of \( X \) in the following sense: For \( \ell \) a prime different from \( p \), there exists a canonical isomorphism

\[
\text{hom}(H^1(X,\mathbb{Z}_\ell),\mathbb{Q}_\ell/\mathbb{Z}_\ell) \cong \text{Alb}_X(k)(\ell),
\]
see [SS03, Cor. 4.3], where \( \text{Alb}_X(k) \{ \ell \} \) is the subgroup of \( \text{Alb}_X(k) \) of all \( \ell \)-power torsion elements. As a semi-abelian variety, \( \text{Alb}_X \) can be written uniquely as an extension of an abelian variety by a torus. The unique abelian variety appearing in this description of \( \text{Alb}_X \) is canonically isomorphic to \( \text{Alb}_X \). By Chevalley’s structure theorem for algebraic groups, the connected component of the origin (with its reduced structure) \( K_0^{\text{red}} \) of the kernel of the morphism of group schemes \( \text{Alb}_Y \to \text{Alb}_X \) is a semi-abelian variety, and since every map from an abelian variety into an affine variety is constant, it follows that the unique abelian variety quotient of \( K_0^{\text{red}} \) is precisely the connected component of the origin (with its reduced structure) of the kernel of \( \text{Alb}_X \to \text{Alb}_X \).

With that in mind, we see that the assumption that \( \pi_1(X)_{\text{ab},(p')} \cong \pi_1(X)_{\text{ab},(p')} \) is a prime different from \( p \) is an isomorphism implies that the map \( \text{Alb}_Y(k) \to \text{Alb}_X(k) \) is an isomorphism on prime-to-\( p \) torsion, and hence \( K_0^{\text{red}} \) is the trivial semi-abelian variety. Then \( \ker(\text{Alb}_Y \to \text{Alb}_X) \) is the trivial abelian variety, which shows that the kernel of \( \text{Alb}_Y(k) \to \text{Alb}_X(k) \) is a finite group.

Writing \( T_\ell \) for the Tate module, we see that if \( \ell \) is a prime different from \( p \), then \( T_\ell \text{Alb}_X \to T_\ell \text{Alb}_Y \) is injective. Finally, since \( X \) and \( Y \) are proper and smooth, we get a morphism of short exact sequences

\[
0 \to \text{hom}(\text{NS}(Y) \ell, \mathbb{Q}/\mathbb{Z}_\ell) \to \pi_1(Y)_{\text{ab},(\ell)} \to T_\ell \text{Alb}_X \to 0 \\
0 \to \text{hom}(\text{NS}(X) \ell, \mathbb{Q}/\mathbb{Z}_\ell) \to \pi_1(X)_{\text{ab},(\ell)} \to T_\ell \text{Alb}_X \to 0.
\]

The groups on the left are finite groups, so we deduce that \( \pi_1(Y)_{\text{ab},(\ell)} \to \pi_1(X)_{\text{ab},(\ell)} \) always has a finite kernel, and is injective for all but finitely many \( \ell \). This implies that the kernel of \( \pi_1(Y)_{\text{ab},(p')} \to \pi_1(X)_{\text{ab},(p')} \) is finite.

**Proposition 3.15.** With the notations and assumptions of Theorem 3.10, pullback along \( f \) induces an isomorphism

\[
\lim_{\leftarrow p} \text{Pic}(X) \cong \lim_{\leftarrow p} \text{Pic}(Y).
\]

**Proof.** By Lemma 3.14 the induced morphism \( \pi_1(Y)_{\text{ab},p'} \to \pi_1(X)_{\text{ab},p'} \) is surjective with finite kernel. Define \( K_X := \ker(\text{Pic}(X) \to \text{Pic}(X)) \) and \( K_Y := \ker(\text{Pic}(Y) \to \text{Pic}(Y)) \). We get a commutative diagram

\[
0 \to K_X \to \text{Pic}(X) \to \text{Pic}(X) \to 0 \\
0 \to K_Y \to \text{Pic}(Y) \to \text{Pic}(Y) \to 0
\]

with \( K_X, K_Y \) finitely generated abelian groups. We have a second exact sequence

\[
0 \to \text{Pic}^0(X) \to \text{Pic}(X) \to \text{NS}(X) \to 0
\]
with $\text{NS}(\overline{X})$ a finitely generated group, according to, e.g., [Kah06, Thm. 3]. Moreover, $\text{Pic}^0(\overline{X})$ is the set of $k$-points of an abelian variety, so $\text{Pic}^0(\overline{X})$ is a divisible abelian group.

Define $\text{Pic}^0(\overline{X})$ as the image of $\text{Pic}^0(\overline{X})$ in $\text{Pic}(\overline{X})$, and $\text{NS}(\overline{X})$ as $\text{Pic}(\overline{X})/\text{Pic}^0(\overline{X})$. Then $\text{Pic}^0(\overline{X})$ still divisible, and $\text{NS}(\overline{X})$ is still finitely generated. Pullback along $f$ induces a morphism $\text{Pic}^0(X) \to \text{Pic}^0(Y)$, and hence so is the third. Since by Lemma 3.14 the morphism $\pi_1(\overline{Y})_{\text{ab},(p')} \to \pi_1(\overline{X})_{\text{ab},(p')}$ is surjective with finite kernel, the same argument as above shows that for almost all $n$ prime to $p$, $\text{Pic}(\overline{X})[n] \twoheadrightarrow \text{Pic}(\overline{Y})[n]$. Since $\text{Pic}^0(\overline{X})$ is divisible, we have a short exact sequence

$$0 \to \text{Pic}^0(\overline{X})[n] \to \text{Pic}(\overline{X})[n] \to \text{NS}(\overline{X})[n] \to 0$$

which shows that for almost all $n$ prime to $p$, $\text{Pic}^0(\overline{X})[n] \twoheadrightarrow \text{Pic}^0(\overline{Y})[n]$. Since $\text{Pic}^0(\overline{X})$ and $\text{Pic}^0(\overline{Y})$ are sets of $k$-points of abelian varieties, and since pullback along $f$ induces a morphism of the underlying abelian varieties, it follows that $\text{Pic}^0(\overline{X}) \to \text{Pic}^0(\overline{Y})$ is surjective with finite kernel. From the morphism of short exact sequences

$$0 \to K_X \cap \text{Pic}^0(\overline{X}) \to \text{Pic}^0(\overline{X}) \to \text{Pic}^0(X) \to 0$$

$$0 \to K_Y \cap \text{Pic}^0(\overline{Y}) \to \text{Pic}^0(\overline{Y}) \to \text{Pic}^0(Y) \to 0$$

We are now in the situation of Lemma 3.13, and check that the conditions (a), (b) and (c) from the lemma are satisfied for diagram (4).

For Lemma 3.13, (a) we have to show that $\text{Pic}^0(\overline{X}) \to \text{Pic}^0(Y)$ is surjective with finite kernel. For every $n$ prime to $p$, Kummer theory shows that there a morphism of short exact sequences of abelian groups

$$0 \twoheadrightarrow H^0(X, \mathcal{O}_X) / H^0(X, \mathcal{O}_X^n) \twoheadrightarrow \text{Hom}(\pi_1(\overline{X})_{\text{ab},(p')}, \mathbb{Z}/n\mathbb{Z}) \to \text{Pic}(X)[n] \twoheadrightarrow 0$$

$$0 \twoheadrightarrow H^0(Y, \mathcal{O}_Y) / H^0(Y, \mathcal{O}_Y^n) \twoheadrightarrow \text{Hom}(\pi_1(\overline{Y})_{\text{ab},(p')}, \mathbb{Z}/n\mathbb{Z}) \to \text{Pic}(Y)[n] \twoheadrightarrow 0$$

where the two left vertical arrows are isomorphisms by the assumptions (b) and (c) of Theorem 3.10, and hence so is the third. Since by Lemma 3.14 the morphism $\pi_1(\overline{Y})_{\text{ab},(p')} \to \pi_1(\overline{X})_{\text{ab},(p')}$ is surjective with finite kernel, the same argument as above shows that for almost all $n$ prime to $p$, $\text{Pic}(\overline{X})[n] \twoheadrightarrow \text{Pic}(\overline{Y})[n]$. Since $\text{Pic}^0(\overline{X})$ is divisible, we have a short exact sequence

$$0 \to \text{Pic}^0(\overline{X})[n] \to \text{Pic}(\overline{X})[n] \to \text{NS}(\overline{X})[n] \to 0$$

which shows that for almost all $n$ prime to $p$, $\text{Pic}^0(\overline{X})[n] \twoheadrightarrow \text{Pic}^0(\overline{Y})[n]$. Since $\text{Pic}^0(\overline{X})$ and $\text{Pic}^0(\overline{Y})$ are sets of $k$-points of abelian varieties, and since pullback along $f$ induces a morphism of the underlying abelian varieties, it follows that $\text{Pic}^0(\overline{X}) \to \text{Pic}^0(\overline{Y})$ is surjective with finite kernel. From the morphism of short exact sequences

$$0 \to K_X \cap \text{Pic}^0(\overline{X}) \to \text{Pic}^0(\overline{X}) \to \text{Pic}^0(X) \to 0$$

$$0 \to K_Y \cap \text{Pic}^0(\overline{Y}) \to \text{Pic}^0(\overline{Y}) \to \text{Pic}^0(Y) \to 0$$

The following commutative diagram with exact rows:

$$
\begin{array}{cccccc}
0 & \to & \text{Pic}^0(X) & \to & \text{Pic}(X) & \to & \text{NS}(X) & \to & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & \text{Pic}^0(Y) & \to & \text{Pic}(Y) & \to & \text{NS}(Y) & \to & 0 \\
\end{array}
$$

(4)
Evidence for a Generalization of Gieseker’s Conjecture

we see that $\text{Pic}^0(X) \to \text{Pic}^0(Y)$ is surjective. To prove that its kernel is finite, it suffices to show that $\text{coker}(K_X \cap \text{Pic}^0(X) \to K_Y \cap \text{Pic}^0(Y))$ is finite. For this, let $\ell$ be a prime different from $p$, and write $X \setminus X =: D_X = \bigcup_{i=1}^r D_{X,i}$ with $D_{X,i}$ smooth divisors. Let $M_X := \bigoplus D_{X,i} \mathbb{Z}$ be the free $\mathbb{Z}$-module of rank $r(X)$, and similarly for $Y$. We have the associated Gysin exact sequence in étale cohomology [Mil80, Ch. VI, Cor. 5.3]:

\[
\begin{array}{ccccccccc}
0 & \to & H^1(X, \mathbb{Z}_\ell(1)) & \to & H^1(X, \mathbb{Z}_\ell(1)) & \to & M_X \otimes \mathbb{Z}_\ell & \to & H^2(X, \mathbb{Z}_\ell(1)) & \\
& & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & H^1(Y, \mathbb{Z}_\ell(1)) & \to & H^1(Y, \mathbb{Z}_\ell(1)) & \to & M_Y \otimes \mathbb{Z}_\ell & \to & H^2(Y, \mathbb{Z}_\ell(1)) & 
\end{array}
\]

As we have seen, for almost all primes $\ell$ the two left vertical arrows are isomorphisms, so $\ker(\epsilon_{1,X}^\ell) \cong \ker(\epsilon_{1,Y}^\ell)$ for almost all $\ell$. But by construction, we have a commutative diagram

\[
\begin{array}{ccc}
\ker(\epsilon_{1,X}^\ell) & \to & (K_X \cap \text{Pic}^0(X)) \otimes \mathbb{Z}_\ell \\
& \downarrow & \downarrow \\
\ker(\epsilon_{1,Y}^\ell) & \to & (K_Y \cap \text{Pic}^0(Y)) \otimes \mathbb{Z}_\ell
\end{array}
\]

This shows that the finitely generated group $\text{coker}(K_X \cap \text{Pic}^0(X) \to K_Y \cap \text{Pic}^0(Y))$ is trivial after tensoring with $\mathbb{Z}_\ell$ for almost all $\ell$ and hence finite. This finishes the proof that condition (a) from Lemma 3.13, holds for diagram (4).

To show that Lemma 3.13, (b) holds for diagram (4) we note that we have seen below (5) that the kernel of $\text{Pic}(X) \to \text{Pic}(Y)$ contains no prime-to-$p$-torsion.

Finally, let’s check that Lemma 3.13, (c) holds. Since $\text{Pic}^0(X)$ is divisible, for every $n$ prime to $p$ we get a commutative diagram

\[
\begin{array}{ccc}
\text{Pic}(X)[n] & \to & \text{NS}(X)[n] \\
\downarrow & & \downarrow \\
\text{Pic}(Y)[n] & \to & \text{NS}(Y)[n]
\end{array}
\]

with surjective horizontal arrows. This shows that $\text{NS}(X)[p'] \to \text{NS}(X)[p']$ is surjective, so Lemma 3.13, (c) holds for diagram (4).

With Proposition 3.15, the proof of Theorem 3.10 becomes simple:

**Proof of Theorem 3.10.** By Proposition 3.4, $f$ induces a morphism of short
exact sequences

\[
0 \rightarrow \lim_{\leftarrow p} \text{Pic}(X) \rightarrow \text{Pic}^\text{Strat}(X) \rightarrow 0
\]

\[
0 \rightarrow \lim_{\leftarrow p} \text{Pic}(Y) \rightarrow \text{Pic}^\text{Strat}(Y) \rightarrow 0
\]

By assumption there exist good compactifications \(\overline{X}\) and \(\overline{Y}\), such that \(f: Y \rightarrow X\). Then the right vertical morphism is an isomorphism by Proposition 3.15. We show that Proposition 3.5 implies that the left vertical arrow is an isomorphism under our assumptions: If \(C := \text{coker}(H^0(X, \mathcal{O}_X) \rightarrow H^0(Y, \mathcal{O}_Y))\), then by assumption \(C\) is a finitely generated abelian \(p\)-group, and hence a finite \(p\)-group. It follows that the diagonal map

\[C \rightarrow \lim_{\leftarrow p} C/p^n\]

is an isomorphism. By the exactness of \(p\)-adic completion of finitely generated abelian groups, we have a canonical isomorphism

\[
\lim_{\leftarrow p} C/p^n \cong \text{coker}(\lim_{\leftarrow p} H^0(X, \mathcal{O}_X)^\wedge /p^n \rightarrow \lim_{\leftarrow p} H^0(Y, \mathcal{O}_Y)^\wedge /p^n),
\]

so looking at the long exact sequence attached to the morphism of short exact sequences

\[
0 \rightarrow H^0(X, \mathcal{O}_X)^\wedge /k^\times \rightarrow H^0(X, \mathcal{O}_X)^\wedge /p^n \rightarrow \mathbb{I}(X) \rightarrow 0
\]

\[
0 \rightarrow H^0(Y, \mathcal{O}_Y)^\wedge /k^\times \rightarrow H^0(Y, \mathcal{O}_Y)^\wedge /p^n \rightarrow \mathbb{I}(Y) \rightarrow 0,
\]

we see that \(\mathbb{I}(X) \rightarrow \mathbb{I}(Y)\) is surjective. But by Remark 3.11, (b) the two left vertical arrows are injective, so \(\mathbb{I}(X) \cong \mathbb{I}(Y)\) as claimed.

To close this section, we show that in the case that \(X = \text{Spec} k\), the assumption that \(Y\) admits a good compactification is not necessary:

**Proposition 3.16.** Let \(Y\) be a smooth, connected, separated, finite type \(k\)-scheme, such that \(\pi^{ab,(\nu)}(Y) = 0\). Then \(\text{Pic}^\text{Strat}(Y) = 0\).

In the proof of Proposition 3.16 we will use Nagata’s theorem on compactifications (see [Lütt93]) to find a normal projective compactification \(\overline{Y}\) of \(Y\). In a preliminary version of this article we used de Jong’s theorem on alterations ([dJ96]) to replace \(\overline{Y}\) with a nice simplicial scheme, and then we studied the attached groups of simplicial line bundles. These arguments were long and technical.
Brian Conrad suggested working directly with the normal projective compactifications instead of using simplicial techniques: He pointed out a theorem by Lang stating that if $\overline{Y}$ is projective and normal, then the group $\text{Cl}^{\text{alg}}(\overline{Y})$ of classes of those Weil divisors modulo linear equivalence which are algebraically equivalent to 0 is the group of $k$-points of an abelian variety; for a modern treatment see [BGS11].

We are grateful to Brian Conrad for bringing this result to our attention.

**Proof of Proposition 3.16.** By Proposition 3.8, we know that $I(Y) = 0$. Hence, according to Proposition 3.4, we only have to show that $\lim_{\leftarrow p} \text{Pic}(Y) = 0$. As in (5), Kummer theory shows that $\text{Pic}(Y)[n] = 0$, whenever $n$ is prime to $p$. Thus, Lemma 3.12 shows that it suffices to prove that $\text{Pic}(Y)$ is a finitely generated group.

To this end, let $\overline{Y}$ be a normal projective compactification of $Y$, and $\text{Cl}(\overline{Y})$ the group of Weil divisors on $\overline{Y}$ modulo linear equivalence. We then have a surjection $\text{Cl}(\overline{Y}) \twoheadrightarrow \text{Pic}(Y)$, and a short exact sequence $0 \to \text{Cl}^{\text{alg}}(\overline{Y}) \to \text{Cl}(\overline{Y}) \to \text{NS}(\overline{Y}) \to 0$ with $\text{NS}(\overline{Y})$ finitely generated by [Kah06, Thm. 3]. By the aforementioned result of Lang, $\text{Cl}^{\text{alg}}(\overline{Y})$ is the set of $k$-points of an abelian variety. More precisely, given a point $y \in Y(k)$, there is an abelian variety $A$ and a rational map $\alpha_y : \overline{Y} \to A$ defined around $y$, such that $\alpha_y(y) = 0$, and such that every rational map $\beta : Y \to B$ defined around $y$, with $B$ an abelian variety and $\beta(y) = 0$, factors through $\alpha_y$. The abelian variety $A$ is the Albanese variety for rational maps, and its dual has the property that $A^\vee(k) = \text{Cl}^{\text{alg}}(\overline{Y})$. For a modern treatment see [BGS11].

Now to finish, note that since the kernel of the surjection $\text{Cl}(\overline{Y}) \twoheadrightarrow \text{Pic}(Y)$ is finitely generated, it follows that $\text{Cl}(\overline{Y})[n] = 0$ for all but finitely many $n$. This implies that $\text{Cl}^{\text{alg}}(\overline{Y})[n] = 0$ for all but finitely many $n$, which shows that $\text{Cl}^{\text{alg}}(\overline{Y}) = 0$, since the underlying abelian variety must have dimension 0. It follows that $\text{Cl}(\overline{Y}) = \text{NS}(\overline{Y})$ is finitely generated and thus that $\text{Pic}(Y)$ is finitely generated.

4 Special case II: Extensions of stratified bundles of rank 1

We continue to write $k$ for an algebraically closed field of characteristic $p > 0$.

**Lemma 4.1.** Let $X$ be a smooth, connected, separated $k$-scheme of finite type which admits a good compactification, and $\bar{x}$ a geometric point. If $\pi_1^{\text{amc}}(X, \bar{x})$ does not have a quotient isomorphic to $\mathbb{Z}/p\mathbb{Z}$, then $\text{Ext}_{\text{Strat}^{s}(X)}^{1}(\mathcal{O}_X, \mathcal{O}_X) = 0$.

**Proof.** Let $\overline{X}$ be a good compactification of $X$ and $E$ a regular singular stratified bundle on $X$, which is an extension of $\mathcal{O}_X$ by $\mathcal{O}_X$ in $\text{Strat}^{\omega}(X)$. The aim is to show that $E$ extends to a stratified bundle $\overline{E} \in \text{Strat}(\overline{X})$; the assumption
on $\pi_1^{\text{tame}}(X, \bar{x})$ implies that $\pi_1(\overline{X}, \bar{x})$ does not have a quotient isomorphic to $\mathbb{Z}/p\mathbb{Z}$, so we can then apply [EM10, Prop. 2.4] to show that $\overline{E}$ is trivial.

To show that $E$ extends to a stratified bundle $\overline{E}$ on $\overline{X}$, by Proposition 2.12 it suffices to show that the exponents of $E$ along every component of the boundary divisor $\overline{X} \setminus X$ are $0$ mod $\mathbb{Z}$. Let $x_0 \in \overline{X}$ be a closed point lying on precisely one component of $\overline{X} \setminus X$. Write $K_{x_0} := \text{Frac}(\mathcal{O}_{\overline{X}, x_0})$. Then, after choosing coordinates $x_1, \ldots, x_n$, $K_{x_0}$ is isomorphic to the fraction field of the ring of formal power series $k[[x_1, \ldots, x_n]]$. Write $\overline{E} := E \otimes K_{x_0}$. The stratification on $E$ gives $\overline{E}$ the structure of a finite dimensional $K_{x_0}$-vector space with an action of the ring of differential operators $k[[\omega^{(m)} \partial_{x_i} \mid i = 1, \ldots, n, m \geq 0]]$, where the usual composition rules hold. Such an object is called an iterative differential module in [MvdP03]. The category $\text{Strat}(K_{x_0})$ of such objects is “almost” a neutral Tannakian category, but there might not exist a $k$-valued fiber functor (for more details see [Kin12b, Ch. 1]). Fortunately, the sub-tensor category $\{\overline{E}\} \subset \text{Strat}(K_{x_0})$ spanned by $\overline{E}$ admits a fiber functor $\omega$ by [Del90, Cor. 6.20], as $k$ is algebraically closed. Composition of $\omega$ with the restriction functor $\langle E \rangle_{\omega} \rightarrow \{\overline{E}\}_{\omega}$ is a fiber functor for $(E)_{\omega}$, hence we get a morphism $G((\overline{E})_{\omega}) \rightarrow G((E)_{\omega})$ of the associated affine $k$-group schemes, and this morphism is a closed immersion by [DM82, Prop. 2.21].

Since $E$ is an extension of $\mathcal{O}_X$ by $\mathcal{O}_X$, its monodromy group $G((E)_{\omega})$ is a closed subgroup scheme of $G_{a,k}$. But $E$ is assumed to be regular singular, so [Gie75, Thm. 3.3] implies that $\overline{E}$ is a direct sum of rank 1 objects of $\text{Strat}(K_{x_0})$ and thus $G((\overline{E})_{\omega})$ is a closed subgroup scheme of $G_{n,k}^2$. We finally conclude that $\overline{E} \cong K_{x_0}$ as an object of $\text{Strat}(K_{x_0})$, since $G_{a,k}$ does not have nontrivial diagonalizable subgroups, and then [Gie75, Thm. 3.3] implies that the exponents of $E$ along the component on which $x_0$ lies are $0$ mod $\mathbb{Z}$. We repeat the same argument for every component of the boundary $\overline{X} \setminus X$, and hence complete the proof.

We can now prove one of the main results of this article:

**Theorem 4.2.** Let $X$ be a smooth, connected, separated $k$-scheme of finite type which admits a good compactification, and $\bar{x}$ a geometric point. If $\pi_1^{\text{tame}}(X, \bar{x})^{\text{ab.}(p')} = 0$, and if $\pi_1^{\text{tame}}(X, \bar{x})$ does not have a quotient isomorphic to $\mathbb{Z}/p\mathbb{Z}$, then every regular singular stratified bundle on $X$ which has abelian monodromy is trivial.

**Proof.** Let $E$ be a regular singular stratified bundle on $X$, and $\omega : \langle E \rangle_{\omega} \rightarrow \text{Vect}_{k}$ a fiber functor. Assume that the associated $k$-group scheme $G((E)_{\omega}, \omega)$ is abelian. By [Wat79, 9.4] every irreducible object of $(E)_{\omega}$ has rank 1. By Proposition 3.16 every rank 1 object of $\text{Strat}(X)$ is trivial, and by Lemma 4.1 there are no nontrivial extensions between trivial objects of rank 1. It follows that every object of $(E)_{\omega}$ is trivial.
5 Special case III: Affine spaces

We continue to denote by $k$ an algebraically closed field of characteristic $p > 0$. It is well known that $\pi_1^{\text{an}}(A^n_k) = 0$ for all $n \geq 0$.

**Theorem 5.1.** Every regular singular stratified bundle on $A^n_k$ is trivial.

**Remark 5.2.**
- A slightly different approach to Theorem 5.1 was sketched in [Esn12, 4.4]
- Note that Theorem 5.1 follows directly from Proposition 3.16 and [Gie75, Thm. 5.3], which states that every regular singular stratified bundle on $\mathbb{P}^n_k \setminus D$ is a direct sum of stratified line bundles, if $D$ is a strict normal crossings divisor. Unfortunately [loc. cit.] is imprecisely stated (it is false for $n = 1$) and its proof is very complicated.

Below we give a simple argument to prove Theorem 5.1, which is certainly implicitly contained in [Gie75].

**Proof of Theorem 5.1.** The case of $A^1_k$ follows from [Gie75, Prop. 4.2] and Proposition 3.16. We proceed by induction; let $n > 1$. Then the $n$-fold product $\mathbb{P}^1_k \times \cdots \times_k \mathbb{P}^1_k$ is a good compactification of $A^n_k$, and if $E$ is a stratified bundle on $A^n_k$, then $E$ is regular singular if and only if it is $(A^n_k, (\mathbb{P}^1_k)^n)$-regular singular by Proposition 2.7.

We compute the exponents of $E$ along the divisor $(\mathbb{P}^1_k)^{n-1} \times_k \{\infty\} \subset (\mathbb{P}^1_k)^n$. To this end let $E$ be a free $\mathcal{O}_{A^n_k \times (\mathbb{P}^1_k \setminus \{0\})} \otimes \mathbb{Q}$-module with $\mathcal{O}_{A^n_k \times (\mathbb{P}^1_k \setminus \{0\})} \log A^n_k \times \{\infty\}$-action extending $E|_{A^n_k \times \mathbb{G}_m}$. Choose coordinates $x_1, \ldots, x_n$ such that $A^n_k \times_k (\mathbb{P}^1_k \setminus \{0\}) = \text{Spec} k[x_1, \ldots, x_{n-1}, x_n^{-1}]$. By Proposition 2.10 there exists a basis $e_1, \ldots, e_r$ of the free module $E$, such that

$$\delta_{x_n^{-1}} (e_i) = \left( \frac{\alpha_i}{m} \right) e_i + x_n^{-1} E$$

with $\alpha_i \in \mathbb{Z}_p$ an exponent of $E$ along $A^n_k \times \{\infty\}$. But the same equation also holds modulo a prime ideal $(x_1 - a_1, \ldots, x_{n-1} - a_{n-1})$, $a_1, \ldots, a_{n-1} \in k$, so $\alpha_i \mod \mathbb{Z}$ is an exponent of the fiber $E|_{(a_1, \ldots, a_{n-1}) \times A^n_k}$ along the divisor $(a_1, \ldots, a_{n-1}, \infty) \subset (a_1, \ldots, a_{n-1}) \times \mathbb{P}^1_k$.

The case $n = 1$ now shows that $\alpha_i \equiv 0 \mod \mathbb{Z}$, and hence the exponents of $E$ along $A^1_k \times_k \{\infty\}$ are 0 mod $\mathbb{Z}$. By Proposition 2.12 this means that $E$ extends to an actual stratified bundle on $A^1_k \times_k \mathbb{P}^1_k$.

But now we are done: The above argument shows that $E$ extends to a stratified bundle on $(\mathbb{P}^1_k)^n$: First to $(\mathbb{P}^1_k)^n$ minus a closed subset of codimension $\geq 2$, and then by [Kin12a, Lemma 2.5] to $(\mathbb{P}^1_k)^n$. But there are only trivial stratified bundles on $(\mathbb{P}^1_k)^n$, as it is birational to $\mathbb{P}^{n-1}_k$, so $E$ is trivial ([Gie75, Thm. 2.2]).
6  SPECIAL CASE IV: UNIVERSAL HOMEOMORPHISMS

We continue to denote by $k$ an algebraically closed field of characteristic $p > 0$. Recall that by [EGA4, 18.12.11], a finite type morphism $f : Y \to X$ of finite type $k$-schemes is a universal homeomorphism if and only if it is finite, purely inseparable (i.e. universally injective) and surjective. It is proven in [SGA1, IX.4.10] that $f$ induces an isomorphism $\pi_1^t(Y) \cong \pi_1^t(X)$ (with appropriate choices of base points); it follows from [Vid01] that the same is true for $\pi_1^{\text{tame}}$. In this section we prove that pull-back along $f$ is an equivalence $\text{Strat}^{rs}(X) \to \text{Strat}^{rs}(Y)$.

For a $k$-scheme $X$, we write $X^{(n)}$ for the base change of $X$ along the $n$-th power of the absolute Frobenius of $k$, and by $F_{X/k}^{(n)} : X \to X^{(n)}$ the associated $k$-linear relative Frobenius. It follows from Theorem 2.3 that pull-back along $F_{X/k}^{(n)}$ induces an equivalence of categories $\text{Strat}(X^{(n)}) \to \text{Strat}(X)$. This remains true in the regular singular case: We first work with respect to one fixed good partial compactification $(X, \overline{X})$.

Proposition 6.1 (“Frobenius descent”). If $(X, \overline{X})$ is a good partial compactification, then the pair $(X^{(n)}, \overline{X}^{(n)})$ also is a good partial compactification, and $F_{X/k}^{(n)}$ induces an equivalence

$$(F_{X/k}^{(n)})^* : \text{Strat}^{rs}((X^{(n)}, \overline{X}^{(n)})) \to \text{Strat}^{rs}((X, \overline{X})).$$

Proof. We may assume that $n = 1$, and write $F_{X/k} = F_{X/k}^{(1)}$.

Write $D := \overline{X} \setminus X$ and $D^{(1)} := \overline{X}^{(1)} \setminus X^{(1)}$. Since the functor $F_{X/k}^*: \text{Strat}(X^{(1)}) \to \text{Strat}(X)$ is an equivalence, it suffices to show that the essential image of $\text{Strat}^{rs}((X^{(1)}, \overline{X}^{(1)}))$ in $\text{Strat}(X)$ is $\text{Strat}^{rs}((X, \overline{X}))$, i.e. it suffices to show that a stratified bundle $E$ on $X^{(1)}$ is $(X^{(1)}, \overline{X}^{(1)})$-regular singular if $F_{X/k}^* E$ is $(X, \overline{X})$-regular singular.

Let $j : X \to \overline{X}$ denote the inclusion. Assume that $F_{X/k}^* E$ is $(X, \overline{X})$-regular singular. Let $E'$ be any torsion free coherent extension of $E$ to $\overline{X}^{(1)}$ and $E$ the $\mathscr{D}_{X^{(1)}/k}(\log D^{(1)})$-module generated by $E'$ in the $\mathscr{D}_{\overline{X}^{(1)}/k}(\log D^{(1)})$-module $j_*^{(1)} E'$. We need to show that $E$ is $\mathcal{O}_{\overline{X}^{(1)}}$-coherent. Since $F_{X/k}^*$ is faithfully flat, it suffices to show that $F_{X/k}^* E$ is $\mathcal{O}_{\overline{X}}$-coherent, [SGA1, Prop. VIII.1.10]. Define $G$ to be the $\mathscr{D}_{\overline{X}/k}(\log D)$-module generated by $F_{X/k}^* E'$ in $j_* F_{X/k}^* E = F_{X/k}^* j_*^{(1)} E$. Then $G$ is $\mathcal{O}_{\overline{X}}$-coherent by assumption, so the proof is complete if we can show that $G = F_{X/k}^* E$.

The $\mathcal{O}_{\overline{X}^{(1)}}$-module $j_*^{(1)} E$ naturally carries a $\mathscr{D}_{\overline{X}^{(1)}/k}(\log D^{(1)})$-action, and similarly for $j_* F_{X/k}^* E$. Then we can describe $F_{X/k}^* E$ as the image of the (pulled-
Evidence for a Generalization of Gieseker’s Conjecture

With the notation \( \delta \) and \( t \)

This is a local question, so we may assume that

where

These morphisms fit in a commutative diagram (writing \( F = F_{X/k} \) for legibility):

\[
\begin{array}{ccc}
F^* \left( \mathcal{D}_{X/k} (\log D) \right) & \longrightarrow & F^* j^*(1) E \\
\downarrow & & \downarrow \\
F^* \mathcal{D}_{X/k} (\log D) & \longrightarrow & j_* F^* E
\end{array}
\]

where

\[ \gamma : \mathcal{D}_{X/k} (\log D) \to F^* \mathcal{D}_{X/k} (\log D) = O_X \otimes_{F-1 \mathcal{O}_{X/k}} F^{-1} \mathcal{D}_{X/k} (\log D) \]

is the canonical morphism coming via restriction from the morphism \( \mathcal{D}_{X/k} \to F^* \mathcal{D}_{X/k} \). It follows that \( G = F^* \mathcal{E}_k \), if \( \gamma \) is surjective.

This is a local question, so we may assume that \( X = \text{Spec} A \) and \( X = \text{Spec} A[t_1^{-1}] \), with \( t_1, \ldots, t_r \) local coordinates on \( A \). Then \( \mathcal{X}^{(1)} = \text{Spec} A \otimes_{F_k} k \), and \( t_1 \otimes 1, \ldots, t_r \otimes 1 \) is a system of local coordinates for \( A \otimes_{F_k} k \). The relative frobenius \( F_{X/k} \) then maps \( t_i \otimes 1 \to t_i^p \), for \( i = 1, \ldots, r \).

With the notation \( \delta_{t_i}^{(p^m)} \) from Definition 2.5, (b), and recalling that \( \delta_{t_i}^{(p^m)} \) “behaves like” \( \frac{\partial^m}{\partial^m} \), we claim that \( \gamma(\delta_{t_i}^{(p^m)}) = 0 \) and

\[ \gamma(\delta_{t_i}^{(p^m)}) = \delta_{t_i \otimes 1}^{(p^{m-1})} \text{ for } m > 1, \]

which shows that the image of \( \gamma \) contains all the generators of the left-\( \mathcal{O}_X \)

algebra \( F^* \mathcal{D}_{X/k} (\log D) \), and thus that \( \gamma \) is surjective. As for the claim, it suffices to observe that for \( s \geq 0 \),

\[
\delta_{t_i}^{(p^m)}((t_j^p)^s) = \begin{cases} 0 & i \neq j \\ (p^m) & \text{otherwise}, \end{cases}
\]
and finally that
\[
\left( \frac{sp}{p^m} \right) \equiv \left( \frac{s}{p^{m-1}} \right) \mod p.
\]  

\[\square\]

**Remark 6.2.** This proof does not show that all \(O_X\)-coherent \(\mathcal{D}_{X/k}(\log D)\)-module descend to \(O_{X^{(1)}}\)-coherent \(\mathcal{D}_{X^{(1)}/k}(\log D^{(1)})\)-modules. In fact, such a statement is false, due to the failure of Cartier’s theorem [Kat70, §5.] for logarithmic connections. On the other hand, in [Lor00] a version of Cartier’s theorem for log-schemes is developed, which is applied in [Mon02, Ch. 4] to construct a generalization of Frobenius descent to the logarithmic setting. For this however, the rings of coefficients have to be enlarged.

**Corollary 6.3.** With the notations from Proposition 6.1, if \(\overline{E}\) is a locally free \(\mathcal{O}_{X^{(1)}}\)-coherent \(\mathcal{D}_{X^{(1)}/k}(\log D^{(1)})\)-module with exponents \(\alpha_1, \ldots, \alpha_n \in \mathbb{Z}_p\), then \(F^*_{X/k}\overline{E}\) is a \(\mathcal{D}_{X/k}(\log D)\)-module with exponents \(p\alpha_1, \ldots, p\alpha_n\).

Consequently, if \(E \in \text{Strat}^s((X^{(1)}, \overline{X}^{(1)}))\), with exponents \(\alpha_1, \ldots, \alpha_n \in \mathbb{Z}_p/\mathbb{Z}\), then \(F^*_{X/k}E\) has exponents \(p\alpha_1, \ldots, p\alpha_n \in \mathbb{Z}_p/\mathbb{Z}\).

**Proof.** The claim follows directly from the formula (6) and the fact that \(\delta_1^{(1)}\) acts on \(F^*_{X/k}\overline{E} = O_X \otimes_{\mathcal{O}_{X^{(1)}}} \overline{E}\) via \(\delta_1^{(1)}(a \otimes e) = \delta_1^{(1)}(a) \otimes e\).  

\[\square\]

**Remark 6.4.** Note that multiplication by \(p\) is an automorphism of the group \(\mathbb{Z}_p/\mathbb{Z}\).

**Theorem 6.5.** Let \((X, \overline{X}), (Y, \overline{Y})\) be good partial compactifications and \(\overline{f} : \overline{Y} \to \overline{X}\) a universal homeomorphism such that \(\overline{f}(X) \subset \overline{Y}\). If we write \(f := \overline{f}|_X\) then \(f\) induces an equivalence
\[
f^* : \text{Strat}^s((X, \overline{X})) \xrightarrow{\sim} \text{Strat}^s((Y, \overline{Y})).
\]

**Proof.** The morphism \(\overline{f}\) is finite of degree \(p^n\) for some \(n\), and thus there is a morphism \(\overline{g} : \overline{X} \to \overline{Y}^{(n)}\), such that \(\overline{g} \overline{f} = F^*_{X/k}\overline{g}\), and such that \(f^{(n)} \overline{g} = F^*_{X^{(1)/k}}\overline{g}\).

Moreover, \(\overline{g}(X) \subset Y^{(n)}\). Write \(g := \overline{g}|_X\). Then \((gf)^*\) is an equivalence by Proposition 6.1, so \(f^*\) is essentially surjective. But \((f^{(n)}g)^*\) also is an equivalence, so \(f^{(n)*}\) is full. This shows that \(f^*\) is full, and since \(f\) is faithfully flat, it follows that \(f\) is faithful as well. This finishes the proof. 

\[\square\]

**Theorem 6.6.** Let \(f : Y \to X\) be a universal homeomorphism of smooth, separated, finite type \(k\)-schemes. Then \(f\) induces an equivalence
\[
f^* : \text{Strat}^s(X) \xrightarrow{\sim} \text{Strat}^s(Y).
\]
Proof. Without loss of generality we may assume that $X,Y$ are connected. The same argument as in the proof of Theorem 6.5 shows that the fact that the relative Frobenius induces an equivalence $\text{Strat}(X^{(n)}) \to \text{Strat}(X)$, implies that the functor $f^* : \text{Strat}(X) \to \text{Strat}(Y)$ is an equivalence.

If follows that we just need to check that $f^*E$ is regular singular for $E \in \text{Strat}(X)$, if and only if $E$ is regular singular.

Assume that $E$ is regular singular and let $(Y,\mathcal{Y})$ be a good partial compactification with $\mathcal{Y} \setminus Y$ smooth with generic point $\eta$. Using that $k(X) \subset k(Y)$ is purely inseparable, one quickly checks that $R := \mathcal{O}_{\mathcal{Y},\eta} \cap k(X)$ is a discrete valuation ring, that $\mathcal{O}_{Y,\eta}$ is the integral closure of $R$ in $k(Y)$, and hence that the residue field of $R$ has transcendence degree $\dim X - 1$ over $k$. This means that $R$ is the local ring of a codimension 1 point on some model of $k(X)$. Hence there exists a good partial compactification $(X,\mathcal{X})$, such that $f$ extends to a morphism $\tilde{f} : \mathcal{Y} \to \mathcal{X}$ (after possibly removing a closed subset of codimension $\geq 2$ from $\mathcal{Y}$). This shows that $f^*E$ is $(Y,\mathcal{Y})$-regular singular. We repeat this for every good partial compactification $(Y,\mathcal{Y})$ to conclude that $f^*E$ is regular singular.

Conversely, assume that $f^*E$ is regular singular. To prove that $E$ is regular singular we need to show that for every good partial compactification $(X,\mathcal{X})$, such that $\mathcal{X} \setminus X$ is smooth with generic point $\xi$, there exists an open neighborhood $\mathcal{U}'$ of $\xi$, such that $E$ is $(\mathcal{U}' \cap X,\mathcal{U}')$-regular singular.

Let $\mathcal{U} := \text{Spec} \mathcal{A}$ be an affine open neighborhood of $\xi$. We may assume that $U := \mathcal{U} \cap X$ is also affine, say $U = \text{Spec} A$. Because $f$ is finite, $V := f^{-1}(U)$ is affine, say $V = \text{Spec} B$, and $f|_V$ is a universal homeomorphism. Let $\mathcal{B}$ be the integral closure of $A$ in $B$, and $\mathcal{V} := \text{Spec} \mathcal{B}$. Then $\tilde{g} : \mathcal{V} \to \mathcal{U}$ is a finite morphism. By construction $V$ is normal, and $\tilde{g}$ is a universal homeomorphism, because $k(X) \subset k(Y)$ is purely inseparable. We may shrink $\mathcal{U}$ around $\xi$ to obtain an open neighborhood $\mathcal{U}' \subset \mathcal{U}$ of $\xi$, such that $\mathcal{V} := \tilde{g}^{-1}(\mathcal{U}')$ is smooth, and $\tilde{g}' : \mathcal{V} \to \mathcal{U}'$ is a universal homeomorphism. Moreover, writing $V' := \mathcal{V} \cap Y$, $U' := \mathcal{U}' \cap X$, we see that $\tilde{g}'|_{V'} = f|_{V'} : V' \to U'$. Since $(f|_{V'})^*(E|_{V'})$ is $(V',\mathcal{V}')$-regular singular by assumption, we can apply Theorem 6.5 to see that $E$ is $(U',\mathcal{U}')$-regular singular.

References


Lars Kindler


Evidence for a Generalization of Gieseker’s Conjecture


THE ZETA FUNCTION OF A FINITE CATEGORY

KAZUNORI NOGUCHI

Received: June 6, 2012
Revised: July 8, 2013

Communicated by Don Blasius and Stephen Lichtenbaum

Abstract. We define the zeta function of a finite category. We prove a theorem that states a relationship between the zeta function of a finite category and the Euler characteristic of finite categories, called the series Euler characteristic. Moreover, it is shown that for a covering of finite categories, $P : E \to B$, the zeta function of $E$ is that of $B$ to the power of the number of sheets in the covering. This is a categorical analogue of the unproved conjecture of Dedekind for algebraic number fields and the Dedekind zeta functions.

2010 Mathematics Subject Classification: Primary 18G30; Secondary 18D30, 30B10, 30B40.
Keywords and Phrases: zeta function of a finite category, Euler characteristic of categories, coverings of small categories, Dedekind conjecture

Contents

1 Introduction 1244

2 Definition and examples 1248
   2.1 Definition .................................................. 1248
   2.2 Groupoids ................................................... 1249
   2.3 Acyclic categories ........................................... 1252
   2.4 Finite categories having one or two objects ................. 1253

Documenta Mathematica 18 (2013) 1243-1274
1 Introduction

Euler characteristics and zeta functions are defined for various mathematical objects; for example, simplicial complexes, algebraic varieties, and graphs. In many cases, we can observe that the zeta function knows the Euler characteristic, as the following three examples suggest.

1. Let $G$ be a finite connected graph. Then, the Ihara zeta function of $G$ is defined by

$$Z_G(u) = \prod_{[C]} \frac{1}{1 - u^\ell[C]}$$

where $[C]$ is an equivalence class of certain paths in $G$ and $\ell$ is the length function. The zeta function $Z_G$ has the determinant expression

$$Z_G(u) = \frac{(1 - u^2)^{1-r}}{|I - A_G u + Q_G u^2|}$$

for some matrices $A_G$ and $Q_G$ where $I$ is the unit matrix and $r$ is the rank of the fundamental group of $G$ (Theorem 2 of [ST96]). It is clear that $1 - r$ is the Euler characteristic $\chi(G)$ of $G$.

2. Let $\Delta$ be a simplicial complex on a vertex set

$$\{1, 2, \ldots, N\}$$

and let $\mathbb{F}_q$ be a finite field. Björner and Sarkaria defined the zeta function of $\Delta$ over $\mathbb{F}_q$ by

$$Z_\Delta(q, t) = \exp \left( \sum_{m=1}^{\infty} \frac{\# V(\Delta, \mathbb{F}_{q^m})}{m} t^m \right)$$

where $V(\Delta, \mathbb{F}_{q^m})$ is the set of points in the projective space $\mathbb{F}_{q^m} P^{N-1}$ whose support belongs to $\Delta$ [BS98]. By Theorem 2.2 of [BS98], the zeta function has a rational expression; that is,

$$Z_\Delta(q, t) = \prod_{k=0}^{d} \frac{1}{(1 - q^k t)^{\ell_k}}$$
for some integers \( f_k^* \) where \( d \) is the dimension of \( \Delta \). Here, we obtain
\[
\sum_{k=0}^{d} f_k^* = \chi(V(\Delta, C)) \quad \text{(Corollary 2.4 of [BS08]).}
\]

3. Let \( X \) be an \( n \)-dimensional smooth projective variety over a finite field \( \mathbb{F}_q \). Then, the zeta function of \( X \) is defined by
\[
Z_X(T) = \exp\left( \sum_{m=1}^{\infty} \frac{N_m(X)}{m} T^m \right)
\]
where \( N_m(X) \) is the number of points in \( X \) over \( \mathbb{F}_{q^m} \). One of the Weil conjectures states that \( Z_X \) has a rational expression of the following form:
\[
Z_X(T) = \frac{P(T)}{Q(T)}
\]
for some polynomials \( P(T) \) and \( Q(T) \) with coefficients in \( \mathbb{Z} \), and we obtain
\[
\chi(X) = \deg Q - \deg P \quad \text{(e.g., see [Har77]).}
\]

These examples tell us that the zeta function knows the Euler characteristic.

In this paper, we define the zeta function of a finite category and we prove a theorem that states a relationship between the zeta function of a finite category and the Euler characteristic of a finite category, called the series Euler characteristic [BL08].

Let \( C \) be a finite category. A finite category is a category having finitely many objects and morphisms. Then, the zeta function of \( C \) is defined by
\[
\zeta_C(z) = \exp\left( \sum_{m=1}^{\infty} \frac{\#N_m(C)}{m} z^m \right)
\]
where
\[
N_m(C) = \{ (x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} \cdots \xrightarrow{f_m} x_m) \text{ in } C \}.
\]
The zeta function of a finite category introduced in this paper is different from the one introduced by Kurokawa [Kur96]. His zeta function is for a large category; for example, the category of Abelian groups.

Next, let us recall the Euler characteristics of categories. The Euler characteristic of a finite category was defined by Leinster [Lei08]. This was the first Euler characteristic for categories. Subsequently, there have emerged the series Euler characteristic by Berger-Leinster [BL08] and the \( L^2 \)-Euler characteristic by Fiore-Lück-Sauer [FLS11] as well as the extended \( L^2 \)-Euler characteristic [Nog] and the Euler characteristic of \( N \)-filtered acyclic categories [Nog11] by the author. In this paper, we often use the series Euler characteristic, so we provide a more detailed explanation for the series Euler characteristic.

For a finite category \( C \) whose set of objects is \( \{x_1, \ldots, x_N\} \), its series Euler characteristic \( \chi_S(C) \) is defined by substituting \( t = -1 \) in
\[
\frac{\sum (\text{adj}(I - (AC - I)t))}{|I - (AC - I)t|}
\]

\[\text{Documenta Mathematica 18 (2013) 1243–1274}\]
if it exists, where $A_C = (\#\text{Hom}(x_i, x_j))$ is called the adjacency matrix of $C$ and sum means to take the sum of all the entries of a matrix. This rational function is a rational expression of the power series $\sum_{m=0}^\infty \# \overline{N}_m(C)t^m$ where $\overline{N}_m(C)$ is the set of nondegenerate chains of morphisms of length $m$ in $C$

$$\overline{N}_m(C) = \{ (x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} \cdots \xrightarrow{f_m} x_m) \mid C \mid f_i \neq 1 \}.$$ 

This Euler characteristic is defined from the viewpoint of the classifying spaces. For a small category $C$, we can construct the topological space (in fact, a CW-complex) $BC$, called the classifying space of $C$. There is a one-to-one correspondence between the set of $m$-dimensional parts ($m$-cells) of $BC$ and $\overline{N}_m(C)$ [Qui73]. The Euler characteristic of a cell-complex is defined by the alternating sum of the number of $m$-cells. Hence, the Euler characteristic of $C$ should be defined by $\sum_{m=0}^\infty \# \overline{N}_m(C)t^m$. For more details, see [BL08].

The following is our main theorem.

**Main Theorem** (Theorem 3.5): Suppose that $C$ is a finite category with Euler characteristic $\chi_\Sigma(C)$ and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the nonzero eigenvalues of $A_C$ whose algebraic multiplicities are $e_1, e_2, \ldots, e_n$. Then,

1. the zeta function of $C$ is

$$\zeta_C(z) = \prod_{k=1}^n \frac{1}{(1 - \lambda_kz)^{e_k}} \exp \left( \sum_{j=1}^{e_k-1} \frac{\beta_{k,j}z^j}{j(1 - \lambda_kz)^j} \right)$$

for some complex numbers $\beta_{k,j}$,

2. the sum of all the indexes $\beta_{k,0}$ is the number of objects of $C$,

3. each $\lambda_k$ is an algebraic integer, and

4. $\sum_{k=1}^n \sum_{j=0}^{e_k-1} (-1)^j \frac{\beta_{k,j}}{\lambda_k^{j+1}} = \chi_\Sigma(C) \in \mathbb{Q}$.

Part 3 is an analogue of the Weil conjecture and, in fact, it does not need the condition that $C$ has Euler characteristic (see Theorem 3.3). Part 4 implies that, although each $\lambda_k$ and $\beta_{k,j}$ is a complex number, this alternating sum is always rational. In this paper, we define $\log z$ and the power functions by the principal value; that is,

$$\log z = \log |z| + i\arg(z) \ (z \in \mathbb{C} - \{ x \in \mathbb{R} \mid x \leq 0 \}, -\pi < \arg(z) < \pi)$$

and

$$z^\alpha = e^{\alpha \log z} \ (z, \alpha \in \mathbb{C}, z \neq 0).$$

If we do not assume the condition that $C$ has Euler characteristic, Part 1 is given by the following theorem.
Theorem 1.1 (Theorem 3.3). Let $C$ be a finite category. Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the nonzero eigenvalues of $A_C$ and their algebraic multiplicities are $e_1, e_2, \ldots, e_n$. Then, the zeta function of $C$ is

$$\zeta_C(z) = \prod_{k=1}^{n} \frac{1}{(1 - \lambda_k z)^{e_k}} \exp \left( Q(z) + \sum_{j=1}^{e_k-1} \frac{\beta_k, j z^j}{j!(1 - \lambda_k z)^j} \right)$$

for some complex numbers $\beta_{k,j}$ and a polynomial $Q(z)$ with $\mathbb{Q}$-coefficients whose constant term is zero.

If we do not assume the condition that $C$ has Euler characteristic, Part 2 fails (see Example 3.7).

Our zeta function is related with coverings of small categories. We show that for a covering of finite categories, $P : E \rightarrow B$, the zeta function of $E$ is that of $B$ to the power of the number of sheets in the covering. This is an analogue of the unproved conjecture of Dedekind. The conjecture is that for a finite extension $K_2$ of an algebraic number field $K_1$ the Dedekind zeta function $\zeta_{K_1}(s)$ of $K_1$ divides that of $K_2$ [Wan75]. An algebraic number field is a finite extension of $\mathbb{Q}$.

A covering of small categories is an analogy of Galois theory. A fundamental theorem of Galois theory is that if $K/F$ is a finite Galois extension, the set of intermediate fields of $K$ and $F$ is bijective to the set of subgroups of the Galois group $\text{Gal}(K/F)$:

$$K \leftarrow \left\{ e \right\} \leftarrow \left\{ e \right\}$$

$$L \rightarrow \left\{ e \right\} \leftarrow \left\{ e \right\}$$

$$F \leftarrow \text{Gal}(K/F).$$

For a covering of small categories $\tilde{P} : \tilde{E} \rightarrow B$, where $\tilde{E}$ is the universal covering of $B$, the set of the isomorphism classes of intermediate coverings of $\tilde{P}$ is bijective to the set of subgroups of the fundamental group $\pi_1(B)$:

$$\tilde{P} \leftarrow \left\{ e \right\} \leftarrow \left\{ e \right\}$$

$$\tilde{P} \leftarrow \left\{ e \right\} \leftarrow \left\{ e \right\}$$

$$B \leftarrow \pi_1(B).$$
K. Noguchi

(see Corollary 2.24 of [Tan]). We have the following correspondences:

- coverings ↔ extensions of fields
- $\pi_1$ ↔ Galois groups
- intermediate coverings ↔ intermediate fields

For an analogy between coverings of spaces and extensions of fields, see [Mor12]. By the diagrams above, we can conclude that the relationship between our zeta functions and coverings is an analogue of the Dedekind conjecture. Graph theoretic analogue of this conjecture was considered in Corollary 1 of §2 of [ST96] ([ST00] and [ST07] are its continuation).

This remainder of this paper is organized as follows: In Section 2, the zeta function of a finite category is defined, and we compute the zeta functions of finite groupoids and finite acyclic categories. We classify the zeta functions of one-object finite categories and two-objects finite categories. In Section 3, we prove our main theorem, and we introduce four zeta functions of finite categories having three-objects. In Section 4, we prove that for a covering of finite categories, $P: E \to B$, the zeta function of $E$ is that of $B$ to the power of the number of sheets in the covering.

Acknowledgment. I wish to thank the referee to give me many useful suggestions.

2 Definition and examples

In this section, we define the zeta function of a finite category, and we compute zeta functions.

2.1 Definition

Before defining the zeta function of a finite category, we review the symbols that are often used in this paper.

Let $C$ be a finite category. Then, let

$$N_m(C) = \{ (x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} \cdots \xrightarrow{f_m} x_m) \text{ in } C \}$$

and

$$\overline{N}_m(C) = \{ (x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} \cdots \xrightarrow{f_m} x_m) \text{ in } C \mid f_i \neq 1 \}.$$ 

The difference between these is merely whether identity morphisms are used or not. For $m = 0$, we set $N_0(C) = \overline{N}_0(C) = \text{Ob}(C)$. In this paper we have the important equality

$$\#N_m(C) = \text{sum}(A^m_m).$$
Indeed, if $\text{Ob}(C) = \{x_1, x_2, \ldots, x_N\}$ and $A_C = (a_{ij})$, then the $(i, j)$-entry of $A_C^m$ is
\[
\sum_{1 \leq k_1, k_2, \ldots, k_{m-1} \leq N} a_{ik_1} a_{k_1k_2} a_{k_2k_3} \cdots a_{k_{m-1}j}.
\]
This is the number of chains of morphisms of length $m$ from $x_i$ to $x_j$. Hence, we obtain the equality.

**Definition 2.1.** Let $C$ be a finite category. Then, we define the zeta function $\zeta_C(z)$ of $C$ by
\[
\zeta_C(z) = \exp \left( \sum_{m=1}^{\infty} \frac{\# N_m(C)}{m} z^m \right).
\]
This function belongs to the power series ring $\mathbb{Q}[[z]]$. If preferable, the zeta function can be considered a function of a complex variable by choosing $z$ to be a sufficiently small complex number. Indeed, for a complex number $z$ such that $|z| < \frac{1}{\sum (A_C)}$, the series absolutely converges; that is,
\[
\sum_{m=1}^{\infty} \frac{\# N_m(C)}{m} |z|^m = \sum_{m=1}^{\infty} \frac{\sum (A_C^m)}{m} |z|^m 
\leq \sum_{m=1}^{\infty} \frac{\{\sum (A_C)\}^m}{m} |z|^m < +\infty.
\]

**Example 2.2.** This is the simplest example. Let $\ast$ denote the terminal category. Then, its zeta function is
\[
\zeta_{\ast}(z) = \exp \left( \sum_{m=1}^{\infty} \frac{\# N_m(\ast)}{m} z^m \right) = \exp \left( \sum_{m=1}^{\infty} \frac{1}{m} z^m \right) = \exp (-\log(1 - z)) = \frac{1}{1 - z}.
\]

### 2.2 Groupoids

In this subsection, we compute the zeta functions of finite groupoids. First, we compute the zeta functions of connected finite groupoids.

A category $C$ is **connected** if $C$ is a nonempty category and there exists a zig-zag sequence of morphisms in $C$:

\[
\begin{array}{ccccccccccccccc}
\ & f_1 & \ & x_1 & & f_2 & & x_2 & & f_3 & & \cdots & & f_n & \ & \ & x & & \ & \ & \ & y
\end{array}
\]
for any objects $x$ and $y$ of $C$. We do not have to consider the direction of the last morphism $f_n$, since we can insert an identity morphism into the sequence. A nonempty groupoid $\Gamma$ is connected if and only if there exists a morphism $f : x \to y$ for any objects $x$ and $y$ of $\Gamma$.

**Proposition 2.3.** Let $\Gamma$ be a connected finite groupoid. Then, its zeta function is

$$
\zeta_\Gamma(z) = \frac{1}{(1 - \#N_0(\Gamma)o(\Gamma))^{\#N_0(\Gamma)}}
$$

where $o(\Gamma)$ is the order of the automorphism group $\text{Aut}(x)$ for some object $x$ of $\Gamma$.

**Proof.** Let

$$\text{Ob}(\Gamma) = \{x_1, x_2, \ldots, x_N\}.$$

We count the chains of morphisms of length $m$ in $\Gamma$. To determine

$$y_0 \xymatrix{ f_1 \ar[r] & y_1 \xymatrix{ f_2 \ar[r] & \cdots \ar[r] & f_m \ar[r] & y_m}$$

we first determine the objects $y_0, y_1, \ldots, y_m$. There are $N^{m+1}$ ways to determine these. There are $o(\Gamma)^m$ ways to determine the morphisms $f_1, f_2, \ldots, f_m$, since we have

$$\#\text{Hom}(x, y) = \#\text{Hom}(x', y') = o(\Gamma)$$

for any objects $x, x', y, y'$ of $\Gamma$. Hence, we obtain $\#N_m(\Gamma) = N^{m+1}o(\Gamma)^m$. Thus, we have

$$
\zeta_\Gamma(z) = \exp \left( \sum_{m=1}^{\infty} \frac{N^{m+1}o(\Gamma)^m}{m} z^m \right)
\quad = \exp \left( N \sum_{m=1}^{\infty} \frac{1}{m} (N o(\Gamma) z)^m \right)
\quad = \exp (-N \log(1 - N o(\Gamma) z))
\quad = \frac{1}{(1 - N o(\Gamma) z)^N}.
$$

**Remark 2.4.** 1. The zeta function of a finite category is not invariant under equivalence of categories. For example, let $\Gamma_N$ be the following groupoid:

$$x_1 \xymatrix{ \ar@{-}[r] & x_2 \xymatrix{ \ar@{-}[r] & \cdots \ar@{-}[r] & x_N}$$

for any natural number $N$. Then, $\Gamma_N$ is equivalent to $\Gamma_M$ for any natural number $M$. Proposition 2.3 implies that their zeta functions are not the same if $N \neq M$. 

---

**Documenta Mathematica 18 (2013) 1243–1274**
2. The zeta function of a finite category depends only on the underlying graph, not on the composition of the finite category, and the zeta function of a finite category is not the same as the zeta function of its underlying graph. For a directed graph $D$, the zeta function $Z_D(u)$ of $D$ is defined by the formal product of certain equivalence classes of paths (see [KS00] and [MS01] for more details). It has a determinant expression of the following form:

$$Z_D(u) = \frac{1}{|I - A_Du|}$$

where $A_D$ is the adjacency matrix of $D$.

For example, the zeta function of $\Gamma_2$ (see above) is

$$\frac{1}{1 - (2z)^2},$$

but the zeta function of its underlying graph is

$$\frac{1}{1 - u^2}.$$

The proposition above can be generalized to the following proposition.

**Proposition 2.5.** Suppose that $C$ is a finite category and its adjacency matrix $A_C = (a_{ij})$ satisfies the condition $\sum_i a_{ij} = \sum_i a_{ij'}$ for any $j$ and $j'$. Then, its zeta function is

$$\zeta_C(z) = \frac{1}{(1 - \sum_i a_{ij}) z^{\# N_0(C)}}.$$

**Proof.** Under the condition that $\sum_i a_{ij} = \sum_i a_{ij'}$ for any $j$ and $j'$, we have

$$\# N_m(C) = \text{sum}(A_C^m) = \# \text{Ob}(C) \left( \sum_i a_{ij} \right)^m.$$

Hence, we obtain the result.

This result is with respect to the columns of $A_C$, but it is clear that there is a similar result with respect to the rows of $A_C$.

**Remark 2.6.** A finite category and its opposite category have the same zeta function. Indeed, we have $A_C = {}^t A_{C^{op}}$, so $\text{sum}(A_C^m) = \text{sum}(A_{C^{op}}^m)$. Hence, their zeta functions are the same.

By the following lemma, computing the zeta function of a finite category is reduced to computing the zeta functions of its connected components.

**Lemma 2.7.** Let $C_1, C_2, \ldots, C_n$ be finite categories. Then, the zeta function of $C = \coprod_{k=1}^n C_k$ is

$$\zeta_C(z) = \prod_{k=1}^n \zeta_{C_k}(z).$$
Proof. Since $N_m(C) = \prod_{k=1}^n N_m(C_k)$, we obtain

$$
\zeta_C(z) = \exp \left( \sum_{m=1}^{\infty} \frac{\#N_m(C)}{m} z^m \right) = \prod_{k=1}^n \exp \left( \sum_{m=1}^{\infty} \frac{\#N_m(C_k)}{m} z^m \right) = \prod_{k=1}^n \zeta_{C_k}(z).
$$

Corollary 2.8. Suppose that $\Gamma$ is a finite groupoid and $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$ are its connected components; that is, $\Gamma = \bigsqcup_{k=1}^n \Gamma_k$ and each $\Gamma_k$ is connected. Then, the zeta function of $\Gamma$ is

$$
\zeta_{\Gamma}(z) = \prod_{k=1}^n \frac{1}{1 - \#N_0(\Gamma_k) \circ o(\Gamma_k) z^{\#N_0(\Gamma_k)}}.
$$

Proof. Lemma 2.7 and Proposition 2.3 directly imply the result.

2.3 Acyclic categories

In this subsection, we compute the zeta functions of finite acyclic categories by using another expression for our zeta function.

Definition 2.9. A small category $A$ is defined to be an acyclic category if all the endomorphisms are only identity morphisms and if there exists a morphism $f: X \to Y$ such that $X \neq Y$, then there does not exist a morphism $g: Y \to X$.

Lemma 2.10. Let $C$ be a finite category. Then, we have

$$
\#N_m(C) = \sum_{j=0}^{m} \binom{m}{j} \#N_j(C)
$$

for any $m \geq 0$.

Proof. Suppose that $0 \leq j \leq m$ and take any $(f_1, f_2, \ldots, f_j)$ of $N_j(C)$. Then, we can make $\binom{m}{j}$-elements of $N_m(C)$ by inserting identity morphisms. Hence, we obtain the result.

Proposition 2.11. Let $C$ be a finite category. Then, we have

$$
\zeta_C(z) = \frac{1}{(1-z)^{\#N_0(C)}} \exp \left( \sum_{j=1}^{\infty} \frac{\#N_j(C) z^j}{j(1-z)^j} \right).
$$
Proof. Lemma 2.10 implies
\[ \zeta_C(z) = \exp \left( \sum_{m=1}^{\infty} \frac{1}{m} \# N_m(C) z^m \right) \]
\[ = \exp \left( \sum_{m=1}^{\infty} \frac{1}{m} \sum_{j=0}^{m} \binom{m}{j} \# N_j(C) z^m \right) \]
\[ = \exp \left( \sum_{j=0}^{\infty} \# N_j(C) \sum_{m=1}^{\infty} \frac{1}{m} \binom{m}{j} z^m \right) \]
\[ = \exp \left( \sum_{m=1}^{\infty} \# N_0(C) \frac{z^m}{m} + \sum_{j=1}^{\infty} \sum_{m=1}^{\infty} \frac{\# N_j(C)}{m} \binom{m}{j} z^m \right) \]
\[ = \frac{1}{(1-z)^{\# N_0(C)}} \exp \left( \sum_{j=1}^{\infty} \frac{\# N_j(C)}{j(1-z)^j} z^j \right). \]
Note that \( \binom{m}{j} = 0 \) if \( m < j \). The last equality is implied by the equality (1.5.5) in [Wil06].

Corollary 2.12. Let \( A \) be a finite acyclic category. Then, the zeta function of \( A \) is
\[ \zeta_A(z) = \frac{1}{(1-z)^{\# N_0(A)}} \exp \left( \sum_{j=1}^{M} \frac{\# N_j(A)}{j(1-z)^j} z^j \right) \]
for a sufficiently large integer \( M \).

Proof. By Lemma 3.5 of [Nog], there exists a sufficiently large integer \( M \) such that \( N_j(A) = \emptyset \) for any \( j > M \). Proposition 2.11 completes this proof.

2.4 Finite categories having one or two objects

In this subsection, we classify the zeta functions of finite categories having one or two objects. In all the zeta functions that we have already seen, only rational numbers appear, but irrational numbers appear in the classification.

First, we compute the zeta functions of one-object finite categories.

Proposition 2.13. Let \( C \) be a one-object finite category. Then, its zeta function is
\[ \zeta_C(z) = \frac{1}{1 - \# N_1(C) z}. \]

Proof. Since \( C \) has only one object, all the morphisms can be composed, so we have
\[ \# N_m(C) = (\# N_1(C))^m. \]
Hence, we obtain the result.
Lemma 2.14. Suppose that $C$ is a finite category having $N$ objects and its adjacency matrix $A_C$ is diagonalizable, with
\[ A_C = P \cdot \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N) \cdot P^{-1}. \]

Then, we have
\[ \zeta_C(z) = \exp \left( \sum \left( P \cdot \text{diag} \left( \frac{1}{1 - \lambda_1 z}, \frac{1}{1 - \lambda_2 z}, \ldots, \frac{1}{1 - \lambda_N z} \right) \cdot P^{-1} \right) \right). \]

Proof. We have
\[
\zeta_C(z) = \exp \left( \sum_{m=1}^{\infty} \frac{A_C^m}{m} z^m \right)
= \exp \left( \sum_{m=1}^{\infty} \frac{1}{m} P \cdot \text{diag} \left( \lambda_1^m, \lambda_2^m, \ldots, \lambda_N^m \right) \cdot P^{-1} z^m \right)
= \exp \left( \sum \left( P \cdot \text{diag} \left( \frac{1}{1 - \lambda_1 z}, \frac{1}{1 - \lambda_2 z}, \ldots, \frac{1}{1 - \lambda_N z} \right) \cdot P^{-1} \right) \right).
\]

\[ \Box \]

Proposition 2.15. Let $C$ be a two-object finite category and let $A_C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.
Then, its zeta function is
\[
\zeta_C(z) = \begin{cases} 
\frac{1}{(1 - az)^2} \exp \left( \frac{bz}{1 - az} \right) & \text{if } a = d, b \neq 0, c = 0 \\
\frac{1}{(1 - az)^2} \exp \left( \frac{cz}{1 - az} \right) & \text{if } a = d, b = 0, c \neq 0 \\
\frac{1}{(1 - \lambda^+ z)^{\beta_+}} \cdot \frac{1}{(1 - \lambda^- z)^{\beta_-}} & \text{otherwise,}
\end{cases}
\]
\[ \lambda^\pm \text{ are the eigenvalues of } A_C \text{ and } \]
\[ \beta_0^\pm = \begin{cases} 
1 & \text{if } a = d, b = c = 0 \\
1 \pm \frac{b \pm c}{\sqrt{\Delta}} & \text{otherwise.}
\end{cases} \]

Here, $\Delta = (a - d)^2 + 4bc$ is the discriminant of the characteristic polynomial of $A_C$. 

Documenta Mathematica 18 (2013) 1243–1274
Proof. If \( a = d, b \neq 0, \) and \( c = 0, \) then we have
\[
\#N_m(C) = a^m + a^{m-1}b + a^{m-2}bd + \cdots + abd^{m-2} + bd^{m-1} + d^m
\]
\[= 2a^m + mba^{m-1}.
\]
Hence, we obtain
\[
\zeta_C(z) = \exp \left( \sum_{m=1}^{\infty} \frac{1}{m} (2a^m + mba^{m-1}) z^m \right)
\]
\[= \frac{1}{(1-az)^2} \exp \left( \frac{bz}{1-az} \right).
\]
If \( a = d, b = 0, \) and \( c \neq 0, \) then we can similarly prove the result.
If \( a = d \) and \( b = c = 0, \) then the category consists of one-object categories, so Lemma 2.7 and Proposition 2.13 imply the result.
In the other cases, \( AC \) is diagonalizable over \( \mathbb{R} \) since \( \Delta \) is nonzero and is a nonnegative real number. We omit the process to compute \( P, \) since the calculation is routine. Lemma 2.14 completes the proof.

Example 2.16. Let
\[
P = x \rightarrow y.
\]
Then, \( AP = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \) so Proposition 2.15 implies that the zeta function is
\[
\zeta_P(z) = \frac{1}{(1-z)^2} \exp \left( \frac{z}{1-z} \right),
\]
which is not a rational function. In the proof of Theorem 3.3 we will find the reason why the zeta function of a finite category has an exponential factor is that a nonzero eigenvalue of its adjacency matrix has algebraic multiplicity.

Example 2.17. Let \( F \) be the following category:
\[
\begin{array}{c}
x \hspace{1cm} i \hspace{1cm} y \hspace{1cm} r \\
\downarrow \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \\
r \hspace{1cm} i \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \\
\end{array}
\]
where \( r \circ i = 1_x, i \circ r \neq 1_y. \) Then, \( AF = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}. \) Proposition 2.15 implies that the zeta function is
\[
\zeta_F(z) = \frac{1}{\left(1 - \left(\frac{3+\sqrt{5}}{2}\right) z\right)^{1+\sqrt{5}}} \frac{1}{\left(1 - \left(\frac{3-\sqrt{5}}{2}\right) z\right)^{1-\sqrt{5}}},
\]
The reason that \( \sqrt{5} \) appears is that the sequence \( (\#N_m(F))_{m \geq 1} \) is a subsequence of the Fibonacci sequence \( (F_m)_{m \geq 1}; \) that is, we have \( \#N_m(F) = F_{m+3} \) and
\[
F_m = \frac{1}{\sqrt{5}} \left( \left(\frac{1+\sqrt{5}}{2}\right)^m - \left(\frac{1-\sqrt{5}}{2}\right)^m \right)
\]
1256  K. NOGUCHI

(see §1.3 of [W106]). Hence, Proposition 2.11 also implies the result. Here, let us confirm that Theorem [3,5] holds for this zeta function.

1. The zeta function of $F$ is

$$\zeta_F(z) = \frac{1}{\left(1 - \left(\frac{3+\sqrt{5}}{2}\right)z\right)^{1+\frac{1}{\sqrt{5}}} \left(1 - \left(\frac{3-\sqrt{5}}{2}\right)z\right)^{1-\frac{1}{\sqrt{5}}}}$$

2. The sum of the indexes is the number of objects in $F$, which is

$$\left(1 + \frac{2}{\sqrt{5}}\right) + \left(1 - \frac{2}{\sqrt{5}}\right) = 2.$$  

3. The numbers $\frac{3\pm\sqrt{5}}{2}$ are algebraic integers. More precisely, they are integers in the real quadratic number field $Q(\sqrt{5})$. The ring of integers in $Q(\sqrt{5})$ is

$$\left\{ \frac{a+b\sqrt{5}}{2} \mid a, b \in \mathbb{Z}, a \equiv b \mod 2 \right\}.$$  

4. We obtain

$$\frac{1 + \frac{2}{\sqrt{5}}}{\frac{3+\sqrt{5}}{2}} + \frac{1 - \frac{2}{\sqrt{5}}}{\frac{3-\sqrt{5}}{2}} = 1 = \chi_\Sigma(F).$$

3  MAIN THEOREM

In this section, we prove our main theorem.

3.1 PREPARATIONS FOR OUR MAIN THEOREM

Throughout this section, we will use the following notation.

1. Unless otherwise stated, $C$ is a finite category having $N$ objects.

2. The two polynomials $|A_C - Iz|$ and $\text{sum(adj}(A_C - Iz))$ that will often be used are expressed in the following forms:

$$|A_C - Iz| = a_0 + a_1z + \cdots + a_Nz^N$$

and

$$\text{sum(adj}(A_C - Iz)) = b_0 + b_1z + \cdots + b_{N-1}z^{N-1}.$$  

3. We denote the codegrees of $|A_C - Iz|$ and $\text{sum(adj}(A_C - Iz))$ by the following:

$$\text{codeg }|A_C - Iz| = r$$

and

$$\text{codeg(sum(adj}(A_C - Iz))) = s.$$
The codegree of a polynomial $f(z)$ is the smallest $n$ such that the coefficient of $z^n$ is nonzero. The coefficients $a_N$, $a_{N-1}$, and $a_0$ are $(-1)^N$, $(-1)^{N-1}\text{Tr}(A_C)$, and $|A_C|$, respectively, and $b_{N-1}$ is $(-1)^{N-1}N$. Hence, the codegree of $|A_C-Iz|$ is less than or equal to $N-1$ if $C$ is a nonempty category, since $\text{Tr}(A_C) \geq N$.

**Remark 3.1.** The category $C$ has Euler characteristic if and only if $s \geq r$. In this case, we have

$$\chi\Sigma(C) = \frac{b_r}{a_r}.$$

(See the bottom of p. 46 in [BL08].)

**Lemma 3.2.** If $C$ has Euler characteristic, then we have

$$\deg\left(\sum(\text{adj}(I - A_C z) A_C)\right) = \deg|I - A_C z| - 1 = N - r - 1.$$

**Proof.** Lemma 2.2 of [NogA] and Remark 3.1 imply this result.

To finish this subsection, we prepare some symbols that are needed to state our main theorem.

Suppose that the characteristic polynomial $|A_C - I z|$ is factored as follows:

$$|A_C - I z| = z^r a_N (z - \lambda_1)^{e_1} (z - \lambda_2)^{e_2} \cdots (z - \lambda_n)^{e_n}$$

where $e_i \geq 1$ for any $i$ and $\lambda_i \neq \lambda_j$ if $i \neq j$. Namely, each $\lambda_k$ is a nonzero eigenvalue of $A_C$ and $e_k$ is its algebraic multiplicity. Then, $|I - A_C z|$ is factored as follows:

$$|I - A_C z| = (-1)^N a_r \left( z - \frac{1}{\lambda_1} \right)^{e_1} \left( z - \frac{1}{\lambda_2} \right)^{e_2} \cdots \left( z - \frac{1}{\lambda_n} \right)^{e_n}.$$

Suppose that

$$\sum(\text{adj}(I - A_C z) A_C) = q(z)|I - A_C z| + r(z),$$

where $r(z)$ and $q(z)$ are polynomials with $\mathbb{Z}$-coefficients and

$$\deg r(z) < \deg|I - A_C z|.$$

Then, $\frac{r(z)}{|I - A_C z|}$ has a partial fraction decomposition to the following form:

$$r(z) \over |I - A_C z| = \frac{(-1)^N}{a_r} \sum_{k=1}^n \sum_{i=1}^{e_k} \frac{A_{k,i}}{(z - \frac{1}{\lambda_k})^i}$$

for some complex numbers $A_{k,i}$. 

---

Documenta Mathematica 18 (2013) 1243–1274
3.2 A proof of our main theorem

In this subsection, we give a proof of our main theorem. The symbols without explanations are explained in the previous subsection.

**Theorem 3.3.** Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the nonzero eigenvalues of $A_C$ and $e_1, e_2, \ldots, e_n$ are their algebraic multiplicities. Then,

1. the zeta function of $C$ is

$$
\zeta_C(z) = \prod_{k=1}^{n} \frac{1}{(1 - \lambda_k z)^{\beta_{k,0}}} \exp \left( Q(z) + \sum_{j=1}^{e_k-1} \frac{\beta_{k,j} z^j}{j(1 - \lambda_k z)^j} \right),
$$

where $\beta_{k,0} = (-1)^{N-1} \frac{\lambda_k + 1}{a_r}$,

$$
\beta_{k,j} = \frac{(-1)^{N-1} e_k-1}{a_r} \sum_{i=j}^{e_k-1} \binom{i-1}{j-1} (-1)^{i+j} \lambda_k^{i+j} A_{k,i+1}
$$

for $j \geq 1$, and $Q(z) = \frac{1}{z} \int q(z) \, dz$ is a polynomial of $\mathbb{Q}[z]$ whose constant term is zero, and

2. each $\lambda_k$ is an algebraic integer.

To prove this theorem, we use the following proposition.

**Proposition 3.4** (Proposition 2.1 of [NogA]). Let $C$ be a finite category. Then, the zeta function of $C$ is

$$
\zeta_C(z) = \exp \left( \int \frac{\text{adj}(I - A_C z) A_C}{|I - A_C z|} \, dz \right).
$$

Proposition 2.1 of [NogA] assumes the invertibility of $A_C$, but that hypothesis is not used in the proof. Hence, we can abandon that hypothesis, and the same proof can be used for this proposition.
Zeta Functions of Categories

Proof of Theorem 3.3. Proposition 3.4 implies

\[
\zeta_C(z) = \exp \left( \int q(z) \, dz + \int \frac{(-1)^N}{a_r} \sum_{k=1}^{n} \sum_{i=1}^{e_k} \frac{A_{k,i}}{(z - \frac{1}{\lambda_k})^i} \, dz \right)
\]

= \exp \left( \int q(z) \, dz + \frac{(-1)^N}{a_r} \int \sum_{k=1}^{n} \frac{A_{k,i}}{(z - \frac{1}{\lambda_k})^i} \, dz \right)

\[
\frac{(-1)^N}{a_r} \int \sum_{k=1}^{n} \sum_{i=2}^{e_k} \frac{A_{k,i}}{(z - \frac{1}{\lambda_k})^i} \, dz
\]

= \exp \left( Q(z) + \frac{(-1)^N}{a_r} \sum_{k=1}^{n} \sum_{i=1}^{A_{k,i}} \frac{A_{k,i+1}}{(i-1) \left( z - \frac{1}{\lambda_k} \right)^{i-1} + B} \right)

\[
= \prod_{k=1}^{n} \frac{1}{\left( z - \frac{1}{\lambda_k} \right)^{(i-1) \left( z - \frac{1}{\lambda_k} \right)^{i-1} + B}}
\]

= \prod_{k=1}^{n} \frac{1}{\left( 1 - \lambda_k z \right)^{(i-1) \left( z - \frac{1}{\lambda_k} \right)^{i-1} + B}}

\[
\text{Lemma 2.7 of [NogA] implies}
\]

\[
\zeta_C(z) = \prod_{k=1}^{n} \frac{1}{(1 - \lambda_k z)^{(i-1) \left( z - \frac{1}{\lambda_k} \right)^{i-1} + B}} \times
\]

\[
\exp \left( Q(z) + \frac{(-1)^N}{a_r} \sum_{k=1}^{n} \sum_{i=1}^{A_{k,i+1}} \frac{A_{k,i+1}}{(i-1) \left( z - \frac{1}{\lambda_k} \right)^{i-1} + B} \right)
\]

where we replaced (and will replace) the constant term by \( B, B', B'' \ldots \)

\[
\text{Lemma 2.7 of [NogA] implies}
\]

\[
\zeta_C(z) = \prod_{k=1}^{n} \frac{1}{(1 - \lambda_k z)^{(i-1) \left( z - \frac{1}{\lambda_k} \right)^{i-1} + B}} \times
\]

\[
\exp \left( Q(z) + \frac{(-1)^N}{a_r} \sum_{k=1}^{n} \sum_{i=1}^{A_{k,i+1}} \frac{A_{k,i+1}}{(i-1) \left( z - \frac{1}{\lambda_k} \right)^{i-1} + B} \right)
\]

Here, we use the boundary condition \( \zeta_C(0) = 1 \). This condition is directly implied by the definition of the zeta function. Hence, we obtain \( B'' = 1 \). By
exchanging $\sum_i$ and $\sum_j$, we have

$$\zeta_C(z) = \prod_{k=1}^{n} \left( \frac{1}{1 - \lambda_k z} \right)^{\frac{1}{a_r \beta_{k,0}}} \exp \left( Q(z) + \frac{(-1)^{N-1}}{a_r} \sum_{k=1}^{n} \sum_{j=1}^{c_k-1} \frac{z^j}{j!} \left( \sum_{i=j}^{c_k-1} \frac{(i-1)! \lambda_k^{i+j} A_{k,i+1}}{(1 - \lambda_k z)^{i+j}} \right) \right).$$

Hence, we obtain the first result.

Since $(-1)^N|A_C - Iz|$ is a monic polynomial with coefficients in $\mathbb{Z}$, it follows that $\lambda_k$ is an algebraic integer, so we obtain the second result.

**Theorem 3.5.** Suppose that $C$ has Euler characteristic and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the nonzero eigenvalues of $A_C$ and $e_1, e_2, \ldots, e_n$ are their algebraic multiplicities. Then, we obtain the following results.

1. The zeta function of $C$ is

$$\zeta_C(z) = \prod_{k=1}^{n} \frac{1}{1 - \lambda_k z} \beta_{k,0} \exp \left( \sum_{j=1}^{c_k-1} \frac{\beta_{k,j} z^j}{j!} \right)$$

where $\beta_{k,0} = (-1)^{N-1} \frac{A_{k,1}}{a_r}$ and

$$\beta_{k,j} = (-1)^{N-1} \frac{A_{k,1}}{a_r} \sum_{i=j}^{c_k-1} \binom{i-1}{j-1} (-1)^i \lambda_k^{i+j} A_{k,i+1}$$

for $j \geq 1$.

2. The sum of all the indexes $\beta_{k,0}$ is the number of objects of $C$; that is,

$$\sum_{k=1}^{n} \beta_{k,0} = N.$$

3. Each $\lambda_k$ is an algebraic integer.

4.

$$\sum_{k=1}^{n} \sum_{j=0}^{c_k-1} (-1)^j \frac{\beta_{k,j} z^j}{\lambda_k^{j+1}} = \chi_{C}(C) \in \mathbb{Q}.$$
Proof. Since $C$ has Euler characteristic, Lemma 3.2 implies
\[ \deg(\text{sum(adj})(I - A_Cz)A_C) < \deg |I - A_Cz|. \]
Hence, we have $q(z) = 0$ and $r(z) = \text{sum(adj)}(I - A_Cz)A_C$, so we obtain the first result by Theorem 3.3 as $Q(z) = 0$.
By elementary calculation, we have
\[ \frac{\text{sum(adj)}(A_C - Iz))}{|A_C - Iz|} = \frac{N}{z} \frac{1}{z^2} \frac{\text{sum(adj)}(I - \frac{1}{2}A_C)A_C}{|I - \frac{1}{2}A_C|}. \]
Since $r(z) = \text{sum(adj)}(I - A_Cz)A_C)$, the partial fraction decomposition (11) tells us that this is equal to
\[ -\frac{N}{z} + \frac{(-1)^{N-1}}{a_r} \sum_{k=1}^n \sum_{k=1}^n \frac{A_{k,i}z^{i-2}}{(1 - \frac{1}{\lambda_k})^r}. \]
This, in turn, is equal to the Laurent series,
\[ \left( \sum_{k=1}^N \beta_{k,0} - N \right) \frac{1}{z} + \frac{(-1)^{N-1}}{a_r} \sum_{k=1}^n \left( \frac{A_{k,1}}{\lambda_k} + A_{k,2} \right) + \sum_{m=1}^\infty c_m z^m, \]
for some complex numbers $c_1, c_2, \ldots$. Since $C$ has Euler characteristic, the rational function $\text{sum(adj)}(I - Iz)$ is defined at zero (see p. 45 of [BL08]), so $\sum_{k=1}^n \beta_{k,0} = N$, proving the second result.
We have already shown the third result in Theorem 3.3.
Finally, we show the fourth result. The left hand side of (2) is
\[ 2 = \sum_{k=1}^n (-1)^{N-1} \frac{A_{k,1}}{\lambda_k a_r} \]
\[ + \frac{(-1)^{N-1}}{a_r} \sum_{k=1}^n \sum_{j=1}^{i-1} (-1)^{i+j+1} \frac{A_{k,i-1}}{\lambda_k^{i-1}} \]
\[ = \sum_{k=1}^n \left( (-1)^{N-1} \frac{A_{k,1}}{\lambda_k a_r} \right. \]
\[ + \left. \frac{(-1)^{N-1}}{a_r} \sum_{j=1}^{i-1} \sum_{i=j}^{i-1} (-1)^{i+j+1} \frac{A_{k,i-1}}{\lambda_k^{i-1}} \right) \]
\[ = \sum_{k=1}^n \left( (-1)^{N-1} \frac{A_{k,1}}{\lambda_k a_r} \right. \]
\[ + \left. \frac{(-1)^{N-1}}{a_r} \sum_{i=1}^n (-1)^{i+1} \lambda_k^{i-1} A_{k,i+1} \right) \]
\[ = \frac{(-1)^{N-1}}{a_r} \left( \sum_{k=1}^n \frac{A_{k,1}}{\lambda_k} \right. \]
\[ + \left. \sum_{k=1}^n \frac{A_{k,2}}{\lambda_k} \right). \]
The Laurent series implies
\[
\chi_\Sigma(C) = \frac{\sum(\text{adj}(A_C - I z))}{|A_C - I z|}_{z=0} = \frac{(-1)^{N-1}}{a_r} \sum_{k=1}^{n} \left( \frac{A_{k,1}}{\lambda_k} + A_{k,2} \right).
\]

Hence, we obtain the result.

We give an interpretation of Part 2 and 4 of Theorem 3.5 by residues. Let \( f \) be a holomorphic function on the whole complex plane with the exception of finitely many poles \( p_1, p_2, \ldots, p_j \). Then the residue of \( f \) at infinity is defined by
\[
\text{Res}(f(z) : \infty) = -\sum_{i=1}^{j} \text{Res}(f(z) : p_i).
\]

**Corollary 3.6.** If \( C \) has Euler characteristic, then we have
\[
\text{Res} \left( \frac{\zeta_C'(z)}{\zeta_C(z)} : \infty \right) = N
\]
and
\[
\text{Res} \left( \frac{z \zeta_C'(z)}{\zeta_C(z)} : \infty \right) = \chi_\Sigma(C).
\]

**Proof.** By Proposition 3.4, the logarithmic derivative of \( \zeta_C(z) \) is
\[
\frac{\sum(\text{adj}(I - A_C z)A_C)}{|I - A_C z|}.
\]

Lemma 3.2, the partial fraction decomposition 11, and Part 2 of Theorem 3.5 imply the first result. Moreover, by elementary calculation, we have
\[
\frac{z \zeta_C'(z)}{\zeta_C(z)} = -N + \frac{(-1)^N}{a_r} \sum_{k=1}^{n} \left( \frac{A_{k,1}}{\lambda_k} + A_{k,2} \right) + \sum_{k=1}^{n} \sum_{i=2}^{c_k} \frac{c_{k,i}}{z - \frac{\lambda_k}{\lambda_e}}
\]
for some complex numbers \( c_{k,i} \). Hence we obtain
\[
\text{Res} \left( \frac{z \zeta_C'(z)}{\zeta_C(z)} : \infty \right) = \frac{(-1)^{N-1}}{a_r} \left( \sum_{k=1}^{n} \frac{A_{k,1}}{\lambda_k} + A_{k,2} \right) = \chi_\Sigma(C).
\]

The last equality follows from one of the equations at the bottom of the proof of Theorem 3.5.
3.3 Examples

In this subsection, we introduce four examples of zeta functions. These are implied, for example, by routine calculations to solve the characteristic polynomial of each adjacency matrix and compute a partial fraction decomposition. Since the calculations are routine, only the results are shown.

Example 3.7. Let $C$ be a finite category whose adjacency matrix is \[
\begin{pmatrix}
2 & 3 & 5 \\
2 & 3 & 5 \\
2 & 1 & 3
\end{pmatrix}.
\] This is Example 4.7 of [BL08]. The existence of such a category is assured by Lemma 4.1 of [BL08]. Then, $\chi_{\Sigma}(C)$ is not defined. Its zeta function is \[
\zeta_C(z) = \frac{1}{1 - 8z^{1/4}}.
\] We note that the index is not the number of objects of $C$; that is, $\frac{13}{4} \neq 12 = 3$. Therefore, we cannot abandon the hypothesis in Theorem 3.5 that $C$ has Euler characteristic.

Example 3.8. Let $C$ be a finite category whose adjacency matrix is \[
\begin{pmatrix}
2 & 2 & 2 \\
2 & 2 & 2 \\
2 & 8 & 5
\end{pmatrix}.
\] This is Example 4.5 of [BL08]. Then, both $\chi_L(C)$ and $\chi_{\Sigma}(C)$ are defined. Here, $\chi_L$ is the Euler characteristic of a finite category by Leinster [Lei08]. We have \[
\chi_L(C) = \frac{1}{2}, \chi_{\Sigma}(C) = \frac{1}{3}.
\] Its zeta function is \[
\zeta_C(z) = \frac{1}{1 - 9z^{3/4}}.
\] Note that $\frac{3}{9} = \chi_{\Sigma}(C)$, but $\frac{3}{9} \neq \chi_L(C)$, so our zeta function does not recover $\chi_L$.

Remark 3.9. Our zeta function also does not recover the $L^2$-Euler characteristic $\chi^{(2)}$ [FLST11], since the zeta function of a finite category does not depend on its composition, but the $L^2$-Euler characteristic does. Indeed, let $C_1$ be a one-object category whose set of morphisms is $\{1, f\}$, where $f \circ f = f$, and let $C_2$ be almost the same category as $C_1$, with the only difference that $f \circ f = 1$ in $C_2$. Then, Proposition 2.13 implies that their zeta functions are \[
\zeta_{C_1}(z) = \zeta_{C_2}(z) = \frac{1}{1 - 2z},
\] but $\chi^{(2)}(C_2) = \frac{1}{2}$ and $\chi^{(2)}(C_1) \neq \frac{1}{2}$ by Example 5.12 and Remark 7.2 of [FLST11].

The zeta functions in the following two examples use nonreal numbers.
Example 3.10. Let $C$ be a finite category whose adjacency matrix is 
\[
\begin{pmatrix}
2 & 3 & 2 \\
1 & 2 & 6 \\
1 & 1 & 2
\end{pmatrix}.
\]
Since $A_C$ has an inverse matrix, Theorem 3.2 of [BL08] implies that $C$ has Euler characteristic given by
\[
\chi_C = \text{sum}(A_C^{-1}) = \frac{5}{6}.
\]
Let us confirm that this zeta function satisfies the statement of Theorem 3.5.

1. The zeta function is
\[
\zeta_C(z) = \frac{1}{(1-6z)^\frac{1}{37} (1-iz)^\frac{2}{37} (1+i\bar{z})^\frac{2}{37}}.
\]

2. The sum of indexes is
\[
\frac{125}{37} + \frac{-7 + 5i}{37} + \frac{-7 - 5i}{37} = 3.
\]

3. The numbers 6 and $\pm i$ are algebraic integers. In particular, they are integers in $\mathbb{Q}(\sqrt{-1})$; that is, they belong to the ring of Gaussian integers $\mathbb{Z}[\sqrt{-1}]$.

4. Moreover, we have
\[
\frac{1}{i} \cdot \frac{125}{37} + \frac{1}{-i} \cdot \frac{-7 - 5i}{37} = \frac{5}{6}.
\]

Example 3.11. Let $C$ be a finite category whose adjacency matrix is
\[
\begin{pmatrix}
4 & 7 & 8 \\
1 & 4 & 5 \\
1 & 1 & 3
\end{pmatrix}.
\]
Since $A_C$ has an inverse matrix, its Euler characteristic is given by
\[
\chi_C = \text{sum}(A_C^{-1}) = 0.
\]
Let us confirm that this zeta function satisfies the statement of Theorem 3.5.

1. The zeta function is
\[
\zeta_C(z) = \frac{1}{(1-9z)^\frac{252}{65} (1-(1+i)z)^\frac{252-i}{130} (1-(1-i)z)^\frac{252+i}{130}}.
\]

2. The sum of indexes is
\[
\frac{252}{65} + \frac{-57 + i}{130} + \frac{-57 - i}{130} = 3.
\]

3. The numbers 6 and $1 \pm i$ belong to the ring of Gaussian integers $\mathbb{Z}[\sqrt{-1}]$.

4. Moreover, we have
\[
\frac{1}{9} \cdot \frac{252}{65} + \frac{1}{1+i} \cdot \frac{-57 + i}{130} + \frac{1}{1-i} \cdot \frac{-57 - i}{130} = 0.
\]
4 Coverings of Small Categories

The aim of this section is to prove that for a covering of finite categories, \( P : E \to B \), the zeta function of \( E \) is that of \( B \) to the power of the number of sheets in the covering. Some examples are given in the last subsection of this section.

4.1 Coverings and Zeta Functions

In this subsection, we show that for a covering of finite categories, \( P : E \to B \), the zeta function of \( E \) is that of \( B \) to the power of the number of sheets in the covering.

Here, let us recall a covering of small categories \[ \text{BH99}. \]

Let \( C \) be a small category. For an object \( x \) of \( C \), let \( S(x) \) be the set of morphisms of \( C \) whose source is \( x \),

\[ S(x) = \{ f : x \to \ast \in \text{Mor}(C) \}, \]

and let \( T(x) \) be the set of morphisms of \( C \) whose target is \( x \),

\[ T(x) = \{ g : \ast \to x \in \text{Mor}(C) \}. \]

For the rest of this section, we assume that \( E \) and \( B \) are small categories and \( B \) is connected. A functor \( P : E \to B \) is a covering if the following two restrictions of \( P \) are bijections for any object \( x \) of \( E \):

\[
P : S(x) \to S(P(x)),
\]

\[
P : T(x) \to T(P(x)).
\]

This condition is an analogue of the condition on an unramified covering of graphs (see \[ \text{ST96}. \]). A functor \( P \) is called a discrete fibration if the restriction \( P : T(x) \to T(P(x)) \) is a bijection for any object \( x \) of \( E \), and \( P \) is called a discrete opfibration if the restriction \( P : S(x) \to S(P(x)) \) is a bijection for any object \( x \) of \( E \). Thus, a functor is a covering if and only if it is both a discrete fibration and a discrete opfibration.

For an object \( b \) of \( B \), the inverse image \( P^{-1}(b) \) of the restriction of \( P \) with respect to objects,

\[
P^{-1}(b) = \{ x \in \text{Ob}(E) \mid P(x) = b \},
\]

is called the fiber of \( b \) by \( P \). The cardinality of \( P^{-1}(b) \) is called the number of sheets in \( P \), and it does not depend on the choice of \( b \) since the base category \( B \) is connected (see Proposition \[ 4.1 \]).

Applying the classifying space functor \( B \) to a covering \( P : E \to B \), we have the covering space \( BP \) in the topological sense (see \[ Tan \]). There has been much work on coverings of small categories; for example, see \[ BH99, CM \], and \[ Tan \]. In particular, coverings of groupoids were studied in \[ May99 \].
Let

It suffices to show that there exists a morphism \( f : b \to b' \). Indeed, if this is proven, then for any objects \( b \) and \( b' \) we have a zig-zag sequence

\[
\begin{array}{c}
b \\
\downarrow \\
b_1 \\
\downarrow \\
b_2 \\
\downarrow \\
\vdots \\
\downarrow \\
b' \\
\end{array}
\]

so we obtain

\[
P^{-1}(b) \cong P^{-1}(b_1) \cong \cdots \cong P^{-1}(b').
\]

Suppose that there exists a morphism \( f : b \to b' \). By the definition of a covering, there exist induced functions

\[
f_* : P^{-1}(b) \to P^{-1}(b'), \quad f^* : P^{-1}(b') \to P^{-1}(b).
\]

Here, \( f_*(x) \) is the target \( x' \) of the unique morphism \( g : x \to x' \) such that \( P(g) = f \), and similarly with \( f^* \). It follows immediately from the uniqueness that \( f^* f_* = 1 \), and similarly with \( f^* \) and \( f_* \), reversed. Hence, \( f_* \) and \( f^* \) are inverse to one another.

**Definition 4.2.** Let \( C \) be a small category and \( x \) be an object of \( C \). Then, let \( N_m(C)_x \) be the set of chains of morphisms of length \( m \) in \( C \) and whose target is \( x \):

\[
N_m(C)_x = \{ \langle x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} \cdots \xrightarrow{f_m} x_m \rangle \mid x_m = x \}.
\]

**Proposition 4.3.** Let \( P : E \to B \) be a covering. Then, \( N_m(E)_x \) is bijective to \( N_m(B)_b \) for any object \( b \) of \( B \), any \( x \) of \( P^{-1}(b) \), and \( m \geq 0 \).

**Proof.** Given a sequence of morphisms in \( B \),

\[
g = \left( b_0 \xrightarrow{g_1} b_1 \xrightarrow{g_2} \cdots \xrightarrow{g_m} b_m = b \right),
\]

there exists a unique morphism \( f_m : x_{m-1} \to x \) such that \( P(f_m) = g_m \) since \( P \) is a covering. If we repeat this process, we get a unique sequence of morphisms in \( E \),

\[
f = \left( x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} \cdots \xrightarrow{f_m} x_m = x \right)
\]

such that \( P(f) = g \). This correspondence gives a bijection between \( N_m(E)_x \) and \( N_m(B)_b \).

**Proposition 4.4.** Let \( P : E \to B \) be a covering and let \( b \) be an object of \( B \). Then, \( N_m(E) \) is bijective to \( P^{-1}(b) \times N_m(B) \) for any \( m \geq 0 \).
Proof. Proposition 4.3 implies

\[
N_m(E) = \prod_{x \in \text{Ob}(E)} N_m(E)_x
= \prod_{b \in \text{Ob}(B)} \prod_{x \in P^{-1}(b)} N_m(E)_x
\cong \prod_{b \in \text{Ob}(B)} \prod_{x \in P^{-1}(b)} N_m(B)_b
\cong P^{-1}(b) \times N_m(B).
\]

The following theorem is an analogue of an unproved conjecture of Dedekind and is the main result of this section. The conjecture is that for algebraic number fields \(K_1 \subset K_2\), the Dedekind zeta function \(\zeta_{K_1}(s)\) of \(K_1\) divides that of \(K_2\) [Wan74]. The graph theoretic analogue of this conjecture was considered in Corollary 1 of §2 of [ST96].

**Theorem 4.5.** Let \(P : E \to B\) be a covering of finite categories and let \(b\) be an object of \(B\). Then, we have

\[
\zeta_E(z) = \zeta_B(z)^{\#P^{-1}(b)}.
\]

Proof. Proposition 4.3 and the definition of the zeta function of a finite category directly imply this fact; that is,

\[
\zeta_E(z) = \exp \left( \sum_{m=1}^{\infty} \frac{\#N_m(E)}{m} z^m \right)
= \exp \left( \sum_{m=1}^{\infty} \frac{\#P^{-1}(b) \#N_m(B)}{m} z^m \right)
= \zeta_B(z)^{\#P^{-1}(b)}.
\]

4.2 Coverings and Euler characteristics

Our main purpose in this section has already been accomplished in Theorem 4.5. Aside from the main topic of this section, we investigate the relationships between coverings and Euler characteristics of categories. Let \(p : X \to Y\) be a topological fibration, which is one of the generalized notions of covering spaces (e.g., see [Hat02] and [May99]). Under a suitable hypothesis, we have the formula

\[
\chi(X) = \chi(F)\chi(Y),
\]

where \(F\) is the fiber of \(p\).
A categorical analogue of this formula was considered in [Lei08] and [FLS11]. Proposition 2.8 of [Lei08] is an analogue for Grothendieck fibrations and the Euler characteristic $\chi_L$. Theorems 5.30 and 5.37 of [FLS11] are analogues for isofibrations, coverings of groupoids, and the $L^2$-Euler characteristic $\chi^{(2)}$.

In this subsection, we consider such analogues for coverings, the Euler characteristic $\chi$, and the $\chi_{fil}$ of $\mathbb{N}$-filtered acyclic categories $\chi_{fil}$ [Nog11].

Here, we recall the Euler characteristic of an $\mathbb{N}$-filtered acyclic category $\chi_{fil}$ [Nog11]. Let $A$ be an acyclic category. We define an order on the set $\text{Ob}(A)$ of objects of $A$ by $x \leq y$ if there exists a morphism $x \to y$. Then, $\text{Ob}(A)$ is a poset; that is, $\text{Ob}(A)$ is acyclic and each hom-set has at most one morphism.

**Definition 4.6.** Let $A$ be an acyclic category. A functor $\mu : A \to \mathbb{N} \cup \{0\}$ satisfying $\mu(x) < \mu(y)$ for $x < y$ in $\text{Ob}(A)$ is called an $\mathbb{N}$-filtration of $A$. A pair $(A,\mu)$ is called an $\mathbb{N}$-filtered acyclic category.

**Definition 4.7.** Let $(A,\mu)$ be an $\mathbb{N}$-filtered acyclic category. For nonnegative integers $i$ and $m$, let $\overline{N}_m(A)_i = \{ f \in \overline{N}_m(A) | \mu(t(f)) = i \}$, where $t(f) = (x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} \cdots \xrightarrow{f_m} x_m)$.

Suppose that each $\overline{N}_m(A)_i$ is finite. We define the formal power series $f_{A}(A,\mu)(t)$ over $\mathbb{Z}$ by

$$f_{A}(A,\mu)(t) = \sum_{i=0}^{\infty} (-1)^i \left( \sum_{m=0}^{i} (-1)^m \# \overline{N}_m(A)_i \right) t^i.$$ 

Then, we define $\chi_{fil}(A,\mu) = f_{A}(A,\mu)|_{t=-1}$ if $f_{A}(A,\mu)(t)$ is rational and has a nonvanishing denominator at $t = -1$.

We first demonstrate the formula for coverings and the Euler characteristic $\chi_{fil}$. Propositions 4.3 and 4.4 hold when nerves are nondegenerate, which means that we do not use identity morphisms. Let $C$ be a small category and let $x$ and $y$ be objects of $C$. We define the following symbols:

$$\overline{S}(x) = S(x) \setminus \{1_x\}, \overline{T}(x) = T(x) \setminus \{1_x\},$$

$$\text{Hom}_{C}(x,y) = \begin{cases} \text{Hom}_{C}(x,y) \setminus \{1_x\} & \text{if } x = y \\ \text{Hom}_{C}(x,y) & \text{if } x \neq y \end{cases}$$

and

$$\overline{N}_m(C)_x = \{ (x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} \cdots \xrightarrow{f_m} x_m) \text{ in } C | f_i \neq 1, x_m = x \}.$$
Proposition 4.8. Let $P : E \to B$ be a covering. Then, $\overline{N_m(E)}_x$ is bijective to $\overline{N_m(B)}_b$ for any object $b$ of $B$, any $x$ of $P^{-1}(b)$, and $m \geq 0$.

Proof. If we replace the symbols in the proof of Proposition 4.3 by the above symbols with bars, we can use the same proof. Note that for any morphism $f$ of $E$ it follows that $f$ is an identity morphism if and only if $P(f)$ is an identity morphism. □

Proposition 4.9. Let $P : E \to B$ be a covering and $b$ be an object of $B$. Then, $\overline{N_m(E)}$ is bijective to $P^{-1}(b) \times \overline{N_m(B)}$ for any $m \geq 0$.

A discrete category consists of only objects and identity morphisms. The fiber of a covering $P : E \to B$ is a discrete category when we regard it as a category.

Proposition 4.10. Let $P : E \to B$ be a covering of finite categories and let $b$ be an object of $B$. Then, $E$ has Euler characteristic if and only if $B$ has Euler characteristic. In this case, we have

$$\chi(E) = \chi(P^{-1}(b)) \chi(B).$$

Proof. Theorem 2.2 of [BL08] and Proposition 4.9 imply

$$\sum_{m=0}^{\infty} \#N_m(E) t^m = \#P^{-1}(b) \sum_{m=0}^{\infty} \#N_m(B) t^m = \#P^{-1}(b) \frac{\text{sum}(\text{adj}(I - (AB - I)t))}{|I - (AB - I)t|}.$$}

Hence, $E$ has Euler characteristic if and only if we can substitute $t = -1$ in

$$\#P^{-1}(b) \frac{\text{sum}(\text{adj}(I - (AB - I)t))}{|I - (AB - I)t|}$$

if and only if we can substitute $t = -1$ in

$$\frac{\text{sum}(\text{adj}(I - (AB - I)t))}{|I - (AB - I)t|}$$

if and only if $B$ has Euler characteristic. Thus, we have proven the first claim. If $E$ has Euler characteristic, then we have

$$\chi(E) = \#P^{-1}(b) \chi(B) = \chi(P^{-1}(b)) \chi(B).$$

Next, we demonstrate the formula for coverings and the Euler characteristic of $\mathbb{N}$-filtered acyclic categories $\chi_{fil}$. 

Documenta Mathematica 18 (2013) 1243–1274
Proposition 4.11. Suppose that $(A, \mu_A)$ and $(B, \mu_B)$ are $\mathbb{N}$-filtered acyclic categories, $b_0$ is an object of $B$, and $P : A \to B$ is a covering whose fiber is finite, satisfying $\mu_A(x) = \mu_B(P(x))$ for any object $x$ of $A$. Then, $(A, \mu_A)$ has Euler characteristic $\chi_{\text{fil}}(A, \mu_A)$ if and only if $B$ has Euler characteristic $\chi_{\text{fil}}(B, \mu_B)$. In this case, we have

$$\chi_{\text{fil}}(A, \mu_A) = \chi_{\text{fil}}(P^{-1}(b_0), \mu)\chi_{\text{fil}}(B, \mu_B)$$

for any $\mathbb{N}$-filtration $\mu$ of $P^{-1}(b_0)$.

Proof. We have

$$\mu_A^{-1}(i) = \prod_{b \in \mu_B^{-1}(i)} P^{-1}(b)$$

for any $i \geq 0$. Propositions 4.1 and 4.8 imply

$$\overline{N_m}(A)_i = \prod_{x \in \mu_A^{-1}(i)} N_m(A)_x$$

$$\cong \prod_{b \in \mu_B^{-1}(i)} \prod_{x \in P^{-1}(b)} N_m(A)_x$$

$$\cong \prod_{b \in \mu_B^{-1}(i)} P^{-1}(b) \times N_m(B)_b$$

$$\cong P^{-1}(b_0) \times \overline{N_m}(B)_i.$$ 

Hence, we have

$$f_X(A, \mu_A)(t) = \sum_{i=0}^{\infty} (-1)^i \left( \sum_{m=0}^{i} (-1)^m \# N_m(A)_i \right) t^i$$

$$= \sum_{i=0}^{\infty} (-1)^i \left( \sum_{m=0}^{i} (-1)^m \# P^{-1}(b_0) \# N_m(B)_i \right) t^i$$

$$= \# P^{-1}(b_0) f_X(B, \mu_B)(t).$$

Accordingly, $\chi_{\text{fil}}(A, \mu_A)$ exists if and only if the power series $f_X(A, \mu_A)(t)$ is rational and we can substitute $t = -1$ in the rational function if and only if the power series $f_X(B, \mu_B)(t)$ is rational and we can substitute $t = -1$ in the rational function if and only if $\chi_{\text{fil}}(B, \mu_B)$ exists. Thus, the first claim has been proven.

If $\chi_{\text{fil}}(A, \mu_A)$ exists, then we have

$$\chi_{\text{fil}}(A, \mu_A) = \# P^{-1}(b_0) \chi_{\text{fil}}(B, \mu_B)$$

$$= \chi_{\text{fil}}(P^{-1}(b_0), \mu) \chi_{\text{fil}}(B, \mu_B).$$

It is clear that $\chi_{\text{fil}}(P^{-1}(b_0), \mu) = \# P^{-1}(b_0)$ for any $\mathbb{N}$-filtration $\mu$. We can provide a filtration to $P^{-1}(b_0)$; for example, we can define $\mu : P^{-1}(b_0) \to \mathbb{N} \cup \{0\}$ by $\mu(x) = 0$ for any $x$ of $P^{-1}(b_0)$.

\[\square\]
4.3 Examples

We give three examples of coverings of small categories.

Example 4.12. Let

\[ \Gamma = \begin{array}{c}
  x \\
  \downarrow \ f \\
  y
\end{array} \]

and \( B = \mathbb{Z}_2 = \{1, -1\} \). A group can be regarded as a category whose object is just one object (denoted by an asterisk), whose morphisms are the elements of \( G \), and whose composition is the operation of \( G \). We define \( P : \Gamma \to B \) by \( P(f) = P(f^{-1}) = -1 \). Then, \( P \) is a covering that was studied in Example 5.33 of [FLS11]. Since \( \Gamma \) and \( B \) are finite connected groupoids, Proposition 2.3 implies

\[ \zeta_\Gamma(z) = \frac{1}{1 - 2z^2}, \quad \zeta_B(z) = \frac{1}{1 - z}. \]

The number of sheets in \( P \) is two. We have \( \zeta_\Gamma(z) = \zeta_B(z^2)^2 \). Example 2.7 of [Lei08] and Theorem 3.2 of [BL08] imply

\[ \chi_\Sigma(\Gamma) = 1, \quad \chi_\Sigma(B) = 1, \quad \chi_\Sigma(P^{-1}(\ast)) = 2, \]

and hence we have

\[ \chi_\Sigma(\Gamma) = \chi_\Sigma(P^{-1}(\ast)) \chi_\Sigma(B). \]

Example 4.13. Let

\[ A = \begin{array}{cccc}
  y_1 \\
  \downarrow f_1 \\
  x_1 \\
  \downarrow f_2 \\
  y_2 \\
  \downarrow f_2 \\
  x_2 \\
  \downarrow f_3 \\
  y_3 \\
  \downarrow f_3 \\
  \vdots \\
  \downarrow f_{n-1} \\
  y_{n-1} \\
  \downarrow f_{n-1} \\
  x_{n-1} \\
  \downarrow f_n \\
  y_n \\
  \downarrow f_n \\
  x_n
\end{array} \]

and

\[ B = \begin{array}{c}
  x \\
  \downarrow g \\
  y
\end{array} \]

We define a functor \( P : A \to B \) by \( P(x_i) = x, \ P(y_i) = y, \ P(f_i) = f, \) and \( P(g_i) = g \) for any \( i \). Then, \( P \) is a covering. By Corollary 2.12 we have

\[ \zeta_A(z) = \frac{1}{(1 - z)^n} \exp \left( \frac{2nz}{1 - z} \right), \quad \zeta_B(z) = \frac{1}{(1 - z)^2} \exp \left( \frac{2z}{1 - z} \right). \]

The number of sheets in \( P \) is \( n \). We have \( \zeta_A(z) = \zeta_B(z^n) \). Since \( A \) and \( B \) are finite acyclic categories, \( \sum_{m=0}^{\infty} \# N(m)(A)t^m \) and \( \sum_{m=0}^{\infty} \# N(m)(B)t^m \) are polynomials by Lemma 3.5 of [Nog]. Hence, we have

\[ \chi_\Sigma(A) = 2n - 2n = 0, \quad \chi_\Sigma(B) = 2 - 2 = 0, \quad \chi_\Sigma(P^{-1}(x)) = n, \]

and then we have

\[ \chi_\Sigma(A) = \chi_\Sigma(P^{-1}(x)) \chi_\Sigma(B). \]
We introduce an example of a covering of infinite categories.

**Example 4.14.** Suppose that

\[
A = x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow \cdots \quad \text{and} \quad y_0 \rightarrow y_1 \rightarrow y_2 \rightarrow \cdots
\]

and

\[
B = b_0 \rightarrow b_1 \rightarrow b_2 \rightarrow \cdots
\]

where \(A\) is a poset. For \(n < m\), \(b_n\) and \(b_m\), we define

\[
\text{Hom}_B(b_n, b_m) = \{\varphi_{n,m}^0, \varphi_{n,m}^1\},
\]

and a composition of \(B\) is defined by \(\varphi_{n,m}^j \circ \varphi_{n,m}^i = \varphi_{n,m}^k\), where \(k = 0\) or \(k = 1\) and \(k \equiv i + j \mod 2\) for \(n < m < \ell\). We define \(P : A \to B\) by \(P(x_i) = P(y_i) = b_i\), with \(P((x_n, x_m)) = P((y_n, y_m)) = \varphi_{n,m}^0\) and \(P((x_n, y_m)) = \varphi_{n,m}^1\) for \(n < m\). Then, \(P\) is a covering. The indexes of objects of \(A\) and \(B\) give \(\mathbb{N}\)-filtrations \(\mu_A\) and \(\mu_B\) to \(A\) and \(B\), respectively. We have

\[
f_\chi(A, \mu_A)(t) = \sum_{i=0}^{\infty} (-1)^i \left( \sum_{m=0}^{i} (-1)^m 2^{m+1} \binom{i}{m} \right) t^i = \frac{2}{1 - t},
\]

so \(\chi^\text{fil}(A, \mu_A) = 1\). We have

\[
f_\chi(B, \mu_B)(t) = \sum_{i=0}^{\infty} (-1)^i \left( \sum_{m=0}^{i} (-1)^m 2^{m+1} \binom{i}{m} \right) t^i = \sum_{i=0}^{\infty} t^i = \frac{1}{1 - t},
\]

so \(\chi^\text{fil}(B, \mu_B) = \frac{1}{2}\). We obtain

\[
\chi^\text{fil}(A, \mu_A) = \chi^\text{fil}(P^{-1}(b_0), \mu) \chi^\text{fil}(B, \mu_B)
\]

for any \(\mathbb{N}\)-filtration \(\mu\) of \(P^{-1}(b_0)\).

In fact, the category \(A\) is the barycentric subdivision of \(\Gamma\) in Example 4.12 and the category \(B\) is that of \(\mathbb{Z}_2\) (see [Nog11] and [Nog]). Hence, Theorem 4.9 of [Nog11] and Example 4.12 directly imply their Euler characteristics \(\chi^\text{fil}(A, \mu_A)\) and \(\chi^\text{fil}(B, \mu_B)\).

**References**


REMARKS ON THE DIGESTIVE SYSTEM.

1. Introduction

1.1. The problem we study

1.2. History

1.3. On characters of $T(\mathcal{O})$

1.4. Satake isomorphisms

1.5. Central families

1.6. Further directions

1.7. Acknowledgements

2. Parabolic, strongly parabolic, and easy characters

2.1. Conventions

2.2. $W$-invariant rational characters

2.3. Easy characters

2.4. Extendable characters

2.5. Comparison between easy and extendable

2.6. On parabolic characters

2.7. Proof of Theorem 3

3. Central families and Satake Isomorphisms

3.1. Recollections on decomposed subgroups

3.2. The subgroup $K$

3.3. Extension of $\hat{G}$

3.4. Proof of Theorem 11

3.5. Proof of Theorem 17

References
1. Introduction

1.1. The problem we study. Let \( F \) be a local non-Archimedean field with ring of integers \( \mathcal{O} \) and residue field \( \mathbb{F}_q \). Let \( G \) be a connected split reductive group over \( F \) with maximal split torus \( T \) and Weyl group \( W = N_G(T)/T \). Let \( \hat{T} \) denote the dual torus. Replacing \( G \) by an isomorphic group, we may, and henceforth will, assume that \( G \) is defined over \( \mathbb{Z} \) (see, for instance, [DG93], [Con11, §5] for the definition of a split reductive group over an arbitrary scheme). Then \( G(\mathcal{O}) \) is a maximal compact (open) subgroup of \( G(\mathbb{F}_q) \).

Let \( \mathcal{H}(G(\mathbb{F}_q), G(\mathcal{O})) \) denote the convolution algebra of compactly supported \( G(\mathcal{O}) \)-bi-invariant complex valued functions on \( G(\mathbb{F}_q) \). A celebrated theorem of Satake [Sat63] states that we have a canonical isomorphism of algebras

\[
\mathcal{H}(G(\mathbb{F}_q), G(\mathcal{O})) \cong \mathbb{C}[\hat{T}/W],
\]

We are interested in generalizing this isomorphism to nontrivial smooth characters \( \bar{\mu} : T(\mathcal{O}) \to \mathbb{C}^* \), as follows. Let \( W_{\bar{\mu}} \subseteq W \) denote the stabilizer of \( \bar{\mu} \) under the action of the Weyl group. Then it is natural to pose:

**Problem 1.** Construct a pair \( (K, \mu) \) consisting of a compact open subgroup \( T(\mathcal{O}) \subseteq K \subseteq G(\mathcal{O}) \) and a character \( \mu : K \to \mathbb{C}^* \) extending \( \bar{\mu} \), such that we have an isomorphism of algebras

\[
\mathcal{H}(G(\mathbb{F}_q), K, \mu) \cong \mathbb{C}[\hat{T}/W_{\bar{\mu}}],
\]

where \( \mathcal{H} \) is the convolution algebra of \( (K, \mu) \)-bi-invariant compactly supported functions on \( G(\mathbb{F}_q) \).

The Satake isomorphism provides a solution for the above problem for \( \bar{\mu} = 1 \). In this paper, we solve the above problem for a large class of characters of \( T(\mathcal{O}) \) which we call “strongly parabolic characters,” which are by definition characters such that \( W_{\bar{\mu}} \) is the Weyl group of a Levi subgroup \( L \subset G \), and moreover such that \( \bar{\mu} \) extends to \( L(\mathbb{F}_q) \). This appears to be the proper generality where the problem has a positive solution. Our construction of \( K \) is tied to \( L \). We think of the isomorphism \( \mathcal{H}(G(\mathbb{F}_q), K, \mu) \cong \mathbb{C}[\hat{T}/W_{\bar{\mu}}] \) as a Satake isomorphism for the (possibly) ramified character \( \bar{\mu} \). Therefore, we call these isomorphisms *ramified Satake isomorphisms*. For characters that are not strongly parabolic, we do not have a reason to expect a positive answer to Problem 1.

1.2. History. Following Satake, R. Howe studied Problem 1 for \( G = \text{GL}_N \) [How73]. Via an isomorphism which he called the \( \bar{\mu} \)-spherical Fourier transform, he completely solved the problem for the general linear group. Howe’s paper went largely unnoticed; however, several cases of Problem 1 were subsequently solved using other methods.

In [Ber84], [Ber92], Bernstein constructed a decomposition of the category of representations of \( G(\mathbb{F}_q) \) using the theory of Bernstein center. Each block admits a projective generator. In particular, for every character \( \bar{\mu} : T(\mathcal{O}) \to \mathbb{C}^* \), one has a block of representations of \( G(\mathbb{F}_q) \), which we denote by \( \mathcal{B}_{\bar{\mu}}(G) \). Bernstein proved that the center of \( \mathcal{B}_{\bar{\mu}}(G) \) is canonically isomorphic to \( \mathbb{C}[\hat{T}/W_{\bar{\mu}}] \);
see, for instance, [Roc09, Theorem 1.9.1.1]. Moreover, he gave an explicit description of a projective generator for each of these blocks; see the RHS of (1.8). When the character \( \bar{\mu} \) is regular; i.e., \( W_{\bar{\mu}} = \{1\} \), then the center is \( C[T] \), and it identifies canonically with the endomorphism ring of Bernstein’s generator. In a fundamental paper [BK98], Bushnell and Kutzko organized the study of representations of \( G(F) \) via compact open subgroups into the theory of types. Namely, they proposed that one should be able to obtain a projective generator for every block of representations of \( G(F) \) by inducing a finite dimensional representation from a compact open subgroup. The pair of the compact open subgroup and its finite dimensional representation, up to a certain equivalence, is called the type. In [BK99] and [BK93], they explicitly construct types for every block of representations of \( \text{GL}_N \). In particular, they construct projective generators for the principal series blocks \( \mathcal{R}_{\bar{\mu}}(\text{GL}_N) \). When the character \( \bar{\mu} \) is regular, their construction provides a pair \((K, \mu)\) satisfying the requirement of Problem 1. We note, however, that Bushnell and Kutzko’s construction of types is technically involved, since they consider all blocks (not merely the principal series blocks); in particular, we were not able to locate exactly where in their papers they construct types for the principal series blocks of \( \text{GL}_N \).

Finally, Roche [Roc98] constructed types for principal series representations of arbitrary reductive groups in good characteristics (which excluded in particular those listed in Convention). In the case that \( \bar{\mu} \) is regular, the type itself is a pair \((K, \mu)\) satisfying the conditions of Problem 1. In this paper, we build on the methods introduced by Bushnell and Kutzko and Roche, and solve the problem for all strongly parabolic characters. We make use of Roche’s type in order to construct a pair \((K, \mu)\) satisfying the conditions of Problem 1.

1.3. On characters of \( T(O) \). A significant part of this paper, which may be of independent interest, is devoted to defining and studying certain smooth characters of \( T(O) \). Recall that a subgroup \( W' \subseteq W \) is parabolic if it is generated by simple reflections. The Levi subgroup \( L \) associated to \( W' \) is the subgroup generated by \( T \) and the the simple root subgroups corresponding to the simple reflections in \( W' \) along with their negatives.

Definition 2. Let \( \bar{\mu} : T(O) \to C^* \) be a smooth character.

(i) \( \bar{\mu} \) is parabolic if the stabilizer \( \text{Stab}_W(\bar{\mu}) \) of \( \bar{\mu} \) in \( W \) is a parabolic subgroup.

(ii) \( \bar{\mu} \) is strongly parabolic if it is parabolic with Levi \( L \) and extends to a character of \( L(F) \).

(iii) \( \bar{\mu} \) is easy if it is parabolic and it extends to a character of \( L(F) \) which is trivial on \([L,L](F)\).

It follows immediately from the definition that the trivial character and all regular characters are easy. Moreover, it is clear that

\[
\text{easy} \implies \text{strongly parabolic} \implies \text{parabolic}.
\]

The reverse implications can all fail; see Examples 19 and 28.
To state our results regarding these characters, we need some notation. Let \( \Phi \) denote the set of roots of \( G \). Let \( X, X^\vee, Q, Q^\vee \) denote the character, cocharacter, root and coroot lattices of \( G \), respectively. Below we will frequently impose the conditions that either \( X/Q \) is free or \( X^\vee/Q^\vee \) is free (or both). We remark that \( X^\vee/Q^\vee \) being free is equivalent to \( [G(\mathbb{C}), G(\mathbb{C})] \) being simply-connected, while \( X/Q \) being free is equivalent to the statement that \( G(\mathbb{C}) \) has connected center.\(^1\)

In the following, given a coweight \( \lambda \in X^\vee \), we view \( \lambda \) as a morphism \( \mathbb{G}_m \to T \), and for every ring \( R \) (e.g., \( R = \mathcal{O} \)), we abusively use \( \lambda \) also to denote the morphism \( \mathbb{G}_m(\mathbb{R}) = \mathbb{R} \to T(\mathbb{R}) \). Thus, given a character \( \bar{\mu} \) of \( T(\mathcal{O}) \), we obtain a character \( \bar{\mu} \circ \lambda \) of \( \mathcal{O}^\times \). We will particularly use this when \( \lambda = \alpha^\vee \) is a coroot.

**Theorem 3.** Let \( \bar{\mu} : T(\mathcal{O}) \to \mathbb{C}^\times \) be a smooth character.

(i) \( \bar{\mu} \) is easy if and only if it is parabolic and can be written as a product \( \chi_1 \cdots \chi_l \), where each \( \chi_i \) is a character \( T(\mathcal{O}) \to \mathbb{C}^\times \) which is a composition of a \( W_\bar{\mu} \)-invariant rational character \( T(\mathcal{O}) \to \mathcal{O}^\times \) and a smooth character \( \mathcal{O}^\times \to \mathbb{C}^\times \).

(ii) The following are equivalent:
(a) \( \bar{\mu} \) is strongly parabolic;
(b) \( \bar{\mu} \circ \alpha^\vee |_{\mathcal{O}^\times} = 1 \), \( \forall \alpha \in \Phi_L \).
Moreover, if \( q > 2 \), then these are also equivalent to:
(c) \( \bar{\mu} \) extends to a character of \( L(\mathcal{O}) \).

(iii) If \( X/Q \) is free or \( \Phi \) has no factors of type \( A_1 \) or \( C_n \), then every parabolic character of \( T(\mathcal{O}) \) is strongly parabolic.

(iv) If \( X^\vee/Q^\vee \) is free, then every strongly parabolic character of \( T(\mathcal{O}) \) is easy.

(v) If \( \Phi \) is simply-laced and \( X/Q \) is free, then every character of \( T(\mathcal{O}) \) is strongly parabolic.

Section 2 is devoted to the proof of the above theorem.

We now indicate what the above theorem implies for characters of various groups. By \( G/Z \) we mean \( G/Z(G) \). The letter \( N \) denotes a positive integer.

We let \( E_n, n = 6, 7, 8 \) (resp. \( F_4 \) and \( G_2 \)) denote the split reductive group whose associated complex group is the connected, simply-connected, simple group of type \( E_n \) (resp. \( F_4 \) and \( G_2 \)).

---

\(^1\)This follows from the fact that if \( G \) is a (connected split) semisimple group, then \( X/Q \) equals the dual of \( Z(G(\mathbb{C})) \) and \( X^\vee/Q^\vee \) equals the dual of \( \pi_1(G(\mathbb{C})) \); see, for example [Con12, Example 6.7]. For example, for \( \text{SL}_2 \), we have \( X \cong \mathbb{Z}, Q = 2X, \) and \( Q^\vee = X^\vee \cong \mathbb{Z} \).
### Reductive group

<table>
<thead>
<tr>
<th>Properties</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>GL(_N), E(_8)</td>
<td>simply-laced, (X/Q) and (X')/(Q') free</td>
</tr>
<tr>
<td>PGL(_N), GO(_2)/(Z), SO(_2)/(Z), (E_6)/(Z), (E_7)/(Z)</td>
<td>simply-laced and (X/Q) free</td>
</tr>
<tr>
<td>SL(_N) ((N ≥ 3)), GSp(_2)/(Z), Spin(_N), (E_N) ((N ≥ 6)), F(_4), G(_2)</td>
<td>(X')/(Q') free, and hypothesis of (iii)</td>
</tr>
<tr>
<td>Sp(_2)/(Z), GO(_N), SO(_N)</td>
<td>hypothesis of (iii)</td>
</tr>
</tbody>
</table>

**Remark 4.** Let \(G\) be a (connected) algebraic group over a field \(k\). Let \(\bar{k}\) denote an algebraic closure of \(k\). Then \(G\) is said to be *easy* if every \(g \in G(\bar{k})\) is in the neutral connected component of its centralizer in \(G \otimes_k \bar{k}\). This definition is due to V. Drinfeld. Based on the discussion in, e.g., [Boy10, §2.2], there appears to be a relationship between Drinfeld’s notion of easy and ours, when \(k\) has characteristic zero. Namely, here we show that, if \([G,G]\) is simply connected and \(Z(G)\) is connected, then every parabolic character is easy (and the parabolic assumption is not needed in the simply-laced case); in [Boy10, §2.2] it is asserted, without proof, that these two assumptions are equivalent (over a field of characteristic zero) to \(G\) being easy in Drinfeld’s sense.

**Remark 5.** To every character \(\bar{\mu} : T(O) \to C^*\), Roche [Roc98, §8] associated a possibly disconnected split reductive group \(\bar{H} = \bar{H}_{\bar{\mu}}\) over \(F\). The connected component of \(\bar{H}\) is an endoscopy group for \(G\). It follows from Theorem 3.(ii) that strongly parabolic characters are exactly those characters for which \(\bar{H}\) is the Levi of a parabolic of \(G\) (and in particular connected). In more detail, by [Roc98, Definition 6.1], the coroots \(\alpha'\) of the connected component \(H\) of the identity of \(\bar{H}\) (as a complex reductive group) are exactly those for which \(\bar{\mu} \circ \alpha'|_{O^*} = 1\), and by [Roc98, Lemma 8.1.(i) ], the stabilizer of \(\bar{\mu}\) equals the Weyl group of \(\bar{H}\) (and is not bigger) if and only if \(\bar{H} = H\). Then, we conclude because the Weyl group of \(H\) is a parabolic subgroup of the Weyl group of \(G\) if and only if \(H\) is a Levi subgroup of \(G\) (i.e., its roots form a closed root subsystem of those of \(G\)).

### 1.4. Satake isomorphisms

As before, \(G\) denotes a connected split reductive group over \(\mathbb{Z}\). Let \(F\) be a local non-Archimidean field with ring of integers \(O\). We impose the following restrictions on the residue characteristic of \(F\).
Convention 6. For every irreducible direct factor of the root system of \( G \), we assume that the residue characteristic of \( F \) is not one of the following primes:

\[
\begin{array}{|c|c|}
\hline
\text{Root system} & \text{Excluded primes} \\
\hline
B_n, C_n, D_n & \{2\} \\
F_4, G_2, E_6, E_7 & \{2, 3\} \\
E_8 & \{2, 3, 5\} \\
\hline
\end{array}
\]

Theorem 7. For every strongly parabolic character \( \bar{\mu} : T(O) \to \mathbb{C}^* \), there exists a compact open subgroup \( K < G(O) \) and an extension \( \mu : K \to \mathbb{C}^* \) such that

\[
(1.6) \quad \mathcal{H}(G(F), K, \mu) \cong \mathbb{C}[\bar{T}/W_{\bar{\mu}}]
\]

As mentioned above, in the case of \( G = \text{GL}_N \), the above theorem is due to Howe [How73], and if \( \bar{\mu} \) is regular, then the above theorem follows by combining results of Bernstein [Ber84], [Ber92], Bushnell-Kutzko [BK98, BK99] and Roche [Roc98]. As far as we know, the generalization to strongly parabolic characters is new.

Example 8. Let \( G = \text{GL}_3 \) and let \( T(O) \simeq (O^*)^3 \) denote the group of diagonal matrices. Write \( \bar{\mu} = (\bar{\mu}_1, \bar{\mu}_2, \bar{\mu}_3) \) where each \( \bar{\mu}_i \) is a smooth character \( O^* \to \mathbb{C}^* \). Suppose \( \bar{\mu}_1 = \bar{\mu}_2 \) and that the conductor \( \text{cond}(\bar{\mu}_1/\bar{\mu}_3) \) equals \( n \geq 2 \). (The conductor of a character \( \chi : O^* \to \mathbb{C}^* \) is the smallest positive integer \( c \) for which \( \chi(1 + p^c) = \{1\} \).) If we follow Howe’s approach, we would take

\[
K = \begin{pmatrix}
O & O & O \\
O & O & O \\
p^n & p^n & O
\end{pmatrix} \cap G(O).
\]

On the other hand, in the present article, following more closely the types of [Roc98], we take instead

\[
K = \begin{pmatrix}
O & O & \mathcal{O}_{p^{[\frac{n}{2}]}} \\
O & \mathcal{O}_{p^{[\frac{n}{2}]}} & \mathcal{O}_{p^{[\frac{n}{2}]}} \\
p^{[\frac{n}{2}]} & \mathcal{O}_{p^{[\frac{n}{2}]}} & \mathcal{O}
\end{pmatrix} \cap G(O).
\]

In both cases, \( \bar{\mu} \) extends to a character \( \mu : K \to \mathbb{C}^* \) and one has an isomorphism \( \mathcal{H}(G(F), K, \mu) \cong \mathbb{C}[\bar{T}] \). This example shows that the subgroup \( K \) of Theorem 7 is not necessarily unique.

To prove Theorem 7 we use Roche’s result on types for principal series representations. Given an arbitrary smooth character \( \bar{\mu} : T(O) \to \mathbb{C}^* \), Roche [Roc98] constructed a compact open subgroup \( J \subset G(F) \) (which depends on the choice of a Borel \( B \)) and an extension \( \mu^J : J \to \mathbb{C}^* \) such that the compactly induced representation

\[
(1.7) \quad \mathcal{W} := \text{ind}_J^{G(F)} \mu^J
\]

is a progenerator for the principal series Bernstein block of \( G \) defined by \( \bar{\mu} \).

More precisely, a combination of results of Bushnell and Kutzko, Dat, and...
Roche implies that in this situation, one has an explicit isomorphism of $G(F)$-modules
\[(1.8) \quad \Psi : \mathcal{W} \xrightarrow{\cong} \Pi := \frac{G(F)}{B(F)} \left( \text{ind}_{T(O)}^{T(F)} \hat{\mu} \right).\]

Here, $\hat{\mu}$ denotes the functor of parabolic induction. See §3.4 for the explicit description of $\Psi$. Note that the endomorphism algebra of $\mathcal{W}$ is canonically isomorphic with $H(G(F), J, \mu^J)$.

Now suppose the character $\bar{\mu}$ is strongly parabolic. Let $L$ denote the corresponding Levi and let $\mu_L^L : L(F) \to \mathbb{C}^*$ denote an extension of $\bar{\mu}$ to $L(F)$. Let $\mu^L := \mu_{L(O)} := \mu^L|_{L(O)}$ denote its restriction to $L(O)$. We prove that $K = JL(O)$ is a subgroup of $G(F)$. Moreover, we show that there exists a canonical character $\mu : K \to \mathbb{C}^*$ which extends $\mu^J$ and $\mu^L$. Theorem 7 states that the Hecke algebra $H(G(F), K, \mu)$, consisting of compactly supported $(K, \mu)$-bi-invariant functions on $G(F)$, is isomorphic to $\mathbb{C}[\hat{T}/W_{\mu}]$. To prove this result, we realize $H(G(F), K, \mu)$ as an endomorphism ring of a family of principal series representations, which we call a central family.

1.5. Central families. In order to prove Theorem 7 we will introduce a certain representation attached to strongly parabolic characters.

**Definition 9.** Let $\bar{\mu}$ be a strongly parabolic character with the corresponding Levi $L$. Let $K = JL(O)$ denote the corresponding compact open subgroup. The central family of principal series representations of $G$ attached to $\bar{\mu}$ is defined by
\[(1.9) \quad \mathcal{V} := \text{ind}_{K}^{G(F)} \mu.\]

We will now give an alternative description of $\mathcal{V}$. Let $P \supseteq B$ be a parabolic subgroup whose Levi is isomorphic to $L$. Let
\[(1.10) \quad \Theta := \text{ind}_{P(F)}^{G(F)} \left( \text{ind}_{L(O)}^{L(F)} \mu^L \right).\]

**Theorem 10.** Under the assumptions of Theorem 7 we have a canonical isomorphism of $G(F)$-modules $\mathcal{V} \xrightarrow{\cong} \Theta$.

We prove the above theorem by identifying $\mathcal{V}$ and $\Theta$ with submodules of $\mathcal{W}$ and $\Pi$, respectively. Then, using the explicit description of $\Psi$ in (1.8), we show that $\Psi|_{\mathcal{V}} : \mathcal{V} \to \Pi$ defines an isomorphism onto $\Theta$. On the other hand, the endomorphism ring of $\mathcal{V}$ identifies with $\mathcal{H}(G(F), K, \mu)$. Thus, to prove Theorem 7 we need to compute the endomorphism algebra of $\Theta$. To this end, we will use a theorem of Roche [Roc02] on parabolic induction of Bernstein blocks.

2Note that here and in §1.10, it does not matter if we use normalized or unnormalized parabolic induction since the representation being induced is isomorphic to its twist by any unramified character.
Remark 11. (i) As mentioned above, in this paper, we construct the pair 
\((K, \mu)\) satisfying requirement of Problem 1 by using Roche’s pair 
\((J, \mu_J)\). In this case, the subgroup \(K\) depends only on the kernel of \(\bar{\mu}\); that is, if \(\ker(\bar{\mu}) = \ker(\bar{\mu}')\) then \(K_{\bar{\mu}} = K_{\bar{\mu}'}\). In fact, it only depends on the conductors of the restrictions of \(\bar{\mu}\) to the coroot subgroups (i.e., the minimal \(c_\alpha \geq 1\) such that \(\bar{\mu}|_{\alpha (1+p^{c_\alpha})}\) is trivial) together with the collection of roots \(\alpha\) such that the entire restriction \(\bar{\mu}|_{\alpha (O)}\) is trivial. This follows immediately from the construction of \(J\); see \(\S 3.2\).

(ii) The pair \((J, \mu_J)\) is a type for the Bernstein block \(\mathcal{B}_\mu(G)\). Types for Bernstein blocks are not, however, necessarily unique. Therefore, it is natural to wonder if our construction could work using a different type \((J', \mu_J')\). In the case \(G = \text{GL}_N\), this is true in view of the results of [How73], as we observed (for \(N = 3\)) in Example 8. We do not, however, pursue this question in the current text.

(iii) Note that \(V\) is a submodule of \(W\) and the latter is a progenerator for the principal series block corresponding to \(\bar{\mu}\). According to Theorem 7, the endomorphism ring \(\mathcal{H}\) of this family identifies with the center of the corresponding Bernstein block (which is isomorphic to the center of \(\mathcal{H}(G(F), J, \mu_J)\), and hence isomorphic to \(\mathbb{C}[T/\mathfrak{W}_{\bar{\mu}}]\); cf. \(\S 1.2\)). Moreover, one can show that, for generic maximal ideals \(m \subset \mathcal{H}\), the \(G(F)\)-module \(V/mV\) is an irreducible principal series representation. (We will neither prove nor use the last statement.)

1.6. Further directions. The proof of Theorem 7 given in this paper is rather indirect; moreover, it relies on nontrivial results of Bernstein, Bushnell and Kutzko, Roche, and Dat. In a forthcoming paper [KS], we hope to give a direct proof of this theorem by writing an explicit support preserving isomorphism \(\mathcal{H}(L(F), L(O), \mu_L) \cong \mathcal{H}(G(F), K, \mu)\). In other words, we hope to prove Theorem 7 using combinatorics and the classical Satake isomorphism. This proof should also make clear the geometric nature of the group \(K\) and some of its double cosets in \(G(F)\); in particular, we expect that it will help with the geometrization program (see below).

In Definition 9, for strongly parabolic characters of the compact maximal torus, we constructed “central families”. The endomorphism ring of the central family identifies canonically with the center of the block defined by the character; moreover, generic irreducible representations in the block appear with multiplicity one in the central family. It would be interesting to find analogous central families for other Bernstein blocks.

It is well-known that the Satake isomorphism allows one to realize the local unramified Langlands correspondence. In more detail, let \(G\) denote the complex reductive group which is the Langlands dual of \(G\). Using the classical Satake isomorphism \(\mathcal{H} \cong \mathcal{H}(G(F), K, \mu)\), one can show that we have a bijection

\[
\text{unramified irreducible representations of } G(F) \leftrightarrow \text{characters of } \mathcal{H}
\]
Combining this with the bijections

\[
\text{characters of } \mathcal{H} \leftrightarrow \text{points of } \hat{T}/W \leftrightarrow \text{semisimple conjugacy classes in } \hat{G}
\]

we obtain a bijection between unramified representations of \( G(F) \) and semisimple conjugacy classes in \( \hat{G} \). It would be interesting to study the role of the ramified Satake isomorphisms (i.e., the ones given by Theorem 7) in the local Langlands program.

In [HR10], a version of the Satake isomorphism for non-split groups is proved. On the other hand, there is also now a Satake isomorphism in characteristic \( p \); see [Her11]. We expect that there is also a version of Theorem 7 for non-split groups and one in characteristic \( p \).

Finally, we expect that there is a geometric version of Theorem 7. The geometric version of the usual Satake isomorphism is proved by Mirkovic and Vilonen [MV07], completing a project initiated by Lusztig, Beilinson and Drinfeld, and Ginzburg. In the case of regular characters; i.e., in the case that the stabilizer of the character in the Weyl group is trivial, a geometric version of Theorem 7 is proved in [KS11]. In [KS11 §1.4], we conjectured the theorems proved in this article; moreover, we formulated precise conjectures for geometrizing these results. We hope to return to this theme in future work.

1.7. ACKNOWLEDGEMENTS. We would like to thank A. Roche for very helpful email correspondence. He sketched proofs of several technical results for us; moreover, he brought to our attention the references [Roc02], [Dat99], and [Bjo05]. We would like to thank J. Adler for reading an earlier draft and several useful discussions. We also thank L. Spice for helpful conversations, and B. Conrad for answering questions on reductive groups (in particular pointing out an elegant proof of Proposition 23(i) via the construction of [24.1]). We thank the Max Planck and Hausdorff Institutes for Mathematics in Bonn for their hospitality. Finally, we thank the anonymous referee for many helpful suggestions.

2. PARABOLIC, STRONGLY PARABOLIC, AND EASY CHARACTERS

2.1. CONVENTIONS. Let \( F \) be a local field with ring of integers \( \mathcal{O} \), unique maximal ideal \( p \), residue field \( \mathbb{F}_p \), and uniformizer \( t \). Let \( G \) be a connected split reductive group over \( Z \) with split maximal torus \( T \). Let \( W = N_G(T)/T \) denote the Weyl group.

Let \( \Phi = \Phi_G \) denote the roots of \( G \) (with respect to \( T \)). For \( \alpha \) a root in \( \Phi \), we write \( \alpha^\vee \) for the corresponding coroot. Let \( X = \text{Hom}(T, \mathbb{G}_m) \) and \( X^\vee = \text{Hom}(\mathbb{G}_m, T) \) denote the the character and cocharacter lattices, respectively.

Let \( Q \subseteq X \) be the root lattice, and let \( Q^\vee \subseteq X^\vee \) denote the coroot lattice. Let \( (Q^\vee)_{\text{sat}} \) be the saturation of \( Q^\vee \) in \( X^\vee \), i.e.,

\[
(Q^\vee)_{\text{sat}} = \{ \lambda \in X^\vee \mid m \cdot \lambda \in Q^\vee, \text{some } m \in \mathbb{Z} \}.
\]

By definition \( X^\vee/(Q^\vee)_{\text{sat}} \) is a torsion free abelian group. To an element \( \lambda \in X^\vee \), we associate \( t^\lambda = \lambda(t) \in T(F) \).
For every $\alpha \in \Phi$, let $u_\alpha : \mathbb{G}_a \to G$ be the one-parameter root subgroup, where $\mathbb{G}_a$ is the additive group. We assume these root subgroups satisfy the conditions specified in [Roc98] §2. Let $U_\alpha < G$ be the image of $u_\alpha$. For all $i \in \mathbb{Z}$, let $U_{\alpha,i} = u_\alpha(p^i) < G(F)$. In particular, $U_{\alpha,0} = u_\alpha(O)$.

Let $H$ and $K$ be topological groups and suppose $H < K$. Let $\chi : H \to \mathbb{C}^\times$ be a character of $H$. We write $\text{ind}_H^K\chi$ for the space of left $(H, \chi)$-invariant relatively compactly supported functions on $K$; that is, those functions $f : K \to \mathbb{C}$ whose support has compact image in $K/H$ and satisfy $f(hk) = \chi(h)f(k)$ for all $h \in H$ and $k \in K$. The group $K$ acts on this space by right translation.

2.2. $W$-invariant rational characters. We start this section with a general lemma which we will repeatedly use below.

**Lemma 12.** Let $H$ be a group and $K < H$ a subgroup. Then a character $\chi : K \to \mathbb{C}^\times$ extends to a character of $H$ if and only if $\chi|_{K\cap[H,H]}$ is trivial. The same is true if $H$ is an l-group (i.e., a locally compact totally disconnected Hausdorff topological group), $K$ is a closed subgroup, and $\chi$ is smooth.

**Proof.** It is clear that the assumption that $\chi$ be trivial on $K \cap [H,H]$ is necessary. Conversely, if this is true, extending the character is the same as extending the induced character of $K/(K \cap [H,H])$ to $H/(H,H)$. Therefore, all the statements of the lemma reduce to the case that $H$ is commutative.

Then, the statement that any character of a subgroup of an abstract (discrete) abelian group extends to the entire group follows from the fact that $\mathbb{C}^\times$ is divisible, and hence injective.

For the locally compact analogue, write $\mathbb{C}^\times \cong S^1 \times \mathbb{R}_{>0}$. For characters to $S^1$, the statement follows from Pontryagin duality. For $\mathbb{R}_{>0}$, note first that, if $H$ is compact, then there are no nontrivial continuous characters to $\mathbb{R}_{>0}$. As an l-group always contains a compact open subgroup, this reduces the problem to the case $H$ is discrete, where it follows as in the previous paragraph, since $\mathbb{R}_{>0}$ is divisible, and hence injective, as a discrete abelian group.

For the algebraic analogue, i.e., where $H$ and $K$ are connected split tori, the statement follows because applying $\text{Hom}(-, \mathbb{G}_m)$ to a short exact sequence $1 \to K \to H \to H/K \to 1$ of split tori is well-known to be an equivalence of short exact sequences of split tori with that of their weight lattices. Hence, the restriction map from characters of $H$ to characters of $K$ is surjective. \(\square\)

**Lemma 13.** Let $G$ be a connected split reductive algebraic group over $\mathbb{Z}$ with split maximal torus $T$. Let $\chi : T \to \mathbb{G}_m$ be a rational character. The following are equivalent:

(1) $\chi$ is trivial on $T \cap [G,G]$.
(2) $\chi$ extends to a character $G \to \mathbb{G}_m$.

---

1. We don’t need to assume that $H$ is totally disconnected, if we use the fact [HIM08] Corollary 7.54 that every locally compact Hausdorff topological group contains a compact subgroup $H'$ such that the quotient $H/H'$ is isomorphic to $\mathbb{R}^n \times D$ for a discrete group $D$. 

Documenta Mathematica 18 (2013) 1275–1300
(3) $\chi$ is $W$-invariant;
(4) $\chi \circ \alpha^\vee : \mathbb{G}_m \to \mathbb{G}_m$ is trivial, for every $\alpha \in \Phi$.

Proof. Lemma [12] implies immediately that (1) $\implies$ (2). Next, it is clear that $[N_G(T), T] \leq [G, G] \cap T$; therefore, if we restrict a character of $G$ to $T$, we obtain a character which is invariant under the conjugation action of $N_T(G)$. This proves (2) $\implies$ (3). Next, suppose $\chi$ is $W$-invariant. Then

$\chi \circ \alpha^\vee = (s_m \chi) \circ \alpha^\vee = \chi \circ (-\alpha^\vee) = (\chi \circ \alpha)^{-1}$

It follows that $(\chi \circ \alpha^\vee)^2 = 1$. Since $G_m$ has no nontrivial character of order 2, it follows that $\chi \circ \alpha^\vee = 1$. Hence, (3) $\implies$ (4). For the final implication, we use the canonical identification

$T \cap [G, G] \cong G_m \otimes \mathbb{Z}(Q^\vee)_{\text{sat}}.$

By the notation on the RHS we mean the group subscheme of $T$ whose $R$ points equals $R^\vee \otimes \mathbb{Z}(Q^\vee)_{\text{sat}}$, where $R$ is a ring over $k$. Now if $\chi \circ \alpha^\vee$ is trivial for every $\alpha \in \Phi$, then $\chi$ is trivial on $T \cap [G, G]$. This proves (4) $\implies$ (1). □

Remark 14. It follows from the above lemma that the group of characters of $T$ which satisfy the above equivalent conditions is canonically isomorphic to $\text{Hom}(T/(T \cap [G, G]), G_m) = X^W \cong \text{Hom}(X^\vee/Q^\vee, \mathbb{Z}) \cong \text{Hom}(X^\vee/(Q^\vee)_{\text{sat}}, \mathbb{Z})$.

The last isomorphism follows from the following: the quotient $X^\vee/(Q^\vee)_{\text{sat}}$ splits, since $X^\vee/(Q^\vee)_{\text{sat}}$ is free, and the resulting pullback maps $\text{Hom}(X^\vee/Q^\vee, \mathbb{Z}) \hookrightarrow \text{Hom}(X^\vee/(Q^\vee)_{\text{sat}}, \mathbb{Z})$ are inverse to each other since the quotient $X^\vee/Q^\vee \to X^\vee/(Q^\vee)_{\text{sat}}$ has finite kernel and $\mathbb{Z}$ is torsion-free. (More generally, for any finite-kernel quotient of finitely-generated abelian groups, the pullback map on $\text{Hom}(-, \mathbb{Z})$ is an isomorphism.)

2.3. Easy characters. Let $G$ be a connected split reductive group defined over $\mathbb{Z}$. Let $\text{Hom}_{\text{sm}}(O^\vee, \mathbb{C}^\times)$ denote the group of smooth characters $O^\vee \to \mathbb{C}^\times$.

Proposition 15. The following conditions are equivalent for a smooth character $\tilde{\mu} : T(O) \to \mathbb{C}^\times$:

(i) The restriction $\tilde{\mu}|_{[G, G] \cap T(O)}$ is trivial;

(ii) The character $\tilde{\mu}$ is a product of compositions of $W$-invariant rational characters $T(O) \to O^\times$ with smooth characters $O^\times \to \mathbb{C}^\times$.

Remark 16. The same statement and proof holds when $O$ is replaced by any (topological) ring.

Definition 17. A smooth character $\tilde{\mu} : T(O) \to \mathbb{C}^\times$ is easy with respect to $G$ if the equivalent conditions of Proposition 15 are satisfied.

Note that, in particular, any easy character with respect to $G$ must be $W$-invariant.

---

4 For a proof of this statement over an algebraically closed field see, for instance, [DM91], §0.20. Note that over an algebraically closed field, one does not need to use $(Q^\vee)_{\text{sat}}$: more precisely, we have $k^\times \otimes \mathbb{Z} Q^\vee = T(k) \cap [G, G](k) = (T \cap [G, G])(k)$. 

---

Documenta Mathematica 18 (2013) 1275–1300
Remark 18. By Lemma 13 and Proposition 15, the group of easy characters of $T(O)$ identifies canonically with
\[
\Hom_{\text{sm}} ( (T/(T \cap [G,G]))(O), \mathbb{C}^*) = \Hom(X^v/(Q^v)_{\text{sat}}, \mathbb{Z}) \otimes_{\mathbb{Z}} \Hom_{\text{sm}}(O^* \cap \mathbb{C}^*, \mathbb{C}^*). \]
The last isomorphism follows from the fact that $X^v/(Q^v)_{\text{sat}}$ is free.

Proof of Proposition 16. The implication (ii) \Rightarrow (i) is immediate from Lemma 13. For the reverse implication, note that by assumption $\bar{\mu}$ is a character of $T(O) = (\mathbb{G}_m \otimes \mathbb{C}^n)(O)$ which is trivial on $([G,G] \cap T)(O) = (\mathbb{G}_m \otimes \mathbb{Z}(Q^v)_{\text{sat}})(O)$. Therefore $\bar{\mu}$ is canonically a character of
\[
\left( \mathbb{G}_m \otimes_{\mathbb{Z}} \mathbb{C}^n \right)(O)/((\mathbb{G}_m \otimes_{\mathbb{Z}} (Q^v)_{\text{sat}})(O) = ((\mathbb{G}_m \otimes_{\mathbb{Z}} \mathbb{C}^n)/((\mathbb{G}_m \otimes_{\mathbb{Z}} (Q^v)_{\text{sat}}))(O) = (\mathbb{G}_m \otimes_{\mathbb{Z}} \mathbb{C}^n)/(Q^v)_{\text{sat}}) \otimes_{\mathbb{Z}} \mathbb{C}^n.
\]
We conclude that $\bar{\mu}$ is a product of compositions of (smooth) characters of $\mathbb{G}_m$ with rational characters $\Hom(X^v/(Q^v)_{\text{sat}}, \mathbb{Z})$. By Remark 14, the group of such rational characters is canonically isomorphic to the sublattice $X^W$ of $W$-invariant rational characters. Therefore, $\bar{\mu}$ has the form claimed in Part (ii).

Note that, if $\bar{\mu}$ is easy, then Lemma 13 implies that $\bar{\mu}$ extends to a character of $G(F)$, and hence of $G(O)$. As the following example illustrates, the converse is not, in general, true.

Example 19. Let $G = PGL_2$. Then the determinant map $GL_2(F) \to F^*$ descends to a map $G(F) \to F^*/|F^*|^2 \cong \{ \pm 1 \} \times \mathbb{Z}/|2\mathbb{Z}|^\times$. Take the composition and the further quotient by the second factor, and view it as a character $G(F) \to \mathbb{C}^*$ (which is trivial on $t\mathbb{C}^\times$). The restriction of this character to $T(O)$ is non-trivial, even though there are no nonzero $W$-invariant rational characters (and hence no nontrivial easy characters).

 Nonetheless, in the next subsection, we give a combinatorial description of all characters of $T(O)$ which extend to characters of $G(F)$, similar to the description of easy characters above.

2.4. Extendable characters.

Proposition 20. The following conditions are equivalent for a smooth character $\bar{\mu} : T(O) \to \mathbb{C}^*$:

(a) $\bar{\mu}$ extends to a character of $G(F)$;
(b) For all $\alpha \in \Phi$, we have
\[
(2.4) \quad \bar{\mu} \circ \alpha^\vee|_{\mathbb{C}^*} = 1.
\]
If in addition $q > 2$, then these are also equivalent to
(c) $\bar{\mu}$ extends to a character of $G(O)$.
Definition 21. A smooth character \( \tilde{\mu} : T(\mathcal{O}) \to \mathbb{C}^* \) is said to be extendable (to \( G(F) \)) if it satisfies the equivalent conditions of the above proposition.

Note, in particular, that any extendable character to \( G(F) \) must be \( W \)-invariant (as is easy to verify from either of the conditions above: for (a), see the proof of Corollary 27 and for (b), see (i') in the proof of Lemma 30).

We will not need this fact in this subsection.

Remark 22. (i) Let \( R \) be an arbitrary ring. Since the span \( \{ \alpha^\vee \}_{\alpha \in \Phi} \) is, by definition, the coroot lattice \( Q^\vee \subseteq X^\vee \), we have \( \langle \alpha^\vee(R^\vee) \rangle = Q^\vee \otimes \mathbb{Z} R \). Also note that \( T = X^\vee \otimes \mathbb{Z} G_m \); therefore, \( T(R) = X^\vee \otimes \mathbb{Z} R^\vee \).

(ii) It follows from the previous remark and the above proposition that the group of extendable (to \( G(F) \)) characters of \( T(\mathcal{O}) \) identifies with

\[
\text{Hom}_{sm}(T(\mathcal{O})/(\alpha^\vee(O^\vee))_{\alpha \in \Phi}, \mathbb{C}^*)^* = \text{Hom}_{sm}(X^\vee/(Q^\vee \otimes \mathbb{Z} O^\vee), \mathbb{C}^*).
\]

The last isomorphism holds because \( (X^\vee \otimes \mathbb{Z} O^\vee)/(Q^\vee \otimes \mathbb{Z} O^\vee) = X^\vee/Q^\vee \otimes \mathbb{Z} O^\vee \) (by definition of \( \otimes \), or by its right-exactness). Note that \( X^\vee/Q^\vee \otimes \mathbb{Z} O^\vee \) has a topology which is induced by the topology on \( O^\vee \). Therefore, we can speak of its smooth characters.

To prove Proposition 20, we will prove the following:

Proposition 23. (i) Let \( R \) be an arbitrary ring. Then, we have an inclusion of groups

\[
\text{Im}(\text{Hom}(\mathcal{O}/(\alpha^\vee(O^\vee))_{\alpha \in \Phi}, \mathbb{C}^*), \mathbb{C}^*))^* \subseteq [G(R), G(R)] \cap T(R) \subseteq \langle \alpha^\vee(R^\vee) \rangle_{\alpha \in \Phi}.
\]

(ii) If \( R = \mathcal{O} \) for \( q > 2 \) or \( R = F \) for arbitrary \( q \), then

\[
\text{Im}(\text{Hom}(\mathcal{O}/(\alpha^\vee(O^\vee))_{\alpha \in \Phi}, \mathbb{C}^*), \mathbb{C}^*))^* \supseteq [G(R), G(R)] \cap T(R) \supseteq \langle \alpha^\vee(R^\vee) \rangle_{\alpha \in \Phi}.
\]

Proof of Proposition 20. In view of Lemma 12, part (i) of Proposition 23 is equivalent to the statement that all characters of \( T(R) \) trivial on \( \{ \alpha^\vee(R^\vee) \}_{\alpha \in \Phi} \) extend to characters of \( G(R) \); the same holds in the context \( R \) is an \( l \)-group and we restrict to smooth characters (because, in both contexts, any nontrivial abelian group admits a nontrivial character). Thus, condition (b) of Proposition 20 implies condition (c). Moreover, note that every character of \( T(O) \) trivial on \( \{ \alpha^\vee(O) \}_{\alpha \in \Phi} \) extends to a character of \( T(F) \) trivial on \( \{ \alpha^\vee(F) \}_{\alpha \in \Phi} \); a unique such extension is given by requiring the character to be trivial on \( tX^\vee \).

Thus, condition (b) of Proposition 20 implies condition (a).

Similarly, part (ii) of Proposition 23 is equivalent to the statement that every character of \( G(R) \) in the cases mentioned must be trivial on \( \{ \alpha^\vee(R) \}_{\alpha \in \Phi} \). Thus, condition (a) of Proposition 20 implies condition (b), and also (c) implies (b).

Thus, Proposition 23 implies Proposition 20.

The rest of this subsection is devoted to proving Proposition 23. To this end, we need to recall some facts about the universal group cover of the commutator subgroup.
2.4.1. Universal cover of $[G,G]$. Let $\tilde{G}$ be the connected reductive algebraic group over $\mathbb{Z}$ such that $G(\mathbb{C})$ is the universal cover of $[G,G](\mathbb{C})$. More precisely, this is the group corresponding to the root datum $(\text{Hom}_\mathbb{Z}(Q',\mathbb{Z}),\Phi,Q',\Phi')$. By [DG93], cf. [Con11] §6, the category of split reductive groups over an arbitrary scheme is equivalent to the category of root data, so one can define $\tilde{G}$ as a split reductive group over $\mathbb{Z}$. Let $\pi$ denote the canonical morphism $\tilde{G} \to G$ of group schemes over $\mathbb{Z}$ (i.e., the composition of the canonical isogeny $\tilde{G} \to [G,G]$, cf. [Con11] Theorem 6.1.6(1)), and the inclusion $[G,G] \subseteq G$.

We claim that the commutator morphism $G \times G \to G$ factors through the map $\tilde{G} \to G$. Let $G^{\text{ad}} := G/Z(G)$, where $Z(G)$ is the center of $G$. This coincides with $(\tilde{G})^{\text{ad}} := \tilde{G}/Z(\tilde{G})$, i.e., the natural map $(\tilde{G})^{\text{ad}} \to G^{\text{ad}}$ is an isomorphism. The commutator map $\tilde{G} \times \tilde{G} \to \tilde{G}$ descends to $(\tilde{G})^{\text{ad}} \times (\tilde{G})^{\text{ad}} = G^{\text{ad}} \times G^{\text{ad}}$. Composing with the quotient $G \times G \to G^{\text{ad}} \times G^{\text{ad}}$, we obtain a morphism $G \times G \to \tilde{G}$ whose composition with $\tilde{G} \to G$ is evidently the commutator morphism, as desired. We note that the above argument that the commutator map lifts to $\tilde{G}$ also appears in [Lev95] §2.0.2. On the other hand, the first author of this paper also studied the group $\tilde{G}$ under the name of “true commutator” of $G$; see [Kam09] §B.2 and Corollary 4.4.

Next, let $T = \pi^{-1}(T)$. Then $\tilde{T}$ is a commutative subgroup of $\tilde{G}$ containing a maximal split torus. Therefore, $\tilde{T}$ is itself a maximal split torus.

**Lemma 24.** For every ring $R$, $\pi(\tilde{T}(R)) = \langle \alpha^\vee(R^\vee) \rangle_{\alpha \in \Phi}$.

**Proof.** The cocharacter lattice of $\tilde{G}$ equals its coroot lattice. Therefore, by Remark 22(i), we have

$$\tilde{T}(R) = X_{\tilde{G}} \otimes_{\mathbb{Z}} R^\times = Q^\vee_{\tilde{G}} \otimes_{\mathbb{Z}} R^\times = \langle \alpha^\vee(R^\vee) \rangle_{\alpha \in \Phi}$$

Now the set of roots for $\tilde{G}$ and $G$ coincide. If, abusively, by $\alpha^\vee$ we also denote the corresponding subgroup of $T$, then we see that $\pi(\tilde{T}(R)) = \langle \alpha^\vee(R^\vee) \rangle_{\alpha \in \Phi}$, as required. \qed

2.4.2. Proof of Proposition 23 Part (i) of the proposition follows easily from Lemma 24, since the commutator map factors through $\pi$, we have $[G(R),G(R)] \subseteq \pi(\tilde{G}(R))$; hence, $[G(R),G(R)] \cap T(R) \subseteq \pi(\tilde{T}(R))$. By Lemma 24, $\pi(\tilde{T}(R))$ equals the RHS of (2.5).

To prove part (ii) of the proposition, we first reduce to the case $G = \text{SL}_2$:

**Lemma 25.** For a fixed ring $R$, the inclusion (2.6) holds for all $G$ if and only if it holds for $G = \text{SL}_2$.

**Proof.** For each $\alpha \in \Phi$, we have $[G(R),G(R)] \supseteq \alpha^\vee(R^\vee)$ if this fact holds when $G$ is replaced by the centralizer of the root $\alpha : T \to \mathbb{G}_m$, which is split connected reductive of semisimple rank one (in fact, it is the subgroup generated by $T$ and the root subgroups $U_{\pm \alpha}$). So we can assume that $G$ has semisimple rank one. Then, the group $\tilde{G}$ is isomorphic to $\text{SL}_2$. In view of the morphism $\tilde{G}(R) \to G(R)$, it suffices to prove (2.7) for $\tilde{G} \cong \text{SL}_2$. So once we establish (2.7) for $\text{SL}_2$, it follows for all split reductive groups $G$. \qed
Thus part (ii) reduces to:

**Lemma 26.** The inclusion $\overset{2.6}{\subseteq}$ holds for $G = \text{SL}_2$ in the case that either $R = \mathcal{O}$ for $q > 2$ or $R$ is a field.

**Proof.** Let $f \in R^*$ and let $\alpha$ be the positive simple root. Then one can verify that

$$(2.7) \quad \left( 1 \begin{array}{c} 1 \\ f \\ \end{array} \right) \left( \begin{array}{cc} (f-1)/f & 0 \\ 1 & 1 \end{array} \right) \left( \begin{array}{c} 1 \\ 0 \\ \end{array} \right) = \left( \begin{array}{c} f \\ 0 \\ \end{array} \right).$$

We now consider the question of when the first and last matrices on the LHS are in $[G(R),G(R)]$. Generally, for $g \in R^*$,

$$(2.8) \quad \left[ \begin{array}{cc} 1 & 0 \\ x & 1 \end{array} \right] \left[ \begin{array}{cc} g & 0 \\ 0 & g^{-1} \end{array} \right] = \left( \begin{array}{cc} 1 \\ x(1-g^2) \\ 1 \end{array} \right).$$

Let $I_R := \{ 1 - g^2 \mid g \in R^* \} = \{ x(1-g^2) \mid x \in R, g \in R^* \}$ be the ideal of elements appearing in the lower-left entry of the final matrix. If this ideal is the unit ideal, then the LHS of $(2.7)$ is in $[G(R),G(R)]$, as desired. This is clearly true if $R$ is a field such that $|R| > 3$. It therefore also true if $R$ is a discrete valuation ring whose residue field has cardinality greater than 3 (then let $g$ be a unit of $R$ whose image $\bar{g}$ in the residue field has the property that $1 - \bar{g}^2$ is invertible, so that $1 - \bar{g}^2$ itself is invertible).

Therefore, we only need to show that, for $R = \mathcal{O}$ for $q = 3$, or $R = \mathbb{F}_q$ for $q \leq 3$, then $(2.6)$ holds.

First, if $R = \mathbb{F}_q$ and $q = 2$, there is nothing to show because now $T(\mathbb{F}_q)$ is trivial. If $R = \mathbb{F}_q$ for $q = 3$, then it is well known that $[G(\mathbb{F}_q),G(\mathbb{F}_q)]$ has index three in $G(\mathbb{F}_q)$; since $T(\mathbb{F}_q)$ has order two, it follows that $T(\mathbb{F}_q)$ must be in the kernel of the abelianization map $G(\mathbb{F}_q) \to G(\mathbb{F}_q)/[G(\mathbb{F}_q),G(\mathbb{F}_q)]$, i.e., that $[G(\mathbb{F}_q),G(\mathbb{F}_q)] \supseteq T(\mathbb{F}_q)$. This completes the proof of the lemma for $R$ equal to a field.

Finally, suppose that $R = \mathcal{O}$ and $q = 3$. Since we already showed that $[G(\mathbb{F}_q),G(\mathbb{F}_q)] \supseteq \alpha^\vee(\mathbb{F}_q^\times)$, $(2.6)$ will follow if we show that $[G(\mathcal{O}),G(\mathcal{O})] \supseteq \alpha^\vee(1 + p)$.

To see this, more generally when $q$ is odd, we claim that $I_\mathcal{O} \supseteq p$. Indeed, the squaring operation is bijective on $1+p$, for all $z \in p$. So, we can take $g \in \mathcal{O}$ such that $g^2 = 1 + z$, and hence $z \in I_\mathcal{O}$.

Hence, we can apply $(2.6)$ to the case $f \in 1+p$, and we conclude that $\alpha^\vee(1+p) \subseteq [G(\mathcal{O}),G(\mathcal{O})]$, as desired.\hfill $\square$

2.5. Comparison between easy and extendable.

**Corollary 27.** Let $\bar{\mu} : T(\mathcal{O}) \to \mathbb{C}^\times$ be a smooth character of $G$. Then

$\bar{\mu}$ is easy for $G \implies \bar{\mu}$ is extendable to $G(F)$

$\implies \bar{\mu}$ is $W$-invariant $\implies (\bar{\mu} \circ \alpha^\vee |_{\mathcal{O}^\times})^2 = 1$, $\forall \alpha \in \Phi$. 
Proof. The first implication was already discussed before Example [19]. The second implication follows from the facts $W = N_G(O(T))/T(O)$ and $[N_G(O(T), T(O))] \subseteq [G(O), G(O)] \cap T(O)$. For the last implication, note that $(\bar{\mu} \circ \alpha^\vee)^2(x) = \bar{\mu}([\alpha^\vee(x), s_\alpha])$, where $s_\alpha$ is any lift to $N_G(O(T))$ of the simple reflection $s_\alpha$. \hfill \box

The reverse implications can all fail. For the first implication, see Example [19].

For the remaining two, we have the following:

Example 28. (i) Let $G = \text{SL}_2$. Let $\bar{\mu} : T(O) \to \mathbb{C}^\times$ denote the composition

$$T(O) = O^\times \to O^\times/(O^\times)^2 \xrightarrow{\theta} \mathbb{C}^\times,$$

where $\theta$ is a nontrivial character. Then $\bar{\mu} : T(O) = O^\times \to \mathbb{C}^\times$ is $W$-invariant; however, it does not extend to $G(F)$ by Proposition 20.

(ii) [Roc98] Example 8.4 Let $G = \text{Sp}_{2n}$, $n \geq 2$. Identify $T(O)$ with $(O^\times)^n$, and let $\bar{\mu} = (\theta, \cdots, \theta)$. Then $\bar{\mu}$ is $W$-invariant; however, it does not extend to $G(F)$. This is because, as observed in [Roc98] Example 8.4, the composition $\bar{\mu} \circ \alpha^\vee$ is not trivial for all $\alpha$ (and in fact, the root subsystem whose coroots have trivial composition produces an endoscopic group $SO_{2n}$, which is not a subgroup of $G$).

Example 29. Let $G = \text{SL}_3$. Define

$$\bar{\mu}(\text{diag}(a, b, a^{-1}b^{-1})) = \theta(a)\theta(b), \quad a, b \in O^\times,$$

where $\theta$ is a nontrivial quadratic character of $O^\times$. By assumption, $(\bar{\mu} \circ \alpha^\vee)^2 = 1$ for both coroots of $G$; however, $\bar{\mu}$ is not invariant under the transformation $(a, b, a^{-1}b^{-1}) \mapsto (a^{-1}b^{-1}, b, a)$; in particular, it is not $W$-invariant.

In certain situations, either (or both) of the first two implications in the above corollary become biconditionals.

Lemma 30. (i) Suppose that $Q^\vee = \langle \lambda - w(\lambda) \mid \lambda \in X^\vee, w \in W \rangle$. Then every $W$-invariant character of $T(O)$ is extendable to $G(F)$.

(ii) The hypothesis of (i) is equivalent to the statement that, for some choice of simple roots $\alpha_i$ (or, equivalently, for any choice of simple roots), there exist cocharacters $\lambda_i \in X^\vee$ such that $\langle \lambda_i, \alpha_i \rangle = 1$. Moreover, this condition is implied by either of the following:

(a) $X/Q$ is free

(b) The root system of $G$ has no factors of type $A_1$ or $C_n$.

(ii) Suppose $X^\vee/Q^\vee$ is torsion-free. Then every extendable character of $T(O)$ (to $G(F)$) is easy.

Proof. (i) If the coroot lattice equals the span of the elements $\lambda - w(\lambda)$ for $w \in W$ and $\lambda \in X^\vee$, then [24] is satisfied. This is because $W$-invariance implies $\bar{\mu}(\lambda(x)) = \bar{\mu}(w(\lambda)(x))$ for all $x \in G_m(O)$, and hence $\bar{\mu}((\lambda - w(\lambda))(x)) = 1$ for all $x \in G_m(O)$.
(i') First, we claim that $Q^\vee \supseteq \{ \lambda - w(\lambda) \mid \lambda \in X^\vee, w \in W \}$. Let $\alpha_i, i \in I$ be a choice of simple roots. Since $W$ is generated by the $s_{\alpha_i}$, 

\[
\{ \lambda - w(\lambda) \mid \lambda \in X^\vee, w \in W \} = \{ \lambda - s_{\alpha_i} \lambda \mid \lambda \in X^\vee, i \in I \} = \{ (\lambda, \alpha_i)_{\alpha^i_j} \mid \lambda \in X^\vee, i \in I \}.
\]

This proves the desired containment. So, we need to show that the opposite inclusion is equivalent to the condition stated in (i').

Given $\lambda_i$ such that $\{ \lambda_i, \alpha_i \} = 1$, we obviously get $\alpha_i^\vee$ in the RHS of the above equation. Conversely, if $\alpha_i^\vee \in \{ (\lambda, \alpha_i)_{\alpha^i_j} \mid \lambda \in X^\vee, j \in I \}$, then there must exist $\lambda_i \in X^\vee$ such that $\{ \lambda_i, \alpha_i \} = 1$. Applying this to all $i$ yields the desired equivalence (since $Q^\vee$ is spanned by the $\alpha_i^\vee$).

(a) If $X/Q$ is torsion-free, then $Q$ must be saturated in $X$, so the condition (i') is satisfied.

(b) For the root system $A_2$, with simple roots $\alpha_1$ and $\alpha_2$, one has $s_{\alpha_1} (\alpha_2^\vee) - \alpha_2^\vee = \alpha_1^\vee$, and similarly with indices 1 and 2 swapped, so that one concludes that $\alpha_1^\vee, \alpha_2^\vee \in \{ \lambda - w(\lambda) \}$ and hence $Q^\vee = \{ \lambda - w(\lambda) \}$. The same argument shows that, for every root system in which every simple root is contained in a root subsystem of type $A_2$, then every coroot is contained in $\{ \lambda - w(\lambda) \}$ and hence (i) is also satisfied.

This takes care of all root systems except for types $A_1, B_n, C_n$, and $G_2$. For type $B_n$ with $n \geq 3$, the above argument shows that, for the standard choice of simple roots $\alpha_1, \ldots, \alpha_n$ where $\alpha_n$ is the short simple root, then $\alpha_i^\vee \in \{ \lambda - w(\lambda) \}$ for $i < n$, since these are incident to a subdiagram of type $A_2$; for $\alpha_n^\vee$, it is still true that $s_{\alpha_n} (\alpha_{n-1}^\vee) - \alpha_{n-1}^\vee = \alpha_n^\vee$, so also $\alpha_n^\vee \in \{ \lambda - w(\lambda) \}$. For type $G_2$, if the simple roots are $\alpha_1$ and $\alpha_2$, we see that $s_{\alpha_1} (\alpha_1^\vee + \alpha_2^\vee) - (\alpha_1^\vee + \alpha_2^\vee) = \pm \alpha_1^\vee$, so $\alpha_1^\vee \in \{ \lambda - w(\lambda) \}$, and the same fact holds (with opposite sign) when indices 1 and 2 are swapped. Note also that $B_2 = C_2$, so we do not need to separately exclude $B_2$.

(ii) The hypothesis is equivalent to the condition that $Q^\vee$ is saturated in $X^\vee$; i.e., $Q^\vee = (Q^\vee)_{\text{sat}}$. The result then follows from Remarks [13] and [22].

2.6. ON PARABOLIC CHARACTERS. Recall that a smooth character $\tilde{\mu} : T(O) \to \mathbb{C}^*$ is parabolic if its stabilizer in $W$ is a parabolic subgroup. Here is an example of a character which is not parabolic.

**Example 31.** (cf. [Roc98, Example 8.3], due to Sanje-Mpacko) Let $N \geq 3$ and $G = SL_N$. Define

\[
\tilde{\mu} = \chi(a_1)\chi^2(a_2)\cdots\chi^{N-1}(a_{N-1}),
\]

where $\chi : O^* \to \mathbb{C}^*$ is a character of order $N$. Then the stabilizer of $\tilde{\mu}$ in $W$ is the subgroup $\mathbb{Z}/N$ of cyclic permutations, which is not parabolic.

On the other hand, as the following proposition illustrates, in certain situations all characters are parabolic.

**Proposition 32.** Let $G$ be a connected simply laced split reductive group over $\mathbb{Z}$. If $X/Q$ is free, then every smooth character of $T(O)$ is strongly parabolic.

If, moreover, $X^\vee/Q^\vee$ is free, then every smooth character of $T(O)$ is easy.
1292. Masoud Kamgarpour and Travis Schedler

Proof. Let $\Phi_\mu$ denote the collection of roots $\alpha \in \Phi$ such that $\overline{\mu} \circ \alpha^\vee = 1$. We claim that $\Phi_\mu$ is a closed root subsystem. Indeed, if $\alpha, \beta \in \Phi_\mu$ and $\langle \alpha, \beta \rangle = -1$, then $(\alpha + \beta)^\vee = \alpha^\vee + \beta^\vee$, and so $\alpha + \beta \in \Phi_\mu$ as well. Let $L$ denote the Levi subgroup corresponding to $\Phi_\mu$. It follows from Proposition 20 along with [Roc98], Lemma 8.1.(i) and the comment at the end of p. 395, that $\overline{\mu}$ is strongly parabolic with Levi $L$ (cf. Remark 5). Alternatively, if we use only from [Roc98] that $\overline{\mu}$ is parabolic, then we can apply Lemma 30 to deduce strong parabolicity. For the final statement, we again apply Lemma 30. □

2.7. Proof of Theorem 4. Parts (i) and (ii) follow from Propositions 15 and 20, respectively. Next, we need a basic fact from the theory of reductive groups.

Lemma 33. Let $G$ be a connected split reductive group over $\mathbb{Z}$ with split maximal torus $T$. Let $L \subset G$ be a Levi containing $T$.

(i) If $X^\vee/Q^\vee$ is torsion free, so is $X^\vee/Q_L^\vee$.
(ii) If $X/Q$ is torsion free, then so is $X/Q_L$.
(iii) If the equivalent conditions (i) or (i') of Lemma 30 are satisfied for $G$ (i.e., $Q^\vee = \{ \lambda - w(\lambda) \mid \lambda \in X^\vee \}$ or, for some choice of simple roots $\alpha_i$, there exist cocharacters $\lambda_i \in X^\vee$ such that $\langle \lambda_i, \alpha_i \rangle = 1$), then they are also satisfied when $G$ is replaced by $L$.

Proof. Parts (i) and (ii) follow from the fact that $Q/Q_L$ and $Q^\vee/Q_L^\vee$ are always torsion free. For part (iii), we consider the condition (i'), i.e., the second condition. But, by definition, one can choose simple roots of $L$ which form a subset of a choice of simple roots of $G$ (and note that the (co)weight lattices are the same for $L$ as for $G$). Hence condition (i') is satisfied for $L$. □

Then, parts (iii) and (iv) both follow from Lemmas 33 and 30. Finally, part (v) follows from Proposition 32 and Lemma 33.

3. Central families and Satake Isomorphisms

3.1. Recollections on decomposed subgroups. We begin this section with some general remarks on compact open subgroups of $G(F)$. Let $P$ be a parabolic subgroup of $G$ with Levi decomposition $LU_P$. Let $P^* = LU_P^*$ denote the opposite of $P$ relative to $L$. (According to [Bor91] Proposition 14.21, the opposite parabolic is unique up to conjugation by a unique element of $U_P^*$. In what follows, by a Levi subgroup of $G$ we always mean a Levi for a parabolic subgroup containing $T$.

Let $J \subset G$ be a compact open subgroup. Let

$$J^+_p = J \cap U_p^*(F), \quad J^+_p = J \cap L(F), \quad J^-_p = J \cap U_p^*(F).$$

For a parabolic $P = LU_P^*$, we let $\Phi^+_p$ denote the set of roots of $U_p^+$. Similarly, we let $\Phi^-_p$ denote the set of roots of $U_p^-$. Note that $\Phi = \Phi_L \cup \Phi^+_p \cup \Phi^-_p$.

Definition 34. (1) The subgroup $J$ is decomposed with respect to $P$ if the product

$$J^+_p \times J^+_p \times J^-_p \to J$$

Documenta Mathematica 18 (2013) 1275–1300
is surjective (and hence bijective).

(2) The group $J$ is totally decomposed with respect to $P$ if it is decomposed, and in addition, the product maps

$$\prod_{\alpha \in \Phi_+} U_{\alpha}(F) \cap J \to J^+_P,$$

are surjective (and hence bijective) for any ordering of the factors on the left hand side.

(3) We say that $J$ is absolutely totally decomposed if it is totally decomposed with respect to all parabolic subgroups $P$.

The above definitions are closely related to the ones given in [BK98, §6] and [Bus01, §1.1]. (Note, however, that similar decompositions appear in [BT72, §6].) The following result, which is immediate from the definitions, is similar to a statement in [Bus01, §1.1].

Lemma 35. Let $J$ be totally decomposed in $G$ with respect to a Borel subgroup $B$. Then $J$ is totally decomposed with respect to every parabolic $P$ containing $B$.

In particular, if $J$ is totally decomposed with respect to all Borels, then it is absolutely totally decomposed. The following lemma is also immediate from the definitions.

Lemma 36. Let $J$ be a compact open subgroup of $G$ which is decomposed with respect to a parabolic $P$. Suppose $L(O)$ normalizes $J^+_P$ and $J^-_P$ and that $J^0_P \subseteq L(O)$. Then the subset $K = JL(O)$ is a subgroup of $G(F)$; moreover, it is decomposed with respect to $P$; that is, $K = K^+_P K^-_P K^0_P$ where $K^+_P = J^+_P$ and $K^0_P = L(O)$.

3.2. The subgroup $K$. Let $f : \Phi \to \mathbb{Z}$ be a function satisfying the properties

(a) $f(\alpha) + f(\beta) \geq f(\alpha + \beta)$, whenever $\alpha, \beta, \alpha + \beta \in \Phi$;
(b) $f(\alpha) + f(-\alpha) \geq 1$.

In particular, $f$ is concave in the sense of Bruhat and Tits (see [BT72, §6.4.3 and §6.4.5]). Let

$$J = J_f := \langle U_{\alpha,f(\alpha)}, T(O) | \alpha \in \Phi \rangle.$$

Using the results of Bruhat and Tits, specifically [BT72 Proposition 6.4.9], Roche proved the following lemma.

Lemma 37. [Roc98 Lemma 3.2] The group $J$ is absolutely totally decomposed in $G$. Moreover, $J \cap U_\alpha(F) = U_{\alpha,f(\alpha)}$ for all $\alpha \in \Phi$.

We are interested in $K = JL(O)$, in the case that it is a group. In view of Lemma 36 to check that it is a group, it is enough to require that $J^0_P \subseteq L(O)$ and to check that $L(O)$ normalizes $J^+_P$ for some choice of parabolic $P$ with Levi component $L$.

Lemma 38. Let $P$ be a parabolic with Levi component $L$. Suppose that

$$f(\beta) = f(\alpha + \beta), \quad \forall \alpha \in \Phi_L, \beta \in \Phi \setminus \Phi_L \text{ such that } \alpha + \beta \in \Phi.$$

Then $L(O)$ normalizes $J^+_P$.  

Documenta Mathematica 18 (2013) 1275–1300
To prove the above, we will make use of the following lemma, which will also be useful later:

**Lemma 39.** Assume that \( g : \Phi \to \mathbb{Z} \) satisfies (3.2), in addition to conditions (a) and (b) (restricting \( \alpha, \beta, \) and \( \alpha + \beta \) to lie in \( \Phi \)). Let \( \Phi^0_L \) be any choice of positive roots, and let \( f : \Phi \to \mathbb{Z} \) be the function defined by
\[
\left\{ \begin{array}{ll}
f|_{\Phi^0_L} = g, & f|_{\Phi^0_L} = 0, \\
f|_{\Phi_L - \Phi^0_L} = 1.
\end{array} \right.
\]
Then \( f \) satisfies conditions (a) and (b).

Note that \( J_{f|_{\Phi_L}} = I_L \) is the Iwahori subgroup of \( L(O) \) corresponding to \( \Phi^0_L \subseteq \Phi_L \).

**Proof.** It is clear (and standard) that \( f|_{\Phi_L} \) satisfies conditions (a) and (b) (where we require in (a) that \( \alpha, \beta, \) and \( \alpha + \beta \) lie in \( \Phi_L \)). By hypothesis, \( f|_{\Phi \setminus \Phi_L} \) satisfies conditions (a) and (b) (requiring \( \alpha, \beta, \) and \( \alpha + \beta \) to be in \( \Phi \setminus \Phi_L \) in (a)). So we only need to check that, if \( \alpha \in \Phi_L \) and \( \beta \in \Phi \setminus \Phi_L \), then condition (a) is satisfied in the case that \( \alpha + \beta \in \Phi \). This is immediate from (3.2). \( \square \)

**Proof of Lemma 38.** Choose a subset \( \Phi^0_L \subseteq \Phi_L \) of positive roots for \( L \). Let \( g = f|_{\Phi \setminus \Phi_L} \), and let \( f' : \Phi \to \mathbb{Z} \) be as in Lemma 39 (i.e., \( f'|_{\Phi \setminus \Phi_L} = f|_{\Phi \setminus \Phi_L} \), \( f'|_{\Phi^0_L} = 0 \) and \( f'|_{\Phi_L - \Phi^0_L} = 1 \)). Let \( I_L = J_{f'|_{\Phi_L}} \subseteq L(O) \) be the corresponding Iwahori subgroup containing \( T(O) \). Then \( I_L \subseteq J_f \), and hence \( I_L \) normalizes \( J_f \). It also normalizes \( U_p^+ \) (since \( L \) normalizes the unipotent radical \( U_p^+ \)), so \( I_L \) normalizes \( J_f \cap U_p^+ = J_f \cap U_p = J_p^+ \). On the other hand, \( L(O) \) is generated by all its Iwahori subgroups, so \( L(O) \) also normalizes \( J_p^+ \). (Note that we could have also used the decomposition \( L(O) = I_L W_L I_L \), for \( W_L \) the Weyl group of \( L \), and the fact that \( W_L \) normalizes \( J_p^+ \) under hypothesis (3.2).) \( \square \)

**Proposition 40.** Let \( L \) be a Levi subgroup of \( G \). Assume that the function \( f : \Phi \to \mathbb{Z} \) satisfies conditions (a) and (b) as well as (3.2), and that \( f(\alpha) \geq 0 \) for all \( \alpha \in \Phi_L \). Set \( J = J_f \). Then \( K = JL(O) \) is a group; moreover, \( K \) is decomposed with respect to every parabolic \( P \) with Levi \( L \).

**Proof.** By Lemma 38 \( L(O) \) normalizes \( J_p^+ \) and \( J_p^- \). Since \( f(\alpha) \geq 0 \) for all \( \alpha \in \Phi_L \), it follows also that \( J_p^+ \subseteq L(O) \). The result then follows from Lemma 36. \( \square \)

The following corollary gives an alternative definition of \( K \).

**Corollary 41.** Let \( L \) be a Levi subgroup of \( G \). Let \( g : \Phi \to \mathbb{Z} \) be a function satisfying the following properties:
\[
\begin{array}{l}
(i) \quad g(\alpha) = 0 \text{ for all } \alpha \in \Phi_L; \\
(ii) \quad g(\alpha) + g(-\alpha) \geq 1 \text{ for all } \alpha \in \Phi \setminus \Phi_L; \\
(iii) \quad g(\alpha) + g(\beta) \geq g(\alpha + \beta), \text{ whenever } \alpha, \beta, \alpha + \beta \in \Phi.
\end{array}
\]
Then \( K = \{ U_{\alpha, g(\alpha), T(O)} \} \) is a compact open subgroup of \( G \). Moreover, \( K \cap U_\alpha = U_{\alpha, g(\alpha)} \). Finally, \( K = L(O)J_f \), where \( f : \Phi \to \mathbb{C}^* \) is defined from \( g|_{\Phi \setminus \Phi_L} \) by (3.2) (for any choice of positive roots \( \Phi^0_L \subseteq \Phi_L \)).
Proof. The inclusion $K \subseteq J_f L(O)$ is clear. For the reverse inclusion, note that it is obvious that $J_f^\sharp \subseteq K$, so we only need to show that $L(O) \subseteq K$. This follows from the fact that $L$ is generated by $T$ and the root subgroups. Thus, $K = J_f L(O)$. In particular, by the above proposition, we have a direct product decomposition $K = K^p_\sigma K^o \cap K^p_\sigma$ for every parabolic with Levi $L$. This implies that for $\alpha \in \Phi^p_\sigma$, $K \cap U_\alpha = U_{\alpha,\alpha(\alpha)}$. On the other hand it is clear that for $\alpha \in \Phi_L$, we have $K \cap U_\alpha = U_{\alpha,0}$ since $U_{\alpha,0} \subseteq L(O)$.

3.3. Extension of $\bar{\mu}$. Let $\bar{\mu} : T(O) \to \mathbb{C}^\times$ be a smooth character. Following Roche [Roc98], we define a compact open subgroup $J$ associated to $\bar{\mu}$. To this end, we have to choose a partition $\Phi = \Phi^+ \cup \Phi^-$. (Note that this amounts to choosing a Borel $B \subseteq G$.) For every $\alpha \in \Phi$, let

$$c_\alpha := \cond(\bar{\mu} \circ \alpha^\vee)$$

denote the conductor of $\bar{\mu} \circ \alpha^\vee$; that is, the smallest positive integer $c$ for which $\bar{\mu}(\alpha^\vee(1 + p^c)) = \{1\}$. Let

$$f_\bar{\mu}(\alpha) = \begin{cases} \lfloor c_\alpha/2 \rfloor, & \text{if } \alpha > 0, \\ \lfloor (c_\alpha + 1)/2 \rfloor, & \text{if } \alpha < 0. \end{cases}$$

**Lemma 42.** [Roc98] §3 Suppose that characteristic of $\mathbb{F}_q$ satisfies the conditions in (3.2). Then $f_\bar{\mu}$ satisfies conditions (a) and (b) of (3.2).

In particular, in view of Lemma 67, we obtain an associated compact open subgroup $J = J_\bar{\mu} = J_{\bar{\mu}}$. Note that the function $f_\bar{\mu}$ and the corresponding group $J_\bar{\mu}$ depend on the partition of $\Phi$ into positive and negative roots (or equivalently, on the chosen Borel $B$). While we ignore this in the notation, the reader should keep in mind that the Borel $B$ is present. In particular, we have a decomposition $J = J^+ J^0 J^-$, where $J^\pm = J^\pm_{\bar{\mu}}$. Let $J^\ast = (J^+, J^-)$.

**Lemma 43.** [Roc98] §3 There exists a unique character $\mu^\ast : J \to \mathbb{C}^\times$ whose restriction to $J^\ast = T(O)$ equals $\bar{\mu}$ and whose restriction to $J^\ast$ is trivial.

Let $\bar{\mu}$ be a strongly parabolic character of $T(O)$ and let $L$ denote the corresponding Levi. Let $P$ be a parabolic for $L$, and $B$ the Borel subgroup of $P$. In terms of $B$, let $f = f_\bar{\mu}$ denote the function associated by Roche, and let $J = J_\bar{\mu}$ denote the corresponding compact open subgroup of $G(F)$.

**Lemma 44.** The set $K = JLL(O)$ is a compact open subgroup of $G(F)$. Moreover, for every parabolic subgroup $P$ containing $L$, we have a decomposition $K = K^p_\sigma K^o \cap K^p_\sigma$ where $K^p_\sigma = J^p_\sigma$ and $K^o_\sigma = L(O)$.

**Proof.** If $\alpha \in \Phi_L$, then $\bar{\mu} \circ \alpha^\vee$ is trivial by (24). Now, if $\beta \in \Phi$ is such that $\alpha + \beta \in \Phi$, then $(\alpha + \beta)^\vee = a\alpha^\vee + b\beta^\vee$ where $a$ and $b$ are relatively prime to $q$ (by our assumption on the characteristic of $\mathbb{F}_q$; see Conventions 9). Therefore, for every $\beta \in \Phi$ such that $\alpha + \beta \in \Phi$, the conductor of $\bar{\mu} \circ (\alpha + \beta)^\vee$ equals the conductor of $\bar{\mu} \circ (\beta^\vee)$; i.e.,

$$f_\bar{\mu}(\alpha + \beta) = f_\bar{\mu}(\beta).$$

The result then follows from Proposition 10. 

□
Let $B_L = B \cap L$ denote the corresponding Borel subgroup of $L$. Let $I_L$ denote the corresponding Iwahori subgroup of $L$. Note that by Proposition 45, we have $P^+_L = J \cap L(O) = I_L$. Let $\mu^{L(F)}$ denote an extension of $\bar{\mu}$ to $L(F)$. Let $\mu^L = \mu^{L(O)} := \mu^{L(F)}|_{L(O)}$ denote its restriction to $L(O)$. Note that $\mu^L$ is automatically trivial on $I_L^L$ and $I_L^O$, since these groups are in $[L(F), L(F)]$. Set $K^*_P = (K^*_P, K^*_P)$.

**Proposition 45.** There exists a unique character $\mu = \mu^K : K \to \mathbb{C}^*$ whose restrictions to $K^*_P$, $J$ and $L(O)$ equal $1$, $\mu^J$, and $\mu^L$, respectively.

**Proof.** We need the following elementary fact: let $H^+, H^0, H^-$ be subgroups of a group $H$ which generate the group. Suppose that $H^0$ normalizes $H^\pm$. Let $\chi$ be a character of $H^0$ which is trivial on $(H^+, H^-) \cap H^0$. Then the map $\tilde{\chi} : H \to \mathbb{C}^*$ defined by $\tilde{\chi}(h^+h^0h^-) = \chi(h^0)$ is a well-defined extension of $\chi$ to $H$.

By assumption the characters $\mu^J$ and $\mu^L$ agree on $J \cap L(O) = I_L$; in particular, $\mu^J$ is trivial on $K^*_P \cap L(O)$ (since $\mu$ is trivial on $J^*$). Applying the above fact, we conclude that there exists a character $\mu : K \to \mathbb{C}^*$ whose restriction to $K^*_P$ is trivial and whose restriction to $L(O)$ equals $\mu^L$. The latter statement implies that the restriction of $\mu$ to $I_L^L$ is trivial; hence, the restriction of $\mu$ to $J^* = K^*_P I_L^L$ is also trivial. Moreover, the restriction of $\mu$ to $T(O)$ equals $\bar{\mu}$. By Lemma [13] the restriction of $\mu$ to $J$ equals $\mu^J$. \hfill \square

### 3.4. Proof of Theorem [13]

Let $\bar{\mu} : T(O) \to \mathbb{C}^*$ be a strongly parabolic character of $T(O)$ with Levi $L$, and extensions $\mu^{L(F)}$ and $\mu^{L(O)} = \mu^L$ as above. Pick a parabolic $P$ containing $L$ and a Borel $B < P$. Let $J = J_H$ denote the compact open subgroup associated to $B$ and $\bar{\mu}$. Let $\mu^J : J \to \mathbb{C}^*$ denote the canonical extension of $\bar{\mu}$ to $J$. Let

$$ (3.7) \quad \mathcal{W} := \text{ind}_{J}^{G(F)} \mu^J. $$

By definition, $\mathcal{W}$ is realized on the space of left $(J, \mu^J)$-invariant compactly supported functions on $G(F)$. The group $G(F)$ acts on this space by right translation.

Note that $J \cap L(O) = I_L$ is the Iwahori subgroup of $J$. Let $\mu^J$ denote the restriction of $\mu^J$ to $I_L$. Let $P^+_J := I_L U^+_P(F)$. The character $\mu^J$ extends uniquely to a character of $P^+_J$ which is trivial on $U^+_P(F)$. By an abuse of notation, we denote this character of $P^+_J$ by $\mu^J$ as well. Let

$$ (3.8) \quad \Pi := \text{ind}_{P^+_J}^{G(F)} \mu^J. $$

Then $\Pi$ is realized as the space of left $(P^+_J, \mu^J)$-invariant smooth functions on $G(F)$ which are compactly supported modulo $P^+_J$. The group $G(F)$ acts by right translation.

**Proposition 46.** The map $\Psi(f)(x) := \frac{1}{|P^+_J|} \int_N f(nx)dn$ is an isomorphism $\mathcal{W} \xrightarrow{\sim} \Pi$. \hfill \footnote{It is easy to check that $\Pi$, thus defined, is isomorphic to the $\Pi$ defined in [13].}
Proof. According to [Roc98, Theorem 7.7], \((J, \mu^J)\) is a cover of \((T(O), \bar{\mu})\), in the sense of [BK98, Definition 8.1] (in [Roc98], the residue characteristic is further restricted so as to obtain a nondegenerate bilinear form on the Lie algebra, but this restriction can be relaxed using the dual Lie algebra as in [Yu01]; see [KS11, §3.1.2, §A.2]). It follows from [BK98, Proposition 8.5] that \((J, \mu^J)\) is also a cover of \((I_L, \mu^L)\). The explicit isomorphism above is given by [Blo05, Theorem 2]. □

Recall from (1.9) that \(V \coloneqq \text{ind}_{G(F)}^G(P(F))\). By definition, this is a submodule of \(W\). On the other hand, let \(P^o = L(O)U^+(F)\). The character \(\mu^L\) extends uniquely to a character of \(P^o\) which is trivial on \(U^+(F)\). By an abuse of notation, we denote this character by \(\mu^L\) as well. Then, recalling the definition of \(\Theta\) in (1.10), we have an isomorphism
\[
\Theta \coloneqq \text{ind}_{L(O)}^U(P(F))\mu^L \cong \text{ind}_{P(F)}^G(P(F))\mu^L.
\]

We identify \(\Theta\) with the \(G(F)\)-module on the RHS of the above isomorphism.

With this convention, it is clear that we have an inclusion \(\Theta \hookrightarrow \Pi\). To establish Theorem 10, we prove that the restriction of \(\Psi\) to \(V\) defines an isomorphism \(G(F)\)-modules \(\Psi : V \rightarrow \Theta\). To this end, we define averaging (or symmetrization) maps \(\Psi \rightarrow \Psi\) and \(\Pi \rightarrow \Theta\) and show that they are compatible with \(\Psi\).

Recall that \(\Psi\) is realized on the space of left \((J, \mu^J)\)-invariant smooth functions on \(G(F)\). Under this identification, the subspace \(\Psi \subset \Psi\) is identified with the space of left \((L(O), \mu^L)\)-invariant functions in \(\Psi\). On the other hand, \(\Theta\) is the subspace of left \((L(O), \mu^L)\)-invariant functions in \(\Pi\).

Choose a Haar measure on \(L\) such that the volume of \(L(O)\) equals 1. For every function \(f : G(F) \rightarrow \mathbb{C}\), define \(f^c\) by
\[
f^c(x) = \int_{L(O)} \mu^L(l)f(l^{-1}x)dl.
\]

Then \(f \mapsto f^c\) defines a splitting of the natural inclusion of left \((L(O), \mu^L)\)-invariant functions on \(G(F)\) into the space of all functions on \(G(F)\). Note that this splitting obviously commutes with the action of \(G(F)\) on the space of all smooth functions by right translation. Therefore, the map \(f \mapsto f^c\) defines a splitting of the natural inclusions of \(G(F)\)-modules \(\Psi \rightarrow \Psi\) and \(\Theta \rightarrow \Pi\). The definition of \(\Psi\), given in Proposition 46, implies that the diagram
\[
\begin{array}{c}
\Psi \\
\downarrow \\
\Theta
\end{array}
\]

commutes. Since the dotted arrows are surjective and the top horizontal arrow is an isomorphism, \(\Psi|_\Psi : \Psi \rightarrow \Theta\) is surjective. Since it is the restriction of \(\Psi\), which is an isomorphism, it is also injective. Thus it is an isomorphism. □
3.5. PROOF OF THEOREM 7. We will continue with the notation of the previous subsection.

**Proposition 47.** We have a canonical isomorphism $\operatorname{End}_{G(F)}(\Theta) \simeq \mathcal{H}(L(F), L(\mathcal{O}), \mu^L)$.

**Proof.** The fact that $\bar{\mu}$ is parabolic with Levi $L$ means that $W_\bar{\mu} = N_G(\bar{\mu})/T = N_L(T)/T$. In particular, $N_G(\bar{\mu}) \subset L(F)$. By [Roc02, Theorem 3.1], parabolic induction with respect to $P$ defines an equivalence of categories between Bernstein block of $L$ corresponding to the pair $(T(\mathcal{O}), \bar{\mu})$ and that of $G$. Under this equivalence, the $L(F)$-module $\operatorname{ind}_{L(O)}^F \mu^L$ is mapped to $\Theta$. Thus, we obtain a canonical isomorphism $\operatorname{End}_{G(F)}(\Theta) \simeq \operatorname{End}_{L(F)}(\operatorname{ind}_{L(O)}^F \mu^L) \simeq \mathcal{H}(L(F), L(\mathcal{O}), \mu^L)$. \qed

Note that the algebra $\mathcal{H}(G(F), K, \mu)$ acts by convolution on the module $\mathcal{V} = \operatorname{ind}_{K}^G(\mu)$. It is a standard fact that $\mathcal{H}(G(F), K, \mu) \simeq \operatorname{End}_{\mathcal{H}(G(F))}(\mathcal{V})$. By Theorem 10, $\mathcal{V}$ is canonically isomorphic to $\Theta = \ell^G_{\mu} \left( \operatorname{ind}_{L(O)}^F \mu^L \right)$. By the preceding paragraph, the endomorphism ring of $\Theta$ is canonically isomorphic to the endomorphism ring of the $L(F)$-module $\operatorname{ind}_{L(O)}^F \mu^L$. Therefore, we obtain a canonical isomorphism $\mathcal{H}(G(F), K, \mu) \simeq \mathcal{H}(L(F), L(\mathcal{O}), \mu^L)$.

Finally, recall that $\mu^L = \mu^L|_{L(O)}$, where $\mu^L : L(F) \to \mathbb{C}^*$ is a character of $L(F)$. Then multiplication by $\mu^L$ defines a canonical isomorphism of algebras $\mathcal{H}(L(F), L(\mathcal{O})) \simeq \mathcal{H}(L(F), L(\mathcal{O}), \mu^L)$. Moreover, by the Satake isomorphism, we have a canonical isomorphism $\mathcal{H}(L(F), L(\mathcal{O})) \simeq \mathbb{C}[T/W_\mu] = \mathbb{C}[T/W_\mu]$. Theorem 7 is established. \qed

**References**


ON THE MULTI-KOSZUL PROPERTY
FOR CONNECTED ALGEBRAS

ESTANISLAO HERSCOVICH

Received: May 1, 2013
Communicated by Max Karoubi

ABSTRACT. In this article we introduce the notion of multi-Koszul algebra for the case of a locally finite dimensional nonnegatively graded connected algebra, as a generalization of the notion of (generalized) Koszul algebras defined by R. Berger for homogeneous algebras. This notion also extends and generalizes the one recently introduced by the author and A. Rey, which was for the particular case of algebras further assumed to be finitely generated in degree 1 and with a finite dimensional space of relations. The idea of this new notion for this generality, which should be perhaps considered as a probably interesting common property for several of these algebras, was to find a grading independent description of some of the more appealing features shared by all generalized Koszul algebras. It includes several new interesting examples, e.g. the super Yang-Mills algebras introduced by M. Movshev and A. Schwarz, which are not generalized Koszul or even multi-Koszul for the previous definition given by the author and Rey in any natural manner. On the other hand, we provide an equivalent description of the new definition in terms of the Tor (or Ext) groups, similar to the existing one for homogeneous algebras, and we show that several of the typical homological computations performed for the generalized Koszul algebras are also possible in this more general setting. In particular, we give a very explicit description of the $A_\infty$-algebra structure of the Yoneda algebra of a multi-Koszul algebra, which has a similar pattern as for the case of generalized Koszul algebras. We also show that a finitely generated multi-Koszul algebra with a finite dimensional space of relations is a $K_2$ algebra in the sense of T. Cassidy and B. Shelton.

2010 Mathematics Subject Classification: 16E05, 16E30, 16E40, 16S37, 16W50
Keywords and Phrases: Koszul algebra, Yoneda algebra, homological algebra, $A_\infty$-algebras.
Koszul algebras were introduced by S. Priddy in [31], motivated by the article [22] of J.-L. Koszul. They seem to have a pervasive appearance in representation theory (cf. [1], [2]), algebraic geometry (cf. [12]), quantum groups (cf. [27]), and combinatorics (cf. [16]), to mention a few. These algebras are necessarily quadratic, i.e. of the form $T(V)/\langle R \rangle$, with $R \subseteq V \otimes^2$. R. Berger generalized in [3] the notion of Koszul algebras (cf. also [15]) for the case of homogeneous algebras, i.e. algebras given by $T(V)/\langle R \rangle$, with $R \subseteq V \otimes^N$ for $N \geq 2$. They were called generalized Koszul, or $N$-Koszul if the mention to the degree of the relations was to be indicated, and the case $N = 2$ of the definition introduced by Berger coincides with the one given by Priddy. The general definition shares a lot of good properties with the one given by Priddy, justifying the terminology (see for example [36]). In particular, the Yoneda algebra of an $N$-Koszul algebra is finitely generated (in degrees 1 and 2), and its structure is easily computed from that of the original algebra. We would like to point out that the new class of algebras satisfying the Koszul property of Berger lacks however of other interesting properties, e.g. they are not closed under taking duals, or under considering graded Ore extensions, the Yoneda algebra of an $N$-Koszul algebra is not formal for $N \geq 3$, etc.

Nevertheless, it is still a natural question to ask if there exists an analogous definition for much general types of algebras, which satisfy some of the good homological properties satisfied by generalized Koszul algebras. In fact, in the recent paper [19], A. Rey and the author have considered an extension of the generalized Koszul property, called multi-Koszul, to the case of finitely generated nonnegatively graded connected algebras generated in degree 1 and with a finite dimensional space of relations, which coincides with the definition given by Berger if the algebra is homogeneous. Even though the definition of that article has several advantages, for it satisfied several homological properties as for the case of generalized Koszul algebras, it seemed to be a little ad hoc, and also too restrictive, for it cannot be applied in much general contexts, where the previous assumptions on the algebras do not hold (e.g. if the algebra is not generated in degree 1). In this article we proceed in a completely different manner to get rid of the hypotheses stated before and to consider all locally finite dimensional nonnegatively graded connected algebras. In fact, our main goal is to provide a collection of algebras which have a very similar homological behaviour to one of the generalized Koszul algebras, for which the Yoneda algebra structure (and even the $A_\infty$-algebra structure) is directly deduced from that of the original algebra, in a similar fashion as the case of generalized Koszul algebras (see Theorem 4.1, Proposition 4.4 and Theorem 4.8). Moreover, the Yoneda algebra of a multi-Koszul algebra is in fact finitely generated if the multi-Koszul algebra in question is finitely generated with a finite dimensional space of relations (see Proposition 3.31). Furthermore, our construction of the multi-Koszul complex, and thus our definition of multi-Koszul algebra, is independent of the (nonnegative) grading of the algebra,
and it restricts to the notion introduced in [19] for the additional assumption that the algebra is finitely generated in degree 1 and it has a finite dimensional space of relations (see Proposition 3.20), which further implies that it coincides with the definition given by Berger for homogeneous algebras, and thus with the one given by Priddy for quadratic algebras. All these properties have precisely given shape to the definition, which in some sense consists of requiring that the Tor (or Ext) groups should have an “optimal” behaviour and relations between them, in such a way that the standard homological techniques of generalized Koszul algebras still hold. This is somehow the main result of the present work and it gives a reason for this new definition, contrary to the one given in [19], which could have seemed to be somehow arbitrary.

We would like to point out that even tough this new definition could still seem to be rather restrictive, it allows several interesting examples, for it includes the super Yang-Mills algebras introduced by M. Movshev and A. Schwarz in [28] and further studied by the author in [18], which were one of the main motivations for the present work (see Example 3.22). These algebras are not generated in degree 1, so they cannot be multi-Koszul in the sense defined in [19], and in particular they cannot be generalized Koszul either. We also admit that our definition is probably not the most general possible and reasonable extension of the Koszul property for all locally finite dimensional nonnegatively graded connected algebras, but all the interesting properties mentioned in the previous paragraph make us believe that any sensible such general definition of Koszul-like algebra in this general context, if it exists, should necessarily include our definition as a special case. That is also one of the main reasons why we have refrained from calling this new family of algebras Koszul.

The contents of the article are as follows. We start by recalling in Section 2 several well-known definitions and results about the category of graded modules over a locally finite dimensional nonnegatively graded connected algebra. Section 3 is devoted to the definition of multi-Koszul algebras and to prove some properties for this family of algebras. In order to simplify the exposition and unify the left and right sides, it seems useful to work directly with the minimal projective resolution of the algebra \( A \) as an \( A \)-bimodule, and to derive from it the minimal projective resolutions of the left or right trivial \( A \)-module \( k \). The definition requires however several preliminary results about the spaces that will provide the generators of these minimal projective resolutions, which are in given in Subsection 3.1 and in the beginning of Subsection 3.2. The first main result, Proposition 3.24 (see also Proposition 3.25) gives a (co)homological description of multi-Koszul algebras in term of their Tor (or Ext) groups, which yields a left-right symmetry of the definition. Moreover, in Subsection 3.4 we also provide several properties satisfied by these algebras, and in particular we show that they are stable under free products (see Proposition 3.30) and they are \( K_2 \) algebras, in the sense defined by B. Cassidy and T. Shelton in [10], if the algebras are further assumed to be finitely generated with finite dimensional spaces of relations (see Proposition 3.31).

Finally, in Section 4 we provide an explicit description of the algebra structure,
and further of $A_\infty$-algebra structure, of the Yoneda algebra of a multi-Koszul algebra (see Theorem 4.1, Proposition 4.4 and Theorem 4.8). In particular, if the multi-Koszul algebra is finitely generated generated in degree 1 and with a finite number of relations, this description coincides with the $A_\infty$-algebra structure of the Yoneda algebra derived from Prop. 3.21 in [19], as explained in Rmk. 3.22, and from Rmk. 3.25 of the same article. The proofs we give here in this more general setting are however completely different. The characterization of both structures given in the previous statements follows a similar pattern to the one given for generalized Koszul algebras in [7], Prop. 3.1, and [17], Thm. 6.5, which gives another indication that the homological behaviour of the algebras satisfying this new definition is parallel to the one of generalized Koszul algebras.

Throughout this article $k$ will denote a field, and all vector spaces will be over $k$. Since there are several (confusing) notations concerning the natural numbers (whether they contain the zero or not), we denote by $\mathbb{N}$ the strictly positive integers (as G. Peano himself defined), and by $\mathbb{N}_0$ the nonnegative integers (as set theoretic definitions may suggest). If $n \in \mathbb{N}_0$, we will also write by $\mathbb{N}_{\geq n}$ (resp. $\mathbb{N}_{>n}$) the set of nonnegative integers strictly greater than $n$ (resp. greater than or equal to $n$). Moreover, $V$ will always be a vector space, and $A$ a nonnegatively graded connected (associative) algebra over $k$ (with unit), to which we will usually just refer as an algebra, with irrelevant ideal $A_{>0} = \bigoplus_{n>0} A_n$. The vector space spanned by a set of elements $\{v_s : s \in S\}$ for some index set $S$, will be denoted by $\text{span}_k \langle v_s : s \in S \rangle$ and the ideal $I$ generated by a set of elements $\{a_s : s \in S\}$ of an algebra $A$ will be denoted by $\langle a_s : s \in S \rangle$. All unadorned tensor products $\otimes$ will be considered over $k$, unless otherwise stated. We shall typically denote the vector space $V^\otimes n$ by $V^{(n)}$ for $n \in \mathbb{Z}$, where we follow the convention $V^{(n)} = 0$, if $n < 0$, and $V^{(0)} = k$, and an elementary tensor $v_1 \otimes \cdots \otimes v_n \in V^{(n)}$ will be usually written by $v_1 \ldots v_n$. Finally, given two chain complexes $(C_\bullet, d_\bullet)_{\in \mathbb{Z}}$ and $(C'_\bullet, d'_\bullet)_{\in \mathbb{Z}}$ of modules over a ring, we say that both complexes coincide up to homological degree $m \in \mathbb{Z}$ if $C_n = C'_n$ and $d_n = d'_n$ for all $n \leq m$.

We would also like to thank the referee for a careful reading of this manuscript.

2 Preliminaries on minimal resolutions of graded algebras

In this section we shall recall some basic facts about the category of (bounded below) graded modules over a nonnegatively graded connected algebra. We refer to [29], Exp. 15, or [4] for all the proofs of the mentioned results. Even though we impose the assumption that $A$ is locally finite dimensional, most of the results of this section hold without that hypothesis.

We recall that a graded vector space is a vector space $W$ provided with a direct sum decomposition of the form $W = \bigoplus_{n \in \mathbb{Z}} W_n$. A nonzero element $w$ in $W_m$ is called homogeneous of degree $m$, and we will write $|w| = m$. A morphism of graded vector spaces of degree $d$ from $W$ to $W'$ is just a linear map $f : W \to W'$ such that $f(W_m) \subseteq W'_{m+d}$ for all $m \in \mathbb{Z}$. If we omit the degree of
the morphism it will be understood that $d = 0$. A bigraded (co)homological vector space is a vector space $W$ together with a direct sum decomposition given by $W = \bigoplus_{n,m} W_{n,m}$ (resp., $W = \bigoplus_{n,m} W^\ast_{n,m}$). We may consider a graded vector space $W = \bigoplus_{m \in \mathbb{Z}} W_m$ as a (co)homological bigraded vector space concentrated in (co)homological degree 0, i.e., $W_{n,m} = W_m$ (resp., $W^\ast_{n,m} = W^\ast_m$) if $n = 0$, and $W_{n,m}$ (resp., $W^\ast_{n,m}$) vanishes otherwise. Given a nonzero element $w \in W_{n,m}$ (resp., $w \in W^\ast_{n,m}$) we will say that it has (co)homological degree $n$ and Adams degree $m$ (this terminology comes from algebraic topology, see for instance [23], p. 38), or, simply, bidegree $(n,m)$. The first will be usually denoted by $\deg(w)$, whereas the second one will be written as $|w|$. A morphism of (co)homological bigraded vector spaces of (co)homological degree $d$ and Adams degree $d'$ (or bidegree $(d,d')$) from $W$ to $W'$ is just a linear map $f : W \to W'$ such that $f(W_{n,m}) \subseteq W'_{n+d,m+d}$ for all $n,m \in \mathbb{Z}$. If we omit the degree of the morphism it will be understood that $d = d' = 0$. If we talk about the degree of an element or a map, it is implicitly assumed that it is homogeneous. If $W = \bigoplus_{n,m} W_{n,m}$ is a homological bigraded vector space, it will be also considered as a cohomological bigraded vector space where $W_{n,m} = W_{-n,-m}$, for all $n,m \in \mathbb{Z}$, and vice versa. By abuse of notation we may thus say that $w \in W_{n,m}$ has cohomological degree $-n$.

If $W$ is a graded vector space, we may consider the graded vector space $W^\#$, called the graded dual, which has $n$-th homogeneous component $(W^\ast)_{-n}$, where $(-)^\ast$ denotes the usual dual vector space operation. We will also consider the analogous graded dual construction $(-)^\ast$ in the category of (co)homological bigraded vector spaces, where in that case $W^\#$ has component of (co)homological degree $n$ and Adams degree $m$ given by $(W_{n,-m})^\ast$ (resp., $(W^\ast_{n,-m})^\ast$). If $f : W \to W'$ is morphism of graded vector spaces of degree $d$, we get another morphism $f^\#$ of graded vector spaces of the same degree whose restriction to $(W^\ast)_{-n}$ is given by $(f_{|W_{-n,-d}})^\ast$. If $f : W \to W'$ is morphism of (co)homological bigraded vector spaces of (co)homological degree $d$, define the morphism $f^\#$ of (co)homological bigraded vector spaces of the same (co)homological degree given by $f^\#(\lambda) = (-1)^{d \deg(\lambda)} \lambda \circ f$ for $\lambda$ homogeneous. Note that $(g \circ f)^\# = f^\# \circ g^\#$ if $f,g$ are morphisms of graded vector spaces such that the composition makes sense, but $(g \circ f)^\# = (-1)^{\deg(f) \deg g} f^\# \circ g^\#$ for $f,g$ morphisms of (co)homological bigraded vector spaces such that the composition makes sense. We will thus apply the Koszul sign rule to the (co)homological degree but not to the Adams degree. This means in particular that, if $W_1, \ldots, W_n$ are locally finite dimensional (co)homological bigraded vector spaces, there is an isomorphism $c_{W_1,\ldots,W_n} : W_1^\# \otimes \cdots \otimes W_n^\# \to (W_1 \otimes \cdots \otimes W_n)^\#$ of the form

$$c_{W_1,\ldots,W_n}(f_1 \otimes \cdots \otimes f_n)(w_1 \otimes \cdots \otimes w_n) = (-1)^s f_1(w_1) \cdots f_n(w_n),$$

where $s = \sum_{i=1}^{n-1} \deg(w_i)(\deg(f_{i+1}) + \ldots + \deg(f_n))$. However, if we refrain from applying the Koszul sign rule to any degree whatsoever of a homological bigraded vector space, we will just call it bigraded vector space.
From now on, $A$ will always denote a nonnegatively graded connected algebra with unit 1. We shall further assume that the underlying graded vector space of $A$ is locally finite dimensional, i.e. $A = \bigoplus_{n \in \mathbb{N}_0} A_n$, where $\dim_k(A_n) < \infty$ for all $n \in \mathbb{N}_0$, and we shall follow the typical convention $A_n = 0$ for $n < 0$. This means that there exists a locally finite dimensional positively graded vector space $V$ and a surjective morphism of graded algebras of the form $\pi : T(V) \to A$, so $A \simeq T(V)/I$, where $I \subseteq T(V)$ is a homogeneous ideal of $T(V)$. Note that for any graded vector space $V$ the tensor algebra $T(V) = \bigoplus_{n \in \mathbb{N}_0} V^{(n)}$ is provided with two compatible gradings, other grading coming from the previous direct sum decomposition, which we shall call special or tensor, and whose $N$-th homogeneous component will be also denoted by $T(V)^N$, and another one coming from the grading on $V$, which we call usual and whose $n$-th homogeneous components will be denoted by $T(V)_n$. To avoid redundancy we will always assume that the vector space $V$ is canonically isomorphic to $A_{>0}/(A_{>0}A_{>0})$ (as graded vector spaces). In this case we further have that $I \subseteq T(V)^{\geq 2} = T(V)_{>0}.T(V)_{>0}$. Let us denote by $R$ a space of relations of $A$, i.e. a graded vector subspace of $I$ which is isomorphic to $I/(T(V)_{>0}.I + I.T(V)_{>0})$ under the canonical projection. It can be equivalently defined as a graded vector subspace $R \subseteq I$ satisfying that the ideal $\langle R \rangle$ of $T(V)$ generated by $R$ coincides with $I$ and that

$$R \cap (T(V)_{>0}.R.T(V) + T(V).R.T(V)_{>0}) = 0. \quad (2.1)$$

Notice that the Hilbert series of $R$ is thus uniquely determined, and the same holds for its first nonvanishing homogeneous component. We may thus suppose that $A = T(V)/(R)$, where $R \subseteq T(V)^{\geq 2}$ is a graded vector subspace for the usual grading. Note that the assumption that $A$ is locally finite dimensional implies that $V$ (hence $T(V)$) is a locally finite dimensional nonnegatively graded vector space, and then $R$ is also so.

A graded left (resp., right) $A$-module $M = \bigoplus_{n \in \mathbb{Z}} M_n$ is a graded vector space together with a left (resp., right) action of $A$ on $M$ such that $A_n M_m \subseteq M_{n+m}$ (resp., $M_m A_n \subseteq M_{m+n}$), and we shall sometimes refer to them simply as left (resp., right) $A$-modules. Moreover, in the rest of the section we will mostly deal with left modules, which may be thus called just modules (or graded modules), even though all the considerations apply verbatim to right modules. We will denote by $A\text{-grMod}$ the abelian category of graded left $A$-modules, where the morphisms are the $A$-linear maps preserving the grading. The space of morphisms in this category between two graded left $A$-modules $M$ and $M'$ will be denoted by $\text{hom}_A(M, M')$. This category is provided with a shift functor $(-)[1]$ defined by $(M[1])_n = M_{n+1}$, where the underlying left $A$-module structure of $M[1]$ is the same as the one of $M$, and the action of the morphisms is trivial. We shall also denote $(-)[d]$ the $d$-th iteration of the shift functor. The graded left $A$-module $M$ is said to be left bounded, or also bounded below, if there exists an integer $n_0$ such that $M_n = 0$ for all $n < n_0$. Notice that the graded left $A$-modules which are left bounded form a full exact subcategory of $A\text{-grMod}$.
Given graded left $A$-modules $N$ and $N'$, we recall the following notation:

$$\text{Hom}_A(N, N') = \bigoplus_{d \in \mathbb{Z}} \text{hom}_A(N, N'[d]).$$

We remark that, if $N$ is finitely generated, then $\text{Hom}_A(N, N') = \text{Hom}_A(N, N')$, where the last morphism space is the usual one for left $A$-modules by forgetting the gradings (see [30], Cor. 2.4.4).

The following result is the graded version of the Nakayama Lemma.

**Lemma 2.1.** Let $M$ be a left bounded graded left $A$-module. If $k \otimes_A M = 0$, then $M = 0$.

**Proof.** See [4], Lemme 1.3 (see also [9], Exp. 15, Prop. 6). □

The left $A$-module $M$ is said to be graded-free if it is isomorphic to a direct sum of shifts $A[-l_i]$ of $A$. We remark that a bounded below graded left $A$-module $M$ is graded-free if and only if its underlying module (i.e. forgetting the grading) is free, if and only if it is projective (as a graded module or not), if and only if $\text{Tor}^A_1(k, M) = 0$ for all $\bullet \geq 1$ (or just $\bullet = 1$). This will follow from the comments on projective covers.

A surjective morphism $f : M \to M'$ in $A$-grMod is called essential if for each morphism $g : N \to M$ in $A$-grMod such that $f \circ g$ is surjective, then $g$ is also surjective. As an application of the Nakayama Lemma we have the following result which characterizes essential surjective maps.

**Lemma 2.2.** Let $f : M \to M'$ be a morphism in the category of graded left (resp., right) $A$-modules. Suppose that $M'$ is left bounded and that $f$ is surjective and essential. Then $\text{id}_k \otimes_A f$ (resp., $f \otimes_A \text{id}_k$) is bijective. Moreover, if $M$ is also left bounded, the converse holds.

**Proof.** See [4], Lemme 1.5 (see also [9], Exp. 15, Prop. 7). □

Let $M$ be a nontrivial object in $A$-grMod. A projective cover of $M$ is a pair $(P, f)$ such that $P \in A$-grMod is projective and $f : P \to M$ is an essential surjective morphism. We remark that every left bounded graded left $A$-module $M$ has a projective cover, which is unique up to (noncanonical) isomorphism (cf. [9], Exp. 15, Thm. 2). Moreover, given $M$ a bounded below left $A$-module, a projective cover may be explicitly constructed as follows. Since $M \neq 0$, the Nakayama lemma tells us that $M/(A_{>0}M) \simeq k \otimes_A M$ is a nontrivial graded vector space. Consider a section $s$ of the canonical projection $M \to M/(A_{>0}M) \simeq k \otimes_A M$. Now, we define $P = A \otimes (k \otimes_A M)$ together with the $A$-linear map $f : P \to M$ given by $f(a \otimes v) = as(v)$, for $a \in A$ and $v \in k \otimes_A M$. Using the previous lemma one directly gets that $(P, f)$ is a projective cover of $M$.

Recall that a (graded) projective resolution $(P_\bullet, d_\bullet)$ of a graded left $A$-module $M$ is minimal if $d_0 : P_0 \to M$ is a projective cover (or equivalently, it is essential) and each of the maps $P_i \to \text{Ker}(d_{i-1})$ induced by $d_i$ is also essential, for
all \( i \in \mathbb{N} \). We want to remark the important fact that, by iterating the process of considering projective covers for bounded below modules, one may easily prove that any bounded below graded left \( A \)-module has a minimal projective resolution (see [4], Thm. 1.11). If the left \( A \)-module \( M \) has a minimal projective resolution \((P_\bullet, \delta_\bullet)\), given any other projective resolution \((Q_\bullet, \delta'_\bullet)\) of \( M \) there exists an isomorphism of (augmented) complexes \( Q_\bullet \simeq P_\bullet \oplus H_\bullet \), where \( H_\bullet \) is acyclic (see [4], Prop. 2.2). Additionally, the minimality assumption on the projective resolution implies that the differential of the induced complex \( k \otimes_A P_\bullet \) vanishes (see [9], Exp. 15, Prop. 10, or [4], Prop. 2.3), so if \((P_\bullet, \delta_\bullet)\) denotes such a minimal projective resolution, one also easily gets that 

\[
P_\bullet \simeq A \otimes \operatorname{Tor}^A_\bullet(k, M).
\]

Combining the results of the two previous sentences, it is trivial to see that if \((Q_\bullet, \delta'_\bullet)\) is a projective resolution of a graded left \( A \)-module \( M \) having a minimal projective resolution, then the former is minimal if and only if the induced differential of \( k \otimes_A Q_\bullet \) vanishes.

If \( N \) is left bounded, let \((P_\bullet, \delta_\bullet)\) be a (minimal) graded projective resolution of \( N \). As usual, we denote

\[
\mathcal{E}xt^i_A(N, N') = H^i(\mathcal{H}om_A(P_\bullet, N')),
\]

If the projective resolution of \( N \) is composed of finitely generated projective left \( A \)-modules, there is a canonical identification \( \mathcal{E}xt^i_A(N, N') \simeq \operatorname{Ext}^i_A(N, N') \).

Moreover, using a very simple duality argument one can see that, if \( M \) is a bounded below graded left \( A \)-module, then there is a canonical isomorphism of graded vector spaces

\[
\mathcal{E}xt^i_A(M, k) \simeq \operatorname{Tor}^A_i(k, M)^\#,
\]

for all \( i \in \mathbb{N}_0 \) (see [4], Eq. (2.15), but cf. also [4], Exp. 15, Prop. 2).

We recall the beginning of the minimal projective resolution of the trivial left \( A \)-module \( k \) for any nonnegatively graded connected algebra \( A \). The analogous statements for the trivial right \( A \)-module \( k \) are immediate. We know that it starts as

\[
A \otimes V \xrightarrow{\delta_1} A \xrightarrow{\delta_0} k \longrightarrow 0,
\]

where \( \delta_0 \) is the augmentation of the algebra \( A \), \( V \simeq A_{>0}/(A_{>0} \cdot A_{>0}) \) is the vector space spanned by a minimal set of (homogeneous) generators of \( A \) indicated before, and \( \delta_1 \) is the restriction of the product of \( A \) (see [4], Exp. 15, end of Section 7). Furthermore, it is also well-known (and follows easily from the definition) that \( \operatorname{Ker}(\delta_1) \simeq I/(I.V) \) (as graded vector spaces), so there is an isomorphism of graded vector spaces \( k \otimes_A \operatorname{Ker}(\delta_1) \simeq I/(T(V)_{>0}.I + I.T(V)_{>0}) \), i.e. we have an isomorphism of graded vector spaces \( k \otimes_A \operatorname{Ker}(\delta_1) \simeq R \) (see [13], Lemma 1, for complete expressions of the graded vector spaces \( \operatorname{Tor}^A_i(k, k) \) for \( i \in \mathbb{N}_0 \) in terms of \( I \) and the irrelevant ideal \( T(V)_{>0} \)). Hence, \( A \otimes R \to \operatorname{Ker}(\delta_1) \) is a projective cover, and the beginning of the minimal projective resolution of the trivial \( A \)-module \( k \) is of the form

\[
A \otimes R \xrightarrow{\delta_2} A \otimes V \xrightarrow{\delta_1} A \xrightarrow{\delta_0} k \longrightarrow 0,
\]
where $\delta_2$ is induced by the usual map $a \otimes v_1 \ldots v_n \mapsto av_1 \ldots v_{n-1} \otimes v_n$ (see [23], Lemme 1.2.11). Moreover, both left and right minimal projective resolutions of $k$ regarded as a left or right $A$-module respectively can be obtained from the minimal projective resolution $(P^*_b, \delta^*_b)$ of $A$ as an $A^e$-module easily. Indeed, the minimal projective resolution of the trivial left (resp., right) $A$-module $k$ is given by $P^*_b \otimes_A k$ (resp., $k \otimes_A P^*_b$). The fact that it is projective resolution of the trivial left (resp., right) $A$-module $k$ follows easily from the standard fact (see for instance [27], Prop. 4.1), that, since the augmented complex $P^*_b \to A$ is an exact and bounded below complex of projective right $A$-modules, it is homotopically trivial as a complex of right $A$-modules, so $P^*_b \otimes_A k$ (resp., $k \otimes_A P^*_b$) is also homotopically trivial, and thus exact. To prove the minimality claim we use the fact we have already explained: a projective resolution $Q_\ast$ of a bounded below graded left (resp., right) module $M$ over a locally finite dimensional nonnegatively graded algebra $B$ is minimal if and only if $k \otimes_B Q_\ast$ (resp., $Q_\ast \otimes_B k$) has vanishing differential. Now, by the minimality of $P^*_b$, we get that $P^*_b \otimes_A k = k \otimes_A P^*_b \otimes_A k$ has vanishing differential, which further implies that $P^*_b = P^*_b \otimes_A k$ is minimal. Moreover, it is easy to see that the beginning of the minimal projective resolution of the $A$-bimodule $A$ is given by

\[ A \otimes R \otimes A \xrightarrow{\delta^*_b} A \otimes V \otimes A \xrightarrow{\delta^*_b} A \otimes A \xrightarrow{\delta^*_b} A \to 0, \]  

where $\delta^*_b$ is given by the multiplication of $A$, $\delta^*_b$ is determined by $\delta^*_b(a \otimes v \otimes a') = av \otimes a' - a \otimes va'$ and $\delta^*_b$ is the linear extension induced by $a \otimes v_1 \ldots v_r \otimes a' \mapsto \sum_{r=1}^{r} av_1 \ldots v_{j-1} \otimes v_j \otimes v_{j+1} \ldots v_r a'$.

### 3 Multi-Koszul algebras

#### 3.1 Auxiliary results on subspaces of the tensor algebra

We recall the following obvious property.

**Fact 3.1.** Let $V, W$ be two graded vector subspaces of a graded vector space $U = \oplus_{n \in \mathbb{Z}} U_n$. Then, the intersection $V \cap W$ is also a graded vector subspace of $U$ such that $(V \cap W)_n = V_n \cap W_n$ for $n \in \mathbb{Z}$.

Let $\mu$ denote the multiplication of the tensor algebra $T(V)$, and, for $n \geq 2$, $\mu^{(n)} : T(V)^{\otimes n} \to T(V)$ the $(n-1)$-th iteration of $\mu$, i.e. it is defined by the recursive process given by $\mu^{(2)} = \mu$ and $\mu^{(i+1)} = \mu \circ (id_{T(V)} \otimes \mu^{(i)})$ for $i \in \mathbb{N}_{\geq 2}$. Note that $\mu^{(n)}$ is surjective (if $V$ is nontrivial) but not injective for all $n \in \mathbb{N}_{\geq 2}$. Given $W_1, \ldots, W_m$ vector subspaces of $T(V)$, we shall denote by $W_1 \ldots W_m$ the vector subspace of $T(V)$ given by the image of the vector subspace $W_1 \otimes \ldots \otimes W_m \subseteq T(V)^{\otimes m}$ under the map $\mu^{(m)}$. If $W_1 = \ldots = W_m = W \subseteq T(V)$, we shall denote $W_1 \ldots W_m$ by $W^{(m)}$. In the next sections we will be particularly interested in the case where $V$ is a locally finite dimensional graded vector space, the tensor algebra $T(V)$ has the
induced grading, and $W_1, \ldots, W_m$ are considered to be graded vector subspaces. Since the product of the tensor algebra is denoted by the $\otimes$ symbol, it is usual to denote this image also by $W_1 \otimes \cdots \otimes W_m$, even though this product differs in principle from the usual (external) tensor product of the vector spaces $W_1, \ldots, W_m$ (e.g. take $W_1 = \text{span}_k \langle x, x^2 \rangle$, and $W_2 = \text{span}_k \langle x^3, x^4 \rangle$). Then, the vector subspace $W_1 \cap W_2$ of $T(V)$ has basis $\{x^3, x^4, x^5\}$, whereas the usual tensor product $W_1 \otimes W_2$ has dimension 4. Even though this notation is coherent in the sense that, given $W_1, \ldots, W_m$ vector subspaces of $T(V)$, $W_1 \otimes \cdots \otimes W_m$ denotes the vector subspace of $T(V)$ that consists of the elements of $T(V)$ given by the sums of terms of the form $w_1 \otimes \cdots \otimes w_m$, where $w_i \in W_i$, and the $\otimes$ symbol indicates the product of the tensor algebra $T(V)$, we shall usually not use it and prefer using $W_1 \ldots W_m$ instead of $W_1 \otimes \cdots \otimes W_m$ to avoid the already mentioned confusion and to emphasize the fact that we are dealing with vector subspaces of $T(V)$. However, we will allow us to denote by $W_1 \otimes \cdots \otimes W_m$ the space $W_1 \ldots W_m$ in the particular case both of them are (canonically) isomorphic as vector spaces (or as graded vector spaces if the $W_1, \ldots, W_m$ are further assumed to be graded vector subspaces of $T(V)$, where the grading of the tensor algebra comes from a grading of $V$), since in this situation there is no ambiguity. This will be the case if $W_1, \ldots, W_m$ are tensor-intersection faithful, which is defined in the following paragraphs (see Remark 3.3).

A word of caution should be stated here. Given $W_1, \ldots, W_m$ vector subspaces of $T(V)$ as before, and linear maps $f_i : W_i \to W'_i$ for $i = 1, \ldots, m$, where $W'_i$ are arbitrary vector spaces, there may be no canonical manner to define a map $f_1 \ldots f_m : W_1 \ldots W_m \to W'_1 \otimes \cdots \otimes W'_m$, since an element of $W_1 \ldots W_m$ could be written inside $T(V)$ in inequivalent manners (see however Corollary 3.9). Let $U \subseteq T(V)$ be a vector subspace of the tensor algebra on a vector space $V$. We say that $U$ is left tensor-intersection faithful (resp., right tensor-intersection faithful) if $(U,W_1) \cap (U,W_2) = U,(W_1 \cap W_2)$ (resp., $(W_1,U) \cap (W_2,U) = (W_1 \cap W_2),U$) for all vector subspaces $W_1,W_2 \subseteq T(V)$. Moreover, $U$ is said to be tensor-intersection faithful if it is left and right tensor-intersection faithful.

We have the following simple characterization of tensor-intersection faithfulness, that we will use extensively in the sequel. We suspect that it is well-known among the experts but we were unable to find any reference whatsoever. In fact, all of the following results seem very natural to us, but we give a detailed proof of them for convenience.

**Proposition 3.2.** Let $U \subseteq T(V)$ be a vector subspace of the tensor algebra. Then, the following conditions are equivalent:

(i) $U$ is left (resp., right) tensor-intersection faithful.

(ii) $U \cap (U,T(V)_{>0}) = 0$ (resp., $U \cap (T(V)_{>0},U) = 0$).

(iii) Given $I$ a finite set of indices, a linearly independent family $\{u_i : i \in I\} \subseteq U$, and arbitrary elements $w_i \in T(V)$ for $i \in I$, if $\sum_{i \in I} u_i \otimes w_i$ (resp., $\sum_{i \in I} w_i \otimes u_i$) vanishes, then $u_i = 0$ for all $i \in I$. 


ON THE MULTI-KOSZUL PROPERTY FOR CONNECTED ALGEBRAS

Proof. The implication (i) ⇒ (ii) is clear. Indeed, it easily follows by considering $W_1 = k$ and $W_2 = T(V)_{>0}$.

We will now prove the implication (ii) ⇒ (iii). We shall only consider the case corresponding to the assumption $U \cap (U.T(V)_{>0}) = 0$, because the other case is analogous. Let $\{v_s\}_{s \in S}$ be a basis of $V$. Given $n \in \mathbb{N}_0$, we will denote by $\bar{s} = (s_1, \ldots, s_n)$ a typical element of $S^n$. Let us write

$$w_i = \sum_{n \in \mathbb{N}_0} \sum_{s \in S^n} c_{i,n,s} v_{s_1} \cdots v_{s_n},$$

where $c_{i,n,s} \in k$, such that the sum is of finite support, i.e. fixed $i \in I$, the coefficients $c_{i,n,s}$ are almost all zero. We have to prove that they are in fact all zero. Let us suppose that this is not the case, and consider $n_0$ the first nonnegative integer satisfying that there exists $i_0 \in I$ and $\bar{s}_0 \in S^{n_0}$ such that $c_{i_0,n_0,\bar{s}_0} \neq 0$. We will write $\bar{s}_0 = (s_0,1,\ldots,s_0,n_0)$. On the one hand, we write

$$\sum_{i \in I} u_i \otimes w_i = \sum_{n \in \mathbb{N}_0} \sum_{s \in S^n} c_{i,n,s} u_i \otimes v_{s_1} \cdots v_{s_n},$$

$$= \sum_{n \in \mathbb{N}_0} \sum_{s \in S^n} c_{i,n,s} u_i \otimes v_{s_1} \cdots v_{s_n},$$

$$= \sum_{n \geq n_0} \sum_{s \in S^n} c_{i,n,s} u_i \otimes v_{s_1} \cdots v_{s_n}.$$

Since this sum vanishes, by the definition of the tensor algebra we also have that the sum

$$\sum_{n \in \mathbb{N}_0} \sum_{s \in S^n} \sum_{i \in I} c_{i,n,n_0,0}(s,\bar{s}_0) u_i \otimes v_{s_1} \cdots v_{s_n} v_{s_0,1} \cdots v_{s_0,n_0}$$

vanishes, where $(\bar{s},\bar{s}_0)$ denotes the $(n + n_0)$-tuple $(s_1, \ldots, s_n, s_0,1,\ldots,s_0,n_0)$. Thus,

$$\sum_{n \in \mathbb{N}_0} \sum_{s \in S^n} \sum_{i \in I} c_{i,n+n_0,0}(s,\bar{s}_0) u_i \otimes v_{s_1} \cdots v_{s_n} = 0.$$

Define

$$\omega_{>0} = \sum_{n \in \mathbb{N}} \sum_{s \in S^n} \sum_{i \in I} c_{i,n+n_0,0}(s,\bar{s}_0) u_i \otimes v_{s_1} \cdots v_{s_n},$$

so

$$\sum_{n \in \mathbb{N}} \sum_{s \in S^n} \sum_{i \in I} c_{i,n+n_0,0}(s,\bar{s}_0) u_i \otimes v_{s_1} \cdots v_{s_n} = (\sum_{i \in I} c_{i,n_0,\bar{s}_0} u_i) + \omega_{>0} = 0.$$

Thus, the sum $\sum_{i \in I} c_{i,n_0,\bar{s}_0} u_i$ belongs to $U$ and to $U.T(V)_{>0}$, so it vanishes. Now, since the $\{u_i\}_{i \in I}$ form a linearly independent set, we get that $c_{i,n_0,\bar{s}_0} = 0$ for all $i \in I$. This is a contradiction, by the assumption that $c_{s_0,n_0,\bar{s}_0} \neq 0$. Hence, all the coefficients $c_{i,n,s}$ vanish, which proves that the corresponding statement of item (iii) holds.
Let us now prove the implication \((iii) \Rightarrow (i)\). As before, we shall only prove the implication for the left side conditions, for the other case is analogous.

We assume thus that \(U\) satisfies that, given \(I\) a finite set of indices, a linearly independent family \(\{u_i : i \in I\} \subseteq U\), and arbitrary elements \(w_i \in T(V)\) for \(i \in I\), if \(\sum_{i \in I} u_i \otimes w_i\) vanishes, then \(w_i = 0\) for all \(i \in I\). We will prove that in this case \((U.W) \cap (U.W') = U.(W \cap W')\). Note that \((U.W) \cap (U.W') \supseteq U.(W \cap W')\) is always true by obvious reasons. We shall prove the reverse inclusion. Consider a element \(\omega \in (U.W) \cap (U.W')\), and fix a basis \(\{u_s\}_{s \in S}\) of \(U\). Then, there exists two finite subsets \(S_1, S_2 \subseteq S\), and elements \(w_s \in W\) for \(s \in S_1\) and \(w'_s \in W'\) for \(s \in S_2\) such that \(\omega = \sum_{s \in S_1} u_s \otimes w_s = \sum_{s \in S_2} u_s \otimes w'_s\). Setting \(w_s = 0\) if \(s \in S \setminus S_1\), and \(w'_s = 0\) if \(s \in S \setminus S_2\), we may write \(\omega = \sum_{s \in S} u_s \otimes w_s = \sum_{s \in S} u_s \otimes w'_s\), so \(\sum_{s \in S} u_s \otimes (w_s - w'_s) = 0\). The assumption implies that \(w_s = w'_s\) for all \(s \in S\), so \(w_s \in W \cap W'\) for all \(s \in S\). Hence, \(\omega \in U.(W \cap W')\), which proves the assertion. The proposition is thus proved. 

\[\square\]

**Example 3.3.** The previous proposition immediately implies that \(V\) is a tensor-intersection faithful vector subspace of the tensor algebra \(T(V)\). Furthermore, it also tells us that a graded vector subspace \(R \subseteq T(V)\) of the tensor algebra on a positively graded vector space \(V\) satisfying condition 2.1 (e.g. a space of relations \(R\) of an algebra \(T(V)/I\)) is tensor-intersection faithful.

**Remark 3.4.** Note that if \(U \subseteq T(V)\) satisfies that \(U \cap (U.T(V)_{>0}) = 0\) (resp., \(U \cap (U.T(V)_{>1}) = 0\)), then any vector subspace \(U'\) of \(U\) trivially satisfies the same condition. The proposition tells us thus that any vector subspace of a left (resp., right) tensor-intersection faithful space is also left (resp., right) tensor-intersection faithful.

Moreover, it is clear that the product \(U_1.U_2\) of two left (resp., right) tensor-intersection faithful vector subspaces \(U_i \subseteq T(V)\) for \(i = 1, 2\), satisfies the same condition. By the previous example we see that \(V^{(i)}\) (for \(i \in \mathbb{N}\)) is a tensor-intersection faithful vector subspace of the tensor algebra \(T(V)\).

**Remark 3.5.** The previous proposition immediately implies that if \(U \subseteq T(V)\) is a left (resp., right) tensor-intersection faithful space, and \(W \subseteq T(V)\) is another subspace, then \(U.W\) (resp., \(W.U\)) is canonically isomorphic to the external tensor product \(U \otimes W\) (resp., \(W \otimes U\)).

In exactly the same manner as in the proof of the implication \((iii) \Rightarrow (i)\) of Proposition 2.2, we may argue to prove the following result.

**Corollary 3.6.** Let \(V\) be a vector space, and \(T(V)\) the tensor algebra. Let us consider two vector subspaces \(U' \subseteq U \subseteq T(V)\) such that \(U\) is left (resp., right) tensor-intersection faithful. Then, given any two vector subspaces \(W' \subseteq W \subseteq T(V)\), we have that \((U'.W) \cap (U.W') = U'.W\) (resp., \((W.U') \cap (W'.U) = W'.U)\), as vector subspaces of \(T(V)\).
Proof. We shall only prove the implication for the left side conditions, for the other case is analogous. Since $U$ is left tensor-intersection faithful, the previous proposition tells us that, given $I$ a finite set of indices, a linearly independent family $\{u_i : i \in I\} \subseteq U$, and arbitrary elements $w_i \in T(V)$ for $i \in I$, if $\sum_{i \in I} u_i \otimes w_i$ vanishes, then $w_i = 0$ for all $i \in I$. We recall that, by the previous Remark, $U'$ satisfies the same property. We will prove that in this case $(U'.W) \cap (U.W') = U'.W'$.

Note that $(U'.W) \cap (U.W') \supseteq U'.W'$ is always true by obvious reasons. We shall prove the reverse inclusion. Consider a element $\omega \in (U'.W) \cap (U.W')$, and fix a basis $\{u_s\} \subseteq S$ of $U$. We furthermore assume that there exists a subset $S' \subseteq S$ such that $\{u_s\}_{s \in S'}$ is a basis of $U'$. Then, there exists two families of almost all zero elements $w_s \in W'$ for $s \in S$ and $w'_{s'} \in W$ for $s' \in S'$, such that $\omega = \sum_{s \in S} u_s \otimes w_s = \sum_{s' \in S'} u_{s'} \otimes w'_{s'}$. Setting $w'_{s'} = 0$ if $s \in S \setminus S'$, we may write $\omega = \sum_{s \in S} u_s \otimes w_s = \sum_{s \in S} u_s \otimes w'_{s}$, so $\sum_{s \in S} u_s \otimes (w_s - w'_{s}) = 0$. The assumption now implies that $w_s = w'_{s}$ for all $s \in S$. Hence, $\omega \in U'.W'$, which proves the assertion.

We have the following interesting corollaries of the previous proposition, the first one of these being a direct consequence of the implication $(ii) \Rightarrow (iii)$.

**Corollary 3.7.** Let $V$ be a vector space, and $T(V)$ the tensor algebra. Let us consider a left (resp., right) tensor-intersection faithful vector subspace $U \subseteq T(V)$. If $\{U_i\}_{i \in I}$ is an arbitrary family of independent vector subspaces of $U$, then the family of vector subspaces of the tensor algebra given by $\{U_i.T(V)\}_{i \in I}$ (resp., $\{T(V).U_i\}_{i \in I}$) is also independent.

**Corollary 3.8.** Let $V$ be a vector space, and $T(V)$ the tensor algebra. Let us consider a left (resp., right) tensor-intersection faithful vector subspace $U \subseteq T(V)$. If $\{W_i\}_{i \in I}$ is an arbitrary family of independent vector subspaces of $T(V)$, then the family of vector subspaces of the tensor algebra given by $\{U.W_i\}_{i \in I}$ (resp., $\{W_i.U\}_{i \in I}$) is also independent.

Proof. We shall prove the statement for the left tensor-intersection faithfulness assumption on $U$, for the right case is completely parallel. Let $w_i \in W_i$ and $u_i \in U_i$ be collections of almost all zero elements such that $\sum_{i \in I} u_i \otimes w_i$ vanishes. We have to prove that each term of the sum vanishes. Without loss of generality we may assume that $I$ is in fact finite, and that $w_i$ is nonvanishing for all $i \in I$. Let $\{\bar{u}_s\} \subseteq S$ be a basis of $U$, and write $u_i = \sum_{s \in S} c_{si} \bar{u}_s$, where $c_{si} \in k$. We then have that

$$\sum_{i \in I} u_i \otimes w_i = \sum_{i \in I} \sum_{s \in S} c_{si} \bar{u}_s \otimes w_i = \sum_{s \in S} \bar{u}_s \otimes (\sum_{i \in I} c_{si} w_i).$$

Since $U$ is left tensor-intersection faithful, we have that $\sum_{i \in I} c_{si} w_i$ vanishes for all $s \in S$. Taking into account that the family $\{W_i\}_{i \in I}$ is independent we get that $c_{si} w_i = 0$ for all $s \in S$ and $i \in I$, which yields that $c_{si} = 0$ for all $s \in S$ and $i \in I$, for the elements $w_i$ are nonzero for all $i \in I$. So, $u_i = 0$ for all
We shall prove the corollary under the left tensor-intersection faithfulness assumption on $U$, because the right case is analogous. Let $\omega \in U.W$ be given in two different manners by the sum $\sum_{i \in I} u_i' \otimes w_i'$, where $u_i' \in U$ and $w_i' \in W$, and $I$ is a finite set of indices, and by the sum $\sum_{j \in J} u_j'' \otimes w_j''$, where $u_j'' \in U$ and $w_j'' \in W$, and $J$ is another finite set of indices. We have to prove that $\sum_{i \in I} f(u_i') \otimes g(w_i') = \sum_{j \in J} f(u_j'') \otimes g(w_j'')$, as elements of the usual tensor product $U'.W'$. Fix a basis $\{ u_s \}_{s \in S}$ of $U$, and write $u_i' = \sum_{s \in S} c_s' u_s$ and $u_j'' = \sum_{s \in S} c_s'' u_s$ for all $i \in I$ and $j \in J$, respectively, where $c_s' \in k$. Then,

$$\omega = \sum_{i \in I} u_i' \otimes w_i' = \sum_{s \in S} u_s \otimes \left( \sum_{i \in I} c_{s,i}' u_i' \right),$$

and

$$\omega = \sum_{j \in J} u_j'' \otimes w_j'' = \sum_{s \in S} u_s \otimes \left( \sum_{j \in J} c_{s,j}'' u_j'' \right).$$

Since $U$ is a left tensor-intersection faithful vector subspace of the tensor algebra, we have that

$$\sum_{i \in I} c_{s,i}' u_i' = \sum_{j \in J} c_{s,j}'' u_j'',$$

for all $s \in S$. Let us denote this element $w_s$. In particular, we have that

$$\sum_{i \in I} f(u_i') \otimes g(w_i') = \sum_{s \in S} \sum_{i \in I} c_{s,i}' f(u_s) \otimes g(w_i') = \sum_{s \in S} f(u_s) \otimes \left( \sum_{i \in I} c_{s,i}' g(w_i') \right)$$

$$= \sum_{s \in S} f(u_s) \otimes g(w_s) = \sum_{s \in S} f(u_s) \otimes \left( \sum_{j \in J} c_{s,j}'' g(w_j'') \right)$$

$$= \sum_{j \in J} \sum_{s \in S} c_{s,j}'' f(u_s) \otimes g(w_j'') = \sum_{j \in J} f(u_j'') \otimes g(w_j''),$$

which was to be shown. 

Finally, we state the last consequence of Proposition 3.2.

**Corollary 3.9.** Let $V$ be a vector space, $T(V)$ the tensor algebra, $U \subseteq T(V)$ a left (resp., right) tensor-intersection faithful vector subspace, and $W \subseteq T(V)$. Given arbitrary vector spaces $U'$ and $W'$, and linear maps $f : U \rightarrow U'$ and $g : W \rightarrow W'$, then the linear map $f \otimes g : U.W \rightarrow U' \otimes W'$ (resp., $g \otimes f : W.U \rightarrow W' \otimes U'$) given by the usual expression $(f \otimes g)(\sum_i u_i \otimes w_i) = \sum_i f(u_i) \otimes g(w_i)$ (resp., $(g \otimes f)\left( \sum_i w_i \otimes u_i \right) = \sum_i g(w_i) \otimes f(u_i)$), where $u_i \in U$, $w_i \in W$, and $I$ is a finite set of indices, is well-defined, where as usual $U.W$ (resp., $W.U$) denotes the product of $U$ and $W$ inside $T(V)$, whereas $U' \otimes W'$ (resp., $W' \otimes U'$) denotes the usual (external) tensor product.

**Proof.** We shall prove the corollary under the left tensor-intersection faithfulness assumption on $U$, because the right case is analogous. Let $\omega \in U.W$ be given in two different manners by the sum $\sum_{i \in I} u_i' \otimes w_i'$, where $u_i' \in U$ and $w_i' \in W$, and $I$ is a finite set of indices, and by the sum $\sum_{j \in J} u_j'' \otimes w_j''$, where $u_j'' \in U$ and $w_j'' \in W$, and $J$ is another finite set of indices. We have to prove that $\sum_{i \in I} f(u_i') \otimes g(w_i') = \sum_{j \in J} f(u_j'') \otimes g(w_j'')$, as elements of the usual tensor product $U'.W'$. Fix a basis $\{ u_s \}_{s \in S}$ of $U$, and write $u_i' = \sum_{s \in S} c_{s,i}' u_s$ and $u_j'' = \sum_{s \in S} c_{s,j}'' u_s$ for all $i \in I$ and $j \in J$, respectively, where $c_{s,i}' \in k$. Then,
Remark 3.10. Under the assumptions of the previous corollary, Remark \[3.5\] tells us that \( U.W \) is canonically isomorphic to \( U \otimes W \) (resp., \( W.U \) is canonically isomorphic to \( W \otimes U \)). Thus, the map \( f \otimes g \) (resp., \( g \otimes f \)) given in the corollary is given by the usual (external) tensor product of maps, which also justifies the tensor notation used.

3.2 The definition of multi-Koszul algebras

Given \( i \in \mathbb{N}_0 \) let us consider the family \( \{ \cap_{j=0}^N V^{(j)}, R^{(i)}, V^{(N-j)} \}_{N \in \mathbb{N}_0} \) of graded vector subspaces of \( T(V) \). If \( i = 0 \), the previous family coincides with \( \{ V^{(N)} \}_{N \in \mathbb{N}_0} \), which is trivially seen to be independent. Moreover, if \( i \in \mathbb{N} \), we claim that this family is also independent. Indeed, this follows from the fact that the family \( \{ V^{(N)}, R^{(i)} \}_{N \in \mathbb{N}_0} \) of graded vector subspaces of \( T(V) \) is independent by Corollary \[3.8\] and the \( N \)-th member of the former family is a vector subspace of the \( N \)-th member of the latter family. Note furthermore that any family of vector subspaces of the previous ones is thus independent.

Set the graded vector subspace of \( T(V) \) given by the direct sum

\[
^iT = \bigoplus_{N \in \mathbb{N}_0} \left( \cap_{j=0}^N V^{(j)}, R^{(i)}, V^{(N-j)} \right).
\] (3.1)

We may thus consider \( ^iT = \bigoplus_{N \in \mathbb{N}_0} T^N \), where \( ^iT^N = \cap_{j=0}^N V^{(j)}, R^{(i)}, V^{(N-j)} \), provided with another grading given by the index \( N \) of the previous direct sum. Since each \( ^iT^N \) is a graded vector space for the grading coming from the grading on \( V \), \( ^iT \) is in fact a bigraded vector space. We shall refer the grading of \( ^iT \) coming from the grading on \( V \) as usual if more indications are necessary, even though we shall call it in general the grading of \( ^iT \), without further specifications. On the other hand, we will refer to the new grading given by (3.1) as special. Note that \( ^0T = T(V) \), and that the usual and special gradings coincide in this case.

Given \( N \in \mathbb{N} \), we recall that a partition \( \bar{n} \) of \( N \) is a sequence of nonnegative integers \( \bar{n} = (n_i)_{i \in \mathbb{N}} \in \mathbb{N}_0^{(\mathbb{N})} \) of finite support (i.e. \( n_i = 0 \) for all but a finite number of indices \( i \in \mathbb{N} \)) such that \( \sum_{i \in \mathbb{N}} n_i = N \). The length of the partition \( \bar{n} \) is defined as the greatest positive integer \( l \in \mathbb{N} \) such that \( n_l \neq 0 \), and it is denoted by \( l(\bar{n}) \). We will denote by \( \text{Par}_l(N) \) the set of partitions of \( N \) of length less than or equal to \( l \), and we will write the set \( \text{Par}_N \) simply by \( \text{Par}(N) \). If \( N = 0 \), we set \( \text{Par}(0) = \emptyset \).

Finally, we will define the family \( \{ J_i \}_{i \in \mathbb{N}_0} \) of graded vector subspaces of \( T(V) \) recursively as follows. We set \( J_0 = k \). We shall first define the spaces indexed by even integers. Suppose we have defined \( J_0, J_2, \ldots, J_{2i} \) for some \( i \in \mathbb{N}_0 \). Then, we set

\[
J_{2(i+1)} = \left( \bigcap_{N \in \mathbb{N}} \left( \bigcap_{m \in \text{Par}(i)} V^{(m_1)}, J_{2n_1}, \ldots, V^{(m_i)}, J_{2n_i}, V^{(m_{i+1})} \right) \right) \cap R^{(i+1)}. \tag{3.2}
\]
On the other hand, we define
\[ J_{2i+1} = (V.J_2) \cap (J_{2i}, V), \tag{3.3} \]
Notice that \( J_1 = V, J_2 = R \) and \( J_3 = (V.R) \cap (R, V) \) as graded vector subspaces of \( T(V) \). On the other hand, note that by construction we have that \( J_{2j} \subseteq R^{(j)} \) for all \( j \in \mathbb{N}_0 \), and thus \( J_{2j+1} \subseteq (V.R^{(j)}) \cap (R^{(j)}, V) \), also for \( j \in \mathbb{N}_0 \). Furthermore, we claim that the sum appearing in the definition (3.2) of \( J_{2i+2} \) is in fact direct. In order to prove so, note that
\[
\bigcap_{J \in \mathbb{N}} V^{(m_1)} \cdot J_{2n_1} \ldots V^{(m_i)} \cdot J_{2n_i} \cdot V^{(m_{i+1})}
\]
is trivially included in
\[
\bigcap_{l=0}^N V^{(l)} \cdot J_{2i} \cdot V^{(N-l)},
\]
as one sees just by considering partitions \( n \in \text{Par}_1(i) \). It suffices to prove that the family given by these graded vector subspaces of \( T(V) \) for \( N \in \mathbb{N} \), is independent. But, since \( J_{2j} \subseteq R^{(j)} \), we see that \( \bigcap_{l=0}^N V^{(l)} \cdot J_{2i} \cdot V^{(N-l)} \subseteq T^N \) for all \( N \in \mathbb{N} \), we conclude that the sum appearing in (3.2) is in fact direct, so
\[
J_{2(i+1)} = \bigcap_{n \in \text{Par}(i) \cap \mathbb{N}} V^{(m_1)} \cdot J_{2n_1} \ldots V^{(m_i)} \cdot J_{2n_i} \cdot V^{(m_{i+1})} \cap R^{(i+1)}. \tag{3.4}
\]
Note that the direct summand appearing in the sum of the first term of the intersection of the right member and corresponding to \( N = 1 \) is included in \((V.J_2) \cap (J_{2i}, V)\), which coincides with \( J_{2i+1} \) by (3.3). Moreover, we claim that
\[
J_{2(i+1)} = \bigcap_{n \in \text{Par}(i) \cap \mathbb{N}} V^{(m_1)} \cdot J_{2n_1} \ldots V^{(m_i)} \cdot J_{2n_i} \cdot V^{(m_{i+1})} \cap R^{(i+1)}. \tag{3.5}
\]
The inclusion of the right member inside the left one is clear, so we only need to prove the reverse inclusion. Let \( \omega \in J_{2(i+1)} \). By (3.4), we have that \( \omega = \omega_1 + \omega_2 \), where \( \omega_1 \in J_{2i+1} \), and \( \omega_2 \) belongs to the first term of the intersection of the right member of (3.5). We have to prove that \( \omega_1 \) vanishes. Since \( J_{2i+1} \subseteq V.R^{(i)} \), we get that \( \omega_1 \in V.R^{(i)} \). Analogously, since
\[
\bigcap_{n \in \text{Par}(i) \cap \mathbb{N}} V^{(m_1)} \cdot J_{2n_1} \ldots V^{(m_i)} \cdot J_{2n_i} \cdot V^{(m_{i+1})}
\]
The inclusion of the statement is equivalent to the two inclusions

\[ \omega \in T(V)^{2, R}, \quad \omega_1 = \omega - \omega_2 \] should be an element of the intersection of \(V.R(\omega)\) and \(T(V)^{2, R}\), which is trivial by Proposition \([3.2]\) so \(\omega_1 = 0\), which was to be shown.

The inclusions \(J_{2j} \subseteq R(i)\), and \(J_{2j+1} \subseteq (V.R(\omega)) \cap (R(i).V)\), together with Example \([3.3]\) and Remark \([3.4]\) imply the following useful result.

**Lemma 3.11.** Let \(\{J_i\}_{i \in \mathbb{N}_0}\) be the collection of graded vector subspaces of \(T(V)\) defined by \(J_0 = k \) and the recursive identities \([3.2]\) and \([3.3]\). For all \(i \in \mathbb{N}_0\), \(J_i\) is a tensor-intersection faithful vector subspace of \(T(V)\).

Given \(j \in \mathbb{N}_0\), let us consider the family of graded vector subspaces \(\{N=0 \}V(\omega), J_{2j+1},N(\omega)\}_{N \in \mathbb{N}_0}\) of the tensor algebra \(T(V)\). Since the \(N\)-th member of this family is trivially included in \(R(i).V(\omega)\), and the family \(\{R(i).V(N\omega)\}_{N \in \mathbb{N}_0}\) is independent by Corollary \([3.8]\), the former family is independent. We define thus

\[ j^1T = \bigoplus_{N \in \mathbb{N}_0} \left( \bigcap_{i=0}^N V(i), J_{2j+1},V(N\omega) \right), \]

which is considered as a bigraded vector space, where the first grading, which will be called *usual*, comes from that of \(V\), and the second one, called *special*, comes from the direct sum decomposition of the definition. Note that \(0^1T = S(T(V)^{\geq 0})\) and \(1^1T = S(T^{\geq 0})\) as bigraded vector spaces, where \(S\) is the shift functor for the special grading, i.e. \(S(E)^N = E^{N+1}\). Furthermore, by \([3.4]\) we have that \(J_{2j} \subseteq j^1T\) for all \(j \in \mathbb{N}\).

We have the following simple result.

**Lemma 3.12.** Let \(\{J_i\}_{i \in \mathbb{N}_0}\) be the collection of graded vector subspaces of \(T(V)\) defined by \(J_0 = k \) and the recursive identities \([3.2]\) and \([3.3]\). For all \(i \in \mathbb{N}_0\), we have the inclusion

\[ J_i \subseteq (R.J_{i-2}) \cap (J_{i-2}.R). \]

**Proof.** The inclusion of the statement is equivalent to the two inclusions

\[ J_i \subseteq R.J_{i-2}, \quad \text{and} \quad J_i \subseteq J_{i-2}.R. \]

We shall only prove the second one, for the other is completely analogous. Let us first suppose that \(i\) is even. In this case, we trivially see from definition \([3.2]\) that

\[ J_i \subseteq (J_{i-2}.T(V)^{\geq 0}) \cap R(i^2), \]
by only considering the partitions of length 1. On the other hand, since \( J_{i-2} \subseteq R^{(i-2)/2} \), Corollary 3.6 and Lemma 3.11 yield that

\[
J_i \subseteq (J_{i-2}.T(V)_{>0}) \cap (R^{(i-3)/2}.R) \subseteq J_{i-2}.R.
\]

If \( i \) is odd, from 3.2 and 3.3 we easily show that

\[
J_i \subseteq (J_{i-2}.T(V)_{>0}) \cap (V.R^{(i-3)/2}.R).
\]

Using that \( J_{i-2} \subseteq V.R^{(i-3)/2} \), Corollary 3.6 and Lemma 3.11 we see that \( J_i \subseteq J_{i-2}.R \), which was to be shown. The lemma is thus proved.

Moreover, the previous lemma can also be strengthen in the next form.

**Corollary 3.13.** Let \( \{ J_i \}_{i \in \mathbb{N}_0} \) be the collection of graded vector subspaces of \( T(V) \) defined by \( \delta_0 = k \) and the recursive identities (3.2) and (3.3). Then, for all \( i \in \mathbb{N}_0 \) and \( j \geq i \) except if \( i \) is odd and \( j \) is even, we have the inclusion

\[
J_j \subseteq (J_{j-1}.J_i) \cap (J_i.J_{j-i}).
\]

On the other hand, if \( i \in \mathbb{N} \) is odd and \( j \geq i \) is even, we have the contention

\[
J_j \subseteq (T(V).J_{j-i}.J_i) \cap (J_{j-i}.T(V).J_i) \cap (J_{j-i}.J_i.T(V)),
\]

and also the same inclusion for \( i \) and \( j-i \) interchanged.

**Proof.** We note that the condition for the first inclusion means that either \( i \in \mathbb{N}_0 \) is even and \( j \geq i \) is arbitrary, or both \( i \in \mathbb{N} \) and \( j \geq i \) are odd. Let us prove the first inclusion of the statement for \( i \in \mathbb{N}_0 \) even and \( j \geq i \). If \( i = 0 \), the inclusion is obvious, so we will assume that \( i \geq 2 \). We shall only show the contention \( J_j \subseteq J_{j-i}.J_i \) for the other is analogous. By the previous lemma we have that \( J_j \subseteq J_{j-i}.J_i \), which in turn implies that \( J_j \subseteq J_{j-i}.J_i \), by Corollary 3.6. If \( j \) is odd, then \( J_j \subseteq V.J_{j-i} \) by definition, and, since \( j-1 \) is even, Lemma 3.12 tells us that \( J_{j-1} \subseteq R^{(j-1)/2}.J_i \). Therefore, \( J_j \subseteq V.R^{(j-1)/2}.J_i \). Corollary 3.6 now gives the desired contention in the case that \( j \) is odd.

Let us now proceed to prove the first inclusion of the statement for \( i \in \mathbb{N} \) odd and \( j \geq i \) also odd, which is the remaining case. We note however that this case follows from the previously analysed case when \( i \in \mathbb{N} \) is even and \( j \geq i \) is arbitrary, by interchanging \( i \) with \( j-i \).

Let us now consider the second inclusion of the statement, which has the assumption that \( i \in \mathbb{N} \) is odd and \( j \geq i \) is even. In this case, we will first show the contention \( J_j \subseteq T(V).J_{j-i}.J_i \). By (3.5) we see that \( J_j \subseteq T(V).J_{j-i}.V \). Since \( i-1 \) is even, the previous lemma tells us that \( J_{j-1} \subseteq J_{j-i}.R^{((i-1)/2)} \), so \( J_j \subseteq T(V).J_{j-i}.R^{((i-1)/2)}.V \). On the other hand, it is trivial to see from the definitions (3.2) and (3.3) that \( J_j \subseteq T(V).J_i \). Now, using the inclusions
By Corollary 3.6 we conclude that \( J_j \subseteq T(V).J_{j-1} \subseteq T(V) \), which was to be shown. The inclusion \( J_j \subseteq J_{j-1}.J_i.T(V) \) is proved in an analogous manner.

Finally, we shall prove the inclusion \( J_j \subseteq J_{j-1}.T(V).J_i \), under the assumption that \( i \in \mathbb{N} \) is odd and \( j \geq i \) is even. The definitions \( 3.2 \) and \( 3.3 \) tell us that \( J_j \subseteq (T(V).J_i) \cap R^{(j/2)} \), and the latter is obviously included in

\[
(T(V).J_i) \cap (R^{((j-1)/2)}.V.T(V).V.R^{((j-1)/2)}).
\]

In the same manner, we have the inclusion \( J_j \subseteq (J_{j-1}.T(V)) \cap R^{(j/2)} \), and the latter is obviously included in

\[
(J_{j-1}.T(V)) \cap (R^{((j-1)/2)}.V.T(V).V.R^{((j-1)/2)}).
\]

By Corollary 3.6 we get the desired contention. The corollary is thus proved. \( \square \)

In order to handle the elements of the graded vector spaces \( J_i \), we will use the following notation. Let us suppose that \( W_1, \ldots, W_m \) are graded vector subspaces of \( T(V)_{>0} \). An element \( \omega \in W_1.\ldots.W_m \subseteq T(V)_{>0} \) can be written as a finite sum

\[
\sum_{(\omega)} \omega^{(1)}|\omega^{(2)}| \cdots |\omega^{(m-1)}|\omega^{(m)},
\]

where the elements \( \omega^{(i)} \in W_i \) for \( i = 1, \ldots, m \). If the vector subspaces \( W_{j_1}, \ldots, W_{j_l} \) of \( T(V)_{>0} \) for \( 1 \leq j_1 < \cdots < j_l \leq m \) coincide with the graded vector subspace \( V \subseteq T(V)_{>0} \), we shall typically denote this by writing a bar over each of the factors \( \omega^{(j_1)}, \ldots, \omega^{(j_l)} \). Moreover, we shall usually omit the implicitly assumed sum, which will be implicitly assumed.

Let \( j \in \mathbb{N}_0 \). By \( 3.3 \), we see that, if \( \omega \in J_{2j+1} \), then we may write it either in the form \( \omega = \bar{\omega}(1)|\omega(2) | \cdots |\omega(m-1)|\omega(m) \), where the sum is implicit as explained before. We may thus consider the elements \( (\pi \otimes \text{id}_{J_{2j}})\omega \otimes 1 \) and \( 1 \otimes (\text{id}_{J_{2j}} \otimes \pi)\omega \) in \( A \otimes J_{2j} \otimes A \), where \( \pi : T(V) \rightarrow A \) denotes the canonical projection. They may be written as \( \pi(\bar{\omega}(1))|\omega(2)|1 \) and \( 1|\omega(1)|\pi(\bar{\omega}(2)) \), respectively. Note that these elements are uniquely defined by Corollary 3.9.

We then get the map \( J_{2j+1} \rightarrow A \otimes J_{2j} \otimes A \) given by

\[
\omega \mapsto \pi(\bar{\omega}(1))|\omega(2)|1 - 1|\omega(1)|\pi(\bar{\omega}(2)),
\]

which we may also simply write as

\[
\omega \mapsto \bar{\omega}(1)|\omega(2)|1 - 1|\omega(1)|\omega(2),
\]

using the obvious identification of \( V \) inside \( A \). We shall denote by \( \delta_{2j+1}^b : A \otimes J_{2j+1} \otimes A \rightarrow A \otimes J_{2j} \otimes A \), the \( A^e \)-linear extension of the former. Note that it is a morphism of graded \( A \)-bimodules. Moreover, notice that the morphism \( \delta_{2j+1}^b \) given before coincides with the corresponding differential (denoted in the same way) given in \( 2.3 \).
On the other hand, for \( j \in \mathbb{N} \), we consider a map \( j^{-1}\hat{T} \to A \otimes J_{2j-1} \otimes A \) of graded vector spaces given by the sum of the maps \( j^{-1}\hat{T}^N \to A \otimes J_{2j-1} \otimes A \) of graded vector spaces for all \( N \in \mathbb{N}_0 \), which we now define. Given \( \omega \in j^{-1}\hat{T}^N \), then for all \( 0 \leq l \leq N \), we may write it in the form \( \omega = \omega_1^l \omega_2^l \omega_3^l \), where \( \omega_1^l \in V^{(0)} \), \( \omega_2^l \in J_{2j-1} \) and \( \omega_3^l \in V^{(N-l)} \), where the sum is explicit as explained before. For each \( 0 \leq l \leq N \) we may thus consider the element
\[
\pi_{1,v}^i (\otimes \text{id}_{J_{2j-1}} \otimes \pi_{1,v}^{(N-l)}) (\omega) = \pi(\omega_1^l) \pi(\omega_2^l) \pi(\omega_3^l),
\]
in \( A \otimes J_{2j-1} \otimes A \). Again, note that these elements are uniquely defined by Corollary \ref{cor}. We have thus obtained a map
\[
j^{-1}\hat{T} \to A \otimes J_{2j-1} \otimes A \]
of graded vector spaces formed by the sum of the maps
\[
j^{-1}\hat{T}^N \to A \otimes J_{2j-1} \otimes A \]
given by
\[
\omega \mapsto \sum_{i=0}^{N} \pi(\omega_1^i) \pi(\omega_2^i) \pi(\omega_3^i).
\]
The inclusion \( j_{2j} \subseteq j^{-1}\hat{T} \) of graded vector spaces in turn induces a map of graded vector spaces
\[
j_{2j} \to A \otimes J_{2j-1} \otimes A.
\]
We shall denote by \( \delta_{2j}^i : A \otimes J_{2j} \otimes A \to A \otimes J_{2j-1} \otimes A \), the \( A^e \)-linear extension of the former. Note that it is a morphism of graded \( A \)-bimodules. Note that \( \delta_{2j}^i : A \otimes J_{2j} \otimes A \to A \otimes J_{2j-1} \otimes A \) coincides with the one given by the same name in \ref{prop}.

We have thus defined a collection of graded \( A \)-bimodules \( \{ K_{L-R}(A) \}_{i} = A \otimes J_i \otimes A \}_{i \in \mathbb{N}_0} \) provided with morphisms of graded \( A \)-bimodules \( \delta_i^b : K_{L-R}(A)_{i} \to K_{L-R}(A)_{i-1} \) for \( i \in \mathbb{N}_0 \).

**Proposition 3.14.** Let \( A \) be a locally finite dimensional nonnegatively graded algebra, and let \( \{ J_i \}_{i \in \mathbb{N}_0} \) be the collection of graded vector subspaces of \( T(V) \) defined by \( J_0 = k \) and the recursive identities \ref{prop} and \ref{prop}. We consider the collection \( \{ K_{L-R}(A) = A \otimes J_i \otimes A \}_{i \in \mathbb{N}_0} \) of graded \( A \)-bimodules provided with morphisms of graded \( A \)-bimodules \( \delta_i^b : K_{L-R}(A)_{i} \to K_{L-R}(A)_{i-1} \) for \( i \in \mathbb{N} \), which were defined in the three previous paragraphs. It gives in fact a complex of graded \( A \)-bimodules, i.e. \( \delta_{i-1}^b \circ \delta_i^b = 0 \) for \( i \in \mathbb{N}_{\geq 2} \).

**Proof.** We have thus to prove that \( \delta_{i-1}^b \circ \delta_i^b = 0 \) for \( i \in \mathbb{N}_{\geq 2} \). By the exactness of the complex \ref{prop}, we get that the previous equality holds for \( i = 2 \). Let us suppose that \( i \geq 3 \). We first suppose that \( i \) is odd, and we further assume that

---

Documenta Mathematica 18 (2013) 1301–1347
On the Multi-Koszul Property for Connected Algebras

3.12 tells us that \( \omega \in J_i \) is given by \( \omega = \sum_{N \in \mathbb{N}_0} \omega^N = \sum_{N \in \mathbb{N}_0} \omega'^N \), where \( \omega^N \in V^{(i-3)/2} \mathcal{T} \), \( \omega'^N \in (i-3)/2 \mathcal{T} V \), and both sums have finite support. Then

\[
(\delta^b_{i-1} \circ \delta^b_i)(1|\omega|1) = \delta^b_{i-1} (\sum_{N \in \mathbb{N}_0} \omega^N) = \delta^b_{i-1} (\sum_{N \in \mathbb{N}_0} \omega'^N)
\]

For Lemma 3.12, \( \omega \in J_i \) is given by

\[
(\delta^b_{i-1} \circ \delta^b_i)(1|\omega|1) = \delta^b_{i-1} (\sum_{N \in \mathbb{N}_0} \omega^N) = \delta^b_{i-1} (\sum_{N \in \mathbb{N}_0} \omega'^N)
\]

On the other hand, let \( i \) be even and and let \( \omega \in J_i \) be of the form \( \omega = \sum_{N \in \mathbb{N}_0} \omega^N \), where \( \omega^N \in (i-2)/2 \mathcal{T} \mathcal{V} \) and the sum has finite support. We have that

\[
(\delta^b_{i-1} \circ \delta^b_i)(1|\omega|1) = \delta^b_{i-1} (\sum_{N \in \mathbb{N}_0} \omega^N) = \delta^b_{i-1} (\sum_{N \in \mathbb{N}_0} \omega'^N)
\]

A telescopic cancellation tells us that the last member coincides with

\[
\sum_{N \in \mathbb{N}_0} \pi(\omega_{(1)})|\omega_{(2)}|^N \neq 0 \quad \text{and both sums have finite support. Then}
\]

By a direct telescopic cancellation, the last member coincides with

\[
\sum_{N \in \mathbb{N}_0} \omega^N \pi(\omega_{(2)}^N)|\omega_{(3)}^N|1 = \sum_{N \in \mathbb{N}_0} \omega_{(1)}^N \pi(\omega_{(2)}^N)|\omega_{(3)}^N|1,
\]

where we have set without loss of generality that \( \omega_{(4)} = 1 \) and \( \omega_{(1)}^N = 1 \). The sum vanishes, for Lemma 3.12 tells us that

\[
\omega = \sum_{N \in \mathbb{N}_0} \omega_{(1)}^N \omega_{(2)}^N \omega_{(3)}^N \in R \cdot J_{i-2},
\]

\[
\omega = \sum_{N \in \mathbb{N}_0} \omega_{(2)}^N \omega_{(3)}^N \omega_{(4)}^N \in J_{i-2} \cdot R.
\]
where we have set without loss of generality that $\omega_{N,N}^{(4)} = 1$ and $\omega_{(1)}^{N,0} = 1$. The previous sum vanishes, since Lemma 3.12 implies that

$$\omega = \sum_{N \in \mathbb{N}_0} \omega_{N,N}^{(1)} \omega_{N,N}^{(2)} \bar{\omega}_{N,N}^{(3)} \in R.J_i - 2,$$

$$\omega = \sum_{N \in \mathbb{N}_0} \omega_{N,0}^{N,0} \omega_{N,0}^{(2)} \bar{\omega}_{N,0}^{N,0} \in J_i - 2.R.$$

This proves the proposition. 

By the previous statement $(K_{L-R}(A)_*, \delta^0_*)$ is a complex of graded $A$-bimodules. It can also be regarded as an augmented complex $\delta^0_* : K_{L-R}(A)_* \to A$ where $\delta^0_*$ is given by the multiplication of $A$, as in (2.3). Hence, the augmented complex $K_{L-R}(A)_*$ coincides with the minimal projective resolution of $A$ as an $A$-bimodule up to homological degree 2. It will be called the multi-Koszul bimodule complex of $A$.

We also define the (augmented) complex $(K(A)_*, \delta_*)$ (resp., $(K(A)'_*, \delta'_*)$) of graded free left (resp., right) $A$-modules given by $(K_{L-R}(A)_* \otimes_A k, \delta_* \otimes_A \text{id}_k)$ (resp., $(k \otimes_A K_{L-R}(A)_*, \text{id}_k \otimes_A \delta'_*)$), which will be called the left (resp., right) multi-Koszul complex of $A$. We note that in either case the differentials are $A$-linear maps preserving the degree. We also remark that the left (resp., right) multi-Koszul complex of $A$ coincides with the minimal projective resolution of the left (resp., right) module $k$ seen at the beginning of this section up to homological degree 2.

**Definition 3.15.** Let $A$ be a locally finite dimensional nonnegatively graded connected algebra. We say that $A$ is left (resp., right) multi-Koszul if the (augmented) left (resp., right) multi-Koszul complex of $A$ defined in the previous paragraphs provides a resolution of the trivial left (resp., right) $A$-module $k$, and in this case it is called the left (resp., right) multi-Koszul resolution for $A$.

**Remark 3.16.** If $A$ is left (resp., right) multi-Koszul, then the (augmented) left (resp., right) multi-Koszul complex of $A$ is in fact a minimal projective resolution of the trivial left (resp., right) $A$-module $k$. Indeed, the left (resp., right) multi-Koszul resolution for $A$ is minimal, by the comments in the antepenultimate paragraph of Section 2, because the induced differential of the complex $k \otimes_A K(A)_*$ (resp., $K(A)' \otimes_A k$) vanishes, due to (3.3) and (3.5). It is also straightforward to see that an algebra is left (resp., right) multi-Koszul if and only if its left (resp., right) multi-Koszul complex defined above is acyclic in positive homological degrees.

We have the following natural result, following the lines of [1], Thm. 4.4.

**Proposition 3.17.** Let $A$ be a locally finite dimensional nonnegatively graded connected algebra. The following statements are equivalent:

(i) $A$ is left multi-Koszul.

(ii) The augmented multi-Koszul bimodule complex $\delta^0_* : K_{L-R}(A)_* \to A$ is exact.
(iii) $A$ is right multi-Koszul.

Proof. We shall prove the equivalence of (i) and (ii), the equivalence between (ii) and (iii) being analogous. Suppose that $A$ is left multi-Koszul. Applying the functor $(-) \otimes_A k$ to the augmented multi-Koszul bimodule complex, we obtain the (augmented) complex $(K(A)_{\bullet}, \delta_{\bullet})$, which is exact when $A$ is left multi-Koszul. Since the $A$-bimodules $K_{L-R}(A)_i$ are graded-free and left bounded for all $i \in \mathbb{N}_0$ [7], Prop. 4.1, (or [4], Lemme 1.6) implies that the augmented complex $\delta^0_{\bullet} : K_{L-R}(A)_\bullet \to A$ is exact.

Assume now that the complex $K_{L-R}(A)_\bullet$ is exact in positive degrees. Since $\delta^0_{\bullet} : K_{L-R}(A)_\bullet \to A$ is exact, it is a projective resolution of $A$ in the category of graded left bounded right $A$-modules. Since $A$ is a projective right $A$-module, the complex $\delta^0_{\bullet} : K_{L-R}(A)_\bullet \to A$ is homotopically trivial as a complex of objects of graded right $A$-modules. Therefore, its image under the functor $(-) \otimes_A k$ is a fortiori homotopically trivial (as a complex of vector spaces). Since this image is the left multi-Koszul complex of $A$, it is exact in positive homological degrees, so $A$ is left multi-Koszul.

By the previous results, we shall usually simply say that an algebra $A$ is multi-Koszul, unless we want to emphasize the use of the corresponding complexes.

Remark 3.18. Suppose that $A$ is left multi-Koszul. Since there is an obvious isomorphism of complexes of the form $k \otimes_A K_{L-R}(A)_\bullet \simeq k \otimes_A K(A)_\bullet$, having in fact vanishing differential, the comments in the antepenultimate paragraph of Section 2 tell us that the complex $(K_{L-R}(A)_\bullet, \delta^0_{\bullet})$ is a minimal projective resolution of $A$ in the category of graded $A$-bimodules.

Remark 3.19. Let $A = T(V)/(R)$ be a locally finite dimensional nonnegatively graded connected algebra. Let us consider a different positive grading on $V$ such that $R$ also remains a graded vector subspace of $T(V)^{\geq 2}$ for the new grading on $V$ (e.g. we double the grading of $V$). Since the condition (2.1) defining a space of relations is also independent of the grading, we see that $R$ is still a space of relations for the algebra $A$ provided with the new grading. This in turn implies that the multi-Koszul complex of $A$ is grading independent. In particular, $A$ is multi-Koszul for the former grading if and only if it is multi-Koszul for the new grading, which we may roughly reformulate as stating that the multi-Koszul property introduced in Definition 3.15 is also grading independent.

Before proceeding further we would like to state some comments on the previous definition. As it may have been noticed, we have used the same terminology (i.e. left or right multi-Koszul) as the one introduced in [19], Def. 3.1 and Rmk. 3.3, where the nonnegatively graded connected algebras were further assumed to be finitely generated in degree one and with a finite number of relations. We claim that the new definition introduced here coincides with the one considered in [19] if the algebras satisfy the assumptions of the latter article.
In order to do so, we shall use the following little variation of the notation used in [19], Def. 3.1. Suppose for the moment that $A = T(V)/\langle R \rangle$, where $V$ is a finite dimensional vector space considered to be concentrated in degree 1, and $R = \oplus_{s \in S} R_s \subseteq T(V)_{\geq 2}$ for $S \subseteq \mathbb{N}_{\geq 2}$, is a finite dimensional graded vector space. For each $s \in \mathbb{N}_{\geq 2}$, we recall that the map $n_s : \mathbb{N}_0 \to \mathbb{N}_0$ is given by 
\[ n_s(2j) = sj, \quad n_s(2j + 1) = sj + 1. \]
If $s \in S$, we will denote
\[ \tilde{J}_s = n_s(i) - s \cap \bigoplus_{l=0}^{s-1} V(l).R_s.V(n_s(i) - s - l), \]
for $i \geq 2$, and $\tilde{J}_i = V(i)$ for $i = 0, 1$. Moreover, we define
\[ \tilde{J} = \bigoplus_{s \in S} \tilde{J}_s, \]
if $i \geq 2$, and $\tilde{J}_i = V(i)$, if $i = 0, 1$.

The differential of the left multi-Koszul complex $(A \otimes \tilde{J} \otimes \mathbb{N}_0)$ introduced in [19], Def. 3.1, will be denoted by $\tilde{\delta}_\bullet$.

**Proposition 3.20.** Let $A$ be a nonnegatively graded connected algebras, which is further assumed to be finitely generated in degree one and with a finite number of relations, i.e. $A = T(V)/\langle R \rangle$, where $V$ is a finite dimensional vector space considered to be concentrated in degree 1, and $R = \oplus_{s \in S} R_s \subseteq T(V)_{\geq 2}$ for $S \subseteq \mathbb{N}_{\geq 2}$, is a finite dimensional graded vector space. Then, the left (resp., right) multi-Koszul complex in the Definition 3.15 coincides with the left (resp., right) multi-Koszul complex introduced in [19], Definition 3.1 (resp., Remark 3.3), which in turn implies that $A$ is left (resp., right) multi-Koszul in the sense of Definition 3.15 if and only if it is left (resp., right) multi-Koszul in the sense introduced in [19], Definition 3.1 (resp., Remark 3.3).

**Proof.** We shall prove the statement for the left multi-Koszul complexes, the right case being analogous. Actually, the equivalence of both right multi-Koszul properties could also follow from the statement for the left case from Proposition 3.17 and [19], Cor. 3.13.

Let us first note that the left multi-Koszul complex introduced in Definition 3.15 trivially coincides with the one considered in [19], Def. 3.1, up to homological degree 2. Note that $J_2 = R = \tilde{J}_2$. They further coincide up to homological degree 3, since in this case
\[ J_3 = (V,R) \cap (R,V) = \bigoplus_{s \in S} (V,R_s) \cap (R_s,V) = \bigoplus_{s \in S} \tilde{J}_s = \tilde{J}_3, \]
where we have used Fact 3.1 since $V$ is concentrated in degree 1 and $R_s$ in degree $s$, and the differentials in homological degree 3 for both complexes clearly coincide.
ON THE MULTI-KOSZUL PROPERTY FOR CONNECTED ALGEBRAS 1325

Let us suppose that both complexes coincide up to homological degree \( d \), i.e.
we have that \( J_i = \tilde{J}_i \) and \( \delta_i = \tilde{\delta}_i \) for \( i \leq d \).
Let us first assume that \( d \) is odd. We claim that
\[
\tilde{J}_{2j+1} = (V, \tilde{J}_{2j}) \cap (\tilde{J}_{2j}, V),
\] (3.7)
for \( j \in \mathbb{N}_0 \). Indeed, if \( j = 0 \) the previous identity is trivial, whereas for \( j \in \mathbb{N} \) it
follows from a rather simple computation
\[
(V, \tilde{J}_{2j}) \cap (\tilde{J}_{2j}, V) = \left( V, \left( \bigoplus_{s \in S} \tilde{J}_{2j}^s \right) \right) \cap \left( \left( \bigoplus_{s \in S} \tilde{J}_{2j}^s \right), V \right)
\]
\[
= \left( \bigoplus_{s \in S} V, \tilde{J}_{2j}^s \right) \cap \left( \bigoplus_{s \in S} \tilde{J}_{2j}^s, V \right)
\]
\[
= \bigoplus_{s \in S} \left( (V, \tilde{J}_{2j}^s) \cap (\tilde{J}_{2j}^s, V) \right)
\]
\[
= \bigoplus_{s \in S} \tilde{J}_{2j+1}^s = \tilde{J}_{2j+1},
\]
where we have used Fact 3.3 in the antepenultimate equality, taking into account that \( V \) is concentrated in degree 1 and each \( \tilde{J}_{2j}^s \) is concentrated in degree \( n_s(2j) = sj \). Now, (3.3) and (3.7) tell us that \( J_{d+1} = \tilde{J}_{d+1} \). It is trivial to verify
that the differentials also satisfy the identity \( \delta_{d+1} = \tilde{\delta}_{d+1} \).
On the other hand, let us suppose that \( d \) is even. In this case we claim that
\[
\tilde{J}_{2j} = \left( \bigoplus_{N \in \mathbb{N} \atop n \in \text{Par}(j-1)} \bigcap_{m \in \text{Par}_N(j)} V^{(m_1)} \cdot \tilde{J}_{2n_1} \cdots V^{(m_{j-1})} \cdot \tilde{J}_{2n_{j-1}} \cdot V^{(m_j)} \right) \cap R^j,
\] (3.8)
for all \( j \in \mathbb{N} \), which we prove as follows. It is trivially verified for \( j = 1 \), since the right term in that case is just \( T(V)_{>0} \cap R \), which coincides with \( \tilde{J}_2 = R \).
Let us consider \( j \geq 2 \). We first show that the right member of (3.8) contains
the left one. Indeed, note that
\[
V^{(m_1)} \cdot \tilde{J}_{2n_1} \cdots V^{(m_{j-1})} \cdot \tilde{J}_{2n_{j-1}} \cdot V^{(m_j)}
\]
trivially includes
\[
V^{(m_1)} \cdot \tilde{J}_s \cdots V^{(m_{j-1})} \cdot \tilde{J}_s \cdot V^{(m_j)},
\]
for all \( s \in S \), so the right member of (3.8) includes
\[
\left( \bigoplus_{N \in \mathbb{N} \atop n \in \text{Par}(j-1)} \bigcap_{m \in \text{Par}_N(j)} V^{(m_1)} \cdot \tilde{J}_s \cdots V^{(m_{j-1})} \cdot \tilde{J}_s \cdot V^{(m_j)} \right) \cap R^j,
\]
for all \( s \in S \), which trivially coincides with
\[
\left( \bigoplus_{N \in \mathbb{N} > \text{Par}(j-2)} \bigcap_{l=0}^N V^{(l)} \cdot R_s \cdot V^{(N-l)} \right) \cap R^j,
\]
for all \( s \in S \). By Fact 3.1, the latter space is given by
\[
\bigoplus_{N \in \mathcal{N}(s)} \left( \bigcap_{l=0}^{N} V^{(l)} . R_{s} . V^{(N-l)} \right) \cap R_{N+s}^{(j)},
\]
for all \( s \in S \).

On the other hand, note that \( (R_{j})_{n} \) is the direct sum of the independent vector subspaces \( \{ R_{s_1} \ldots R_{s_j} \}_{s_1 + \ldots + s_j = n} \). This follows from a simple recursive argument using Corollary 3.8. In particular, \( \omega \in (R_{j})_{n} \) has to be written in a unique manner as a sum of unique elements \( \omega_{s_1, \ldots, s_j} \in R_{s_1} \ldots R_{s_j} \) where we assume that \( s_1 + \cdots + s_j = n \). It is thus trivial that \( \bigcap_{l=0}^{N} V^{(l)} . R_{s} . V^{(N-l)} \) is included in \( R_{N+s}^{(j)} \) if \( N = s(j-1) \). Moreover, another application of Corollary 3.8 tells us that the intersection of them vanishes otherwise. This in turn implies that
\[
\bigoplus_{N \in \mathcal{N}(s)} \left( \bigcap_{l=0}^{N} V^{(l)} . R_{s} . V^{(N-l)} \right) \cap R_{N+s}^{(j)},
\]
coincides with \( \tilde{J}_{j}^{s} \) for all \( s \in S \), and we then conclude that the right member of (3.8) contains the left one.

For the reverse inclusion, note first that the right member of (3.8) is trivially included in
\[
\left( \bigoplus_{N \in \mathcal{N}} \bigcap_{l=0}^{N} V^{(l)} . \tilde{J}_{j-1}^{s} . V^{(N-l)} \right) \cap R_{j}^{(j)}
\]

\( = \bigoplus_{N \in \mathcal{N}} \bigcap_{l=0}^{N} V^{(l)} . \tilde{J}_{j-1}^{s} . V^{(N-l)} \cap R_{j}^{(j)} \)

\( = \bigoplus_{N \in \mathcal{N}} \bigcap_{s \in S} \bigcap_{l=0}^{N+n_{j}(2j-3)} V^{(l)} . R_{s} . V^{(N+n_{j}(2j-3)-l)} \cap R_{j}^{(j)} \),

where we used Fact 3.1 in the second equality, for \( V \) is concentrated in degree 1 and \( \tilde{J}_{j-1}^{s} \) is concentrated in degree \( n_{j}(2j-1) = sj - s + 1 \). Using Fact 3.1 once more in the last member, we get that the latter should be the direct sum of the intersection of the \( n \)-th homogeneous components of each corresponding term, i.e. the intersection of the \( n \)-th direct summand of \( R_{j}^{(j)} \) and
\[
\bigoplus_{s \in S \cap S_{j,n}} \bigcap_{l=0}^{n-s} V^{(l)} . R_{s} . V^{(n-s-l)} \quad (3.9)
\]
where \( S_{j,n} = \{ s \in S : n_{j}(2j-1) < n \} \). Note that if \( s \in S_{j,n} \), then \( n > s \) for \( j \geq 2 \). As explained in the previous paragraph, a direct application of
Corollary \[5.7\] tells us that the former intersection should coincide with
\[\bigoplus_{s \in S} \left( \bigcap_{i=0}^{n-s} (V(i^1) \cdot R_s \cdot V(n-s-l)) \cap (R_s^{(j)})_n \right),\]
which is directly seen to be equal to
\[\bigoplus_{s \in S} \left( \bigcap_{i=0}^{n-s} (V(i^1) \cdot R_s \cdot V(n-s-l)) \cap (R_s^{(j)})_n \right) = \bigoplus_{s \in S} (\tilde{J}^s_{d+1})_n = (\tilde{J}_{d+1})_n.\]

This proves that the left member of \[5.8\] contains the right one, and so the
equality of the assertion \[5.8\] holds.

Finally, \[3.3\] and \[5.8\] imply that \(J_{d+1} = \tilde{J}_{d+1}\). It is also direct to check that the
differentials \(\delta_{d+1}\) and \(\tilde{\delta}_{d+1}\) coincide. The proposition is thus proved. \(\square\)

**Remark 3.21.** Since the previous definition of left or right multi-Koszul property
coincides with the one considered in \[19\] if the algebras satisfy the assumptions of the
latter article, by \[19,\] Rmk. 3.4, we get that it also coincides with the corresponding
one given in \[3,\] Def. 2.10 (or \[5,\] Section 5), if the algebra is homogeneous. Thus, a
homogeneous algebra is left (resp., right) multi-Koszul if and only if it is generalized
left (resp., right) Koszul.

**Example 3.22.** We will provide a collection of examples of multi-Koszul algebras,
which was one of the main motivations of this article. The space of generators \(V\)
of these graded algebras does not lie in degree 1, so they cannot be considered as multi-
Koszul algebras for the definition given in \[19\] in any natural manner.

Given two nonnegative integers \(n, s \in \mathbb{N}_2 \setminus \{(0, 0)\}\), and a collection of symmetric
\((s \times s)\)-matrices \((\Gamma^{i}_{a,b})\), for \(i = 1, \ldots, n\) \((a, b = 1, \ldots, s)\), the (associative)
Yang-Mills algebra \(\text{YM}(n, s)^\Gamma\) over an algebraically closed field \(k\) of characteristic
zero is defined as follows. Take \(V = V_2 \oplus V_3\) be a graded vector space over \(k\), where
\(\dim_k(V_2) = n\) and \(\dim_k(V_3) = s\), and choose in fact a (homogeneous) basis \(B = B_2 \cup B_3\) of \(V\),
where \(B_2 = \{x_1, \ldots, x_n\}\) and \(B_3 = \{z_1, \ldots, z_s\}\), with \(|x_i| = 2\) for all
\(i = 1, \ldots, n\), and \(|z_a| = 3\) for all \(a = 1, \ldots, s\). We suppose further that the matrices
\((\Gamma^{i}_{a,b})\) satisfy the nondegeneracy assumption explained in the third paragraph before
Rmk. 1 of \[18\].

The graded algebra \(\text{YM}(n, s)^\Gamma\) is given by the quotient of the graded free algebra
\(T(V)\), by the homogeneous relations given by
\[r_{0,i} = \sum_{j=1}^{n} [x_j, [x_j, x_i]] - \frac{1}{2} \sum_{a, b=1}^{s} \Gamma^{i}_{a,b}[z_a, z_b],\]
\[r_{1,a} = \sum_{i=1}^{n} \sum_{b=1}^{s} \Gamma^{i}_{a,b}[x_i, z_b],\]
for \(i = 1, \ldots, n\) and \(a = 1, \ldots, s\), respectively. The bracket \([,]\) denotes the graded
commutator, i.e. \([a, b] = ab - (-1)^{|a||b|}ba\) for \(a, b \in B\). They have been previously
Let us consider the graded algebras $A$ and $B$ as given below, we see that these graded algebras are multi-Koszul.

3.3 An equivalent description of multi-Koszul algebras

We would like to make some comments on the left multi-Koszul complex of $A$. The obvious statements for the right multi-Koszul complex trivially hold.

First, given $i \in \mathbb{N}_0$, note that the map of graded vector spaces $J_{i+1} \to \text{Ker}(\delta_i)$ given by the restriction of $\delta_{i+1}$ is injective. This can be proved as follows. The cases $i = 0, 1$ are immediate, so we will suppose that $i \geq 2$. Furthermore, the kernel of the restriction of $\delta_{i+1}$ to $J_{i+1}$ is easily seen to be $J_{i+1} \cap (I.J_i)$. The first term of this intersection is included in $R.J_i$ by Lemma 3.12 whereas the second is included in $I.V.J_{i-1}$ if $i$ is odd, by (3.3), and it is included in $I.T(V)_{\geq 0}.J_{i-1}$ if $i$ is even, by (3.5). In both situations we have thus that $I.J_i \subseteq I.T(V)_{\geq 0}.J_{i-1}$. Then, the kernel of the restriction of $\delta_{i+1}$ to $J_{i+1}$ is contained in the intersection

$$(R.J_{i-1}) \cap (I.T(V)_{\geq 0}.J_{i-1}) = (R \cap (I.T(V)_{>0})).J_{i-1},$$

where we have used Proposition 3.2 and Lemma 3.11. By the defining property (2.1) of the space of relations the last space vanishes, and we thus get that the restriction of $\delta_{i+1}$ to $J_{i+1}$ is injective. This proves the claim.

For each $i \in \mathbb{N}_0$, we now consider the map $J_{i+1} \to k \otimes_A \text{Ker}(\delta_i)$ given by the composition of $J_{i+1} \to \text{Ker}(\delta_i)$ and the canonical projection $\text{Ker}(\delta_i) \to k \otimes_A \text{Ker}(\delta_i)$. We claim that this composition is in fact injective if $i$ is even. This can be proved as follows. By the comments in the last paragraph of Section 2 we know that the mentioned map is in fact an isomorphism for $i = 0$ (and also for $i = 1$). We shall suppose thus that $i \geq 2$. Since $i$ is even, the image of the map $J_{i+1} \to \text{Ker}(\delta_i)$ is contained in $V.J_i$, so one sees that the kernel of $J_{i+1} \to k \otimes_A \text{Ker}(\delta_i)$ vanishes if and only if $J_i \cap \text{Ker}(\delta_i) = 0$, which follows from the previous paragraph.

We have thus the following result.

**Lemma 3.23.** Let $A = T(V)/(R)$ be a locally finite dimensional nonnegatively graded algebra and let $(K(A)_\bullet, \delta_i)$ be its left multi-Koszul complex. Given $i \in \mathbb{N}_0$, the map of graded vector spaces $J_{i+1} \to \text{Ker}(\delta_i)$ given by the restriction of $\delta_{i+1}$ is injective. Consider now the map of graded vector space given by the composition of the previous morphism and the canonical projection $\text{Ker}(\delta_i) \to k \otimes_A \text{Ker}(\delta_i)$. If $i$ is even or $i = 1$, it is injective. If $i$ is odd and $i \geq 3$, we note that the restriction to the $n$-th homogeneous components of the previous composition map is injective if there are no nontrivial homogeneous components of $(\text{Ker}(\delta_i))_m$ for $m < n$. 

*Estanislao Herscovich*

Documenta Mathematica 18 (2013) 1301–1347
The corresponding formulation of the lemma for the right multi-Koszul complex of \( A \) is obvious, and we shall refer to the lemma whether we are considering the left or the right version. We may in fact prove one of the main result of this section (cf. \cite{19}, Prop. 3.12):

**Proposition 3.24.** Let \( A = T(V)/(R) \) be a locally finite dimensional nonnegatively graded algebra. Then \( A \) is left (resp., right) multi-Koszul if and only if there is an isomorphism of graded vector spaces \( \text{Tor}^A_i(k, k) \cong J_i \) for all \( i \in \mathbb{N}_0 \).

**Proof.** We shall prove the statement for the left multi-Koszul property, since the right one is analogous. Moreover, we will only show the “if” part, since the converse follows immediately from the minimality of the (left) multi-Koszul complex.

Assume the existence of the isomorphism of graded vector spaces in the statement. We will prove that the left multi-Koszul complex is in fact a minimal projective resolution of the trivial left \( A \)-module \( k \). In fact, we will show that \( K(A)_* \) is a minimal projective resolution of \( k \) up to homological degree \( i \) for all \( i \in \mathbb{N} \). Since the former coincides with such a minimal projective resolution up to homological degree \( 2 \), we suppose that the statement is true for \( i \geq 2 \).

By the comments on the construction of projective covers in Section 2 and the assumption \( \text{Tor}^A_{i+1}(k, k) \cong J_{i+1} \), there is an isomorphism of graded vector spaces \( h_{i+1} : J_{i+1} \to k \otimes_A \ker(\delta_i) \).

If \( i \geq 2 \) and \( i \) is even, the previous lemma tells us that the composition

\[
J_{i+1} \leftrightarrow \ker(\delta_i) \leftrightarrow k \otimes_A \ker(\delta_i),
\]

where the first map is the restriction of \( \delta_{i+1} \), is injective. Hence, the composition of this map with the inverse of \( h_{i+1} \) is an injective endomorphism of graded vector spaces of \( J_{i+1} \), so an isomorphism, since the latter is a locally finite dimensional graded vector space. Hence, the composition map (3.10) is also an isomorphism, so \( \delta_{i+1} \) is in fact a projective cover of \( \ker(\delta_i) \).

We now assume that \( i \geq 2 \) and \( i \) is odd. We consider as before the map of graded vector spaces given by (3.10), which we denote by \( f_{i+1} \). Let us denote by \( (f_{i+1})_n : (J_{i+1})_n \to (k \otimes_A \ker(\delta_i))_n \) the restriction to the \( n \)-th homogeneous components. We shall prove that \( (f_{i+1})_n \) is an isomorphism for all \( n \in \mathbb{N}_0 \) by induction on \( n \). Note that, by the isomorphism \( \text{Tor}^A_{i+1}(k, k) \cong J_{i+1} \) of graded vector spaces, it suffices to prove that \( (f_{i+1})_n \) is injective, for an injective map between finite dimensional vector spaces is automatically an isomorphism, since the corresponding \( n \)-th homogeneous components are finite dimensional. Let \( n_{\text{min}} \in \mathbb{N} \) be the first positive integer such that \( (J_{i+1})_{n_{\text{min}}} \neq 0 \). Note that this in particular implies that \( (f_{i+1})_n \) is injective for \( n < n_{\text{min}} \). The assumption \( \text{Tor}^A_{i+1}(k, k) \cong J_{i+1} \) implies that \( \ker(\delta_i) \) is concentrated in degrees greater that or equal to \( n_{\text{min}} \). This in turn implies that \( (f_{i+1})_{n_{\text{min}}} \) is injective by Lemma 3.25. Let us thus assume that \( (f_{i+1})_n \) is injective for \( n \leq m \) for some \( m \geq n_{\text{min}} \), so, again by the hypothesis \( \text{Tor}^A_{i+1}(k, k) \cong J_{i+1} \), they should be in fact isomorphisms, since the corresponding \( n \)-th homogeneous components
are finite dimensional. We shall prove that \((f_{i+1})_{m+1}\) is also injective. If this is not the case, by the definition of the map \((f_{i+1})_{m+1}\) and the inductive assumption, we must have that the intersection \(J_{i+1+m}\) \(\subseteq R_{i+1}^{(i+1)/2}\), the previous intersection is included in \(R_{i+1}^{(i+1)/2}\), which vanishes by the defining property \((2.1)\) of the space of relations. The proposition is thus proved. 

We now have the following immediate consequence of Proposition \((2.24)\) and the isomorphism \((2.2)\).

**Proposition 3.25.** For a locally finite dimensional nonnegatively graded algebra connected \(A\) with space of relations \(R\), the following statements are equivalent:

1. \(A\) is multi-Koszul,
2. \(\text{Ext}^i_A(k, k) \simeq J_i^\#\) for all \(i \in \mathbb{N}_0\), where \((-)^\#\) denotes the graded dual of a graded vector space.

**Remark 3.26.** Suppose further that \(A\) is a finitely generated multi-Koszul algebra with a finite dimensional space of relations. Then, the multi-Koszul resolution for \(A\) is composed of finitely generated projective \(A\)-modules, for each vector space \(J_i\) is finite dimensional, so, by the comments in the penultimate paragraph of Section \(2\) there is a canonical identification \(\text{Ext}^i_A(k, k) \simeq \text{Ext}^i_A(k, k)\).

We may also mention an easy corollary of the previous Lemma.

**Corollary 3.27.** Let \(A = T(V)/\langle R \rangle\) be a locally finite dimensional nonnegatively connected graded algebra. Suppose that its left multi-Koszul complex \((K(A)\bullet, \delta\bullet)\) is exact in homological degrees \(\bullet = 1, \ldots, N - 1\) for \(N \in \mathbb{N}_{\geq 2}\), and that \(\delta_N\) is injective. Then \(A\) is left multi-Koszul.

**Proof.** By the proof of Proposition \((3.24)\), the exactness hypothesis is equivalent to say that \(\text{Tor}^A_i(k, k) \simeq J_i\) for all \(i = 1, \ldots, N - 1\). Moreover, the injectivity of \(\delta_N\) implies that that \(J_{N+1}\) vanishes, for the restriction of \(\delta_{N+1}\) to \(J_{N+1}\) is injective, by Lemma \((3.23)\) and its image lies in the Kernel of \(\delta_N\). This in turn implies that \(J_i\) also vanishes for \(i \geq N + 1\), by the definitions \((3.22)\) and \((3.3)\). In consequence, the left multi-Koszul complex \((K(A)\bullet, \delta\bullet)\) is exact in positive degrees, so \(A\) is left multi-Koszul. 

### 3.4 Properties of multi-Koszul algebras

We have a direct consequence of the Proposition \((3.17)\) Consider the (unique) anti-morphism of (unitary) algebras \(\tau : T(V) \rightarrow T(V)\) such that \(\tau |_V = \text{id}_V\). It is in fact an anti-automorphism of \(T(V)\), and it further induces an algebra anti-isomorphism \(\tilde{\tau} : A \rightarrow T(V)/\langle \tau(R) \rangle\). In other words, it induces an isomorphism between the (usual) opposite algebra \(A^{op}\) of \(A\) and \(A^o = T(V)/\langle \tau(R) \rangle\). Note that \(\tau(R) \subseteq T(V)^{\geq 2}\) is clearly a space of relations of \(A^o\). By the previous isomorphism, we may thus say that \(A^o\) is (also) the opposite algebra of \(A\).
Corollary 3.28. The algebra $A^\circ$ is multi-Koszul if and only if $A$ is multi-Koszul.

Proof. It is an immediate consequence of Proposition 3.17. □

Since the length of a minimal projective resolution of $k$ gives the global dimension of $A$, the following proposition is an immediate consequence of Corollary 3.27.

Corollary 3.29. Let $A$ be a locally finite dimensional nonnegatively graded algebra. If the global dimension of $A$ is 2, then $A$ is multi-Koszul.

We also have the following result, which is a generalization of [19], Prop. 3.7, and which shows that the multi-Koszul property is stable under free products. The proof is completely parallel but we provide it for completeness.

Proposition 3.30. Let $\{B^s : s \in S\}$, where $S$ is an index set, be a finite collection of locally finite dimensional nonnegatively graded connected algebras such that $B^s$ is multi-Koszul for each $s \in S$. Then, the free product (i.e. the coproduct in the category of graded algebras) $A = \coprod_{s \in S} B^s$ of the collection $\{B^s : s \in S\}$ is a multi-Koszul algebra.

Proof. Suppose $B^s = T(V^s)/(R^s)$ for $s \in S$, is a multi-Koszul algebra, where $V^s$ and $R^s \subseteq T(V^s)_{\geq 2}$ are locally finite dimensional positively graded vector spaces. By the definition of the free product, we may consider that $A = T(V)/(R)$, where $V = \bigoplus_{s \in S} V^s$ and $R = \bigoplus_{s \in S} R^s$. The canonical inclusion $B^s \hookrightarrow A$ is a morphism of graded algebras, and it makes $A$ a free graded (left or right) $B^s$-module. For $s \in S$, denote by $J_i$ the graded vector space defined by the recursive equations (3.2) and (3.3) for the algebra $B^s$, and by $J_i$ the corresponding one defined for $A$.

Since the graded vector spaces $V^s$ are independent, by the definition of the tensor algebra it is trivial to verify that $J_i = \bigoplus_{s \in S} J^s_i$ for $i \in \mathbb{N}$. If $(K(B^s)_*, \delta^s_*)$ is the multi-Koszul complex of $B^s$, which is acyclic in positive homological degrees by assumption, we have that $A \otimes_{B^s} K(B^s)_* = A \otimes J^s_*$ is also acyclic in positive homological degrees, as we now show. By the Künneth spectral sequence $E^2_{p,q} = \operatorname{Tor}^b_p(A, H_q(K(B^s)_*)) \Rightarrow H_{p+q}(A \otimes_{B^s} K(B^s)_*)$ (see [33], Application 5.6.4). The exactness of the Koszul complex of $B^s$ and the freeness of the $B^s$-module $A$ imply that $E^2_{p,q} = 0$ if $(p,q) \neq (0,0)$, so $H_n(A \otimes_{B^s} K(B^s)_*) = 0$ for $n \geq 1$.

We now note that the multi-Koszul complex $(K(A)_*, \delta_*)$ of the algebra $A$ can be decomposed as $K(A)_* = \bigoplus_{s \in S} A \otimes_{B^s} K(B^s)_*$, for $\bullet \geq 1$, and $\delta_* = \bigoplus_{s \in S} \delta^s_*$ for $\bullet \geq 2$, the exactness of $A \otimes_{B^s} K(B^s)_*$ in positive homological degrees tells us that $K(A)_*$ is acyclic in homological degrees greater than or equal to 2. On the other hand, the exactness of the multi-Koszul complex in homological degree 1 is automatically satisfied for a nonnegatively graded connected algebra. We have thus that $K(A)_*$ is exact in positive homological degrees, so $A$ is multi-Koszul.

□

Another interesting property for this class of algebras is the following.
Proposition 3.31. Let \( A \) be a finitely generated nonnegatively graded connected algebra such that its space of relations \( R \) is finite dimensional, and assume that \( A \) is multi-Koszul. Then, the graded algebra \( \mathcal{E}xt_A^\bullet(k, k) = \mathcal{E}xt_A^\bullet(k, k) \) is generated by \( \mathcal{E}xt_A^1(k, k) \) and \( \mathcal{E}xt_A^2(k, k) \), i.e. \( A \) is \( K_2 \) (in the sense of Cassidy and Shelton).

Proof. Let \( E_i \) (following the notation of the article [10], Section 4) be the matrix with entries in \( I/(T(V) > 0.I + I.T(V) > 0) \) given by the class of the matrix with entries in \( T(V) \) which is a lift of the composition \( \delta_{i-1} \circ \delta_i \). By Lemma 3.12 we have the inclusion \( J_i \subseteq R.J_{i-2} \), which implies that the matrix \( E_i \) represents an injective linear transformation of the form \( T(V).J_i \rightarrow T(V).J_{i-1} \), so the rows \( E_i \) are linearly independent over \( k \). The statement is now a direct consequence of [10], Thm. 4.4.

\[ \square \]

Remark 3.32. As explained in [19], Rmk. 3.24, the converse of the previous proposition is not true in general (e.g. see the algebra \( B \) in [11], which is not multi-Koszul, but it is 2-3-Koszul in the sense of [14]).

4 The \( A_\infty \)-algebra structure of the Yoneda algebra of a multi-Koszul algebra

In this section, we shall provide a direct procedure to compute the complete \( A_\infty \)-algebra structure of the Yoneda algebra \( \mathcal{E}xt_A^\bullet(k, k) \) of a multi-Koszul algebra \( A \). We will first compute the (plain) algebra structure by explicitly providing quasi-isomorphisms in both directions between the cochain complexes \( \text{Hom}_A(K(A)_* k) \) and \( \text{End}_A(K(A)_* k) \). Then we shall compute the remaining higher multiplications. In order to do so, it will be useful to profit from the theory of \( A_\infty \)-algebras and \( A_\infty \)-coalgebras. Even though we refer for further references to [24], or [32], we will provide a short introduction, in particular for stating our (sign) conventions and notation. A more intensive study of the \( A_\infty \)-algebra structure of the Yoneda algebra may be found in [26], to which we also refer for further reading.

We recall that, since the cochain complex \( \text{End}_A(K(A)_* k) \) is a differential graded algebra (or dg algebra for short), its cohomology has an algebra structure, and further a structure of minimal \( A_\infty \)-algebra, defined via the theorem of T. Kadeishvili in [21]. Following [26], any of these \( A_\infty \)-algebra structures on the cohomology of \( \text{End}_A(K(A)_* k) \) is called a model. Kadeishvili also proved in the mentioned article that any of these models on the cohomology ring are in fact quasi-isomorphic. It can be shown that the algebra structure of \( \mathcal{E}xt_A^\bullet(k, k) \) is in fact independent of the projective resolution of the trivial module \( k \) used to compute it, and furthermore, the \( A_\infty \)-structure is also unique up to quasi-isomorphism (see [26], Lemma 4.2, (a)). Moreover, the endomorphism dg algebra of any projective resolution of the trivial \( A \)-module \( k \) is quasi-isomorphic to the graded dual \( B^+(A)^\# \) of the bar construction of \( A \), as \( A_\infty \)-algebras (see [26], Lemma 4.2, (b)).
4.1 The Multi-Koszul Property for Connected Algebras

We will first compute the Yoneda product for the cohomology space \( \text{Ext}^*_{\mathcal{A}}(k, k) \). Just for convenience we recall that \( \text{End}_{\mathcal{A}}(K(A)_{\bullet}) \) is the graded algebra whose \( i \)-th homogeneous component \( \text{End}_{\mathcal{A}}(K(A)_{\bullet})_i \) is given by

\[
\prod_{j \in \mathbb{Z}} \text{Hom}_{\mathcal{A}}(K(A)_j, K(A)_{j-i}),
\]

for \( i \in \mathbb{Z} \), together with the multiplication induced by the composition and the differential \( \partial \) given as follows. If \( (f_j)_{j \in \mathbb{Z}} \in \text{End}_{\mathcal{A}}(K(A)_{\bullet}) \), then the \( j \)-th component of \( \partial((f_j)_{j \in \mathbb{Z}}) \in \text{End}_{\mathcal{A}}^{j+1}(K(A)_{\bullet}) \) is given by

\[
\delta_{j-i} \circ f_j - (-1)^i f_{j-i} \circ \delta_i,
\]

where we remark that, by definition of the complex \( (K(A)_{\bullet}, \delta_{\bullet}) \), \( \delta_i \) vanishes for \( l \leq 0 \).

On the one hand, the augmentation map \( \delta_0 : K(A)_{\bullet} \to k \) induces a quasi-isomorphism, that we will denote \( \phi \), from \( \text{End}_{\mathcal{A}}(K(A)_{\bullet}) \) to \( \text{Hom}_{\mathcal{A}}(K(A)_{\bullet}, k) \). So \( \phi \) is given explicitly by the linear map which sends \( (f_j)_{j \in \mathbb{Z}} \in \text{End}_{\mathcal{A}}(K(A)_{\bullet}) \) to \( (-1)^i \delta_0 \circ f_i \). We remark that the differential of \( \text{Hom}_{\mathcal{A}}(K(A)_{\bullet}, k) \) is given by \( g \mapsto (-1)^i g \circ \delta_{i+1} \) for \( g \in \text{Hom}_{\mathcal{A}}(K(A)_{\bullet}, k) \). It is trivial to see that this is a map of complexes, i.e. it commutes with the differentials, and furthermore it respects the grading. That this is a quasi-isomorphism is a standard fact on homological algebra (see [3], §5.2, Prop. 4; and the first paragraph of §7.1, where it is done for the endomorphism dg algebra of an injective resolution, but the analogous considerations hold for projective resolutions). Note that the procedure described in the paragraph can be applied to any projective resolution of \( k \).

We will now show how to construct a linear map from \( \text{Hom}_{\mathcal{A}}(K(A)_{\bullet}, k) \) to \( \text{End}_{\mathcal{A}}(K(A)_{\bullet}) \), that will be denoted by \( \psi \). We first note that, since the differential of the complex \( \text{Hom}_{\mathcal{A}}(K(A)_{\bullet}, k) \) vanishes, we have the obvious identification \( \text{Hom}_{\mathcal{A}}(K(A)_{\bullet}, k) \cong J^\# \) of graded vector spaces, where \( (-)^\# \) denotes the usual graded dual of a graded vector space. Let us consider an element \( f \in \text{Hom}_{\mathcal{A}}(K(A)_{\bullet}, k) \), where \( i \in \mathbb{N}_0 \), and we denote by \( f \in J^\# \) the corresponding element of the graded dual. We will construct elements \( f_j \in \text{Hom}_{\mathcal{A}}(K(A)_j, K(A)_{j-i}) \) for all \( j \in \mathbb{Z} \), such that \( (f_j)_{j \in \mathbb{Z}} \) lies in fact in the kernel of the differential \( \partial \) of the dg algebra \( \text{End}_{\mathcal{A}}(K(A)_{\bullet}) \) and such that \( (-1)^i \delta_0 \circ f_i = f \). In order to do so, we will consider two cases.

Since the graded vector spaces \( J_i \subseteq T(V) \) for \( i \in \mathbb{N}_0 \) are tensor-intersection faithful, given \( i_1, \ldots, i_m \in \mathbb{N}_0 \), throughout this section we will use the identification \( J_{i_1} \otimes \cdots \otimes J_{i_m} \cong J_{i_1} \cdots J_{i_m} \) (see Remark 5.13). If the nonnegative integer \( i \) is even, then we define \( f_j : K(A)_{j-i} \to K(A)_{j-i} \) as follows. If \( j < i \), we set \( f_j = 0 \). We now consider the case \( j \geq i \). We first note that, by Corollary 5.13, \( J_j \subseteq J_{j-i} \otimes J_i \). Hence, we have a map of graded vector spaces \( p^f_{j,i} : J_j \to J_{j-i} \) defined as the composition of the inclusion \( J_j \subseteq J_{j-i} \otimes J_i \) and the map \( \text{id}_{J_{j-i}} \otimes f \). We then set \( f_j \) to be \( \mathcal{A} \)-linear map given by \( f_j(a \otimes \omega) = a \otimes p^f_{j,i}(\omega) \).
If the nonnegative integer \( i \) is odd, the map \( f_j : K(A)_j \to K(A)_{j-i} \) is given as follows. If \( j < i \), we also set \( f_j = 0 \), so let us consider the case \( j \geq i \). If \( j \) is odd, Corollary 3.13 tells us that \( J_j \subseteq J_{j-i} \otimes J_i \), so we consider again the map \( p_{j,i}^f : J_j \to J_{j-i} \) of graded vector spaces given by the composition of the inclusion \( J_j \subseteq J_{j-i} \otimes J_i \) and the map \( \text{id}_{J_{j-i}} \otimes f \). We then set \( f_j \) to be \( A \)-linear map defined as \( f_j(a \otimes \omega) = -a \otimes p_{j,i}^f(\omega) \). If \( j \) is even, by Corollary 3.13 we have that \( J_j \subseteq T(V) \otimes J_{j-i} \otimes J_i \). We have in this case the map \( p_{j,i}^f : J_j \to A \otimes J_{j-i} \) of graded vector spaces defined as the composition of the inclusion \( J_j \subseteq T(V) \otimes J_{j-i} \otimes J_i \) and \( \pi \otimes \text{id}_{J_{j-i}} \otimes f \), where \( \pi : T(V) \to A \) is the canonical projection. Set \( f_j \) to be \( A \)-linear map given by \( f_j(a \otimes \omega) = a p_{j,i}^f(\omega) \).

It is straightforward, but rather tedious, to show that the element \((f_j)_{j \in \mathbb{Z}}\) is in the kernel of the differential of \( \text{End}_A(K(A)_\bullet) \) and such that \((-1)^i \delta_0 \circ f_i = f \). We define thus the map \( \psi : \text{Hom}_A(K(A)_\bullet, k) \to \text{End}_A(K(A)_\bullet) \) via \( \psi(f) = (f_j)_{j \in \mathbb{Z}} \). It is trivially verified that this is a morphism of graded vector spaces, which commutes with the differentials (for the image of \( \psi \) lies in the kernel of the differential of its codomain and the domain has vanishing differential), and the composition \( \phi \circ \psi \) is the corresponding identity map. Since \( \phi \) is a quasi-isomorphism, \( \psi \) satisfies the same property. Furthermore, the map induced by \( \psi \) at the level of cohomology spaces is in fact the inverse of the corresponding map induced by \( \phi \).

These maps allow us to explicitly compute the algebra structure of the Yoneda algebra \( \mathcal{E}x^\bullet_A(k, k) \), since given two elements in the Yoneda algebra which are represented by \( f \in \text{Hom}_A(K(A)_i, k) \) and \( g \in \text{Hom}_A(K(A)_{i'}, k) \), or more concretely, by \( f \in J^\#_i \) and \( g \in J^\#_{i'} \), the product is just given by \( \psi(\phi(f) \otimes g) \in \text{Hom}_A(K(A)_{i+i'}, k) \), or simply by the induced element in \( J^\#_{i+i'} \). This can be written down in the following very explicit manner. In order to do so, we recall that we will consider \( \mathcal{E}x^\bullet_A(k, k) \) as a (cohomological) bigraded algebra (i.e. a cohomological bigraded vector space provided with a multiplication, unit and augmentation which respect both gradings), with one grading coming from the cohomological degree, which will be called the cohomological grading, and another one coming from the original grading of the modules over \( A \), which will be called the Adams grading.

In the same manner, we consider that the space \( J_i \) is concentrated in homological degree \( i \) and the Adams grading coincides with the one induced by the grading of \( V \), which was the only one we considered before. We moreover define the homological bigraded vector space \( J = \bigoplus_{i \in \mathbb{N}_0} J_i \), with the homological and Adams gradings induced from the ones of the homological bigraded vector spaces \( J_i^\# \) for \( i \in \mathbb{N}_0 \). Note that \( J_i^\# = \bigoplus_{j \in \mathbb{N}_0} J^\#_i \) is a cohomological bigraded vector space, where \( J^\#_i \) is concentrated in cohomological degree \( i \).

As explained before, we will apply the Koszul sign rule to the cohomological grading but not to the Adams grading.

Let \( i \) and \( i' \) be two nonnegative integers. If either \( i \) or \( i' \) is even, we shall denote by \( \iota_{i,i'} \) the inclusion map \( J_{i+i'} \subseteq J_i \otimes J_{i'} \) given in Corollary 3.13.
If both $i$ and $i'$ are odd, we shall denote by $t_{i,i'}$ the composition of the inclusion map $J_{i+i'} \subseteq T(V) \otimes J_i \otimes J_{i'}$, given in Corollary 3.1, together with the map $\epsilon_{T(V)} \otimes \text{id}_{J_i} \otimes \text{id}_{J_{i'}}$, where $\epsilon_{T(V)} : T(V) \to k$ denotes the augmentation of the tensor algebra. In any case, $t_{i,i'}$ is a map of graded vector spaces from $J_{i+i'}$ to $J_i \otimes J_{i'}$. This in turn induces a map of graded vector spaces $t_{i,i'}^# : (J_i \otimes J_{i'})# \to J_{i+i'}^#$. Recall that, given $V$ and $W$ two locally finite dimensional bigraded vector spaces, where the first one is the cohomological grading and the second one is the Adams grading, we denote by finite dimensional bigraded vector spaces, where the first one is the cohomological grading and the second one is the Adams grading, we denote by $V \otimes W$ the obvious isomorphism of bigraded vector spaces $c_{V,W} : V# \otimes W# \to (V \otimes W)#$ the obvious isomorphism of bigraded vector spaces $c_{V,W}(f \otimes g)(v \otimes w) = (-1)^{\deg(v) \deg(g)} f(v)g(w)$, where $f, g, v, w$ are homogeneous elements, and $\deg(\cdot)$ denotes the cohomological degree of the corresponding element. Now, by making use of the usual identification $\operatorname{Ext}^i_A(k, k) \simeq J_i^#$ of cohomological bigraded vector spaces and the previous comments, we obtain the following result, which provides a description of the structure of augmented cohomological bigraded algebra (i.e. a cohomological bigraded vector space provided with a multiplication, unit and augmentation which respect both gradings) of the Yoneda algebra.

**Theorem 4.1.** Let $A$ be a multi-Koszul algebra, and let $\{J_i\}_{i \in \mathbb{N}_0}$ be the collection of graded vector subspaces of $T(V)$ defined by $J_0 = k$ and the recursive identities 3.2 and 3.3. We utilize the usual identification $\operatorname{Ext}^i_A(k, k) \simeq J_i^#$ for $i \in \mathbb{N}_0$, coming from the use of the multi-Koszul resolution for $A$. Given $i, i' \in \mathbb{N}_0$, let $t_{i,i'} : J_{i+i'} \to J_i \otimes J_{i'}$ be the map of graded vector spaces introduced in the previous paragraph, and $c_{J_i, J_{i'}} : J_i^# \otimes J_{i'}^# \to (J_i \otimes J_{i'})^#$ be the usual identification of graded vector spaces also recalled in the previous paragraph. Then, the restriction of the Yoneda product of $\operatorname{Ext}^i_A(k, k)$ to $J_i^# \otimes J_{i'}^#$ is given by $t_{i,i'} \circ c_{J_i, J_{i'}}$.

**Remark 4.2.** It is easy to check that the algebra algebra structure for $\operatorname{Ext}^i_A(k, k)$ given in the theorem coincides with the one deduced from 19, Prop. 3.21 and Rmk. 3.22, under the further assumptions that the algebra $A$ is finitely generated in degree 1 with a finite dimensional space of relations. In particular, the previous algebra structure also coincides with the one given in 7, Prop. 3.1, for the Yoneda algebra of a generalized Koszul (homogeneous) algebra.

### 4.2 The $A_\infty$-algebra structure

We will now describe the $A_\infty$-algebra structure of the Yoneda algebra. First, we note that the graded dual of a locally finite dimensional augmented cohomological bigraded algebra is a coaugmented homological bigraded coalgebra (i.e. a homological bigraded vector space provided with a comultiplication, counit and coaugmentation which respect both gradings). Indeed, if $E = \oplus_{(n,m) \in \mathbb{Z}^2} E_{n,m}^#$ is a locally finite dimensional augmented cohomological bigraded algebra, then $E^#$ is a coaugmented homological bigraded coalgebra, where the comultiplication of $E^#$ is the composition of the graded dual of the multiplication of $E$ and $c_{E,E}$, the counit of $E^#$ is the graded dual of $E$. 

---

*Documenta Mathematica* 18 (2013) 1301–1347
the unit of $E$, and the coaugmentation of $E^#$ is the graded dual of the augmentation of $E$. The corresponding result given by interchanging the terms “augmented cohomological bigraded algebra” and “coaugmented homological bigraded coalgebra” (without the locally finite dimensional assumption) also holds, with the obvious analogous definitions of multiplication, unit and augmentation. Hence, one finds that the homological bigraded vector space $J$ is naturally a coaugmented homological bigraded coalgebra. The counit of $J$ is the canonical projection onto the $J_0$ and the unit is the inclusion of $J_0$ inside $J$. Moreover, the comultiplication $\Delta$ of $J$ is given by $\Delta|_{J_i} = \sum_{l=0}^{i} \iota_{l,i} - l$. We will extend the coaugmented coalgebra structure on $J$ to a minimal Adams graded coaugmented $A_\infty$-coalgebra on $J$. For the following definitions we refer to [32], Chapitre 3, Section 3.1 (or also [24], Déf. 1.2.1.1, 1.2.1.8, using the obvious equivalences between non(co)unitary objects and (co)augmented ones), even though we do not follow the same sign conventions and they do not consider any Adams grading (see for instance [26] for several uses of Adams grading in $A_\infty$-algebra theory). Moreover, our definitions will be somehow more restricted with respect to the gradings that the usual ones. We first recall that an Adams graded augmented $A_\infty$-algebra structure on a cohomological bigraded vector space $A$ such that the cohomological grading is nonnegative and the component of zero cohomological degree is just $k$ (also lying in Adams degree zero), is the following data:

(i) a collection of maps $m_i : A^{\otimes i} \to A$ for $i \in \mathbb{N}$ of cohomological degree $2 - i$ and Adams degree zero satisfying the Stasheff identities given by

$$\sum_{(r,s,t) \in I_n} (-1)^{r+s+t} m_{r+1+s+t} \circ (\text{id}_A^{\otimes r} \otimes m_s \otimes \text{id}_A^{\otimes t}) = 0, \quad (4.1)$$

for $n \in \mathbb{N}$, where $I_n = \{(r,s,t) \in \mathbb{N}_0 \times \mathbb{N} \times \mathbb{N}_0 : r + s + t = n\}$.

(ii) a map $\eta : k \to A$ of bidegree $(0,0)$ such that

$$m_i \circ (\text{id}_A^{\otimes r} \otimes \eta \otimes \text{id}_A^{\otimes t})$$

vanishes for all $i \neq 2$ and all $r, t \geq 0$ such that $r + 1 + t = i$, and

$$m_2 \circ (\eta \otimes \text{id}_A) = \text{id}_A = m_2 \circ (\eta \otimes \eta) = 0$$

for all $i \in \mathbb{N} \setminus \{2\}$.

(iii) a map $\epsilon : A \to k$ of bidegree $(0,0)$ such that $\epsilon \circ \eta = \text{id}_k$, $\epsilon \circ m_2 = \epsilon \otimes 2$, and $\epsilon \circ m_i = 0$ for all $i \in \mathbb{N} \setminus \{2\}$.

It is further called minimal if $m_1$ vanishes.

We recall that a family of linear maps $\{f_i : C \to C_i\}_{i \in \mathbb{N}}$, where $C$ and $C_i$ for $i \in \mathbb{N}$, are vector spaces, is called locally finite if, for all $c \in C$, there exists a finite subset $S \subseteq \mathbb{N}$, which depends on $c$, such that $f_i(c)$ vanishes for all $i \in \mathbb{N} \setminus S$. An Adams graded coaugmented $A_\infty$-coalgebra structure on a homological bigraded vector space $C$ such that the homological grading is nonnegative and the component of zero homological degree is just $k$ (also lying in Adams degree zero) is the following data:
(i) a locally finite collection of maps $\Delta_i : C \to C^\otimes i$ for $i \in \mathbb{N}$ of homological degree $i - 2$ and Adams degree zero satisfying the following identities

$$\sum_{(r,s,t) \in I_n} (-1)^{r+s+t} (\text{id}_C^\otimes r \otimes \Delta_s \otimes \text{id}_C^\otimes t) \circ \Delta_{r+1+t} = 0, \quad (4.2)$$

for $n \in \mathbb{N}$.

(ii) a map $\epsilon : C \to k$ of bidegree $(0,0)$ such that

$$(\text{id}_C^\otimes r \otimes \epsilon \otimes \text{id}_C^\otimes t) \circ \Delta_i$$

vanishes for all $i \neq 2$ and all $r, t \geq 0$ such that $r + 1 + t = i$, and

$$(\text{id}_C \otimes \epsilon) \circ \Delta_2 = \text{id}_C = (\epsilon \otimes \text{id}_C) \circ \Delta_2.$$

(ii) a map $\eta : k \to C$ of bidegree $(0,0)$ such that $\epsilon \circ \eta = \text{id}_k$, $\Delta_2 \circ \eta(1) = \eta(1)^{\otimes 2}$, and $\Delta_i \circ \eta(1) = 0$ for all $i \in \mathbb{N} \setminus \{2\}$.

An Adams graded coaugmented $A_\infty$-coalgebra $C$ is called minimal if $\Delta_1 = 0$. Note that the condition that the family $\{\Delta_n\}_{n \in \mathbb{N}}$ is locally finite follows from the other data if we further suppose that $\text{Ker}(\epsilon)$ is positively graded for the Adams degree.

We will not introduce the definitions of morphisms between (augmented) $A_\infty$-algebras, neither for (coaugmented) $A_\infty$-coalgebras, and refer to [32], Sections 3.2 and 3.3 (Déf. 3.3, 3.4, and 3.11), or [24], Sections 1.2 and 1.3. We remark however that these morphisms are further supposed to preserve the Adams degree (cf. [26], Section 2).

Notice that the graded dual $C^\#$ (as homological bigraded spaces) of an Adams graded coaugmented $A_\infty$-coalgebra $C$ is an Adams graded augmented $A_\infty$-algebra. Indeed, the unit of $C^\#$ is the graded dual of the counit of $C$ and the augmentation of $C^\#$ is the dual of the coaugmentation of $C$. Moreover, for $n \in \mathbb{N}$, the multiplication $m_n$ of $C^\#$ is the composition of the canonical map $c_{C,\ldots,C} : (C^\#)^{\otimes n} \to (C^\otimes n)^\#$ and the graded dual of $\Delta_n$.

We shall now proceed to describe the $A_\infty$-coalgebra structure of $J$. In order to do so, we first construct the following maps. Let $n \geq 3$ and $j = (j_1, \ldots, j_n) \in \mathbb{N}_0^n$. We write $j_l = 2j_l' + r_l$ for all $l = 1, \ldots, n$, where $j_l' \in \mathbb{N}_0$ and $r_l \in \{0, 1\}$, and consider the direct sum decomposition

$$\bigoplus_{N \in \mathbb{N}_{\geq 2}} \bigcap_{m \in \text{Par}_{n+1}(N)} V^{(m_1)} \otimes J_{2j_1'} \otimes \cdots \otimes V^{(m_n)} \otimes J_{2j_n'} \otimes V^{(m_n+1)}.$$

(4.3)

Let $p_j^N$ be the projection of the former vector subspace of the tensor algebra onto the $N$-th direct summand for $N \in \mathbb{N}_{\geq 2}$.

If $j$ satisfies that $j_l$ is even for all $1 \leq l \leq n$ except for two integers $1 \leq a < b \leq n$, for which $j_a, j_b$ are odd, (3.5) tells us that $J_{j_1+\cdots+j_n}$ is included.
Moreover, it is trivial to see (using definition 3.3, Lemma 5.11 and Proposition 3.2) that the \( n \)-th direct summand of the former vector space is contained in \( J_{j_1+1} \otimes \cdots \otimes J_{j_n+1} \). Let us denote by \( \iota_{j_1+1,\ldots,j_n+1} \) the map of graded vector spaces from \( J_{j_1+1} \otimes \cdots \otimes J_{j_n+1} \) given by the composition of the inclusion of \( J_{j_1+1} \otimes \cdots \otimes J_{j_n+1} \) in 4.3, the projection \( \rho^N_{jn} \), and the inclusion of the \( n \)-th direct summand of 4.3 in \( J_{j_1+1} \otimes \cdots \otimes J_{j_n+1} \). In other words, given \( \iota = (i_1, \ldots, i_n) \in \mathbb{N}^n \) such that \( i_j \) is odd for all \( 1 \leq l \leq n \), we have defined a map \( \iota_{i_1,\ldots,i_n} \) of graded vector spaces from \( J_{i_1+1,\ldots,i_n+2-n} \) to \( J_{i_1} \otimes \cdots \otimes J_{i_n} \). Suppose further that \( a = 1 \), that is, \( j_1 \) is odd. Then, the \( N \)-th direct summand of 4.3 is in fact contained in the \( N \)-th direct summand of

\[
\bigoplus_{N \in \mathbb{N}_{\geq 2}} \left( \bigcap_{m \in \text{Par}(N-1)} J_{j_1} \otimes V^{(m_1)} \otimes J_{j_2} \otimes \cdots \otimes V^{(m_{n-1})} \otimes J_{j_n} \otimes V^{(m_n)} \right),
\]

for all \( N \in \mathbb{N}_{\geq 2} \), which coincides with

\[
J_{j_1} \otimes \bigoplus_{N \in \mathbb{N}_{\geq 2}} \left( \bigcap_{m \in \text{Par}(N-1)} V^{(m_1)} \otimes J_{j_2} \otimes \cdots \otimes V^{(m_{n-1})} \otimes J_{j_n} \otimes V^{(m_n)} \right),
\]

by Proposition 3.2 and Lemma 3.11. Hence, we see that the restriction of \( \rho^N_{jn} \) to 4.4 coincides with the restriction of \( \text{id}_{J_{j_1}} \otimes p^N_{j_2,\ldots,j_n} \) to the same space. On the other hand, if we assume that \( b = n \), i.e. \( j_n \) is odd, the restriction of \( \rho^N_{jn} \) to 4.3 coincides with the restriction of \( p^N_{j_1,\ldots,j_{n-1}} \otimes \text{id}_{J_{j_n}} \) to the same space.

**Remark 4.3.** Note that if \( \iota = (1, \ldots, 1) \in \mathbb{N}^n \), \( \iota_{1,\ldots,1} \) is just the composition of the inclusion of \( R \) inside the tensor algebra with the canonical projection \( \pi_n : T(V) \rightarrow V^{(n)} \). More generally, for \( i \in \mathbb{N} \), we consider \( \iota_i = (i-1,1,\ldots,1) \in \mathbb{N}^n \) (the last \( n-1 \) integers are 1’s). It is easy to check that \( \iota_i : J_i \rightarrow J_{i-1} \otimes V^{(n-1)} \) coincides with the composition of the inclusion \( J_i \subseteq J_{i-1} \otimes T(V)_{>0} \) with the map \( \text{id}_{J_{i-1}} \otimes \pi_{n-1} \). Indeed, for the case \( \iota = (i-1,1,\ldots,1) \), the direct sum decomposition 4.3 gives

\[
\bigoplus_{N \in \mathbb{N}_{\geq 2}} \left( \bigcap_{m=0}^N V^{(m)} \otimes J_{i-2} \otimes V^{(N-m)} \right).
\]

Moreover, the \( N \)-th direct summand of the former is trivially included in the \( N \)-th direct summand of

\[
\bigoplus_{N \in \mathbb{N}_{\geq 2}} J_{i-1} \otimes V^{(N-1)},
\]

for all \( N \in \mathbb{N}_{\geq 2} \), by Proposition 3.2 and Lemma 3.11. Hence, the restriction of \( \text{id}_{J_{i-1}} \otimes \pi_{n-1} \) to the first direct sum decomposition coincides with the projection \( p^N_{jn} \), so the restriction of \( \text{id}_{J_{i-1}} \otimes \pi_{n-1} \) to \( J_i \) coincides with \( \iota_i \).
The $A_{\infty}$-(co)algebra structure of $J^\#$ (resp., $J$) is described in the next statement.

**Proposition 4.4.** Let $A$ be a multi-Koszul algebra, and let $\{J_i\}_{i \in \mathbb{N}_0}$ be the collection of graded vector subspaces of $T(V)$ defined by $J_0 = k$ and the recursive identities \eqref{eq1} and \eqref{eq2}. Define $J = \bigoplus_{i \in \mathbb{N}_0} J_i$, which is considered as a homological bigraded vector space for the grading coming from the index $i$. By Theorem \ref{thm1} it has the structure of a conaugmented bigraded coalgebra, and we denote $\Delta_2$ the comultiplication coming from the multiplication given in the previous theorem. We consider that $\Delta_2$ is onto $J_i$ such that $\Delta_2$ is a coassociative coalgebra for the comultiplication $\Delta_2$, and we denote $\Delta_2$ the map of graded vector spaces introduced in the paragraph before Remark \ref{rem1} and $\Delta_n|_{J_i} = \sum t_1, \ldots, t_n$, where the sum is indexed over all $n$-tuples of odd integers $(i_1, \ldots, i_n)$ such that $i_1 + \cdots + i_n + 2 - n = i$. Then $J$ is a minimal Adams graded conaugmented $A_{\infty}$-coalgebra. This in turn implies that $J^\#$ is a minimal Adams graded augmented $A_{\infty}$-algebra, where the multiplication $m_n : (J^\#)^{\otimes n} \to J^\#$ is the map whose restriction to $J^\#_i \otimes \cdots \otimes J^\#_n$ is zero, if there is $1 \leq l \leq n$ such that $i_l$ is even, and is $i^\#_1, \ldots, i^\#_n \otimes cJ^\#_1, \ldots, J^\#_n$ otherwise.

**Proof.** We first note that $\Delta_n$ has bidegree $(n-2, 0)$ for all $n \in \mathbb{N}$. Moreover, it is easy to check that the family $\{\Delta_n\}_{n \in \mathbb{N}}$ is locally finite, since $\text{Ker}(\epsilon)$ is positively graded for the Adams degree.

Since $\Delta_1$ is zero, the first, second and third of the identities \eqref{eq1} simply mean that $J$ is a coassociative coalgebra for the comultiplication $\Delta_2$, which is a direct consequence of the previous theorem.

Suppose now that $n \geq 3$. The assumption that $\Delta_n$ vanishes on $J_i$ if $i$ is odd, and that its image is included in the sum of tensors of $n$ factors, each of which has odd homological degree, tell us that the $(n+1)$-th defining identity \eqref{eq2} simplifies to give

$$
\sum_{r=1}^{n-1} (-1)^r (\text{id}_J^{\otimes r} \circ \Delta_2 \otimes \text{id}_J^{\otimes (n-r-1)}) \circ \Delta_n = (\text{id}_J \otimes \Delta_n) \circ \Delta_2 - (-1)^n (\Delta_n \otimes \text{id}_J) \circ \Delta_2.
$$

\hspace{1cm} (4.5)

It suffices to show the previous identity restricted to each subspace $J_i$. Note that the image of each term of the previous identity must be thus in the direct sum of (independent) subspaces $J_{i_1} \otimes \cdots \otimes J_{i_{n+1}}$, where $i = i_1 + \cdots + i_{n+1} + 2 - n$. Then, it suffices to prove \eqref{eq3} for the special case that we restrict to $J_i$, and we compose with the projection $p_{i_1, \ldots, i_{n+1}}$ onto $J_{i_1} \otimes \cdots \otimes J_{i_{n+1}}$, where $i = i_1 + \cdots + i_{n+1} + 2 - n$. We shall refer to this new identity also as the $(i_1, \ldots, i_{n+1})$-specialization of \eqref{eq3}. By the assumption on the higher multiplications $m_n$ for $n \geq 3$, the image of any of the terms of \eqref{eq3} must be in the direct sum of the subspaces $J_{i_j} \otimes \cdots \otimes J_{i_{n+1}}$, where there exists at most one $1 \leq j \leq n + 1$ such that $i_i$ is even. If we compose the restriction of \eqref{eq3} to $J_i$ with the projection onto $J_{i_1} \otimes \cdots \otimes J_{i_{n+1}}$, the only cases left to consider are thus:

(i) all $i_1, \ldots, i_{n+1}$ are odd,
(ii) $i_1$ is even and $i_2, \ldots, i_{n+1}$ are odd,

(iii) $i_n$ is even and $i_1, \ldots, i_{n}$ are odd,

(iv) there is $1 < j < n + 1$ such that $i_j$ is the only even integer among $i_1, \ldots, i_{n+1}$.

Let us write how the composition of $(n+1)$-th defining identity \[4.5\] restricted to $J_1$ and the projection onto $J_1 \otimes \cdots \otimes J_{n+1}$ further simplifies in each case, and how this identity follows from the construction of the higher comultiplication $\Delta_n$.

In case $(i)$, the $(i_1, \ldots, i_{n+1})$-specialization of \[4.5\] simplifies to give

\[ p_{i_1,\ldots, i_{n+1}} \circ (\id \otimes \Delta_n) \circ \Delta_2 |_{J_1} = (-1)^n p_{i_1,\ldots, i_{n+1}} \circ (\Delta_n \otimes \id) \circ \Delta_2 |_{J_1}, \]

for all the terms of the left member of \[4.5\] vanish. By the Koszul sign rule this identity is equivalent to the commutativity of the square

\[
\begin{array}{ccc}
J_{i_1+\cdots+i_{n+1}+2-n} & \overset{i_{i_1+\cdots+i_{n+1}+2-n}}{\longrightarrow} & J_{i_1} \otimes J_{i_2+\cdots+i_{n+1}+2-n} \\
\downarrow & & \downarrow \\
J_{i_1+\cdots+i_{n+1}+2-n} \otimes J_{i_{n+1}} & \overset{i_{i_1+\cdots+i_{n+1}} \otimes \id}{\longrightarrow} & J_{i_1} \otimes \cdots \otimes J_{i_{n+1}} \\
\end{array}
\]

Note that, since all $i_1, \ldots, i_{n+1}$ are odd the upper horizontal and the left vertical maps are the plain inclusions. The commutativity easily follows from the definitions. Indeed, consider the direct sum decomposition of the form

\[ \bigoplus_{N \in \mathbb{N}_{\geq 2}} \bigcap_{m \in \Par_{n+2}(N)} V^{(m_1)} \otimes J_{i_1} \otimes \cdots \otimes V^{(m_{n+1})} \otimes J_{i_{n+1}} \otimes V^{(m_{n+2})}, \] \[4.6\]

which trivially includes $J_{i_1+\cdots+i_{n+1}+2-n}$. Proposition \[3.2\] and Lemma \[5.11\] tell us that the $N$-th direct summand of the former is trivially included in the $N$-th direct summand of

\[ \bigoplus_{N \in \mathbb{N}_{\geq 2}} \bigcap_{m \in \Par_n(N-2)} J_{i_1} \otimes V^{(m_1)} \otimes J_{i_2-1} \otimes \cdots \otimes V^{(m_{n-1})} \otimes J_{i_{n-1}} \otimes V^{(m_n)} \otimes J_{i_{n+1}} \otimes V^{(m_{n+2})}, \]

for all $N \in \mathbb{N}_{\geq 2}$, which, also by Proposition \[3.2\] and Lemma \[5.11\] coincides with

\[ \bigoplus_{N \in \mathbb{N}_{\geq 2}} J_{i_1} \otimes \bigcap_{m \in \Par_n(N-2)} V^{(m_1)} \otimes J_{i_2-1} \otimes \cdots \otimes J_{i_{n-1}} \otimes V^{(m_n)} \otimes J_{i_{n+1}}. \]

Hence, by the comments at the end of the paragraph before Remark \[4.3\] we see that the restriction of $\id_{J_{i_1}} \otimes p_{(i_2,\ldots, i_n)} \otimes \id_{J_{i_{n+1}}}$ to \[4.6\] coincides with the
restrictions of \( p_n^{(i_1,\ldots,i_n)} \otimes \text{id}_{J_{i_{n+1}}^+} \) and of \( \text{id}_{J_{i_1}^+} \otimes p_n^{(i_2,\ldots,i_{n+1})} \) to the same space. This proves the claimed commutativity.

In case (ii), the \((i_1,\ldots,i_{n+1})\)-specialization of (4.5) simplifies to give
\[
p_{i_1,\ldots,i_{n+1}} \circ (\Delta_2 \otimes \text{id}_J^{(n-1)}) \circ \Delta_n \mid_{J_i} = p_{i_1,\ldots,i_{n+1}} \circ (\text{id}_J \otimes \Delta_n) \circ \Delta_2 \mid_{J_i},
\]
for all the other terms trivially vanish. This identity is equivalent to the commutativity of the square
\[
\begin{array}{c}
J_{i_1+\cdots+i_{n+1}+2-n} \\
i_{i_1+i_2+i_{n+1}} \\
\text{id}_{J_{i_1}^+} \otimes \cdots \otimes \text{id}_{J_{i_{n+1}}^+} \\
\end{array}
\quad \begin{array}{c}
J_{i_1+i_2} \otimes \cdots \otimes J_{i_{n+1}^+} \\
\end{array}
\]
\[
\quad \begin{array}{c}
J_{i_1+i_2+i_{n+1}} \\
\text{id}_{J_{i_1}^+} \otimes \cdots \otimes \text{id}_{J_{i_{n+1}}^+} \\
\end{array}
\quad \begin{array}{c}
J_{i_1+i_2} \otimes \cdots \otimes J_{i_{n+1}^+} \\
\end{array}
\]

The assumption on the integers \( i_1,\ldots,i_{n+1} \) tells us that the horizontal maps are the plain inclusions. The commutativity follows directly from the definitions, but we provide a proof. Consider the direct sum decomposition
\[
\bigoplus_{N \in \mathbb{N}_{\geq 2}} \left( \bigcap_{m \in \text{Par}_{n+2}(N)} V^{(m_1)} \otimes J_{i_1} \otimes V^{(m_2)} \otimes J_{i_2-1} \otimes \cdots \otimes J_{i_{n+1}-1} \otimes V^{(m_{n+1})} \right) \\
\cap \left( \bigcap_{m \in \text{Par}_{n+1}(N)} V^{(m_1)} \otimes J_{i_1+i_2-1} \otimes V^{(m_2)} \otimes J_{i_3-1} \otimes \cdots \otimes J_{i_{n+1}-1} \otimes V^{(m_{n+1})} \right).
\]

(4.7)

It contains \( J_{i_1+\cdots+i_{n+1}+2-n} \). On the one hand, Proposition 3.2 and Lemma 3.11 tell us that the \( N \)-th direct summand of the former is trivially included in the \( N \)-th direct summand of
\[
\bigoplus_{N \in \mathbb{N}_{\geq 2}} J_{i_1} \otimes \left( \bigcap_{m \in \text{Par}_{n+1}(N)} V^{(m_1)} \otimes J_{i_1+i_2-1} \otimes V^{(m_2)} \otimes J_{i_3-1} \otimes \cdots \otimes J_{i_{n+1}-1} \otimes V^{(m_{n+1})} \right),
\]
for all \( N \in \mathbb{N}_{\geq 2} \). On the other hand, it is obvious that the \( N \)-th direct summand of (4.7) is contained in the \( N \)-th direct summand of
\[
\bigoplus_{N \in \mathbb{N}_{\geq 2}} \left( \bigcap_{m \in \text{Par}_{n+1}(N)} V^{(m_1)} \otimes J_{i_1+i_2-1} \otimes V^{(m_2)} \otimes J_{i_3-1} \otimes \cdots \otimes J_{i_{n+1}-1} \otimes V^{(m_{n+1})} \right),
\]
for all \( N \in \mathbb{N}_{\geq 2} \). It is trivial to see that each independent family given by the \( N \)-th direct summands of each of the two previous direct sum decompositions is a subfamily of the independent family
\[
\left\{ \bigcap_{m \in \text{Par}_n(N)} J_{i_1+i_2-1} \otimes V^{(m_1)} \otimes J_{i_3-1} \otimes V^{(m_2)} \otimes \cdots \otimes J_{i_{n+1}-1} \otimes V^{(m_{n+1})} \right\}_{N \in \mathbb{N}_{\geq 2}}.
\]
The comments at the end of the paragraph before Remark 4.3 tell us that the restriction of \(\id_{J_i} \otimes p_i\) to \(i_{n+1}\) coincides with the restrictions of \(p_i\) to the same space. This proves the claim. Dually to (ii), under the assumptions of (iii), the \((i_1, \ldots, i_{n+1})\)-specialization of (4.5) simplifies as
\[
p_{i_1, \ldots, i_{n+1}} \circ (\id_j^{(n-1)} \otimes \Delta_2) \circ \Delta_n|_{J_i} = p_{i_1, \ldots, i_{n+1}} \circ (\Delta_n \otimes \id_j) \circ \Delta_2|_{J_i},
\]
for all the other terms are easily seen to vanish. This identity is equivalent to the commutativity of the square
\[
\begin{array}{c}
\begin{array}{c}
J_{i_1+\ldots+i_{n+1}+2-n} \\
\downarrow \\
J_{i_1+\ldots+i_{n-1}+i_n+i_{n+1}} \\
\downarrow \\
J_{i_1} \otimes \cdots \otimes J_{i_{n-1}} \otimes J_{i_n+i_{n+1}} \\
\end{array} \\
\begin{array}{c}
J_{i_1+\ldots+i_{n+1}+2-n} \\
\downarrow \\
J_{i_1} \otimes \cdots \otimes J_{i_{n+1}} \\
\end{array}
\end{array}
\]
Notice that the hypotheses on the integers \(i_1, \ldots, i_{n+1}\) imply that the horizontal maps are the plain inclusions. The commutativity is clear from the definitions and follows by an analogous argument to the one given in the previous paragraph. Finally, in the case (iv), the \((i_1, \ldots, i_{n+1})\)-specialization of (4.5) gives
\[
p_{i_1, \ldots, i_{n+1}} \circ (\id_j^{(j-2)} \otimes \Delta_2 \otimes \id_j^{(n-j+1)}) \circ \Delta_n|_{J_i} = p_{i_1, \ldots, i_{n+1}} \circ (\id_j^{(j-1)} \otimes \Delta_2 \otimes \id_j^{(n-j)}) \circ \Delta_n|_{J_i},
\]
for all the other terms are easily seen to vanish. This identity is equivalent to
\[
(\id_j \otimes \cdots \otimes \id_{J_{i_{j-2}}} \otimes \id_{J_{i_{j-1}}} \otimes \id_{J_{i_j+1}} \otimes \cdots \otimes \id_{J_{i_{n+1}}})
\circ \kappa_{i_1, \ldots, i_{j-2}, i_{j-1}+i_j, i_{j+1}, \ldots, i_{n+1}}
= (\id_j \otimes \cdots \otimes \id_{J_{i_{j-1}}} \otimes \id_{J_{i_j+1}} \otimes \id_{J_{i_{j+2}}} \otimes \cdots \otimes \id_{J_{i_{n+1}}})
\circ \kappa_{i_1, \ldots, i_{j-1}, i_j+i_{j+1}, i_{j+2}, \ldots, i_{n+1}}, \tag{4.8}
\]
By the assumption on the integers \(i_1, \ldots, i_{n+1}\) we have that the left maps of the left and right compositions are inclusions. This identity is proved as follows. Consider the direct sum decomposition
\[
\bigoplus_{\mathop{n \geq 2}} \bigcap_{\mathop{m \in \Par_n(N-2)}} V^{(m_1)} \otimes J_{i_1-1} \otimes V^{(m_2)} \otimes \cdots \otimes J_{i_{j-2}-1} \otimes V^{(m_{j-1})} \otimes J_{i_{j-1}} \\
\otimes J_{i_j} \otimes J_{i_{j+1}} \otimes V^{(m_j)} \otimes J_{i_{j+2}-1} \otimes V^{(m_{j+1})} \otimes \cdots \otimes V^{(m_{n+1})} \otimes J_{i_{n+1}-1} \otimes V^{(m_n)}, \tag{4.9}
\]
which trivially includes $J_{i_1+\cdots+i_{n+1}+2-n}$. Furthermore, its $N$-th direct summand includes the $N$-th direct summand of the decomposition \((4.3)\) corresponding to $j = (i_1, \ldots, i_{j-2}, i_{j-1} + i_j, i_{j+1}, \ldots, i_{n+1})$ and $j = (i_1, \ldots, i_{j-1}, i_j + i_{j+1}, i_{j+2}, \ldots, i_{n+1})$ for all $N \in \mathbb{N}_{\geq 2}$. So, the restriction of the projection of this decomposition onto the $n$-th direct summand to the space \((4.3)\) for $j = (i_1, \ldots, i_{j-2}, i_{j-1} + i_j, i_{j+1}, \ldots, i_{n+1})$ coincides with $p_{i_j}^n$, and the same applies if $j = (i_1, \ldots, i_{j-1}, i_j + i_{j+1}, i_{j+2}, \ldots, i_{n+1})$. This proves identity \((4.3)\).

The properties satisfied by the counit and the augmentation are also clear. The proposition is thus proved.

\[\Box\]

Remark 4.5. Note that the pattern of the proof is, roughly, (the dual of) the one given by \cite{15}, Thm. 4.5, even though our case is much involved. This similarity is for us, however, another indication that this definition of multi-Koszul property resembles the notion of generalized Koszul algebras given by Berger.

Remark 4.6. It is easy to check that the $A_\infty$-algebra structure for $J^\#$ given in the proposition coincides with the one given in \cite{19}, Rmk. 3.25, in the particular case that $A$ is finitely generated in degree 1 and has a finite dimensional space of relations.

The following definitions are the duals to the ones given by Kadeishvili in \cite{20}, p. 235, and we refer to \cite{32}, Chapitre 3, Section 5, even though the sign convention is different. We recall that a linear map $\tau : C \to A$ of cohomological degree 1 and Adams degree zero from a coaugmented Adams graded $A_\infty$-coalgebra $C$ to an augmented Adams graded dg algebra $A$ is called a twisting cochain (sometimes called homotopical or generalized twisting cochain) if the composition of $\tau$ with the augmentation of $A$ vanishes, the composition of the unit of $C$ with $\tau$ is also zero, and if

\[d \circ \tau = \sum_{n \in \mathbb{N}} \mu^{(n)} \circ \tau^\otimes n \circ \Delta_n,\]

where $d$ is the differential of $A$ and $\mu^{(n)}$ denotes the $(n - 1)$-th iteration of the product of $A$, as explained for the tensor algebra at the beginning of Subsection \cite{33}. Note that the previous sum is finite when evaluated at $c \in C$, for the family $\{\Delta_i\}_{i \in \mathbb{N}}$ is locally finite. We denote by $B^+(\cdot)$ the bar construction of an augmented $A_\infty$-algebra and by $\Omega^+(\cdot)$ the cobar construction of an augmented $A_\infty$-coalgebra. Moreover, $\tau_C : C \to \Omega^+(C)$ indicates the universal twisting cochain, given by the composition of the projection $C \to C/\text{Im}(\eta)$, the shift and the canonical inclusion, where $\eta$ denotes the coaugmentation of $C$ (see \cite{32}, Déf. 3.14). Dually, for an augmented Adams graded dg algebra $A$, its universal twisting cochain is the map $\tau_A : B^+(A) \to A$, given by the composition of the minus canonical projection onto $\text{Ker}(\epsilon)$, the shift and canonical inclusion, where $\epsilon$ denotes the augmentation of $A$ (see \cite{32}, (2.23)). Then, twisting cochains $\tau : C \to A$ are in correspondence with morphisms of dg algebras $f_\tau : \Omega^+(C) \to A$ via $f_\tau \circ \tau_C = \tau$ (see \cite{32}, Lemma 3.17). In this case we may consider the twisted tensor product $C \otimes^\tau A$, whose underlying vector space
Let us show that there is a twisting cochain $\tau$ given by the complex of graded right $A$-modules for the regular right action of $A$ provided with a differential $d_\tau$ defined as
\[
\sum_{n\in\mathbb{N}} (\text{id}_C \otimes \mu^{(n)}) \circ (\text{id}_C \otimes \tau^{(n-1)} \otimes \text{id}_A) \circ (\Delta_n \otimes \text{id}_A).
\]

The following theorem must be well-known to the experts, but we sketch a proof. It was announced by B. Keller at the X ICRA of Toronto, Canada, in 2002.

**Theorem 4.7.** Let $C$ be a minimal Adams graded coaugmented $A_\infty$-coalgebra and $A$ be a nonnegatively graded connected algebra. Then, the following are equivalent:

1. There is a quasi-isomorphism of Adams graded augmented minimal $A_\infty$-algebras
   \[\text{Ext}^\bullet_A(k, k) \to C^\# .\]

2. There is a twisting cochain $\tau : C \to A$ such that the twisted tensor product $C \otimes_\tau A$ is a minimal projective resolution of the trivial right $A$-module $k$.

**Proof.** Let us show that (ii) $\Rightarrow$ (i). First note that we have a morphism of augmented dg algebras $f_\tau : \Omega^+(C) \to A$, which in turn induces a morphism of augmented dg coalgebras $B^+(f_\tau) : B^+(\Omega^+(C)) \to B^+(A)$. By [22], Thm. 3.25, we know that $f_\tau$ is a quasi-isomorphism if and only if $B^+(f_\tau)$ is so, if and only if $B^+(\Omega^+(C)) \otimes_{f_\tau \circ \tau_\Omega^+(C)} A$ is acyclic (in the sense given by [22], Déf. 1.26). Using the adjunction quasi-isomorphism $\psi : C \to B^+(\Omega^+(C))$ (see [22], Thms. 2.28, or [24], Lemme 1.3.2.3) we get that $C \otimes_\tau A$ is quasi-isomorphic to $B^+(\Omega^+(C)) \otimes_{f_\tau \circ \tau_\Omega^+(C)} A$ (see [22], Thm. 3.23 for the existence of the map. The fact that it is a quasi-isomorphism follows easily, see for instance [22], Thm. 2.8, (ii)). Hence, the assumption that $C \otimes_\tau A$ is acyclic yields that the map $B^+(f_\tau) : B^+(\Omega^+(C)) \to B^+(A)$ is a quasi-isomorphism of coaugmented dg coalgebras. If we consider the composition of the adjunction quasi-isomorphism $\psi : C \to B^+(\Omega^+(C))$ of coaugmented $A_\infty$-coalgebras with the quasi-isomorphism $B^+(f_\tau)$ we get thus a quasi-isomorphism of coaugmented $A_\infty$-coalgebras $C \to B^+(A)$, which in turn induces a quasi-isomorphism of augmented $A_\infty$-algebras $B^+(A)^\# \to C^\#$. Using the quasi-isomorphism $\text{Ext}^\bullet_A(k, k) \to B^+(A)^\#$ of augmented $A_\infty$-algebras given by the theorem of Kadeishvili, we get the desired quasi-isomorphism of augmented $A_\infty$-algebras. It further preserves the Adams degree for all the involved maps preserve it.

We will now prove the converse. Suppose thus that $f : \text{Ext}^\bullet_A(k, k) \to C^\#$ is a quasi-isomorphism of minimal Adams graded augmented $A_\infty$-algebras. This implies that we have a quasi-isomorphism of augmented $A_\infty$-algebras $g : B^+(A)^\# \to C^\#$. Define $\tau$ as the twisting cochain composition of $g^\#$ and...
the universal twisting cochain $\tau_A : B^+(A) \to A$ of $A$ (see [32], Déf. 3.21 and Lemme 3.22). Since $g^\#$ is a quasi-isomorphism, we get that the map $g^\# \otimes \text{id}_A$ defines a quasi-isomorphism from $C \otimes \tau_A$ to $B^+(A) \otimes \tau_A A$, which is always acyclic (see for example [32], Thm. 3.19). See [32], Thm. 3.23 for the proof that this map is a morphism of complexes of $A$-modules. The fact that it is a quasi-isomorphism follows easily (see for instance [32], Thm. 2.8, (i)). Hence $C \otimes \tau_A$ is also acyclic. □

For the coaugmented $A_\infty$-coalgebra structure defined on $J$, we now set $\tau : J \to A$ as the linear map given by the composition of the canonical projection $J \to J_1 = V$ and the inclusion $V \subseteq A$. It is easy to check that $\tau$ is a twisting cochain, and furthermore $J \otimes \tau$ coincides with the right multi-Koszul complex of $A$. Indeed, we first note that the underlying homological bigraded vector space of $J \otimes \tau$ coincides with the corresponding one of $K(A)^\bullet$. Moreover, the differential $d_i$ restricted to $J_i \otimes A$ for $i$ odd is trivially seen to coincide with the differential $b_i'$ of $K(A)^\bullet$. The case $i$ even follows from Remark 4.3. Hence, by Proposition 4.4 and the previous theorem we have the following result.

**Theorem 4.8.** Let $A$ be a multi-Koszul algebra, and let $\{J_i\}_{i \in \mathbb{N}_0}$ be the collection of graded vector subspaces of $T(V)$ defined by $J_0 = k$ and the recursive identities (3.2) and (3.3). Then the Adams graded augmented $A_\infty$-algebras $\text{Ext}_A^\bullet(k, k)$ and $J^\#$ are quasi-isomorphic, where the structure of Adams graded augmented $A_\infty$-algebra of $J^\#$ was given in Proposition 4.4.

**References**


On the Multi-Koszul Property for Connected Algebras


Estanislao Herscovich
Institut Joseph Fourier
Université Grenoble I
France
and
Departamento de Matemática, FCEyN
Universidad de Buenos Aires
Argentina
Reduced Operator Algebras of Trace-Preserving Quantum Automorphism Groups

Michael Brannan

Received: January 10, 2013
Communicated by Joachim Cuntz

Abstract. Let $B$ be a finite dimensional $C^*$-algebra equipped with its canonical trace induced by the regular representation of $B$ on itself. In this paper, we study various properties of the trace-preserving quantum automorphism group $G$ of $B$. We prove that the discrete dual quantum group $\hat{G}$ has the property of rapid decay, the reduced von Neumann algebra $L^\infty(G)$ has the Haagerup property and is solid, and that $L^\infty(G)$ is (in most cases) a prime type $\text{II}_1$-factor. As applications of these and other results, we deduce the metric approximation property, exactness, simplicity and uniqueness of trace for the reduced $C^*$-algebra $C_r(G)$, and the existence of a multiplier-bounded approximate identity for the convolution algebra $L^1(G)$.

2000 Mathematics Subject Classification: Primary 46L65, 20G42; Secondary 46L54.
Keywords and Phrases: Quantum automorphism groups, approximation properties, property of rapid decay, $\text{II}_1$-factor, solid von Neumann algebra, Temperley-Lieb algebra.

1 Introduction

Consider a finite set $X_n$ consisting of $n$ distinct points. An elementary and well known fact from group theory is that the automorphism group of $C(X_n)$, the commutative algebra of complex valued functions over $X_n$, is precisely the permutation group $S_n = \text{Aut}(X_n)$. In other words, the permutation group

\textsuperscript{1}Research partially supported by an NSERC Postdoctoral Fellowship.
on $n$ points is the universal object within the category of groups acting on $C(X_n)$. In [40], Wang made the remarkable discovery that if one replaces the category of groups acting on $C(X_n)$ by the category of compact quantum groups acting on $C(X_n)$, then there still exists a universal object within this new category, called the quantum permutation group $S^+_n$. $S^+_n$ is a C*-algebraic compact quantum group which contains the classical permutation group $S_n$ as a quantum subgroup. Moreover, the inclusion $S_n \subseteq S^+_n$ is proper when $n \geq 4$. That is, finite sets with at least four elements admit genuinely “quantum permutations”.

More generally, given a (non-commutative) finite dimensional C*-algebra $B$, one can ask about the existence of a universal object within the category of compact quantum groups acting on $B$. This question was also considered by Wang [40], who showed that such an object exists provided we fix a faithful state $\psi : B \to \mathbb{C}$ and restrict to the subcategory of compact quantum groups acting on $B$ which preserve the state $\psi$. The resulting object is called the quantum automorphism group of the pair $(B, \psi)$, and is denoted by $G_{\text{aut}}(B, \psi)$. When $B = C(X_n)$ and $\psi$ is the uniform probability measure on $X_n$, we recover the quantum permutation group $S^+_n$. See Example 3.4.

Since the quantum groups $G_{\text{aut}}(B, \psi)$ generally depend a great deal on the initial choice of state $\psi$, a natural first step in attempting to develop a structure theory for these objects is to restrict attention to certain “canonical” choices of states on $B$ (which should ideally correspond to uniform probability measures in the commutative case). In [3, 4], Banica showed that one very natural choice of state $\psi : B \to \mathbb{C}$, which ensures that the generators of the category of finite dimensional representations of $G = G_{\text{aut}}(B, \psi)$ satisfy the fewest relations (or are as “free” as possible), is to take $\psi$ to be the restriction to $B$ of the unique trace on $L(B)$, where $B \hookrightarrow L(B)$ via the left regular representation. In particular, with the above choice of $\psi$ (or more generally by taking $\psi$ to be a $\delta$-form on $B$ – see Definition 3.5), Banica showed that when $\dim B \geq 4$, the finite dimensional representation theory of $G$ admits a nice description using Temperley-Lieb algebras. Using this link with Temperley-Lieb algebras, he obtained a complete description of the fusion rules for the irreducible unitary representations of $G$ and from this deduced many interesting structure results, including the fact that $G$ is co-amenable if and only if $\dim B \leq 4$.

The goal of this paper is to investigate the operator algebraic structure of the quantum groups $G = G_{\text{aut}}(B, \psi)$, where $\psi : B \to \mathbb{C}$ is the canonical trace described above. Our focus will be on properties of the reduced C*-algebras $C_r(G)$ and reduced von Neumann algebras $L^\infty(G) = C_r(G)'$, as well as the quantum convolution algebras $L^1(G) = L^\infty(G)_*$. Note that in the co-amenable (i.e., $\dim B \leq 4$) case, the structure of $G$ is already well understood. If $\dim B = n \leq 3$, then $B = C(X_n)$ and $G = S_n$. If $\dim B = 4$, then either $B = M_2(\mathbb{C})$ or $C(X_4)$. In the first case, $G$ turns out to be isomorphic to $SO(3)$, the classical *-automorphism group of $M_2(\mathbb{C})$ (see [3]). In the second case, $G = S^+_4$, which has been extensively studied in [5] and [7]. In particular, Banica and Collins [7] have obtained a concrete model for the reduced (= full, Documenta Mathematica 18 (2013) 1349–1402
Quantum Automorphism Groups

due to co-amenability) C∗-algebra $C_r(S^+_4)$ in the form of an explicit embedding $C_r(S^+_4) \hookrightarrow M_4(\mathbb{C}) \otimes C(SU(2))$.

In the non-co-amenable (i.e., dim $B \geq 5$) regime, much less is known about the structure of $G$ and its operator algebras. Among the known results, we note that De Rijdt and Vander Vennet [20] have obtained a detailed description of the probabilistic boundaries associated to the discrete dual quantum group $\hat{G}$. In addition, combinatorial “Weingarten” formulas for the Haar state on the quantum permutation groups $S^+_n$ were obtained in [6]. The large $n$ asymptotics of these formulas have found many interesting applications in free probability theory. See for example [8, 9, 28]. Perhaps most striking of these applications is Köstler and Speicher’s free analogue of de Finetti’s theorem [28], which asserts that an infinite sequence of random variables in a $W^*$-probability space is identically distributed and free with amalgamation if and only if its joint $*$-distribution is invariant under the action of the quantum permutation groups $\{S^+_n\}_{n \geq 1}$.

Returning now to a fixed quantum automorphism group $G = G_{\text{aut}}(B, \psi)$ with dim $B \geq 5$, basic results in co-amenability theory imply that $C_r(G)$ is non-nuclear, $L^\infty(G)$ is non-injective and $L^1(G)$ is a Banach algebra which fails to have a bounded approximate identity. However, based on the above connections between $S^+_n$ and free independence, as well as some compelling evidence provided by work of Vaes and Vergnioux [36] on free orthogonal quantum groups, more is conjectured to be true. Namely, it is expected that $L^\infty(G)$ is a full, prime type II$_1$-factor and $C_r(G)$ is a simple C∗-algebra with unique trace when dim $B \geq 5$ (see [7], Introduction).

In this paper we show that if dim $B \geq 8$, then $L^\infty(G)$ is a full type II$_1$-factor (Theorem 5.1) and $C_r(G)$ is simple with unique trace (Corollary 5.12) – verifying the above conjecture in all but six cases. We also prove that $L^\infty(G)$ is a solid von Neumann algebra (Corollary 5.14): the relative commutant of any diffuse injective von Neumann subalgebra of $L^\infty(G)$ is injective. Combining this with the above factoriability result, we conclude that $L^\infty(G)$ is a prime factor as soon as dim $B \geq 8$. The solidity of $L^\infty(G)$ is established by proving that $L^\infty(G)$ has property (AO)$^+$ (Theorem 5.13), which implies solidity by a celebrated result of Ozawa [31]. See Section 5.4 for the relevant details and definitions. Our proof of property (AO)$^+$ follows very closely the method of Vergnioux [37], where property (AO)$^+$ was established for universal discrete quantum groups.

In addition to considering the algebraic structure of $L^\infty(G)$ and $C_r(G)$, we also investigate various approximation properties for these operator algebras and the convolution algebra $L^1(G)$. We prove that when dim $B \geq 5$, $L^\infty(G)$ has the Haagerup property (Theorem 4.2), $C_r(G)$ (respectively $L^\infty(G)$) has the (respectively weak*) metric approximation property, and that $L^1(G)$ admits a central approximate identity which is contractive in the multiplier-norm (Theorem 4.15). To obtain these approximation properties, we first construct a suitable family of $L^2$-compact completely positive multipliers on $L^\infty(G)$, establishing the Haagerup property. We then prove that the discrete dual quantum
group $\hat{G}$ has the property of rapid decay, which amounts to proving a type of Haagerup inequality for $L^\infty(G)$ (Theorem 4.10). The norm control provided by the property of rapid decay for $\hat{G}$ allows us to truncate our completely positive multipliers down to finite rank contractions in such a way as to yield the (weak$^*$) metric approximation property. (We also use the property of rapid decay in the proof of simplicity for $C_r(G)$). The existence of the claimed approximate identity for $L^1(G)$ then follows by a standard duality argument. We also observe that $C_r(G)$ is an exact C$^*$-algebra (Corollary 4.17), by piecing together some results on monoidal equivalence for compact quantum groups from [20] and [36].

The remaining parts of this paper are organized as follows: Section 2 fixes some notation and contains the basic facts on compact and discrete quantum groups that will be needed. In Section 3 we recall the definition of the quantum automorphism groups $G = G_{\mathrm{aut}}(B, \psi)$, $\delta$-forms and describe the representation theory of $G$ when $\psi$ is a $\delta$-form. Section 4 deals with approximation properties for the reduced operator algebras of $G$ (assuming $\psi$ is the canonical trace on $B$) and the property of rapid decay for $\hat{G}$. In Section 5 we prove our factoriality, fullness and simplicity results for the reduced operator algebras of $G$. We end Section 5 by establishing property $(AO)^+$ for $L^\infty(G)$, deducing solidity and primeness results from this. The final section (Section 6) is an appendix containing a proof of Theorem 5.4, which is a technical result required in our study of factoriality and fullness for $L^\infty(G)$.

Acknowledgements

The author wishes to thank T. Banica, B. Collins, P. Fima, M. Junge, J. A. Mingo, R. Speicher and R. Vergnioux for useful conversations and comments at various stages of this project.

2 Preliminaries

2.1 Notation

For a Hilbert space $H$, we take its inner product to be conjugate-linear in the first variable and write $\mathcal{B}(H)$ for the C$^*$-algebra of bounded linear operators on $H$. Given $\xi, \eta \in H$, we write $\omega_{\xi, \eta} \in \mathcal{B}(H)_+$ for the $\sigma$-weakly continuous linear functional $T \mapsto \langle \eta | T \xi \rangle$. For tensor products, we write $\otimes$ for the minimal tensor product of C$^*$-algebras or Hilbertian tensor product of Hilbert spaces, $\boxtimes$ for the spatial tensor product of von Neumann algebras and $\otimes_{\text{alg}}$ for the algebraic tensor product. If $X$ is an operator space, we write $\mathcal{CB}(X)$ for the algebra of completely bounded linear maps on $X$. We take the natural numbers $\mathbb{N}$ to include 0 and for $m \geq 1$ we write $[m] := \{1, 2, \ldots, m\}$. We also use standard leg numbering notation for operators on multiple tensor products. For example, if $H_1, H_2, H_3$ are Hilbert spaces and $T \in \mathcal{B}(H_1 \otimes H_2)$, then
Every compact quantum group $G_{a,b} \in L$ of reduced von Neumann algebra $a$ normal faithful $^\ast$-homomorphism $\Delta$ of the Haar state respectively. Let $H$ be a finite dimensional Hilbert space. A representation $\rho : \mathfrak{A} \to \mathcal{B}(L^2(H))$ of $\mathfrak{A}$ is $\ast$-homomorphism $\Delta$ satisfying the coassociativity condition

$$(\text{id}_A \otimes \Delta) \circ \Delta = (\Delta \otimes \text{id}) \circ \Delta$$

and the cancellation property

$$\text{span}(\Delta(A)(1_A \otimes A)) = A \otimes A = \text{span}(\Delta(A)(A \otimes 1_A)).$$

Every compact quantum group $\mathcal{G} = (A, \Delta)$ admits a Haar state, which is a state $h : A \to \mathbb{C}$ determined uniquely by the following bi-invariance condition with respect to the coproduct:

$$(h \otimes \text{id}_A)\Delta(a) = (\text{id}_A \otimes h)\Delta(a) = h(a)1_A \quad (a \in A).$$

Let $L^2(\mathcal{G}) := L^2(A, h)$ denote the GNS Hilbert space associated to the state $h$ and let $\lambda : A \to \mathcal{B}(L^2(\mathcal{G}))$ the left regular representation. That is, $L^2(\mathcal{G})$ is the Hilbert space obtained from $A$ by separation and completion with respect to the sesquilinear form $\langle a|b \rangle_h = h(a^* b)$, and $\lambda$ is defined by $\lambda(a) = \Delta_h(A)$, where $a, b \in A$ and $\Delta_h : A \to L^2(\mathcal{G})$ is the GNS map. The reduced $C^\ast$-algebra and reduced von Neumann algebra of $\mathcal{G}$ are the concrete operator algebras acting on $L^2(\mathcal{G})$ given by

$$C_r(\mathcal{G}) := \lambda(\mathfrak{A}) \subseteq \mathcal{B}(L^2(\mathcal{G})) \quad \text{and} \quad L^\infty(\mathcal{G}) := C_r(\mathcal{G})'',$$

respectively.

Observe that the $\Delta$-invariance of the Haar state $h$ implies that there exists a normal faithful $\ast$-homomorphism $\Delta_r : L^\infty(\mathcal{G}) \to L^\infty(\mathcal{G}) \subseteq L^\infty(\mathcal{G})$ given by $\Delta_r \circ \lambda = (\lambda \otimes \lambda) \circ \Delta$, turning $(C_r(\mathcal{G}), \Delta_r)$ into a compact quantum group (the reduced version of $\mathcal{G}$). The pre-adjoint of $\Delta_r$ induces a completely contractive Banach algebra structure on the predual $L^1(\mathcal{G}) := L^\infty(\mathcal{G})_\ast$. $L^1(\mathcal{G})$ is called the convolution algebra of $\mathcal{G}$. $\mathcal{G}$ is said to be of Kac-type if the Haar state $h$ is tracial. In this case, $L^\infty(\mathcal{G})$ is a finite von Neumann algebra with faithful Haar trace $h$.

Let $H$ be a finite dimensional Hilbert space. A representation of $\mathcal{G}$ on $H$ is an element $U \in \mathcal{B}(H) \otimes A$ such that

$$(\text{id} \otimes \Delta)U = U_{12}U_{13}. \quad (2.1)$$

2.2 Compact and Discrete Quantum Groups

We present here a very brief summary of the basic facts on compact and discrete quantum groups that will be needed for this paper. Our main reference for this will be [43] and the excellent book [35].

**Definition 2.1.** A **compact quantum group** is a pair $\mathcal{G} = (A, \Delta)$, where $A$ is a unital $C^\ast$-algebra and $\Delta : A \to A \otimes A$ is a unital $\ast$-homomorphism (called a **coproduct**) satisfying the coassociativity condition
By fixing an orthonormal basis \( \{ e_i \}_{i=1}^d \) for \( H \) and writing \( U = [u_{ij}] \in M_d(A) \) relative to this basis, \((2.1)\) is equivalent to requiring that \( \Delta(u_{ij}) = \sum_{k=1}^d u_{ik} \otimes u_{kj} \) for \( 1 \leq i, j \leq d \). \( U \) is called a unitary representation if \( U \) is, in addition, a unitary element of \( M_d(A) \). If \( U^1 = [u_{ij}^{(1)}(1)] \in M_{d(1)}(A) \) and \( U^2 = [u_{ij}^{(2)}(2)] \in M_{d(2)}(A) \) are (unitary) representations, then their tensor product \( U^1 \otimes U^2 = [u_{ij}^{(1)}(1)u_{ij}^{(2)}(2)] \in M_{d(1)d(2)}(A) \) and direct sum \( U^1 \oplus U^2 \in M_{d(1)+d(2)}(A) \) are also (unitary) representations. The vector space

\[
\text{Mor}(U^1, U^2) = \{ T \in B(H_1, H_2) \mid (T \otimes 1_A)U^1 = U^2(T \otimes 1_A) \},
\]

is called the space of morphisms (or intertwiners) between \( U^1 \) and \( U^2 \). \( U^1 \) and \( U^2 \) are called (unitarily) equivalent if there exists an invertible (unitary) operator \( T \in \text{Mor}(U^1, U^2) \), and in this case we write \( U^1 \cong U^2 \). If \( U = [u_{ij}] \in M_d(A) \) is a representation, then so is \( \overline{U} = [\overline{u}_{ij}] \). The representation \( \overline{U} \) is called the conjugate of \( U \). Note that the passage from \( U \) to its conjugate \( \overline{U} \) does not preserve unitarity in general. A representation \( U \) is called irreducible if \( \text{Mor}(U, U) = \text{Cld} \). Note that \( U \) is irreducible if and only if \( \overline{U} \) is irreducible.

The following is a fundamental result in the theory of compact quantum groups.

**Theorem 2.2.** Every irreducible representation of a compact quantum group \( G = (A, \Delta) \) is finite dimensional and equivalent to a unitary one if it is invertible. Moreover, every unitary representation is unitarily equivalent to a direct sum of irreducibles.

Let \( \{ U^\alpha = [u_{ij}^\alpha]_{1 \leq i,j \leq d_\alpha} : \alpha \in \text{Irr}(G) \} \) be a maximal family of pairwise inequivalent finite-dimensional irreducible unitary representations of \( G = (A, \Delta) \), with \( 0 \in \text{Irr}(G) \) denoting the index corresponding to the trivial representation \( 1_A \in A \) and \( U^0 \) denoting the representative of the equivalence class of \( \overline{U^0} \). If \( \mathcal{A} \subseteq A \) denotes the linear span of \( \{ u_{ij}^\alpha : 1 \leq i,j \leq d_\alpha, \alpha \in \text{Irr}(G) \} \), then \( \mathcal{A} \) is a Hopf *-algebra on which \( h \) is faithful, \( \mathcal{A} \) is norm dense in \( A \), and the set \( \{ u_{ij}^\alpha : 1 \leq i,j \leq d_\alpha, \alpha \in \text{Irr}(G) \} \) is a linear basis for \( \mathcal{A} \). The coproduct of \( \mathcal{A} \) arises as the restriction of the coproduct of \( A \), so in particular \( \Delta(A) \subseteq \mathcal{A} \otimes_{\text{alg}} \mathcal{A} \).

The coinverse \( \kappa : \mathcal{A} \to \mathcal{A} \) is the antihomomorphism given by \( \kappa(u_{ij}^\alpha)^* = (u_{ij}^{\alpha^*})^* \), and the counit \( \epsilon : \mathcal{A} \to \mathbb{C} \) is the *-character given by \( \epsilon(u_{ij}^\alpha) = \delta_{ij} \). The Hopf *-algebra \( \mathcal{A} \) is unique in the sense that if \( B \subseteq A \) is any other dense Hopf *-subalgebra, then \( B = \mathcal{A} \) (see [10, Theorem 5.1]).

For each \( \alpha \in \text{Irr}(G) \), fix an isometric morphism \( t_\alpha \in \text{Mor}(1, U^\alpha \otimes \overline{U^0}) \) (which exists and is unique up to scalar multiplication by \( T \)). Let \( j_\alpha : H_\alpha \to H_{\overline{\alpha}} \) be the invertible conjugate-linear map defined by \( j_\alpha(\xi) = (\xi^* \otimes \text{id})t_\alpha(\xi \in H_\alpha) \). Renormalize \( j_\alpha \) so that the positive operator \( Q_\alpha := j_\alpha^*j_\alpha > 0 \) satisfies \( \text{Tr}(Q_\alpha) = \text{Tr}(Q_\alpha^{-1}) \) and \( j_\alpha^*j_\alpha = \pm \text{id}_{H_\alpha} \). The operators \( \{ Q_\alpha \}_{\alpha \in \text{Irr}(G)} \) can then be used to describe the Schur orthogonality relations for the matrix elements of irreducible unitary representations of \( G \), which are given by

\[
h((u_{ij}^\alpha)^*u_{kl}^\beta) = \frac{\delta_{\alpha\beta}\delta_{ik}(Q_\alpha^{-1})_{kj}}{\text{Tr}Q_\alpha} \quad \text{and} \quad h(u_{ij}^\alpha(u_{kl}^\beta)^*) = \frac{\delta_{\alpha\beta}\delta_{jk}(Q_\alpha)_{ij}}{\text{Tr}Q_\alpha}, \quad (2.2)
\]
where \( \alpha, \beta \in \text{Irr}(G) \) and \( 1 \leq i, j \leq d_\alpha, 1 \leq k, l \leq d_\beta \). See [43] for details. In particular, setting \( L_\alpha^2(G) = \text{span}\{\Lambda_h(u_{ij}^\alpha) : 1 \leq i, j \leq d_\alpha\} \), we obtain the \( L^2(G) \)-decomposition

\[
L^2(G) = \bigoplus_{\alpha \in \text{Irr}(G)} L_\alpha^2(G) \cong \bigoplus_{\alpha \in \text{Irr}(G)} H_\alpha \otimes H_{\overline{\alpha}},
\]

(2.3)

where the second isomorphism is given on each direct summand \( L_\alpha^2(G) \) by

\[
A_h((\omega_{\alpha, \xi} \otimes \text{id}_A)U^\alpha) \mapsto \xi \otimes (1 \otimes \eta^*)u_{\overline{\alpha}} \quad (\alpha \in \text{Irr}(G), \; \xi, \eta \in H_\alpha).
\]

(2.4)

Note that when \( G \) is of Kac-type, then \( Q_\alpha = \text{id}_{H_\alpha} \) for each \( \alpha \in \text{Irr}(G) \) and \( \{d_{i/2}A_h(u_{ij}^\alpha) : 1 \leq i, j \leq d_\alpha\} \) is then an orthonormal basis for \( L_\alpha^2(G) \). For future reference, we will always write \( P_\alpha \) for the orthogonal projection from \( L_\alpha^2(G) \) onto \( L_\alpha^2(G) \cong H_\alpha \otimes H_{\overline{\alpha}} \).

Denote by \( C_u(G) \) the universal enveloping \( C^* \)-algebra of the Hopf \( * \)-algebra \( A \) and by \( \pi_u : A \to C_u(G) \) the universal representation. Then there is a \( * \)-homomorphism \( \Delta_u : C_u(G) \to C_u(G) \otimes C_u(G) \) defined by \( \Delta_u \circ \pi_u = (\pi_u \otimes \pi_u) \circ \Delta \), making \( (C_u(G), \Delta_u) \) a compact quantum group (the universal version of \( G \)). In the universal setting, we write \( C_{\text{alg}}(G) \) for the canonical dense Hopf \( * \)-subalgebra \( \pi_u(A) \subseteq C_u(G) \). We say that \( G \) is co-amenable if the regular representation \( \lambda : C_u(G) \to C_r(G) \) is an isomorphism. The compact quantum groups encountered in this paper will all appear canonically in universal form. I.e., given by \( G = (A, \Delta) \) where \( A = C_u(G) \) and \( \Delta = \Delta_u \).

Following [32], we view a discrete quantum group as the dual of a compact quantum group: Let \( G = (A, \Delta) \) be a compact quantum group and let \( \{U^\alpha\}_{\alpha \in \text{Irr}(G)} \) (with \( U^\alpha \in \mathcal{B}(H_\alpha) \otimes A \)) be a maximal family of pairwise inequivalent irreducible unitary representations of \( G \). Define

\[
\mathbb{C}[\hat{G}] = \bigoplus_{\alpha \in \text{Irr}(G)} \mathcal{B}(H_\alpha), \quad c_0(\hat{G}) = c_0 - \bigoplus_{\alpha \in \text{Irr}(G)} \mathcal{B}(H_\alpha),
\]

\[
\ell^\infty(\hat{G}) = \prod_{\alpha \in \text{Irr}(G)} \mathcal{B}(H_\alpha),
\]

and let \( \hat{P}_\alpha \in \ell^\infty(\hat{G}) \) denote the minimal central projection corresponding to the factor \( \mathcal{B}(H_\alpha) \in \ell^\infty(\hat{G}) \).

The unitary operator

\[
\mathbb{V} = \bigoplus_{\alpha \in \text{Irr}(G)} U^\alpha \in M(c_0(\hat{G}) \otimes A),
\]

(2.5)

where \( M(B) \) denotes the multiplier algebra of a \( C^* \)-algebra \( B \), implements the duality between \( G \) and \( \hat{G} \). There is a dual coproduct \( \hat{\Delta} : \ell^\infty(\hat{G}) \to \ell^\infty(\hat{G}) \otimes \ell^\infty(\hat{G}) \) determined by

\[
(\hat{\Delta} \otimes \text{id}_A)\mathbb{V} = \mathbb{V}_{13}\mathbb{V}_{23}.
\]
or equivalently by the relation $\Delta(x)S = Sx$ for each $x \in \ell^\infty(\hat{G})$, $S \in \text{Mor}(U^\alpha, U^\beta \otimes U^\gamma)$. The left and right invariant Haar weights on $\ell^\infty(\hat{G})$, $h_L$ and $h_R$, are given by

$$h_L(x) = \sum_{\alpha \in \text{Irr}(G)} m_\alpha \text{Tr}(Q_\alpha \hat{P}_\alpha x), \quad h_R(x) = \sum_{\alpha \in \text{Irr}(G)} m_\alpha \text{Tr}(Q^{-1}_\alpha \hat{P}_\alpha x)$$

(2.6)

for each $x \in \mathbb{C}[\hat{G}]$, where $m_\alpha = \text{Tr}(Q_\alpha) = \text{Tr}(Q^{-1}_\alpha)$ denotes the quantum dimension of $H_\alpha$. Finally, we define the positive invertible unbounded multiplier $Q = \sum_{\alpha \in \text{Irr}(G)} \hat{P}_\alpha Q_\alpha$ of $c_0(\hat{G})$. Note that $Q$ satisfies the relation $\hat{\Delta}(Q) = Q \otimes Q$ and plays the role of the Haar modular function on a locally compact group.

3 Quantum Automorphism Groups of Finite Dimensional C$^*$-Algebras.

We now turn to the central objects of this paper – quantum automorphism groups of finite dimensional C$^*$-algebras. To define these objects, we first recall the notion of an action of a compact quantum group on a unital C$^*$-algebra.

**Definition 3.1.** Let $B$ be a unital C$^*$-algebra and $G = (A, \Delta)$ a compact quantum group.

1. A right action of $G$ on $B$ is a unital $*$-homomorphism $\alpha : B \to B \otimes A$ satisfying

$$(\alpha \otimes \text{id})\alpha = (\text{id} \otimes \Delta)\alpha \quad \text{and} \quad \text{span}\{\alpha(B)(1 \otimes A)\} = B \otimes A.$$ 

2. If $\psi : B \to \mathbb{C}$ is a positive linear functional on $B$, a right action $\alpha$ of $G$ on the pair $(B, \psi)$ is a right action of $G$ on $B$ which is also $\psi$-invariant. That is,

$$(\psi \otimes \text{id}_A)\alpha(b) = \psi(b)1_A, \quad (b \in B).$$

**Definition 3.2.** ([40, 3, 4]) Let $(B, \psi)$ be a finite dimensional C$^*$-algebra equipped with a faithful state $\psi$. The quantum automorphism group of $(B, \psi)$, denoted by $G = G_{\text{aut}}(B, \psi)$, is the universal quantum group defined by a right action of $G$ on $(B, \psi)$, say $\alpha : B \to B \otimes C_u(G)$, with the following properties:

1. $C_u(G)$ is defined as the universal C$^*$-algebra generated by

$$\{((\omega \otimes \text{id})\alpha(a) : \omega \in B^*, a \in B\}.$$ 

2. The action $\alpha$ is universal in the sense that if $\beta : B \to B \otimes A'$ is an action of another compact quantum group $G' = (A', \Delta')$ on $(B, \psi)$, then there exists a unital $*$-homomorphism $\pi : C_u(G) \to A'$ such that $\beta = (\text{id}_B \otimes \pi)\alpha$.
Remark 3.3. As was shown in [40, 3, 4], it is possible to give a more "concrete" description of the quantum group $G = G_{\text{aut}}(B, \psi)$ in terms of generators and relations. Let $n = \dim B$ and fix an orthonormal basis $\{b_i\}_{i=1}^n$ for $B$ relative to the inner product $\langle \cdot | \cdot \rangle_\psi$ induced by $\psi$. Let $m : B \otimes B \to B$ be the multiplication map and $\nu : \mathbb{C} \to B$ the unit map. Then $C_u(G)$ is the universal $C^*$-algebra with $n^2$ generators $\{u_{ij} : 1 \leq i, j \leq n\}$ subject to the relations which make

$$U = [u_{ij}] \in M_n(C_u(G)) \text{ unitary, } \quad m \in \text{Mor}(U^{\otimes 2}, U) \quad \text{and} \quad \nu \in \text{Mor}(1, U).$$

(3.1)

The universal property of $C_u(G)$ then allows us to define a coproduct $\Delta : C_u(G) \to C_u(G) \otimes C_u(G)$ by the formula

$$\Delta(u_{ij}) = \sum_{k=1}^n u_{ik} \otimes u_{kj} \quad (1 \leq i, j \leq n),$$

which turns $(C_u(G), \Delta)$ into a compact quantum group with fundamental representation $U$ acting on the Hilbert space $H = (B, \langle \cdot | \cdot \rangle_\psi)$. (Recall that a fundamental representation of a compact quantum group $G = (A, \Delta)$ is a representation $U$ whose matrix elements generate $A$ as a $C^*$-algebra.) The universal action $\alpha$ of $G$ on $(B, \psi)$ is uniquely determined by $\alpha(b_i) = \sum_{j=1}^n b_j \otimes u_{ji}$, $(1 \leq i \leq n)$.

Example 3.4. Consider the simplest situation where $X_n = \{1, \ldots, n\}$, $B = C(X_n)$ and $\psi$ is the uniform probability measure on $X_n$. Taking as an orthogonal basis the $n$ standard projections in $C(X_n)$, it can be shown using Remark 3.3 that $C_u(G)$ is the universal $C^*$-algebra generated by $n^2$ projections $\{u_{ij} : 1 \leq i, j \leq n\}$ with the property that the rows and columns of the matrix $U = [u_{ij}]$ form partitions of unity. (An $n \times n$ matrix $U$ over a $C^*$-algebra $A$ satisfying the above properties is called a magic unitary.) On the other hand, $C_u(S_n^+)\otimes C_u(S_n^+)$, the universal $C^*$-algebra associated to the quantum permutation group $S_n^+$, admits the same description by [40], Theorem 3.1. Therefore $G_{\text{aut}}(C(X_n), \psi) = S_n^+$.

As mentioned in the introduction, this paper deals almost exclusively with finite dimensional $C^*$-algebras equipped with certain canonical tracial states, called tracial $\delta$-forms [3, 4].

Definition 3.5. Let $B$ be a finite dimensional $C^*$-algebra equipped with a faithful state $\psi : B \to \mathbb{C}$ and let $\delta > 0$. We call $\psi$ a $\delta$-form if for the inner products on $B$ and $B \otimes B$ implemented by $\psi$ and $\psi \otimes \psi$, respectively, the multiplication map $m : B \otimes B \to B$ satisfies $mm^* = \delta^2 \text{id}_B$.

The reason for considering $\delta$-forms is two-fold. On the one hand, the class of $\delta$-forms yields many of the interesting concrete examples of quantum automorphism groups (including quantum permutation groups – see Example 3.7 below). On the other hand, it turns out that the concrete monoidal $W^*$-category of finite dimensional unitary representations of $G_{\text{aut}}(B, \psi)$ admits a
very tractable description in this setting (see Section 3.1). For a brief discussion of the situation where \( \psi \) is not a \( \delta \)-form, we refer the reader to the concluding remark at the end of the paper in Section 5.5.

**Remark 3.6.** Let \( \psi \) be a \( \delta \)-form on a finite dimensional C*-algebra \( B \). One can readily verify (see for example [4], Theorem 1) that in addition to usual relations

\[
\begin{align*}
mm^* &= \delta^2 \text{id}_B, \\
\nu^* \nu &= 1, \\
m(m \otimes \text{id}_B) &= m(\text{id}_B \otimes m), \\
m(\text{id}_B \otimes \nu) &= m(\nu \otimes \text{id}_B) = \text{id}_B
\end{align*}
\]

satisfied by the multiplication map \( m : B \otimes B \to B \) and the unit map \( \nu : \mathbb{C} \to B \), the additional useful relation

\[
m^*m = (m \otimes \text{id}_B)(\text{id}_B \otimes m^*),
\]

also holds.

We now consider some examples.

**Example 3.7.**

1. If \( B = \mathbb{C}(X_n) \) where \( X_n = \{1,2,\ldots,n\} \) and \( \psi \) is the uniform probability measure on \( X_n \), then \( \psi \) is a \( \delta \)-form with \( \delta^2 = \dim B = n \). Note that by Example 3.4, \( \mathbb{G}_{\text{aut}}(B,\psi) = S_n^+ \).

2. If \( B = M_n(\mathbb{C}) \) and \( \psi = \text{Tr}(Q \cdot) \) with \( Q > 0 \), then \( \psi \) is a \( \delta \)-form with \( \delta^2 = \text{Tr}(Q^{-1}) \).

3. If \( (B,\psi) \) is arbitrary, then \( B \) decomposes as the direct sum \( B = \bigoplus_{i=1}^k M_{n_i}(\mathbb{C}) \) and \( \psi \) can be written as \( \psi = \bigoplus_{i=1}^k \text{Tr}(Q_i \cdot) \) with \( Q_i > 0 \). Then \( \psi \) is a \( \delta \)-form if and only if \( \text{Tr}(Q_i^{-1}) = \delta^2 \) for \( 1 \leq i \leq k \). See [4].

Using the notation from Example 3.7(3) and the inequality \( \dim B \leq \sum_{1 \leq i \leq k} \text{Tr}(Q_i)\text{Tr}(Q_i^{-1}) \), it follows that \( \delta^2 \geq \dim B \) for any \( \delta \)-form \( \psi \). If \( \psi \) is moreover a trace, then \( \delta^2 = \dim B \) and \( \psi \) is precisely the restriction to \( B \) of the unique trace on \( \mathcal{L}(B) \), described in Section 1. In this case, we call \( \psi \) the \( \delta \)-trace or tracial \( \delta \)-form on \( B \). Note that when \( \psi \) is a trace, \( \mathbb{G}_{\text{aut}}(B,\psi) \) is a compact quantum group of Kac-type [40].

We remind the reader that for \( n = \dim B \leq 3 \), \( \mathbb{G} = \mathbb{G}_{\text{aut}}(B,\psi) \) is just the usual permutation group \( S_n \) [40], so for the remainder of the paper we will assume that \( \dim B \geq 4 \).

### 3.1 Representation Theory and the 2-cabled Temperley-Lieb Category

Let \( \delta \geq 2 \) and \( \psi \) be a \( \delta \)-form on a finite dimensional C*-algebra \( B \). In this section we review the description of the representation theory of \( \mathbb{G} = \mathbb{G}_{\text{aut}}(B,\psi) \) obtained by Banica in [3, 4]. In the following paragraphs it will be convenient to use the language of (concrete) monoidal W*-categories. We refer to [42] and [23] for the appropriate definitions.
In [4] (see [3] for the tracial case) it is shown that the fundamental representation \( U \) of \( G \) is equivalent to \( \mathbb{U} \), and therefore the concrete monoidal \( W^* \)-category \( R \) of all finite dimensional unitary representations of \( G \) is the completion of the concrete monoidal \( W^* \)-category

\[ R_0 = \{ \{ U^{\otimes k} \}_{k \in \mathbb{N}}, \{ \text{Mor}(U^{\otimes k}, U^{\otimes l}) \}_{k, l \in \mathbb{N}} \}, \]

whose objects are the tensor powers \( U^{\otimes k} \) of \( U \) and whose morphisms are the associated spaces of intertwiners. Moreover, \( R_0 \) is generated as a monoidal \( W^* \)-category by the morphisms \( m \in \text{Mor}(U^{\otimes 2}, U), \nu \in \text{Mor}(1, U) \) and \( \text{id}_B \in \text{Mor}(U, U) \).

Consider the universal graded \( C^* \)-algebra \( (\text{TL}_{k,l}(\delta))_{(k,l) \in \mathbb{N} \times \mathbb{N}} \) given by \( \text{TL}_{k,l}(\delta) = \{ 0 \} \) if \( k-l \) is odd, and generated by elements \( t(k, l) \in \text{TL}_{k+1,l+2}(\delta) \) with the following relations (where \( 1_n \) denotes the unit of the \( C^* \)-algebra \( \text{TL}_n(\delta) := \text{TL}_{n,n}(\delta) \)):

\[
\begin{align*}
t(k, l)^* t(k, l) &= 1_{k+l} \\
t(k, l + 1)^* t(k + 1, l) &= 1_{k+l+1} \\
t(r, k + l + 2)^* t(r + k, l) &= t(r + k + 2, l) t(r, k + l) \\
t(r, k + l + 2)^* t(r + k + 2, l) &= t(r + k, l)^* t(r, k + l + 2).
\end{align*}
\]

It is well known that each (necessarily finite dimensional!) vector space \( \text{TL}_{k,l}(\delta) \) has the following planar-diagrammatic interpretation: \( \text{TL}_{k,l}(\delta) \) is the \( \mathbb{C} \)-vector space spanned by a basis of \((k, l)\)-Temperley-Lieb diagrams \( \{ D_\pi \} \in \text{NC}_2(k+l) \).

(That is, diagrams \( D_\pi \) consisting of two parallel rows of points \( - k \) points on the bottom row and \( l \) points on the top row \( - \) which are connected by a non-crossing pairing \( \pi \in \text{NC}_2(k+l) \)). We refer to the book [27] for details on the spaces \( TL_{k,l}(\delta) \). Note, however, that our choice of notation and terminology differs slightly from theirs. In terms of diagrams, the generators \( t(k, l) \in \text{TL}_{k+1,l+2}(\delta) \) are given by the (scaled) diagrams

\[
\begin{align*}
t(k, l) &= \delta^{-1/2} \left| \begin{array}{c} \cdot && \cdot \\
\cdot & \cdot & \cdots \\
\cdot & \cdot & \cdots \\
\cdot & \cdot & \cdots \end{array} \right|, \\
t(k, l)^* &= \delta^{-1/2} \left| \begin{array}{c} \cdot && \cdot \\
\cdot & \cdot & \cdots \\
\cdot & \cdot & \cdots \\
\cdot & \cdot & \cdots \end{array} \right|.
\end{align*}
\]

The composition \( D_\pi D_\sigma \in \text{TL}_{k,l}(\delta) \) of diagrams \( D_\pi \in \text{TL}_{k,l}(\delta), D_\sigma \in \text{TL}_{k,l}(\delta) \) is obtained by the following procedure. First stack the diagram \( D_\pi \) on top of \( D_\sigma \), connecting the bottom row of \( s \) points on \( D_\pi \) to the top row of \( s \) points on \( D_\sigma \). The result is a new planar diagram, which may have a certain number \( c \) of internal loops. By removing these loops, we obtain a new diagram \( D_\rho \in \text{TL}_{k,l}(\delta) \). The product \( D_\pi D_\rho \) is then defined to be \( \delta^{c} D_\rho \). The involution \( D_\pi \mapsto D_\pi^* \) is just the conjugate linear extension of the operation of turning diagrams upside-down.

The collection \( \text{TL}(\delta) = \{ \mathbb{N}, \{ \text{TL}_{k,l}(\delta) \}_{k, l \in \mathbb{N}} \} \) has a natural structure as a monoidal \( W^* \)-category and is called the \textit{Temperley-Lieb category} (with parameter \( \delta \)). Related to this is the \textit{2-cabled Temperley-Lieb category} \( \text{TL}_2^2(\delta) = \)}
\{N, \{TL_{k,l}^2(\delta)\}_{k,l \in \mathbb{N}}\}, \text{ where } TL_{k,l}^2(\delta) := TL_{2k,2l}(\delta) \text{ for each } k,l \in \mathbb{N}. \text{ A fundamental result of [4] is that there is an isomorphism of monoidal } W^*-\text{categories}

\[ \pi : R_0 \to TL^2(\delta) \text{ such that } \pi(U_{\otimes k}) = k, \]

\[ \pi(\text{Mor}(U_{\otimes k}, U_{\otimes l})) = TL_{2k,2l}(\delta) \quad (k,l \in \mathbb{N}). \]

This is defined on the generating morphisms \(\nu, m, \text{id}_B\) of \(R_0\) by

\[ \pi(\nu) = \iota(0,0) = \delta^{-1/2} \bigcup TL_{0,2}(\delta), \]

\[ \pi(m) = \delta\iota(1,1)^* = \delta^{1/2} \bigcap TL_{4,2}(\delta), \quad \pi(\text{id}_B) = 1_2 = 1 \bigcap TL_2(\delta). \]

From now on we will omit \(\pi\) in our notation and simply view \(\text{id}_B, \nu\) and \(m\) as generators of (a concrete faithful representation of) \(TL^2(\delta)\).

Using the above isomorphism, a description of the irreducible unitary representations of \(\mathbb{G}\) and their fusion rules can be obtained. See [4] and Theorem 4.1 of [3] for the tracial case.

**Theorem 3.8.** There exists a bijection \(\text{Irr}(\mathbb{G}) \cong \mathbb{N}\) and a maximal family \(\{U^k\}_{k \in \mathbb{N}}\) of pairwise inequivalent irreducible finite dimensional unitary representations of \(\mathbb{G}\) such that:

1. \(\overline{\cup}^k \cong U^k\) for each \(k \in \mathbb{N}\).

2. \(U^0 = 1_{C_4(\mathbb{G})}\) and the fundamental representation \(U\) decomposes as \(U \cong U^0 \oplus U^1\).

3. \(U^n \otimes U^k \cong \bigoplus_{r=0}^{\min(n,k)} U^{k+n-r}\) for all \(n,k \in \mathbb{N}\). In particular, \(\mathbb{G}\) has the same fusion rules as \(SO(3)\).

4. The sequence of representation dimensions \(\{d_k = \dim(U^k)\}_{k \in \mathbb{N}}\) is given by the recursion

\[ d_0 = 1, \quad d_1 = \dim B - 1, \quad d_1d_k = d_{k+1} + d_k + d_{k-1} \quad (k \geq 1). \]

### 3.2 Explicit Models

To facilitate our investigation of certain structural properties \(\mathbb{G}\), it will be useful to have explicit models for the irreducible representations \(\{U^k\}_{k \in \mathbb{N}}\) of \(\mathbb{G}\), as well as concrete expressions for certain morphisms between their tensor products.

Consider the \(k\)th tensor power \(U_{\otimes k}\) of the fundamental representation of \(\mathbb{G}\). Using the isomorphism \(R_0 = TL^2(\delta)\), we can consider the Jones-Wenzl projection \(p_{2k} \in \text{Mor}(U_{\otimes 2k}, U_{\otimes 2k}) = TL_{2k,2l}(\delta)\). Put \(H_k = p_{2k}B_{\otimes k}\) and define \(U^k\) to be the subrepresentation of \(U_{\otimes k}\) obtained by restricting to the invariant subspace \(H_k \subset B_{\otimes k}\). One can inductively show that \(\{U^k\}_{k \in \mathbb{N}}\) forms a complete family of pairwise inequivalent irreducible unitary representations of \(\mathbb{G}\) satisfying the hypotheses of Theorem 3.8. For a good introduction to the basic properties of the Jones-Wenzl projections \(\{p_y\}_{y \in \mathbb{R}}\), we refer the reader to the book [27].
Using the notation from Section 3.1, note that \( p_y \in TL_q(\delta) \) is given explicitly by \( p_0 = 1, p_1 = 1, p_2 = 1 - 2\nu \nu^* = 1 - (0,0)t(0,0)^* \), and (more generally)
\[
p_y = 1_y - \bigvee_{0 \leq r \leq y-2} t(r,y-r-2)t(r,y-r-2)^*
= \cdots - \bigvee_{0 \leq r \leq y-2} \left( \delta^{-1} \cdots \cdots \right) \ (y \geq 2).
\]
Here, \( \vee \) denotes the supremum of a family of projections. Note that a useful consequence of the above formulae for \( \{ \lor \} \) is the following absorption property:
\( (p_x \otimes p_y)p_{x+y} = p_{x+y} \forall x, y \in \mathbb{N} \). This property will be used throughout the paper.

Given \( n, k, l \in \mathbb{N} \) such that \( U^l \subset U^n \otimes U^k \) (i.e., \( U^l \) is equivalent to a sub-representation of \( U^n \otimes U^k \)), we now proceed to construct explicit morphisms \( \rho_{l}^{n,k} \in \text{Mor}(U^l, U^n \otimes U^k) = (p_{2n} \otimes p_{2k})TL_{2l,2(n+k)}(\delta)p_{2l} \). To do this we first require some notation.

**Notation 3.9.**
1. Let \( \delta \geq 2 \) and let \( 0 < q \leq 1 \) be such that \( \delta = q + q^{-1} \).
   Recall that the \( q \)-numbers and \( q \)-factorials are defined for each \( a \in \mathbb{N} \) by
   \[
   [a]_q = q^a - q^{-a} \quad \text{and} \quad [a]_q! = \frac{q^{a+1}(1-q^{2a})}{1-q^2}, \quad [a]_q! = [a]_q[a-1]_q \cdots [1]_q.
   \]
   Of course when \( q = 1 \), the above formulas reduce to \( [a]_q = a \) and \( [a]_q! = a! \). Observe that \( \delta = [2]_q, [3]_q = \delta^2 - 1 \), and \( [2k+1]_q = m_k \) - the quantum dimension of \( U^k \).

2. Given \( a \in \mathbb{N} \), \( 1_a \) will denote the unit of the algebra \( TL_a(\delta) \). Thus \( 1_{2a} = \text{id}_{B_2^\infty} \) via the isomorphism \( R_0 \cong TL^2(\delta) \). More generally, if \( x \leq a \), \( 1_x \) will denote any collection of \( x \) parallel through-strings in a Temperley-Lieb diagram belonging to \( TL_a(\delta) \).

3. For \( r \geq 1 \), let \( t_r = [r+1]_q^{-1/2} (p_r \otimes p_r) \cup (r) \), where \( \cup (r) \) denotes \( r \) nested cups in \( TL_{0,2r}(\delta) \). In terms of planar diagrams,
   \[
t_r = [r+1]_q^{-1/2} \quad \text{decorated with} \quad p_r \quad \text{and} \quad p_r \in TL_{0,2r}(\delta).
   \]

Now if \( U^l \subset U^n \otimes U^k \), then by Theorem 3.8 there is a unique \( 0 \leq r \leq 2 \min \{n, k\} \) such that \( l = n + k - r \). Define \( \rho_{l}^{n,k} \in \text{Mor}(U^l, U^n \otimes U^k) \) by setting
\[
\rho_{l}^{n,k} = (p_{2n} \otimes p_{2k})(1_{2n-r} \otimes t_r \otimes 1_{2k-r})p_{2l}.
\]

**Documenta Mathematica** 18 (2013) 1349–1402
Using the results of [27], Section 9.10, it follows that $(\rho^k_1)^* \rho^k_1 = C_{(n,k,t)}p_t$, where

$$C_{(n,k,t)} = \frac{[2n + 2k - r + 1]_q^2 [2n - r]_q! [2k - r]_q!}{[r + 1]_q [2l + 1]_q [2n]_q [2k]_q! [2l]_q!} > 0. \quad (3.3)$$

Therefore $C_{(n,k,t)}^{-1/2} \rho_1^k$ is an isometric morphism from $H_1$ onto the subspace of $H_n \otimes H_k$ equivalent to $H_t$. Moreover, this isometry is unique (up to scalar multiplication by $T$) since the inclusion $U^j \subset U^n \otimes U^k$ is multiplicity-free by Theorem 3.8.

When we study the property of rapid decay and factoriality, it will be useful for us to have a description of the quantity $1_{2n-r} \otimes t_r \otimes 1_{2k-r}$ in formula (3.2) in terms of the basic morphisms $m, \nu, l$ and the Jones-Wenzl projections $\{ p_{y} \}_{y \in \mathbb{N}}$. This can be done recursively as follows: if $r = 2s$ is even, then $t_{2s}$ is given by the initial condition and recursion

$$t_0 = 1 \in \mathbb{C}, \quad t_2 = [3]_q^{-1/2} (p_2 \otimes p_2)(m^* \nu), \quad p_2 = 1_2 - \nu \nu^*, \quad (3.4)$$

$$t_{2(k+1)} = \left( \frac{[2k + 2][2l + 1]}{[2k + 2l + 1]} \right)^{1/2} (p_{2k+2l} \otimes p_{2l+2k})(1_{2k} \otimes t_{2l} \otimes 1_{2k})t_{2k}. \quad (3.5)$$

If $r = 2s + 1$ is odd, then

$$1_{2n-r} \otimes t_r \otimes 1_{2k-r} = \left( \frac{[2s + 1]_q}{[2s + 2]_q [2]_q} \right)^{1/2} (1_{2n-2s-1} \otimes p_{2s+1} \otimes p_{2s+1} \otimes 1_{2k-2s-1})$$

$$\times (1_{2n-2s-2} \otimes (1_2 \otimes t_{2s} \otimes 1_2)) m^* \otimes 1_{2k-2s-2}, \quad (3.6)$$

which can be readily verified using the fact that $m^* = [2]_q (1_1 \otimes t_1 \otimes 1_1)$. Note furthermore that $C_{(k,k,0)} = 1$ for each $k \in \mathbb{N}$ by equation (3.3). Thus each $t_{2k} \in \text{Mor}(1, U^k \otimes U^k)$ is an isometric morphism, and $\rho^k_1$ is a contraction (being a composition of contractions).

To conclude this section let us also fix once and for all an orthonormal basis $\{ e_i \}_{i=1}^{d_1}$ for $H_1 = B \otimes C_1$ and write $U^1 = [u_{ij}]$ relative to this basis. The equivalence $U^1 \cong U^1$ given by Theorem 3.8 ensures that there is an invertible matrix $F_1 \in \text{GL}(d_1, \mathbb{C})$ such that $U^1 = (F_1 \otimes 1) U U^T (F_1^{-1} \otimes 1)$. Moreover, since $U^1$ is irreducible, we can assume that $F_1 F_1 = c1$ where $c = \pm 1$ (see [1], Page 242). By possibly replacing $F_1$ by $z F_1$ for some $z \in \mathbb{T}$, we may also assume that

$$t_2 = \text{Tr}(F_1^* F_1)^{-1/2} \sum_{i=1}^{d_1} e_i \otimes F_1 e_i \in \text{Mor}(1, U^1 \otimes U^1). \quad (3.7)$$

This follows because both sides of equation (3.7) are evidently isometries in the one dimensional space $\text{Mor}(1, U^1 \otimes U^1)$. Finally, note that when $\psi$ is the $\delta$-trace on $B$, then $G_{\text{ann}}(B, \psi)$ is of Kac-type. We therefore can and will assume...
that $F_1 \in U(d_1)$ is unitary. Also note that when $\psi$ is the $\delta$-trace, formula (3.4) yields $\sigma t_2 = t_2$, where $\sigma$ is the tensor flip-map. On the other hand, formula (3.7) yields $\sigma t_2 = ct_2$. Consequently $F_1 F_1 = 1$. We leave the easy details to the reader.

4 Approximation Properties and The Property of Rapid Decay

This section is devoted to proving some approximation properties for the reduced operator algebras and convolution algebras associated to the quantum groups $G_{\text{aut}}(B, \psi)$, where $\psi$ is the $\delta$-trace on $B$.

4.1 The Haagerup Property

Our first goal is to prove that $L^\infty(G)$ has the Haagerup property. We begin by recalling this approximation property.

**Definition 4.1.** A finite von Neumann algebra $(M, \tau)$ has the Haagerup property if there exists a net $\{\Phi_t\}_{t \in \Lambda}$ of normal unital completely positive $\tau$-preserving maps on $M$ such that

1. For each $t \in \Lambda$, the $L^2$-extension $\hat{\Phi}_t : L^2(M) \to L^2(M)$ is a compact operator.

2. For each $\xi \in L^2(M)$, $\lim_{t \in \Lambda} \|\hat{\Phi}_t \xi - \xi\|_{L^2(M)} = 0$.

Note that the Haagerup property does not actually depend on the choice of faithful normal trace $\tau : M \to \mathbb{C}$ [26].

**Theorem 4.2.** Let $B$ be a finite dimensional $C^*$-algebra with $\delta$-trace $\psi$ and let $G = G_{\text{aut}}(B, \psi)$. Then $L^\infty(G)$ has the Haagerup property.

To prove Theorem 4.2, we begin by recalling a standard method for constructing normal completely positive maps on $L^\infty(G)$ from states on $C_u(G)$. See for example [14], Lemma 3.4.

**Proposition 4.3.** Let $G$ be a compact quantum group and $\varphi \in C_u(G)^*$ be a state. Then there exists a unique normal unital completely positive $h$-preserving map $M_\varphi \in CB(L^\infty(G))$ defined by

$$M_\varphi \lambda(a) = \lambda((\varphi \otimes \text{id}_{C_u(G)}) \Delta a) \quad (a \in C_u(G)).$$

In particular, $M_\varphi C_r(G) \subseteq C_r(G)$.

**Remark 4.4.** It is easily verified that the map $M_\varphi$ defined in Proposition 4.3 satisfies the relation $\Delta_\varphi M_\varphi = (M_\varphi \otimes \text{id}_{L^\infty(G)}) \Delta_\varphi$. A map $T \in CB(L^\infty(G))$ such that $\Delta_\varphi T = (T \otimes \text{id}_{L^\infty(G)}) \Delta_\varphi$ (respectively $\Delta_\varphi T = (\text{id}_{L^\infty(G)} \otimes T) \Delta_\varphi$) is called a completely bounded left (respectively right) multiplier of $L^\infty(G)$. This terminology originates from the classical situation where $L^\infty(G) = VN(\Gamma)$ is the reduced von Neumann algebra of a discrete group $\Gamma$. In this case, the above
Now let \( \{ U^\alpha = [u^\alpha_{ij}] : \alpha \in \operatorname{Irr}(G) \} \) be a maximal family of irreducible unitary representations of a compact quantum group \( G \) and let \( \chi_\alpha = (\operatorname{Tr} \otimes \text{id}_{C_u(G)}) U^\alpha \in C_u(G) \) be the character of the representation \( U^\alpha \). We say that \( T \in \operatorname{CB}(L^\infty(G)) \) is a central completely bounded multiplier if \( T \) is both a left and right completely bounded multiplier of \( L^\infty(G) \). This is equivalent to saying that for each \( \alpha \in \operatorname{Irr}(G) \), there is a constant \( c_\alpha \in \mathbb{C} \) such that \( T \lambda(u^\alpha_{ij}) = c_\alpha \lambda(u^\alpha_{ij}) \) for all \( 1 \leq i, j \leq d_\alpha \). The following result (proved in [14]) will be our main tool for proving Theorem 4.2. It shows that when \( G \) is of Kac type, central completely positive multipliers can be obtained from states on the (generally much smaller and more tractable) subalgebra \( C^*(\chi_\alpha : \alpha \in \operatorname{Irr}(G)) \subset C_u(G) \) generated by the irreducible characters of \( G \).

**Theorem 4.5** ([14], Theorem 3.7). Let \( G \) be a compact quantum group of Kac type and let \( \varphi \in C^*(\chi_\alpha : \alpha \in \operatorname{Irr}(G))^* \) be a state. Then there exists a unique normal unital completely positive \( h \)-preserving map \( T_\varphi \in \operatorname{CB}(L^\infty(G)) \) defined by

\[
T_\varphi \lambda(u^\alpha_{ij}) = \frac{\varphi(\chi_\alpha^*)}{d_\alpha} \lambda(u^\alpha_{ij}) \quad (\alpha \in \operatorname{Irr}(G), 1 \leq i, j \leq d_\alpha).
\]

Moreover, \( T_\varphi C_r(G) \subset C_r(G) \).

**Remark 4.6.** Our proof in [14] of Theorem 4.5 uses the traciality of the Haar state in an essential way. Indeed, the proof relies on both the boundedness of the coinverse \( \kappa : A \to A \) and the existence of an \( h \)-preserving conditional expectation \( E : L^\infty(G) \otimes L^\infty(G) \to \Delta_r(L^\infty(G)) \). Each of these conditions imply that \( h \) is tracial.

Let us now return to the quantum automorphism groups \( G = \mathbb{G}_{\text{aut}}(B, \psi) \). Theorem 4.5 suggests that we study the (state space of the) \( C^* \)-algebra \( C^*(\chi_k : k \in \mathbb{N}) \subset C_u(G) \), where \( \chi_k \) is the character of the irreducible representation \( U^k \) given by Theorem 3.8. Since \( G \) has commutative fusion rules and each irreducible representation \( U^k \) is equivalent to \( U^k \) by Theorem 3.8, it follows from general character theory that \( C^*(\chi_k : k \in \mathbb{N}) \) is commutative and that each character \( \chi_k \) is self-adjoint. In the following paragraphs we will obtain a more precise description of this \( C^* \)-algebra.

Let \( \{ S_k \}_{k=0}^\infty \) denote the (dilated) Chebyshev polynomials of the second kind, which are defined by the recursion

\[
S_0(x) = 1, \quad S_1(x) = x, \quad S_1(x)S_k(x) = S_{k+1}(x) + S_{k-1}(x) \quad (k \geq 1).
\]

These are the monic orthogonal polynomials for Wigner’s semicircle law, which is the probability measure supported on \([-2, 2] \) with density \( \frac{\sqrt{4-x^2}}{\pi} \). Following Section 4 of [29], we let \( \{ \Pi_k \}_{k=0}^\infty \) be the family of polynomials defined by...
Proof. By Theorem 3.8, let \( \Pi_0(x) = 1 \) and \( \Pi_k(x) = S_k(x-2) + S_{k-1}(x-2) \) for each \( k \geq 1 \). Then we have the recursion

\[
\Pi_1(x) \Pi_k(x) = \Pi_{k+1}(x) + \Pi_k(x) + \Pi_{k-1}(x) \quad (k \geq 1).
\]

The polynomials \( \{\Pi_k\}_{k \in \mathbb{N}} \) are the monic orthogonal polynomials for the free Poisson law (with parameter 1), which is the probability measure supported on \([0, 4]\) with density given by \( \frac{\sqrt{4x-x^2}}{2\pi} \). Observe that one can equivalently define the family \( \{\Pi_k\}_{k \in \mathbb{N}} \) by setting \( \Pi_k(x) = S_{2k}(\sqrt{x}) \). To see this, note that \( S_{2k}(x) \) is a polynomial of degree 2 which contains only terms of even degree. Therefore we can define a degree \( k \) polynomial \( \Pi_k \) by setting \( \Pi_k(x) = S_{2k}(\sqrt{x}) \).

It is then easy to verify that \( \Pi_0 = 1, \Pi_1(x) = x-1 = \Pi_1(x) \), and that the \( \Pi_k \)'s satisfy the recursion (4.1).

**Proposition 4.7.** Let \( \chi \in C_u(\mathbb{G}) \) be the character of the fundamental representation \( U \) of \( \mathbb{G} = \mathbb{G}_{\text{aut}}(B, \psi) \). Then \( \chi = \chi^* \) and there is a *-isomorphism

\[
\Psi : C^*(\chi_k : k \in \mathbb{N}) \to C(\sigma) \quad \text{given by} \quad \Psi : \chi_k \mapsto \Pi_k|_\sigma,
\]

where \( \sigma := \sigma(\chi) \subset \mathbb{R} \) is the spectrum of \( \chi \).

**Proof.** By Theorem 3.8, \( U \cong 1 \oplus U^1 \) and each \( \chi_k \) is self-adjoint. Therefore \( \chi = 1 + \chi_1 = (1 + \chi_1)^* = \chi^* \). Moreover, the fusion rules for \( \{U^k\}_{k \in \mathbb{N}} \) yield the character relations \( \chi_1 \chi_k = \chi_{k+1} + \chi_k + \chi_{k-1} \) for all \( k \geq 1 \). Thus \( C^*(\chi_k : k \in \mathbb{N}) = C^*(1, \chi) \). Let \( \Psi : C^*(\chi_k : k \in \mathbb{N}) = C^*(1, \chi) \to C(\sigma) \) be the Gelfand isomorphism. Then for any \( t \in \sigma \) and \( k \geq 1 \), we have \( \Psi_{\chi_1}(t) \Psi_{\chi_k}(t) = \Psi_{(\chi_1 \chi_k)}(t) = \Psi_{\chi_{k+1}}(t) + \Psi_{\chi_k}(t) + \Psi_{\chi_{k-1}}(t) \), with the initial conditions \( \Psi_{\chi_0}(t) = \Psi(1_{c_u(\mathbb{G})}(g)) = 1 \) and \( \Psi_{\chi_1}(t) = \Psi(\chi - \chi_0)(t) = t - 1 \). Comparing with the initial conditions and recursion (4.1) for the polynomials \( \{\Pi_k\}_{k \in \mathbb{N}} \), we obtain \( \Psi_{\chi_k} = \Pi_k|_\sigma \ (k \in \mathbb{N}) \).

This next proposition gives us some useful information about the spectrum \( \sigma \subset \mathbb{R} \) of \( \chi \in C_u(\mathbb{G}) \), at least in the \( \delta \)-trace case.

**Proposition 4.8.** Let \( \psi \) be the \( \delta \)-trace on \( B \) and let \( \sigma = \sigma(\chi) \subset \mathbb{R} \) be as in Proposition 4.7. Then \([0, \dim B] \subset \sigma\).

**Remark 4.9.** We suspect that the equality \([0, \dim B] = \sigma\) always holds in Proposition 4.8, but we are unable to show that \( \sigma \geq 0 \) in general.

**Proof.** Let \( \mu \) be the spectral measure of \( \chi \) relative to the Haar state \( h \). Since the irreducible characters of a compact quantum group form an orthonormal family with respect to the \( L^2(\mathbb{G}) \)-inner product, Proposition 4.7 yields

\[
\int_{\text{supp}(\mu)} \Pi_k(t) \Pi_l(t) d\mu(t) = h(\Pi_k(\chi)^* \Pi_l(\chi)) = h(\chi_k^* \chi_l) = \delta_{kl} \quad (k, l \in \mathbb{N}).
\]

I.e., \( \{\Pi_k\}_{k \in \mathbb{N}} \) are the monic orthogonal polynomials for \( \mu \). Therefore \( \mu \) must be the free Poisson law, and in particular \([0, 4] = \text{supp}(\mu) \subset \sigma\).
We now show that \([4, \dim B] \subset \sigma\). Write \(B\) as a direct sum of matrix algebras, say
\[
B = C(X_0) \bigoplus \bigoplus_{1 \leq i \leq k, n_i \geq 2} M_{n_i}(\mathbb{C}),
\]
where \(C(X_0)\) denotes the maximal commutative ideal of \(B\) (which may be zero). In what follows, we will make extensive use of the fact that \(\mathcal{G}\) contains \(\mathbb{Z}_2^{+\lceil \log_2(\dim B) \rceil} \times \Pi_{i=1}^k SO(n_i)\) as a quantum subgroup. Let \(n_0 = \lceil \dim B \rceil\) so that \(\dim B = n_0 + \sum_{i=1}^k n_i^2\). We will assume for the remainder of the proof that both of the components \(C(X_0)\) and \(\bigoplus_{1 \leq i \leq k, n_i \geq 2} M_{n_i}(\mathbb{C})\) are non-zero. (The trivial modification of our argument for the remaining two cases is left to the reader.)

Let \(\mathcal{G}_0 = S_{n_0}^+\) and let \(\mathcal{G}_i = SO(n_i)\) for \(1 \leq i \leq k\). For each \(i\), let \(\beta_i : B_i \to B_i \otimes C_u(\mathcal{G}_i)\) be the canonical action of \(\mathcal{G}_i\) on \((B_i, \psi_i)\), where \(B_i\) is the \(i\)th direct summand of \(B\) and \(\psi_i = \psi|_{B_i}\). (In other words, \(\beta_0\) is the fundamental action of \(S_{n_0}^+\) on \((X_0, \chi)\) and \(\beta_i\) is the standard action of \(SO(n_i)\) on \(M_{n_i}(\mathbb{C})\) by inner automorphisms.) Now let \(\mathcal{G}' = \bigotimes_{i=0}^k \mathcal{G}_i\) be the tensor product quantum group (see [35], Chapter 6) and consider the diagonal action
\[
\beta : B \to B \otimes C_u(\mathcal{G}'); \quad \beta = \bigoplus_{0 \leq i \leq k} (\text{id}_B \otimes \iota_i)\beta_i,
\]
of \(\mathcal{G}'\) on \((B, \psi)\), where \(\iota_i : C_u(\mathcal{G}_i) \hookrightarrow C_u(\mathcal{G}')\) is the canonical embedding.

Invoking Definition 3.2, we see that there exists a unital *-homomorphism \(\pi : C_u(\mathcal{G}) \to C_u(\mathcal{G}')\) such that \(\beta = (\text{id}_B \otimes \pi)\alpha\), where \(\alpha : B \to B \otimes C_u(\mathcal{G})\) is the universal action of \(\mathcal{G}\) on \((B, \psi)\). Moreover, since \(\beta\) is faithful in the sense of [40], it follows that \(\pi\) is surjective. In particular, if \(U'\) denotes the representation of \(\mathcal{G}'\) associated to the action \(\beta\) and \(\chi'\) is its character, then \(\chi' = \pi(\chi)\) and \(\pi(\chi') \subseteq \sigma\). It therefore suffices to show that \([4, \dim B] \subseteq \sigma(\chi')\).

Let \(\chi(0) \in C_u(\mathcal{G}_0)\) denote the fundamental character of \(\mathcal{G}_0\), and let \(\chi(i) \in C(SO(n_i))\) denote the character of the standard representation of \(SO(n_i)\) on \(\mathbb{C}^{n_i}\). A simple calculation using the definition of \(\mathcal{G}'\) and the action \(\beta\) shows that
\[
\chi' = \chi(0) \otimes 1 \otimes \ldots \otimes 1 + \sum_{i=1}^k \underbrace{1 \otimes \ldots \otimes 1}_{k \text{ times}} \otimes \chi(i) \otimes \underbrace{1 \otimes \ldots \otimes 1}_{i \text{ times}},
\]
and therefore (by tensor independence) \(\sigma(\chi') = \sigma(\chi(0)) + \sum_{i=1}^k \sigma(\chi^2(i))\). Moreover, an elementary exercise in linear algebra shows that for \(i \geq 1\), \(\|\chi(i)\| = n_i\) and \([0, n_i] \subset \sigma(\chi(i))\). Therefore
\[
\sigma(\chi') = \sigma(\chi(0)) + \sum_{i=1}^k [0, n_i^2] = \sigma(\chi(0)) + [0, \dim B - n_0]. \tag{4.2}
\]
To finish the proof, we analyze the spectrum \(\sigma(\chi(0))\) of \(\chi(0) \in C_u(\mathcal{G}_0)\). If \(1 \leq n_0 \leq 4\), then \(4 \geq n_0 = \epsilon(\chi(0)) \in \sigma(\chi(0))\), and (4.2) yields \([4, \dim B] \subset \sigma(\chi')\).
here is a unique state $\phi$. We now show that this net of normal unital completely positive maps is $L^*$ compact and converges pointwise in $L^2$-norm to the identity map. Using the decomposition $L^2(G) = \bigoplus_{k \in \mathbb{N}} L^2_k(G)$ given by (2.3), each $L^2$-extension $\Phi_t \in \mathcal{B}(L^2(G))$ is given by $\Phi_t = \sum_{k \in \mathbb{N}} \frac{\Pi_k(t)}{\Pi_k(d_k \dim B)} P_k$. Since $\lim_{t \to \dim B} \frac{\Pi_k(t)}{\Pi_k(d_k \dim B)} = 1$ for each $k$ and $\|\Phi_t\| = 1$ for each $t$, it follows that $\lim_{t \to \dim B} \|\Phi_t - \xi\|_{L^2(G)} = 0$ for each $\xi \in L^2(G)$. To see that each $\Phi_t$ is compact, observe that since $\Phi_t$ is the

\[ \{0, \dim B\} \subset \sigma(\chi'). \]  If $n_0 \geq 5$, write $n_0 = 2s + r$ where $s \geq 2$ and $r \in \{0, 1\}$. Since $C_u(G_0) = C_u(S_{n_0}^+)$ is the universal $C^*$-algebra generated by an $n_0 \times n_0$ magic unitary matrix $[v_{ij}]$, there exists a unique surjective $*$-homomorphism $\pi_0 : C_u(G_0) \to \ast_{t=1}^n C^*(\mathbb{Z}_2)$ which sends $[v_{ij}]$ to the block-diagonal matrix $W \in M_{n_0} \ast_{t=1}^n C^*(\mathbb{Z}_2)$ given by

$$ W = \begin{cases} \bigoplus_{j=1}^s \left( \begin{array}{cc} p_j & 1 - p_j \\ 1 - p_j & p_j \end{array} \right) & \text{if } r = 0, \\ 1 \oplus \bigoplus_{j=1}^s \left( \begin{array}{cc} p_j & 1 - p_j \\ 1 - p_j & p_j \end{array} \right) & \text{if } r = 1. \end{cases} $$

Here $p_j$ is the projection $\frac{1}{2}(1 - g_j)$, where $g_j$ is the canonical unitary generator of the $j$th copy of $C^*(\mathbb{Z}_2)$ in the free product $\ast_{t=1}^n C^*(\mathbb{Z}_2)$. We then have $\pi_0(\chi(0)) = r1 + \sum_{j=1}^s 2p_j$. Since the spectrum of a sum of $s$ free projections is well known to equal $[0, s]$ (see [18], Example 4), we obtain $\sigma(\pi_0(\chi(0))) = [r, 2s + r] \subset \sigma(\chi)$. Using this last inclusion in (4.2), we conclude that $\sigma(\chi') \supset [r, \dim B] \supset [4, \dim B]$.

We are now ready to prove Theorem 4.2.

**Proof of Theorem 4.2.** If $1 \leq \dim B \leq 4$ then $G$ is co-amenable by [3], Corollary 4.2, and it follows from [11], Theorem 1.1, that $L^\infty(G)$ is injective. In particular, $L^\infty(G)$ has the Haagerup property.

Assume now that $\dim B \geq 5$ and fix $4 < t_0 < 5$. From the isomorphism $C^*(\chi_k : k \in \mathbb{N}) \cong C(\sigma)$ given by Proposition 4.7 together with the fact that $[0, \dim B] \subset \sigma$ from Proposition 4.8, it follows that for each $t \in [t_0, \dim B)$, there is a unique state $\varphi_t \in C^*(\chi_k : k \in \mathbb{N})^*$ given by

$$ \varphi_t(\chi_k) = \Pi_k(t) \quad (k \in \mathbb{N}). $$

Consider the irreducible representation $U^k = [u^k_{ij}]$ of $G$ with label $k$. From part (4) of Theorem 3.8 and the recursion (4.1), we have $d_k = \Pi_k(\dim B)$ for all $k \in \mathbb{N}$. Applying Theorem 4.5 to the family of states $\{\varphi_t\}_{t \in [t_0, \dim B)} \subset C^*(\chi_k : k \in \mathbb{N})^*$, we obtain a net of central completely positive multipliers $\{\Phi_t\}_{t \in [t_0, \dim B)} \subset CB(L^\infty(G))$ defined by

$$ \Phi_t(\lambda(u^k_{ij})) = \frac{\Pi_k(t)}{\Pi_k(d_k \dim B)} \lambda(u^k_{ij}) \quad (k \in \mathbb{N}, 1 \leq i, j \leq d_k). $$

(4.3)

Here $[t_0, \dim B) \subset \mathbb{R}$ is directed in the usual way.

We now show that this net of normal unital completely positive maps is $L^2$-compact and converges pointwise in $L^2$-norm to the identity map. Using the decomposition $L^2(G) = \bigoplus_{k \in \mathbb{N}} L^2_k(G)$ given by (2.3), each $L^2$-extension $\Phi_t \in \mathcal{B}(L^2(G))$ is given by $\Phi_t = \sum_{k \in \mathbb{N}} \frac{\Pi_k(t)}{\Pi_k(d_k \dim B)} P_k$. Since $\lim_{t \to \dim B} \frac{\Pi_k(t)}{\Pi_k(d_k \dim B)} = 1$ for each $k$ and $\|\Phi_t\| = 1$ for each $t$, it follows that $\lim_{t \to \dim B} \|\Phi_t - \xi\|_{L^2(G)} = 0$ for each $\xi \in L^2(G)$. To see that each $\Phi_t$ is compact, observe that since $\Phi_t$ is the
orthogonal direct sum of the finite rank projections \( \{ P_k \}_{k \in \mathbb{N}} \), it suffices to show that
\[
\left\{ \frac{\Pi_k(t)}{\Pi_k(\dim B)} \right\}_{k \in \mathbb{N}} \in C_0(\mathbb{N})
\]
for each \( t \in [t_0, \dim B] \). To check this last fact, note that
\[
0 < \frac{\Pi_k(t)}{\Pi_k(\dim B)} = \frac{S_{2k}(\sqrt{t})}{S_{2k}(\sqrt{\dim B})},
\]
and an easy exercise (see for example [14], Proposition 4.4) shows that there is a constant \( A(t_0) > 0 \) (depending only on \( t_0 \)) such that
\[
S_{2k}(\sqrt{t}) S_{2k}(\sqrt{\dim B}) \leq A(t_0) \left( \frac{t}{\sqrt{\dim B}} \right)^k \to 0 \quad (k \to \infty).
\]
(4.4)

4.2 The Property of Rapid Decay

Given a finitely generated discrete group \( \Gamma \) with symmetric generating set \( S = \{ \gamma_1, \ldots, \gamma_s \} \), there is a natural length function \( \ell = \ell_S : \Gamma \to \mathbb{N} \) defined by assigning to each \( \gamma \in \Gamma \) the length of its shortest representation as a reduced word in the elements of \( S \). The group \( \Gamma \) is said to have the property of rapid decay (with respect to \( \ell \)), or just property RD, if there exists a polynomial \( P \in \mathbb{R}_+[x] \) such that for each \( n \in \mathbb{N} \) and each function \( f : \{ \gamma \in \Gamma \mid \ell(\gamma) = n \} \to \mathbb{C} \) supported on words in \( \Gamma \) of length \( n \), the following inequality holds
\[
\| \lambda(f) \|_{VN(\Gamma)} \leq P(n) \| f \|_{L^1(\Gamma)}.
\]

It is a standard fact that property RD is essentially independent of the initial generating set \( S \) for \( \Gamma \). Perhaps the most famous example of a group with property RD is the non-abelian free group \( F_d \) on \( d \geq 2 \) generators, which was shown by Haagerup [24] to have property RD with \( P(x) = x + 1 \).

The notion of property RD for discrete quantum groups was defined and studied by Vergnioux in [38]. There he showed that property RD can only occur for unimodular discrete quantum groups, and he also exhibited the first (and what currently appear to be the only) truly “quantum” examples having this property. The goal in this section is to prove that if \( B \) is a finite dimensional \( C^* \)-algebra with \( \delta \)-trace \( \psi \), then the dual unimodular discrete quantum group \( \hat{\Gamma} \) of \( \Gamma = G_{\text{aut}}(B, \psi) \) has property RD. We will use this result to obtain some finite rank approximation properties for \( L^\infty(\hat{\Gamma}), C_r(\hat{\Gamma}) \) and \( L^1(\hat{\Gamma}) \) in Theorem 4.15. Property RD will also be a key tool in our analysis of simplicity and uniqueness of trace for \( C_r(\hat{\Gamma}) \) in Section 5.3.

For our purposes, property RD for \( \hat{\Gamma} \) will be most conveniently formulated in dual terms, as an equivalent property of \( L^\infty(\Gamma) \). More precisely, we will prove the following theorem.

**Theorem 4.10.** Let \( \psi \) be the \( \delta \)-trace on a finite dimensional \( C^* \)-algebra \( B \) and let \( C_u(\hat{\Gamma}) = \text{span}\{ (\omega_{\xi,\eta} \otimes id)U^n : \xi, \eta \in H_n \} \subset C_u(\hat{\Gamma}) \) be the subspace of Documenta Mathematica 18 (2013) 1349–1402
matrix elements of the $n$th irreducible representation of $G$. Then there exists a constant $D > 0$ (depending only on $\dim B$) such that

$$
\|\lambda(a)\|_{L^\infty(G)} \leq D(2n + 1)\|\Lambda_h(a)\|_{L^2(G)} \quad (n \in \mathbb{N}, a \in C_n(G)).
$$

In particular, each orthogonal projection $P_n : L^2(G) \to L^2_n(G)$ is bounded as a map from $L^2(G)$ into $C_r(G)$ with

$$
\|P_n\|_{L^2 \to L^\infty} \leq D(2n + 1).
$$

Remark 4.11. Note that the $O(2n + 1)$ growth rate of the constants in Theorem 4.10 is optimal. To see this, consider the character $\chi_n \in C_n(G)$ of the irreducible representation $U^n$. Using the notation and arguments from the proof of Proposition 4.8, it follows that $\|\lambda(\chi_n)\| = \sup_{t \in [0,4]}|\Pi_n(t)| = S_{2n}(\sqrt{7}) = S_{2n}(2) = 2n + 1$. Noting that $\|\Lambda_k(\chi_n)\|_{L^2(G)} = 1$, we see that the $O(2n + 1)$ bound is actually obtained by the sequence $\{\lambda(\chi_n)\}_{n \in \mathbb{N}}$.

Our proof of Theorem 4.10 is based on the ideas in Section 4 of [38], where property RD is proved for the duals of free orthogonal and free unitary quantum groups. The main difference between the present situation and the one in [38] is that for quantum automorphism groups, the fusion rules $U^k \boxtimes U^k \cong \bigoplus_{r=0}^{2\min\{n,k\}} U^{k+n-r}$ imply that each tensor product $U^n \boxtimes U^k$ contains both even and odd length subrepresentations, whereas in the orthogonal case all subrepresentations of $U^n \boxtimes U^k$ have the same parity. This has the effect of complicating some norm estimates.

To begin the proof of Theorem 4.10, note that it suffices to show that there is a constant $D > 0$ (depending only on $\dim B$) such that for each $n \in \mathbb{N}$ and each $a \in C_n(G)$,

$$
\|P_k\lambda(a)P_l\|_{B(L^2(G))} \leq D\|\Lambda_h(a)\|_{L^2(G)} \quad (k, l \in \mathbb{N}). \quad (4.5)
$$

To prove this, we follow along the lines of Haagerup [24], where an analogous reduction was obtained for free groups. Assume that (4.5) holds and take $n \in \mathbb{N}$, $a \in C_n(G)$ and $\xi \in L^2(G)$. Using the fusion rules from Theorem 3.8, it is not difficult to see that the sets $A_{l,n} = \{k \in \mathbb{N} : U^k \subset U^n \boxtimes U^k\}$ and
\[ B_{k,n} = \{ l \in \mathbb{N} : U^l \subset U^n \mathbb{B}U^k \} \] both have cardinality at most \( 2n+1 \). Therefore
\[
\| \lambda(a) \xi \|^2 = \sum_{l \in \mathbb{N}} \| P_l \lambda(a) \xi \|^2 = \sum_{l \in \mathbb{N}} \left( \sum_{k \in \mathbb{N} : U^l \subset U^n \mathbb{B}U^k} P_l \lambda(a) P_k \xi \right) \]
\[
\leq \sum_{l \in \mathbb{N}} \left( \sum_{k \in \mathbb{N} : U^l \subset U^n \mathbb{B}U^k} \| P_k \lambda(a) P_l \xi \| \right) \]
\[
\leq \sum_{l \in \mathbb{N}} \left( \sum_{k \in \mathbb{N} : U^l \subset U^n \mathbb{B}U^k} D \| \Lambda_h(a) \|_2 \| P_k \xi \|_2 \right) \]
\[
\leq D^2 \| \Lambda_h(a) \|^2 \sum_{l \in \mathbb{N}} \left( \sum_{k \in \mathbb{N} : U^l \subset U^n \mathbb{B}U^k} \| P_k \xi \|_2 \right) \]
\[
\leq D^2 \| \Lambda_h(a) \|^2 (2n+1) \sum_{k \in \mathbb{N}} \| P_k \xi \|_2 \]
\[
= D^2 \| \Lambda_h(a) \|^2 (2n+1) \sum_{k \in \mathbb{N}} B_{k,n} \| P_k \xi \|_2 \leq D^2 \| \Lambda_h(a) \|^2 (2n+1)^2 \| \xi \|^2,
\]
which yields \( \| \lambda(a) \|_r (G) \leq D(2n+1) \| \Lambda_h(a) \|_{L^2(G)} \).

The remainder of this section is devoted to proving (4.5). To do this, we utilize the Fourier transform defined by Podleś and Woronowicz in Section 2 of [32]. Let \( G \) be a compact quantum group and let \( V \in M(c_0(\hat{G}) \otimes C_u(G)) \) be the unitary given by (2.5). Define
\[
F : C_{\text{alg}}(G) \rightarrow \mathcal{C}[\hat{G}]; \quad F(a) = (\text{id} \otimes ah)^* \quad (a \in C_u(G)),
\]
where \( ah \in C_u(G)^* \) is given by \( ah(b) := h(ba), (b \in C_u(G)) \). Using the descriptions of \( h \) and \( h_L \) given by (2.2) and (2.6), it is readily checked that \( F : C_{\text{alg}}(G) \rightarrow \mathcal{C}[\hat{G}] \) is a linear bijection with inverse
\[
F^{-1} : \mathcal{C}[\hat{G}] \rightarrow C_{\text{alg}}(G); \quad F^{-1} (x) = (xh_L \otimes \text{id})V \quad (x \in \mathcal{C}[\hat{G}]),
\]
where \( xh_L(y) := h_L(xy), (y \in \ell^\infty(\hat{G})) \). Indeed, a simple calculation shows that \( F(c_{ij}^a) = m_{ij}^{-1} \sum_{1 \leq r < a} (Q_{ij}^{-1}) r c_{ij}^a \), where \( \{ c_{ij}^a \} \) denotes the system of matrix units for \( B(H_a) \) associated to the matrix representation \( U^a = [u_{ij}^a] \). This calculation also shows that \( F \) extends to a unitary isomorphism \( F : L^2(G) \rightarrow \ell^2(\hat{G}) \).

Recall that one can define a convolution product \( \text{Conv} : \mathcal{C}[\hat{G}] \otimes_{\text{alg}} \mathcal{C}[\hat{G}] \rightarrow \mathcal{C}[\hat{G}] \) using the dual coproduct \( \hat{\Delta} \) as follows. Given \( x, y \in \mathcal{C}[\hat{G}] \), \( \text{Conv}(x \otimes y) \in \mathcal{C}[\hat{G}] \) is defined to be the unique element, denoted by \( x \ast y \), such that \((x \ast y)\hat{h} = (xh_L \otimes yh_L) \circ \hat{\Delta} \). With respect to the usual algebra structure on \( C_{\text{alg}}(G) \) and the convolution product on \( \mathcal{C}[G] \), the Fourier transform \( F : C_{\text{alg}}(G) \rightarrow (\mathcal{C}[\hat{G}], \text{Conv}) \) becomes an algebra isomorphism. Indeed, since \( (\hat{\Delta} \otimes \text{id})V = \mathcal{V}_{13} \mathcal{V}_{23} \), we have
\[
F^{-1}(x \ast y) = (xh_L \otimes yh_L \otimes \text{id})(\hat{\Delta} \otimes \text{id})V = (xh_L \otimes yh_L \otimes \text{id})(\mathcal{V}_{13} \mathcal{V}_{23}) \]
\[
= (xh_L \otimes \text{id})V \cdot (yh_L \otimes \text{id})V = F^{-1}xF^{-1}y.
\]
As a final useful observation, we note that $\mathcal{F}\mathcal{P}_l\mathcal{F}^{-1} = \hat{P}_l$, where $\{\hat{P}_l\}_{l \in \mathbb{N}}$ are the canonical minimal central projections in $\ell^\infty(\hat{G}) = \prod_{l \in \mathbb{N}} B(l^2(\hat{G}))$.

In view of the above properties of the Fourier transform, we can transfer our required condition (4.5) to an equivalent one on the dual quantum group $\hat{G}$. Namely, we must find a $D > 0$ (depending only on $\dim B$) such that

$$\|\hat{P}_l(x \ast y)\|_{\mathcal{F}(\hat{G})} \leq D\|x\|_{\mathcal{F}(\hat{G})}\|y\|_{\mathcal{F}(\hat{G})} \quad (x \in \hat{P}_n C[\hat{G}], \ y \in \hat{P}_m C[\hat{G}], \ l \in \mathbb{N}).$$

(4.8)

The following useful result provides an alternate expression for the left-hand side of (4.8) and can be found in [38], Lemma 4.6. For the convenience of the reader we include a proof.

**Lemma 4.12.** Let $G = G_{aut}(B, \psi)$, where $\psi$ is a $\delta$-form on $B$, and let $w \in \hat{P}_n C[\hat{G}] \otimes \hat{P}_k C[\hat{G}]$. Then for any $l \in \mathbb{N}$ such that $U^l \subset U^n \boxtimes U^k$, we have

$$\|\hat{P}_l \text{Conv}(w)\|_{\mathcal{F}(\hat{G})} = (m_n m_k m_l^{-1})^{1/2} \|\hat{\Delta}(\hat{P}_l) w \hat{\Delta}(\hat{P}_l)\|_{\mathcal{F}(\hat{G}) \otimes \mathcal{F}(\hat{G})}.$$  

**Proof.** Since the inclusion $U^l \subset U^n \boxtimes U^k$ is multiplicity-free by Theorem 3.8, there exists a unique $z \in \hat{P}_n C[\hat{G}]$ such that $\Delta(\hat{P}_l) w \Delta(\hat{P}_l) = \Delta(z)(\hat{P}_n \otimes \hat{P}_k)$. (I.e., $z = C^{-1}_{(n,k,l)}(\rho_l^{G\boxtimes k})^* \rho_l^{G\boxtimes k}$ using the notation of (3.2)–(3.3)). Then from the definition of convolution, we have $\hat{P}_l \text{Conv}(w) = \text{Conv}(\hat{\Delta}(\hat{P}_l) w \hat{\Delta}(\hat{P}_l)) = \text{Conv}(\Delta(z)(\hat{P}_n \otimes \hat{P}_k)) = z(\hat{P}_n \ast \hat{P}_k) = \lambda z$ for some $\lambda \in \mathbb{C}$. The last equality follows because $\hat{P}_n \ast \hat{P}_k$ is evidently central. This yields

$$\|\hat{P}_l \text{Conv}(w)\|_{\mathcal{F}(\hat{G})}^2 = \hat{h}_L((\hat{P}_l \text{Conv}(w))^* \hat{P}_l \text{Conv}(w)) = \lambda \hat{h}_L(z^* \text{Conv}(w))$$

$$= \lambda \hat{h}_L(\Delta(z)^* w) = \lambda \hat{h}_L(\hat{P}_n \otimes \hat{P}_k)(\Delta(z)^* w \Delta(\hat{P}_l))$$

$$= \lambda \hat{h}_L(\Delta(\hat{P}_l) w \Delta(\hat{P}_l))^* w \Delta(\hat{P}_l)) = \lambda \|\Delta(\hat{P}_l) w \Delta(\hat{P}_l)\|_{\mathcal{F}(\hat{G}) \ast \mathbb{C}}.$$  

Finally, the appropriate value of $\lambda$ is obtained from the equation

$$\lambda m_l^2 = \lambda h_L(\hat{P}_l) = h_L(\hat{P}_l(\hat{P}_n \ast \hat{P}_k)) = m_n m_k (\text{Tr}_n \otimes \text{Tr}_k)((Q_n \hat{P}_n \otimes Q_k \hat{P}_k) \Delta(\hat{P}_l))$$

$$= m_n m_k (\text{Tr}_n \otimes \text{Tr}_k)\Delta(Q_l \hat{P}_l)) = m_n m_k \text{Tr}_l(Q_l) = m_n m_k m_l.$$  

We now restrict to the unimodular case, so that $Q_k = \text{id}_{H_k}$ and $m_k = d_k = \dim U^k$ for all $k$. In this case, we have, $\|\hat{P}_l(x \ast y)\|_{\mathcal{F}(\hat{G})} = d_n^{1/2}\|x\|_{HS(H_n)}$ and $\|\Delta(\hat{P}_l)(x \otimes y)\Delta(\hat{P}_l)\|_{\mathcal{F}(\hat{G}) \otimes \mathcal{F}(\hat{G})} = (d_n d_k)^{1/2}\|\Delta(\hat{P}_l)(x \otimes y)\Delta(\hat{P}_l)\|_{HS(H_n \otimes H_n)}$ for each $x \in \hat{P}_n C[\hat{G}]$, $y \in \hat{P}_k C[\hat{G}]$ (where $\| \cdot \|_{HS(\cdot)}$ denotes the usual Hilbert-Schmidt norm). Combining this fact with Lemma 4.12, we can recast our required inequality (4.8) in terms of Hilbert-Schmidt norms. Thus to prove Theorem 4.10, it suffices to establish the following proposition.
Suppose first that $r$ equals 2.
We now consider two possible cases: either $r$ is not a subrepresentation of $U^1 \boxtimes U^k$, then the lower bound can be obtained by analyzing the expansion of $\hat{C}(\hat{F}_1)$ relative to the orthonormal bases $\rho_{i,j}^{(k)}$, $\rho_{i,j}^{(l)}$, $\rho_{i,j}^{(s)}$, and $\rho_{i,j}^{(\infty)}$. To do this, we first need the following lemma, which gives a uniform estimate on the size of $C_{(n,k,l)}$.

**Lemma 4.14.** With the notation above, there exists a constant $D_0 > 0$ (independent of $n, k, l$ and only depending on $\dim B$) such that $D_0 \leq C_{(n,k,l)} \leq 1$.

**Proof.** Since $\rho_{i,j}^{(k)}$ is a contraction, the upper bound $C_{(n,k,l)} \leq 1$ is immediate. The lower bound can be obtained by analyzing the expansion of $C_{(n,k,l)}$ in terms of the $q$-numbers. This, however, has already been done by Vaes and Vergnioux in [36], Lemma A.6. Note that in [36], the authors work with a concrete representation of the Temperley-Lieb algebras which is generally different from the one considered here. However, since all representations under consideration are faithful, their estimates transfer directly to our context.

**Proof of Proposition 4.13.** With $U^1 \subset U^n \boxtimes U^k$ fixed, let $0 \leq r \leq 2 \min\{n, k\}$ be such that $l = k + n - r$. From Lemma 4.14 and the discussion preceding it, it follows that

$$\|\hat{C}(\hat{F}_1)(x \otimes y)\|_{\bar{H}(H_n \otimes H_k)} = D_{(n,k,l)}^{-1}\|\rho_{i,j}^{(k)} \otimes (x \otimes y)\|_{\bar{H}(H_n \otimes H_k)} \leq D_{(n,k,l)}^{-1}\|\rho_{i,j}^{(k)} \otimes (x \otimes y)\|_{\bar{H}(H_n \otimes H_k)}.$$

We now consider two possible cases: either $r$ is odd or even. Suppose first that $r = 2s + 1$ is odd. Identify $H_t$ with the highest weight subspaces $\hat{\Delta}(\hat{F}_1)/(H_{n-s-1} \otimes H_{k-s})$, and $H_n, H_k$ with the highest weight subspaces of $H_{n-s-1} \otimes H_{1}^{\otimes(s+1)}$ and $H_{1}^{\otimes s} \otimes H_{k-s}$, respectively. Now recall Section 3.2 where we fixed an orthonormal basis $\{e_i\}_{i=1}^{d_1}$ for $H_1$, a unitary $F_1$ such that $(F_1 \otimes 1)T(F_1 \otimes 1) = U^1$ and an isometric morphism $t_2 = d_1^{-1/2} \sum e_i \otimes F_1 e_i \in \text{Mor}(1, U^1 \boxtimes U^1)$. Let $e_i$ and $f_i$ denote the standard matrix units for $B(H_1)$ relative to the orthonormal bases $\{e_i\}$ and $\{f_i = F_1 e_i\}$, respectively. Given functions $i, j : [s] \to [d_1]$, we will also write $e_i$ and $f_i$ for the tensors $e_i(1) \otimes \ldots \otimes e_i(s)$ and $f_i(1) \otimes \ldots \otimes f_i(s) \in H_{1}^{\otimes s}$, respectively, and write $e_{ij}$ and $f_{ij}$
for the matrix units $e_{i(1)}e_{j(1)}\otimes \ldots \otimes e_{i(s)}e_{j(s)}$ and $f_{i(1)}f_{j(1)}\otimes \ldots \otimes f_{i(s)}f_{j(s)} \in \mathcal{B}(H_1^\otimes s)$, respectively. With these identifications, we can uniquely express

\[ x = \sum_{1 \leq i_0, j_0 \leq d_1} x_{i_0, i, j_0} \otimes e_{i_0 j_0} \otimes e_{i j}, \quad y = \sum_{u, v : [s] \rightarrow [d_1]} f_{u v} \otimes y_{u v}, \]

where $\{x_{i_0, i, j_0}\}_{i_0, i, j_0} \subseteq \mathcal{B}(H_{n-s-1})$ and $\{y_{u, v}\}_{u, v} \subseteq \mathcal{B}(H_{k-s})$. Note that $\|x\|_{HS(H_{n-s-1})}^2 = \sum_{i_0, i, j_0} \|x_{i_0, i, j_0}\|_{HS(H_{n-s-1})}$ and $\|y\|_{HS(H_{n-s-1})}^2 = \sum_{u, v} \|y_{u, v}\|_{HS(H_{n-s-1})}^2$. By repeatedly using the expression (3.7) for $t_{2s}$ in the recursion formula (3.5), it follows that $[2s + 1]_q^{1/2} t_{2s} = (p_{2s} \otimes p_{2s}) \sum_{i : [s] \rightarrow [d_1]} e_i \otimes f_i$; where for any multi-index $\hat{i} = (i(1), \ldots, i(s - 1), i(s))$, $\hat{i} := (i(s), i(s - 1), \ldots, i(1))$. Plugging this expression for $t_{2s}$ into (3.6), we obtain

\[ 1_1 \otimes t_{2s+1} \otimes 1_1 = [2s + 2]_q^{-1/2} [2s + 2]_q^{-1/2} (1_1 \otimes p_{2s+1} \otimes p_{2s+1} \otimes 1_1) (1_2 \otimes \sum_{i : [s] \rightarrow [d_1]} e_i \otimes f_i \otimes 1_2) m^*, \]

which then yields

\[
(p_1^{\otimes k})^* (x \otimes y) p_1^{\otimes k} = p_{2l} (1_{2n-2s-1} \otimes t_{2s+1} \otimes 1_{2k-2s-1}) \\
\times \sum_{i, j, u, v : [s] \rightarrow [d_1]} (x_{i_0, i, j_0} \otimes e_{i_0 j_0} \otimes e_{i j} \otimes f_{u v} \otimes y_{u v}) \\
\times (1_{2n-2s-1} \otimes t_{2s+1} \otimes 1_{2k-2s-1}) p_{2l} \\
= [2]_q^{-1} [2s + 2]_q^{-1} p_{2l} \\
\times \sum_{i, j : [s] \rightarrow [d_1]} (x_{i_0, i, j_0} \otimes (m \otimes 1_{2k-2s-2}) (e_{i_0 j_0} \otimes y_{i, j}) (m^* \otimes 1_{2k-2s-2}) p_{2l} \\
= [2]_q [2s + 2]_q^{-1} p_{2l} \sum_{i, j : [s] \rightarrow [d_1]} (x_{i_0, i, j_0} \otimes (\rho_1^{\otimes (k-s)})^* (e_{i_0 j_0} \otimes y_{i, j}) \rho_1^{\otimes (k-s)} p_{2l}. \]

Note that in the last equality above, we have also used that $(p_{2} \otimes p_{2k-2s})(m^* \otimes 1_{2k-2s-2}) p_{2k-2s} = [2]_q^{1/2} \rho_1^{\otimes (k-s)}$, which follows from (3.2). If we now take the Hilbert-Schmidt norm of the above expression and use the triangle and Cauchy-
Schwarz inequalities, we obtain
\[\|\rho_1^{G_k} * (x \otimes y) \rho_1^{G_k}\|_{HS(H_1)}\]
\[\leq \frac{[2]^q}{[2s+2]^q}\]
\[\times \left\| \sum_{i,j: [s] \to [d_1]} x_{i, j} \otimes \left( \rho_{k-s}^{1G_k} \right)^*(e_{i, j} \otimes y_{i, j}) \rho_{s}^1 \right\|_{HS(H_{n-s-1} \otimes H_{k-s})}\]
\[\leq [2]^q [2s+2]^{-1}\left( \sum_{i,j: [s] \to [d_1]} \|x_{i, j}\|^2_{HS(H_{n-s-1})} \right)^{1/2}\]
\[\times \left( \sum_{i,j: [s] \to [d_1]} \|\left( \rho_{k-s}^{1G_k} \right)^*(e_{i, j} \otimes y_{i, j}) \rho_{s}^1 \|^2_{HS(H_{k-s})} \right)^{1/2}\]
\[\leq [2]^q [2s+2]^{-1}\|x\|_{HS(H_n)} \left( \sum_{i,j: [s] \to [d_1]} \|e_{i, j} \prod y_{i, j}\|^2_{HS(H_{h} \otimes H_{k-s})} \right)^{1/2}\]
\[\leq [2]^q [2s+2]^{-1}\|x\|_{HS(H_n)} \left( \sum_{i,j: [s] \to [d_1]} \|y_{i, j}\|^2_{HS(H_{k-s})} \right)^{1/2}\]
\[\leq [2]^q [2s+2]^{-1}\|x\|_{HS(H_n)} \|y\|_{HS(H_k)}\].

Now consider the second case where \(r = 2s\) is even. We will only sketch this situation, as it is similar to (and in fact simpler than) the previous case. We now identify \(H_1\) with the highest weight subspace \(\Delta(\hat{P})_{H_{n-s} \otimes H_{k-s}}\), and \(H_n, H_k\) with the highest weight subspaces of \(H_{n-s} \otimes H_{r}^{1G_k} \otimes H_{r}^{1G_k} \otimes H_{k-s}\), respectively. Using the notation from the previous case, write
\[x = \sum_{i,j: [s] \to [d_1]} x_{i, j} \otimes e_{i, j}, \quad y = \sum_{u,v: [s] \to [d_1]} f_{uv} \otimes y_{u, v},\]
where \(\{x_{i, j}\}_{i, j} \subset B(H_{n-s})\) and \(\{y_{u, v}\}_{u, v} \subset B(H_{k-s})\). Since \([2s+1]^q t_{2s} = (p_{2s} \otimes p_{2s}) \sum_{i: [s] \to [d_1]} e_i \otimes f_i\), it is easy to verify that
\[(\rho_1^{G_k} * (x \otimes y) \rho_1^{G_k}) = p_{2l}(1_{2n-2s} \otimes t_{2s}^* \otimes 1_{2k-2s})\]
\[\times \left( \sum_{i,j: [s] \to [d_1]} x_{i, j} \otimes e_{i, j} \otimes f_{uv} \otimes y_{u, v} \right)(1_{2n-2s} \otimes t_{2s} \otimes 1_{2k-2s})p_{2l}\]
\[= [2s+1]^{-1} p_{2l} \left( \sum_{i,j: [s] \to [d_1]} x_{i, j} \otimes y_{i, j} \right)p_{2l}.\]
Taking the Hilbert-Schmidt norm then yields
\[ \| (\rho_t^{\mathcal{G}k})^* (x \otimes y) \rho_t^{\mathcal{G}k} \|_{HS(H_\omega)} \leq [2s + 1]^{-1} q \| x \|_{HS(H_\omega)} \| y \|_{HS(H_{\mathcal{G}})}. \]

From the preceding two estimates we finally obtain the inequality
\[ \| \hat{\Delta}(\hat{P}) (x \otimes y) \hat{\Delta}(\hat{P}) \|_{HS(H_\omega \otimes H_{\mathcal{G}})} \leq [r + 1]^{-1} q \| x \|_{HS(H_\omega)} \| y \|_{HS(H_{\mathcal{G}})}, \]
where \( l = n + k - r \). To complete the proof, we therefore need to show that there exist constants \( 0 < D' \leq D'' \) (depending only on \( q = q(\dim B) \)) such that
\[ D' \leq \frac{[r + 1]^2 d l_n^{-1} d_k^{-1}}{[2n + 1] q [2k + 1] q} \leq D''. \]

But since
\[ \frac{[r + 1]^2 d l_n^{-1} d_k^{-1}}{[2n + 1] q [2k + 1] q} = \frac{1 - q^{2r+2} (1 - q^{2(2n+2k-2r)+2})}{(1 - q^2)(1 - q^{4n+2})(1 - q^{4k+2})}, \]
the existence of the constants \( D', D'' \) is clear.

\[ \square \]

### 4.3 Further Approximation Properties

We will now use the property of rapid decay and the Haagerup property for \( L^\infty(\mathcal{G}) \) to prove some additional approximation properties for \( L^\infty(\mathcal{G}), \ C_r(\mathcal{G}) \) and the convolution algebra \( L^1(\mathcal{G}) \). In this section we will assume that \( \dim B \geq 5 \) (i.e., \( \mathcal{G} \) is not co-amenable), so that \( L^\infty(\mathcal{G}) \) is non-injective, \( C_r(\mathcal{G}) \) is non-nuclear, and \( L^1(\mathcal{G}) \) does not have a bounded approximate identity (see [11], Theorem 1.1 and [12], Theorem 3.1).

Recall that a Banach space \( X \) has the **metric approximation property** if there exists a net of finite rank contractions \( \{ \Phi_t : X \to X \}_{t \in \Lambda} \) such that \( \lim_t \Phi_t = \text{id}_X \) pointwise in norm. If \( X \) is a dual Banach space, we say that \( X \) has the **weak* metric approximation property** if there is a net of weak* continuous finite rank contractions \( \{ \Phi_t : X \to X \}_{t \in \Lambda} \) such that \( \lim_t \Phi_t = \text{id}_X \) pointwise in the weak* topology. Finally, recall that the **left multiplier norm** on \( L^1(\mathcal{G}) \) is the norm
\[ \| \omega \|_{M(L^1(\mathcal{G}))} := \sup \{ \| \omega * \omega' \|_{L^1(\mathcal{G})} : \| \omega' \|_{L^1(\mathcal{G})} = 1 \} \quad (\omega \in L^1(\mathcal{G})), \]
where \( * := (\Delta_c)_* \) denotes the convolution product on \( L^1(\mathcal{G}) \). Obviously \( \| \omega \|_{M(L^1(\mathcal{G}))} = \| L_\omega \|_{B(L^1(\mathcal{G}))} \leq \| \omega \|_{L^1(\mathcal{G})} \), where \( L_\omega \) is the left-multiplication operator induced by \( \omega \in L^1(\mathcal{G}) \).

**Theorem 4.15.** Let \( \psi \) be the \( \delta \)-trace on a finite dimensional C*-algebra \( B \) and let \( \mathcal{G} = \mathcal{G}_{out}(B, \psi) \). Then the following approximation properties hold for the quantum groups \( \mathcal{G} \):

1. \( C_r(\mathcal{G}) \) has the metric approximation property.
2. $L^1(G)$ has a central approximate identity $\{\omega_n\}_{n \geq 1}$ such that $\sup_n \|\omega_n\|_{M_1(L^1(G))} \leq 1$.

3. $L^\infty(G)$ has the weak* metric approximation property.

Our proof of this theorem uses standard truncation techniques for completely positive maps using property RD. Compare with [24] and [14].

Proof of (1). We will use the notation and results from Theorem 4.2 and its proof. Let $\{\Phi_t\}_{t \in [t_0, \dim B]} \subset CB(L^\infty(G))$ be the net of normal unital completely positive maps in (4.3), whose $L^2$-extensions are given by

$$\hat{\Phi}_t = \sum_{k \in \mathbb{N}} \frac{\Pi_k(t)}{\Pi_k(\dim B)} P_k.$$ 

Recall that by Theorem 4.5, $\Phi_t$ restricts to a unital completely positive map on $C_r(G)$. Moreover, since $\lim_{t \to \dim B} \frac{\Pi_k(G)}{\Pi_k(\dim B)} = 1$ and $\|\hat{\Phi}_t\|_{B(C_r(G))} = 1$ for all $t$, it follows that $\lim_{t \to \dim B} \Phi_t = \text{id}_{C_r(G)}$ pointwise in norm.

Now, for each $n \in \mathbb{N}$ consider the finite rank truncation

$$\hat{\Phi}_{t,n} = \sum_{k \leq n} \frac{\Pi_k(t)}{\Pi_k(\dim B)} P_k \in B(L^2(G)),$$

and let $\Phi_{t,n} : L^\infty(G) \to C_r(G)$ be the associated finite rank map determined by $\hat{\Phi}_{t,n} \circ \Lambda_h = \Lambda_h \circ \Phi_{t,n}$. Using Theorem 4.10, we have $\|\Pi_k\|_{B(C_r(G))} \leq \|\Pi_k\|_{L^\infty(L^\infty)} \leq D(2k+1)$ for each $k \in \mathbb{N}$. Therefore

$$\|\Phi_t - \Phi_{t,n}\|_{B(C_r(G))} = \left\| \sum_{k \geq n+1} \frac{\Pi_k(t)}{\Pi_k(\dim B)} P_k \right\|_{B(C_r(G))} \leq \sum_{k \geq n+1} \frac{\Pi_k(t)}{\Pi_k(\dim B)} \|P_k\|_{B(C_r(G))} \leq D \sum_{k \geq n+1} \frac{\Pi_k(t)}{\Pi_k(\dim B)} (2k+1)$$

$$\leq A(t_0)D \sum_{k \geq n+1} (2k+1) \left( \frac{t}{\dim B} \right)^k \to 0 \quad (n \to \infty, \ t \in [t_0, \dim B]),$$

where in the last line we have used (4.4). Consequently $\Phi_t = \lim_{n \to \infty} \Phi_{t,n}$ in operator norm and $\lim_{n \to \infty} \|\Phi_{t,n}\| = \|\Phi_t\| = 1$. Therefore if we set $\hat{\Phi}_{t,n} = \|\Phi_{t,n}\|^{-1} \Phi_{t,n}$, we obtain a family of finite rank contractions $\{\hat{\Phi}_{t,n}\}_{t,n}$ which also satisfies $\lim_{t \to \dim B} \hat{\Phi}_{t,n} = \hat{\Phi}_t$ in norm. Since we already know that $\lim_{t \to \dim B} \Phi_t = \text{id}_{C_r(G)}$ in the point-norm topology, we conclude that the set $\{\hat{\Phi}_{t,n}\}_{t,n}$ contains $\text{id}_{C_r(G)}$ in its point-norm closure. We can now easily extract a sequence of finite rank contractions from $\{\hat{\Phi}_{t,n}\}_{t,n}$ which yields the metric approximation property. Indeed, just choose any sequence of numbers $\{t(n)\}_{n \in \mathbb{N}} \subset [t_0, \dim B)$ such that

$$\lim_{n \to \infty} t(n) = \dim B \quad \text{and} \quad \lim_{n \to \infty} \sum_{k \geq n+1} (2k+1) \left( \frac{t(n)}{\dim B} \right)^k = 0.$$
Then the sequence \( \{ \Psi_n \}_{n \in \mathbb{N}} \subset B(C_r(\mathcal{G})) \) where \( \Psi_n = \tilde{\Phi}_{t(n)} \) does the job.

**Proof of (2).** Identify \( L^1(\mathcal{G}) \) with the closure of \( C_{\text{alg}}(\mathcal{G}) \) with respect to the norm
\[
\| \omega \|_{L^1(\mathcal{G})} := h(\| \omega \|) = \sup\{ |h(\omega x)| : x \in C_{\text{alg}}(\mathcal{G}), \| \lambda x \|_{L^\infty(\mathcal{G})} = 1 \}.
\]
Consider the sequence \( \{ \omega_n \}_{n \in \mathbb{N}} \subset Z(L^1(\mathcal{G})) \), where \( \omega_n = \| \Phi_{t(n)} \|^{-1} \sum_{k \leq n} \Pi_k(t(n)) \chi_k \) and \( \{ t(n) \}_{n \in \mathbb{N}} \) is the sequence chosen in the proof of part (1). Since \( \lim_n t(n) = \dim B \), \( \lim_n \| \Phi_{t(n)} \| = 1 \) and \( \chi_k * u_{ij} = u_{ij} * \chi_k = \frac{\delta_{k,l}}{\dim B} u_{ij} \) for each irreducible matrix element \( u_{ij} \in C_{\text{alg}}(\mathcal{G}) \subset L^1(\mathcal{G}) \), it follows that
\[
L_{\omega_n} \omega = \omega_n * \omega = \omega * \omega_n \longrightarrow \omega \quad (n \to \infty, \omega \in C_{\text{alg}}(\mathcal{G}) \subset L^1(\mathcal{G})). \quad (4.9)
\]
To prove that (4.9) holds for each \( \omega \in L^1(\mathcal{G}) \), the usual density argument shows that it suffices to observe that \( \| \omega_n \|_{M_1(L^1(\mathcal{G}))} \) is uniformly bounded. This, however, is true because an easy duality calculation shows that \( (L_{\omega_n})^* = \Psi_n \in B(L^\infty(\mathcal{G})) \) and therefore \( \| \omega_n \|_{M_1(L^1(\mathcal{G}))} = \| L_{\omega_n} \| = \| \Psi_n \|_{B(L^\infty(\mathcal{G}))} = \| \Psi_n \|_{B(C_r(\mathcal{G}))} = 1 \). The result now follows.

**Proof of (3).** Since the maps \( \Psi_n = (L_{\omega_n})^* \) are \( \sigma \)-weakly continuous finite rank contractions, and \( \| L_{\omega_n} \omega - \omega \|_{L^1(\mathcal{G})} \to 0 \) for all \( \omega \in L^1(\mathcal{G}) = L^\infty(\mathcal{G})_\ast \), it follows by duality that \( \lim_n \Psi_n = \text{id}_{L^\infty(\mathcal{G})} \) pointwise \( \sigma \)-weakly.

### 4.4 A Remark on Exactness

We close this section with a few remarks on the exactness of the reduced C*-algebras \( C_r(\mathcal{G}) \). We suspect that Corollary 4.17 below may already be known by the experts, but we could not find a reference.

In [13], the notion of *monoidal equivalence* for compact quantum groups was introduced, and in [36], Vaes and Vergnioux showed that exactness of reduced C*-algebras of compact quantum groups is preserved by monoidal equivalence.

**Theorem 4.16 ([36], Theorem 6.1).** Let \( \mathcal{G} \) and \( \mathcal{G}_1 \) be monoidally equivalent compact quantum groups. Then \( C_r(\mathcal{G}) \) is exact if and only if \( C_r(\mathcal{G}_1) \) is exact.

Now let \((B, \psi)\) be a finite dimensional C*-algebras equipped with a (possibly non-tracial) \( \delta \)-form \( \psi : B \to \mathbb{C} \) and consider the quantum automorphism group \( \mathcal{G} = \mathbb{G}_{\text{aut}}(B, \psi) \). In [20], Section 9.3, it is shown that \( \mathcal{G} \) is monoidally equivalent to \( \mathcal{G}_1 = \mathbb{G}_{\text{aut}}(M_2(\mathbb{C}), \text{Tr}(Q)) \), where \( Q \in M_2(\mathbb{C}) \) is any positive invertible matrix satisfying \( \text{Tr}(Q^{-1}) = \delta^2 \). Using [34], we can assume that \( \mathcal{G}_1 \) is isomorphic to \( SO_q(3) \) if \( 0 < q \leq 1 \) is such that \( q + q^{-1} = \delta \). Since \( SO_q(3) \) is well known to be \( \sigma \)-amenable, we have \( C_r(\mathcal{G}_1) \cong C_{\sigma}(\mathcal{G}_1) \) is nuclear (therefore exact) by [11], Theorem 1.1. This yields the following corollary.

**Corollary 4.17.** Let \( \psi \) be a \( \delta \)-form on a finite dimensional C*-algebra \( B \) and let \( \mathcal{G} = \mathbb{G}_{\text{aut}}(B, \psi) \). Then \( C_r(\mathcal{G}) \) is an exact C*-algebra.
5 Algebraic Structure

We now turn to some algebraic questions concerning the operator algebras $L^\infty(\mathbb{G})$ and $C_\tau(\mathbb{G})$ associated to the trace-preserving quantum automorphism groups $\mathbb{G} = G_{\text{aut}}(B, \psi)$. We will first prove a factoriality and fullness result for $L^\infty(\mathbb{G})$, followed by a simplicity and uniqueness of trace result for $C_\tau(\mathbb{G})$. We conclude this section by studying property (AO) for $L^\infty(\mathbb{G})$, and use this to prove that $L^\infty(\mathbb{G})$ is always a solid von Neumann algebra.

5.1 Factoriality and Fullness

Let $(M, \tau)$ be a II$_1$-factor (with unique faithful normal trace $\tau$). Given a sequence $\{x_n\}_{n=0}^\infty \subset M$, we say that $\{x_n\}_{n=0}^\infty$ is asymptotically central if $\|x_ny - yx_n\|_2 \to 0$ for each $y \in M$. We say that $\{x_n\}_{n=0}^\infty$ is asymptotically trivial if $\|\tau(x_n)\|_2 \to 0$. Recall that $M$ is said to be a full factor if every bounded asymptotically central sequence in $M$ is asymptotically trivial. Concerning factoriality and fullness for $L^\infty(\mathbb{G})$, we obtain the following result.

**Theorem 5.1.** Let $\psi$ be the $\delta$-trace on a finite dimensional $C^*$-algebra $B$ and let $\mathbb{G} = G_{\text{aut}}(B, \psi)$. If $\dim B \geq 8$, then $L^\infty(\mathbb{G})$ is a full type II$_1$-factor.

**Remark 5.2.** Note that there are six cases corresponding to $\dim B = 5, 6, 7$ (i.e., $B = C(X_n)$ with $n = 5, 6, 7$ and $B = C(X_n) \oplus M_2(\mathbb{C})$ with $n = 1, 2, 3$) that are not addressed by Theorem 5.1. The question of factoriality remains open in these cases. See Remark 5.10 for more details.

Note also that when $\dim B \leq 4$ and $\mathbb{G} \neq S^+_4$, $L^\infty(\mathbb{G})$ is commutative. If $\mathbb{G} = S^+_4$, the matrix model of Banica and Collins [7] (mentioned in Section 1) shows that $L^\infty(S^+_4) \hookrightarrow M_4(\mathbb{C}) \otimes L^\infty(SU(2))$ as an infinite dimensional subalgebra. Based on this explicit embedding, it seems likely that $Z(L^\infty(S^+_4))$ is infinite dimensional, but we are currently unable to verify this.

The arguments in our proof of Theorem 5.1 are based on the ideas of [36], Section 7, where factoriality for free orthogonal quantum groups is studied. A fundamental difference between the present situation and the one for free orthogonal quantum groups is that the fundamental representation of $\mathbb{G}$ is reducible in our case. This means that this representation does not provide an optimal set of generators for $L^\infty(\mathbb{G})$ with which to check commutators (i.e., factoriality). We remedy this by working instead with the matrix elements of the irreducible representation $U^1$ of $\mathbb{G}$ labeled by $1 \in \mathbb{N}$.

Recall that in Section 3.1 we fixed an orthonormal basis $\{e_i\}_{i=1}^{d_1}$ for the Hilbert space $H_1$, with respect to which the irreducible representation $U^1 \in B(H_1) \otimes C_\tau(\mathbb{G})$ can be written as $U^1 = [u_{ij}^1]$. Observe that by Theorem 3.8, every irreducible representation of $\mathbb{G}$ is contained in some tensor power of $U^1$ and therefore $L^\infty(\mathbb{G}) = \{\lambda(u_{ij}^1) : 1 \leq i, j \leq d_1\}''$. To study factoriality and fullness,
we therefore consider the bounded linear map
\[ T : L^2(G) \to H_1 \otimes L^2(G) \otimes H_1 \]
\[ T(yz_0) = \sum_{i,j=1}^{d_1} F_1 e_j \otimes \left( \lambda(u_{ij}) yz_0 - y \lambda(u_{ij})z_0 \right) \otimes e_i \quad (y \in L^\infty(G)), \]
where \( z_0 = \Lambda_0(1) \in L^2(G) \) is the cyclic and separating vector for \( L^\infty(G) \) and \( F_1 \in \mathcal{U}(d_1) \) is the unitary matrix from (3.7).

**Remark 5.3.** The connection between the operator \( T \) and the factoriality or fullness of \( L^\infty(G) \) is obvious: \( L^\infty(G) \) is a factor if and only if \( \text{ker} \, T = Cz_0 \).

If \( T |_{Cz_0} \) is bounded below, then \( L^\infty(G) \) is moreover a full factor. Indeed, if \( C > 0 \) is such that \( \| T \xi \|_{H_1 \otimes L^2(G) \otimes H_1} \geq C \| \xi \|_{L^2(G)} \) for all \( \xi \in Cz_0 \) and \( \{ x_n \}_{n=1}^\infty \subset L^\infty(G) \) is an asymptotically central sequence, then
\[ \| x_n - h(x_n)1 \|_{L^2(G)} \leq C^{-1} \| T(x_n - h(x_n)1) \|_{H_1 \otimes L^2(G) \otimes H_1} \to 0 \quad (n \to \infty). \]

**5.2 Analysis of the operator \( T \)**

The commutator operator \( T \) compares the left and right action of the operators \( \{ \lambda(u_{ij}) \}_{1 \leq i,j \leq d_1} \) on \( L^2(G) \). At the level of representation theory, this can be interpreted in terms of comparing certain decompositions of tensor products of irreducible representations. This leads us to consider, for each \( k \geq 1 \) and \( \alpha \in \{0, \pm 1\} \), the isometric morphisms \( \phi^{(\alpha)}_{k,L} : H_k \to H_1 \otimes H_{k+\alpha} \) and \( \phi^{(\alpha)}_{k,R} : H_k \to H_{k+\alpha} \otimes H_1 \) defined using formulas (3.2)–(3.3) as follows:

\[
\phi^{(1)}_{k,L} = C^{-1/2}_{(1,1,1,1)k} \bigotimes_{1}^{2k} \bigotimes_{1}^{2k} = \left( \frac{[2k]!q}{[2k+2]!q} \right)^{1/2} (p_2 \otimes p_{2k+2})(t_2 \otimes 1_{2k})p_{2k}, \\
\phi^{(1)}_{k,R} = C^{-1/2}_{(1,1,1,1)k} \bigotimes_{1}^{2k} \bigotimes_{1}^{2k} = \left( \frac{[2k]!q}{[2k+2]!q} \right)^{1/2} (p_{2k+2} \otimes p_2)(1_{2k} \otimes t_2)p_{2k}, \\
\phi^{(0)}_{k,L} = C^{-1/2}_{(1,1,1,1)k} \bigotimes_{1}^{2k} \bigotimes_{1}^{2k} = \left( \frac{[2k]!q}{[2k+2]!q} \right)^{1/2} (p_2 \otimes p_{2k})(m^* \otimes 1_{2k-2})p_{2k}, \\
\phi^{(0)}_{k,R} = C^{-1/2}_{(1,1,1,1)k} \bigotimes_{1}^{2k} \bigotimes_{1}^{2k} = \left( \frac{[2k]!q}{[2k+2]!q} \right)^{1/2} (p_{2k} \otimes p_2)(1_{2k-2} \otimes m^*)p_{2k}, \\
\phi^{(1)}_{k,L} = C^{-1/2}_{(1,1,1,1)k} \bigotimes_{1}^{2k} \bigotimes_{1}^{2k} = (p_2 \otimes p_{2k-2})p_{2k}, \\
\phi^{(1)}_{k,R} = C^{-1/2}_{(1,1,1,1)k} \bigotimes_{1}^{2k} \bigotimes_{1}^{2k} = (p_{2k-2} \otimes p_2)p_{2k}.
\]

The following theorem shows that, relative to the decomposition \( L^2(G) \cong \bigoplus_{k \in \mathbb{N}} H_k \otimes H_k \) given by (2.3)–(2.4), the commutator operator \( T \) is block-tridiagonal and built quite naturally from the above list of isometric morphisms.

**Theorem 5.4.** There is a decomposition \( T = T^{(+1)} + T^{(0)} + T^{(-1)} \), where for each \( \alpha \in \{0, \pm 1\} \),
\[
T^{(\alpha)}z_0 = 0, \quad T^{(\alpha)}(H_k \otimes H_k) \subseteq H_1 \otimes H_{k+\alpha} \otimes H_{k+\alpha} \otimes H_1 \quad (k \geq 1),
\]
and
\[
T^{(+1)}|_{H_k \otimes H_k} = \left( \frac{2k+3}{2k+1} \right)^{1/2} \left( \phi_{k,L}^{(+1)} \otimes \phi_{k,R}^{(+1)} - \sigma \phi_{k,R}^{(+1)} \otimes \sigma^* \phi_{k,L}^{(+1)} \right),
\]
\[
T^{(0)}|_{H_k \otimes H_k} = \phi_{k,L}^{(0)} \otimes \phi_{k,R}^{(0)} - \sigma \phi_{k,R}^{(0)} \otimes \sigma^* \phi_{k,L}^{(0)},
\]
\[
T^{(-1)}|_{H_k \otimes H_k} = \left( \frac{2k-1}{2k+1} \right)^{1/2} \left( \phi_{k,L}^{(-1)} \otimes \phi_{k,R}^{(-1)} - \sigma \phi_{k,R}^{(-1)} \otimes \sigma^* \phi_{k,L}^{(-1)} \right),
\]
where \( \sigma : H_k \otimes H_1 \to H_1 \otimes H_k \) is the tensor flip map.

The proof of Theorem 5.4 is rather long and tedious, so we delay it to the Appendix (Section 6).

Before commencing the proof of Theorem 5.1, we need one more remark.

**Remark 5.5.** At various places during the proofs of Theorems 5.1 and 5.4, it will be useful to have slight modifications of the recursions (3.4)–(3.7) for the isometries \( t_{2k} \in \text{Mor}(1, U^k \boxtimes U^k) \). Namely, we claim that \( t_2 = |q|^{1/2} (1_2 \otimes p_2) m^* \nu = |q|^{1/2} (p_2 \otimes l_2) m^* \nu \) and more generally,
\[
t_{2k} = \left( \frac{2k-1}{2k+1} \right)^{1/2} (1_{2k} \otimes p_{2k})(1_{2k-2} \otimes t_{2k-2})t_{2k-2} \quad (5.1)
\]
\[
= \left( \frac{2k-1}{2k+1} \right)^{1/2} (p_{2k} \otimes 1_{2k})(1_{2k-2} \otimes t_{2k-2})t_{2k-2}. \quad (5.2)
\]

While the above equalities are perhaps “diagrammatically” obvious, we can nonetheless algebraically verify these claims as follows. If \( k \geq 2 \), let \( \tilde{t}_{2k} \in \text{Mor}(1, U^{(k-1)} \boxtimes U^1 \boxtimes U^k) \) and \( \tilde{t}_{2k} \in \text{Mor}(1, U^k \boxtimes U^1 \boxtimes U^{(k-1)}) \) be the morphisms appearing on the right sides of (5.1) and (5.2), respectively. Since \( \text{Mor}(1, U^{(k-1)} \boxtimes U^1 \boxtimes U^k) \) is one-dimensional and \( t_{2k} \in \text{Mor}(1, U^k \boxtimes U^1 \boxtimes U^{(k-1)}) \subseteq \text{Mor}(1, U^{(k-1)} \boxtimes U^1 \boxtimes U^k) \), we must have \( \tilde{t}_{2k} = zt_{2k} \) for some \( z \in \mathbb{C} \). Moreover, since \( \mathfrak{C} = t_{2k} \tilde{t}_{2k} = t_{2k}^* \tilde{t}_{2k} = 1 \), we must have \( z = 1 \). A similar argument gives \( \tilde{t}_{2k} = t_{2k} \). The case of \( k = 1 \) goes similarly.

We will now use the tridiagonal decomposition of \( T \) given by Theorem 5.4 to show that if the dimension of \( B \) is sufficiently large, then \( T \) is bounded below on \( \mathbb{C} \xi_0^\perp \subset L^2(\mathbb{G}) \). By Remark 5.3, this will prove that \( L^\infty(\mathbb{G}) \) is a full factor. To obtain this bound, we actually prove that \( T^{(+1)}|_{\mathbb{C} \xi_0^\perp} \) is bounded below by some constant which tends to \( \infty \) as \( \text{dim} \ B \to \infty \), while \( (T^{(0)} + T^{(-1)})|_{\mathbb{C} \xi_0^\perp} \) is bounded above by a constant which is uniformly bounded as a function of \( \text{dim} \ B \). An application of the triangle inequality will then show that \( T|_{\mathbb{C} \xi_0^\perp} \) is bounded below for \( \text{dim} \ B \) sufficiently large.

**Notation 5.6.** Given \( \xi \in L^2(\mathbb{G}) \), denote by \( \xi_k \) the component of \( \xi \) belonging to the subspace \( L^2_k(\mathbb{G}) \cong H_k \otimes H_k \), and let \( \xi^0 = \xi - \langle \xi_0, \xi \rangle \xi_0 \) be the orthogonal projection of \( \xi \) onto \( \mathbb{C} \xi_0^\perp \). We then have \( T \xi = T \xi^0 = T^{(+1)} \xi^0 + (T^{(0)} + T^{(-1)}) \xi^0 \). Similarly, given \( \eta \in H_1 \otimes L^2(\mathbb{G}) \otimes H_1 \), let \( \eta_k \) be the component of \( \eta \) belonging to the subspace \( H_1 \otimes H_k \otimes H_k \).
We start by finding a uniform upper bound for the norm of \( (T^{(0)} + T^{(-1)})|_{C_2} \).

**Lemma 5.7.**
\[
\| (T^{(0)} + T^{(-1)})|_{C_2} \| \leq 2(1 + q).
\]

**Proof.** Theorem 5.4 implies that for any \( \xi \in L^2(G) \),
\[
\|T^{(0)}\xi^0\|^2 = \sum_{k=1}^{\infty} \|T^{(0)}\xi_k\|^2 = \sum_{k=1}^{\infty} \|T^{(0)}H_k \xi_k\|^2 \|\xi_k\|^2 \\
\leq \sum_{k=1}^{\infty} 4\|\xi_k\|^2 = 4\|\xi^0\|^2,
\]
\[
\|T^{(-1)}\xi^0\|^2 = \sum_{k=1}^{\infty} \|T^{(-1)}\xi_k\|^2 = \sum_{k=1}^{\infty} \|T^{(-1)}H_k \xi_k\|^2 \|\xi_k\|^2 \\
\leq \sum_{k=1}^{\infty} 4\|\xi_k\|^2 \leq 4q^2\|\xi^0\|^2,
\]
where in the last line we have used the fact that

\[
k \mapsto \frac{[2k - 1]_q}{[2k + 1]_q} = q^2 \left( \frac{1 - q^{4k-2}}{1 - q^{4k+2}} \right)
\]
is increasing and \( \sup_k \frac{[2k - 1]_q}{[2k + 1]_q} = q^2 \).

The result now follows from the triangle inequality. \(\square\)

Next we deal with more difficult task of bounding \( T^{(+1)}|_{C_2} \) from below. Fix \( \xi \in L^2(G) \). Then \( \|T^{(+1)}\xi^0\|^2 = \sum_{k=1}^{\infty} \|T^{(+1)}\xi_k\|^2 \) and for each \( k \geq 1 \), Theorem 5.4 gives
\[
\|T^{(+1)}\xi_k\|^2 = \frac{[2k + 3]_q}{[2k + 1]_q} \left( \left\| (\phi^{(+1)}_{k,L} \otimes \phi^{(+1)}_{k,R}) - \sigma \phi^{(+1)}_{k,L} \otimes \sigma^* \phi^{(+1)}_{k,L} \right\| \xi_k \right)^2 \\
= \frac{[2k + 3]_q}{[2k + 1]_q} \left( \left\| (\phi^{(+1)}_{k,L} \otimes \phi^{(+1)}_{k,R}) + ((\phi^{(+1)}_{k,R})^* \sigma^* \phi^{(+1)}_{k,L}) \right\| \xi_k \right)^2 \\
- 2\text{Re} \left( (\sigma \phi^{(+1)}_{k,R} \otimes \sigma^* \phi^{(+1)}_{k,L})^*(\phi^{(+1)}_{k,L} \otimes \phi^{(+1)}_{k,R}) \xi_k \right) \\
\geq \frac{[2k + 3]_q}{[2k + 1]_q} \left( 2\|\xi_k\|^2 - 2\text{Re} \left( ((\phi^{(+1)}_{k,R})^* \sigma^* \phi^{(+1)}_{k,L}) \right) \right)
\geq \frac{[2k + 3]_q}{[2k + 1]_q} \left( 2\|\xi_k\|^2 - 2\|\phi^{(+1)}_{k,R}\|^2 \|\xi_k\|^2 \right).
\]

Therefore
\[
\|T^{(+1)}\xi^0\|^2 \geq 2 \sum_{k=1}^{\infty} \frac{[2k + 3]_q}{[2k + 1]_q} \left( 1 - \|\phi^{(+1)}_{k,R}\|^2 \|\xi_k\|^2 \right)
\geq 2 \min_{k \geq 1} \left\{ \frac{[2k + 3]_q}{[2k + 1]_q} \left( 1 - \|\phi^{(+1)}_{k,R}\|^2 \|\xi_k\|^2 \right) \right\}, \quad (5.3)
\]
showing that we need to find some useful upper bounds for the norms \( \| (\phi_{k,L}^{(1)})^* \sigma \phi_{k,R}^{(1)} \| \), where

\[
(\phi_{k,L}^{(1)})^* \sigma \phi_{k,R}^{(1)} = \frac{[3]_q[2k+1]_q}{[2k+3]_q} p_{2k}(t_*^2 \otimes 1_{2k})(p_2 \otimes p_{2k+2}) \sigma(1_{2k} \otimes t_2)p_{2k} \quad (k \geq 1). \tag{5.4}
\]

As a first observation, note that the naïve estimate \( \| (\phi_{k,L}^{(1)})^* \sigma \phi_{k,R}^{(1)} \| \leq [3]_q[2k+1]_q \) is of no use since the right hand side of this inequality always exceeds 1. To obtain a sharper estimate, we need to unravel the right-hand side of (5.4) using the recursive structure of the Jones-Wenzl projections \( \{ p_y \}_{y \in \mathbb{N}} \). We will do this by first finding a suitable two-step recursion formula which expresses the projection \( p_{2k+2} \) in terms of the projection \( p_{2k} \). Plugging this recursion into the right-hand side of (5.4) will yield an expression whose norm can be more sharply bounded from above. The required two-step recursion for \( p_{2k+2} \) will be obtained by iterating twice the following (one-step) recursion for the Jones-Wenzl projections \( \{ p_y \}_{y \in \mathbb{N}} \) due to Frenkel and Khovanov ([21], Equation 3.8):

\[
p_y = \left( 1_y - \sum_{r=1}^{y-1} (-1)^{y-r-1} \frac{[2]_q[y]_q}{[y]_q} (1_{r-1} \otimes t_1 \otimes 1_{y-r-1} \otimes t_1^* \otimes 1) \right) \times (p_{y-1} \otimes 1_1) \quad (y \geq 2). \tag{5.5}
\]

Setting \( y = 2k+2 \) in (5.5) and then applying (5.5) again to the resulting \( p_{2k+1} \) term yields

\[
p_{2k+2} = \left( 1_{2k+2} - \sum_{r=1}^{2k+1} (-1)^{2k+1-r} \frac{[2]_q[r]_q}{[2k+2]_q} (1_{r-1} \otimes t_1 \otimes 1_{2k+1-r} \otimes t_1^* \otimes 1) \right) \times \left( p_{2k+2} - \sum_{l=1}^{2k} (-1)^{2k-l} \frac{[2]_q[l]_q}{[2k+1]_q} (1_{l-1} \otimes t_1 \otimes 1_{2k-l} \otimes t_1^* \otimes 1_1) \right) \times (p_{2k} \otimes 1_2).
\]

Multiplying this expression by \( 1_2 \otimes p_{2k} \) on the left, observing that \( p_{2k+2} = (1_2 \otimes p_{2k})p_{2k+2} \) and \( p_{2k} \text{TL}_{y,2k}(\delta) = \text{TL}_{2k,y}(\delta)p_{2k} = 0 \) for all \( y < 2k \), we see that most terms in the above sums vanish. That is,
After collecting like terms, this gives

\[
p_{2k+2} = (1_2 \otimes p_{2k}) \left( 1_{2k+2} - \frac{[2]_q}{[2k+2]_q} (t_1 \otimes 1_{2k} \otimes t_1^*) \right.
\]

\[
+ \frac{[2]^2_q}{[2k+2]_q} (1_1 \otimes t_1 \otimes 1_{2k-1} \otimes t_1^*) \right) \times (p_{2k} \otimes 1_2)
\]

\[
= (1_2 \otimes p_{2k}) \left[ 1_{2k+2} + \frac{[2]_q}{[2k+1]_q} (t_1 \otimes 1_{2k-1} \otimes t_1^* \otimes 1_1) \right.
\]

\[
- \frac{[2]_q}{[2k+1]_q} (1_1 \otimes t_1 \otimes 1_{2k-2} \otimes t_1^* \otimes 1_1) \left. - \frac{[2]_q}{[2k+2]_q} (t_1 \otimes 1_{2k} \otimes t_1^*) \right] \times (p_{2k} \otimes 1_2).
\]

After collecting like terms, this gives

\[
p_{2k+2} = (1_2 \otimes p_{2k}) \left[ 1_{2k+2} + \left( \frac{[2]_q}{[2k+1]_q} + \frac{[2]_q [2k]_q}{[2k+1]_q [2k+2]_q} \right) (t_1 \otimes 1_{2k-1} \otimes t_1^* \otimes 1_1) \right.
\]

\[
- \frac{[2]_q}{[2k+1]_q} (1_1 \otimes t_1 \otimes 1_{2k-2} \otimes t_1^* \otimes 1_1) \left. - \frac{[2]_q}{[2k+2]_q} (t_1 \otimes 1_{2k} \otimes t_1^*) \right] \times (p_{2k} \otimes 1_2).
\]
If we now insert the above expression for $p_{2k+2}$ into equation (5.4) and use the fact that $p_2t_1 = 0$, we see that the contributions coming from (b), (d) and (e) are all zero. Thus,

$$
\begin{align*}
&= \left[ 3 \right]_q \left[ 2k + 1 \right]_q \left[ 2k + 3 \right]_q p_{2k}(t_2^* \otimes p_{2k}) \sigma (1_{2k} \otimes t_2)p_{2k} \\
&\quad - \left( \left[ 2 \right]_q^2 + \left[ 2 \right]_q^2 \left[ 2k + 2 \right]_q \right) \\
&\quad \times p_{2k}(t_2^* \otimes p_{2k}) \left( 1_2 \otimes (1_1 \otimes t_1 \otimes 1_{2k-2} \otimes t_1^* \otimes 1_1) \right) \sigma (1_{2k} \otimes t_2)p_{2k} \\
&\quad + \left[ 2 \right]_q^3 \left[ 2k + 1 \right]_q \left[ 2k + 2 \right]_q \text{Ad}(p_{2k}) \\
&\quad \times (t_2^* \otimes p_{2k}) \left( 1_2 \otimes (1_1 \otimes t_1 \otimes 1_1) \right) \left( 1_{2k-2} \otimes t_1^* \otimes 1_1 \right) \sigma (1_{2k} \otimes t_2) \\
&\quad = \left[ 2k + 1 \right]_q \left[ 2k + 3 \right]_q p_{2k} \sigma^* p_{2k} - \left( 1 + \left[ 2k \right]_q \left[ 2k + 2 \right]_q \left[ 2k + 3 \right]_q \right) p_{2k}(1_{2k-2} \otimes m) \sigma^*(m^* \otimes 1_{2k-2})p_{2k} \\
&\quad + \left[ 2 \right]_q \left[ 2k + 3 \right]_q \left[ 2k + 2 \right]_q p_{2k} \sigma^* p_{2k},
\end{align*}
$$

where, to obtain the last equality, we have substituted the following three identities for the quantities $\alpha, \beta$ and $\gamma$:

$$
[3]_q \alpha = p_{2k} \left( \sum_{i=1}^{d_1} e_i^* \otimes (F_1 e_i)^* \otimes p_{2k} \right) \left( \sum_{j=1}^{d_1} e_j \otimes p_{2k} \otimes F_1 e_j \right) \\
= p_{2k} \sum_{i=1}^{d_1} ((F_1 e_i)^* \otimes p_{2k})(p_{2k} \otimes F_1 e_i) = p_{2k} \sigma^* p_{2k},
$$
Returning to inequality (5.3) with Lemma 5.8 now at hand, we obtain the lower bound

\[ [2_q^2]^{[3_q][1]} = [3_q]^{[2_p^2]}(t_2^* \otimes p_{2k})(1_2 \otimes m^* \otimes 1_{2k-4} \otimes m)\sigma(1_{2k} \otimes t_2)p_{2k} \]

(using (3.6))

\[
= p_{2k} \left( \sum_{i=1}^{d_1} e_i^* \otimes (F_i e_i)^* \otimes p_{2k} \right) (1_2 \otimes m^* \otimes 1_{2k-4} \otimes m)
\]

\[
\times \left( \sum_{j=1}^{d_2} e_j \otimes p_{2k} \otimes F_1 e_j \right)
\]

\[
= p_{2k} \sum_{i=1}^{d_1} ((F_i e_i)^* \otimes p_{2k})(1_2 \otimes m)((m^* \otimes 1_{2k-2})p_{2k} \otimes F_i e_i)
\]

\[
= p_{2k}(1_{2k-2} \otimes m) \sum_{i=1}^{d_1} ((F_i e_i)^* \otimes 1_{2k-2+2})(m^* \otimes 1_{2k-2})p_{2k} \otimes F_i e_i
\]

\[
= p_{2k}(1_{2k-2} \otimes m)\sigma^*(m^* \otimes 1_{2k-2})p_{2k},
\]

and

\[ [3_q][2_q^2]^2 = [3_q]^2 p_{2k}(t_2^* \otimes p_{2k})(1_2 \otimes t_2 \otimes 1_{2k-2} \otimes t_2^*)\sigma(1_{2k} \otimes t_2)p_{2k} \]

(using the fact that \( t_2 = [3_q]^{-1/2}[2_q](p_2 \otimes p_2)(1_1 \otimes t_1 \otimes 1_1))t_1 \)

\[
= [3_q]^2 p_{2k}((t_2^* \otimes 1_2)(1_2 \otimes t_2) \otimes 1_{2k-2} \otimes t_2^*)\sigma(1_{2k} \otimes t_2)p_{2k}
\]

\[
= [3_q] p_{2k}(p_2 \otimes 1_{2k-2} \otimes t_2^*)\sigma(1_{2k} \otimes t_2)p_{2k}
\]

(since \((t_2^* \otimes 1_2)(1_2 \otimes t_2) = [3_q]^{-1}p_2 \)

\[
= p_{2k} \left( p_2 \otimes 1_{2k-2} \otimes \sum_{i=1}^{d_1} e_i^* \otimes (F_i e_i)^* \right) \left( \sum_{j=1}^{d_2} e_j \otimes p_{2k} \otimes F_1 e_j \right)
\]

\[
= p_{2k}(p_2 \otimes 1_{2k-2}) \sum_{i=1}^{d_1} (1_{2k} \otimes e_i^*)(e_i \otimes p_{2k}) = p_{2k}\sigma p_{2k}. \]

A simple application of the triangle inequality using the fact that \( \|m\| = \delta = [2]_q \) now yields our desired norm estimate.

**Lemma 5.8.** For each \( k \geq 1 \),

\[
\| (\phi_{k,L}^{(+1)})^* \sigma \phi_{k,L}^{(+1)} \| \leq \frac{[2k + 1]_q + [2]_q[1 + \frac{[2]_q}{[2k+2]_q}] + \frac{[2]_q}{[2k+2]_q}}{[2k + 3]_q}. \]

(5.6)

Returning to inequality (5.3) with Lemma 5.8 now at hand, we obtain the lower bound...
Proof. From the definition of \( \min \) we have

\[
\parallel T^{(1+)\xi^0} \parallel^2 \geq 2 \parallel \xi^0 \parallel^2 \min_{k \geq 1} \left\{ \frac{[2k+3]_q}{[2k+1]_q} - \left( \frac{[2k+1]_q + [2]_q^2 \left( 1 + \frac{[2]_q}{[2k+2]_q} \right) + \frac{[2]_q}{[2k+2]_q} \right)^2 \right\}
\]

By expanding the square in the previous line and using the fact that \( \min_k \frac{[2k+3]_q}{[2k+1]_q} = q^{-2} \), we obtain

\[
\parallel T^{(1+)\xi^0} \parallel^2 \geq 2 \left( q^{-2} - q^2 - \frac{[2]_q^4 (1 + q^2)^2}{[3]_q [5]_q} - \frac{[2]_q^2}{[4]_q [3]_q [5]_q} - \frac{2 [2]_q^2 (1 + q^2)}{[5]_q} \right) \parallel \xi^0 \parallel^2 =: C(q) \parallel \xi^0 \parallel^2,
\]

If we now combine the preceding inequality with Lemma 5.7, we have

\[
\parallel T \xi \parallel = \parallel T^{(1+)\xi^0} + (T^{(0)} + T^{(1-)}\xi^0) \parallel \geq \parallel T^{(1+)\xi^0} \parallel - \parallel (T^{(0)} + T^{(1-)}\xi^0) \parallel \geq \left( C(q)^{1/2} - 2 (1 + q) \right) \parallel \xi^0 \parallel \quad (\xi \in L^2(G)).
\]

We are now ready to prove Theorem 5.1. The remaining ingredient we need is the following lemma, which will also be useful when studying the simplicity of \( C_r(G) \) in Section 5.3. Recall that \( 0 < q < 1 \) was defined so that \( \delta = \sqrt{\dim B} = q + q^{-1} \). Note that \( q = q(\delta) = \frac{\sqrt{\delta^2 - 1}}{2} \) is a decreasing function of \( \delta \).

**Lemma 5.9.** Consider the function \( f(\delta) := \frac{1}{2} \left( C(q(\delta))^{1/2} - 2 (1 + q(\delta)) \right) \). Then \( f \) is an increasing function on the interval \([\sqrt{\delta}, \infty)\) and \( f(\sqrt{\delta}) > 0 \).

Proof. From the definition of \( C(q) \), it follows that \( f(\delta) = q(\delta) \) where

\[
g(q) = \sqrt{2} \left[ q^{-2} - q^2 - \frac{[2]_q^4 (1 + q^2)^2}{[3]_q [5]_q} - \frac{[2]_q^2}{[4]_q [3]_q [5]_q} - \frac{2 [2]_q^2 (1 + q^2)}{[5]_q} \right]^{1/2} - \frac{2 (1 + q)}{[3]_q^{1/2}}.
\]

When \( \delta = \sqrt{\delta}, \ q = \frac{\sqrt{\delta^2 - 1}}{2} \approx 0.4142 \) and direct substitution yields \( f(\delta) \approx 0.1111 > 0 \). It is now just an elementary calculus exercise to verify that \( g \) is a decreasing function of \( q \in (0, \frac{\sqrt{\delta^2 - 1}}{2}) \) and therefore \( f \) is an increasing function of \( \delta \in (\sqrt{\delta}, \infty) \).

**Proof of Theorem 5.1.** Since \( f(\sqrt{\delta}) > 0 \) and \( f \) is increasing by Lemma 5.9, inequality (5.7) shows that \( T \parallel \xi \parallel^2 \) is bounded below by \( \left[ \frac{1}{2} \right]_{q(\delta)} f(\delta) > 0 \) whenever \( \dim B = \delta^2 \geq 8 \). Thus \( L^\infty(G) \) is a full factor by Remark 5.3. \( \square \)
Remark 5.10. Note that if $S^2 = 5, 6$ or 7, one can check that $|3|^{1/2}_q f(\delta) \leq -0.4386 < 0$. So the preceding calculations yield no information about the factoriality or fullness of $L^\infty(\mathbb{G})$ when $5 \leq \dim B \leq 7$.

5.3 Simplicity of $C_r(G)$

We now consider the reduced C*-algebra $C_r(G)$ and show that, at least in most cases, it is simple with unique trace. To do this, we adapt the “conjugation by generators” method used in [36], Section 7.

Write $W^1 = (id \otimes \lambda)U^1$ where $U^1$ is the irreducible representation of $G$ labeled by $1 \in \mathbb{N}$ and consider the unital completely positive map $\Phi : C_r(G) \to C_r(G)$ defined by

$$\Phi(x) = \frac{1}{2|3|_q} (\text{Tr} \otimes id) \left( (W^1)^*(1 \otimes x)W^1 + W^1(1 \otimes x)(W^1)^* \right).$$

Observe that $\Phi$ is $\tau$-preserving for any tracial state $\tau$ on $C_r(G)$ and that $\Phi(\mathcal{I}) \subseteq \mathcal{I}$ for any closed two-sided ideal $\mathcal{I} \subseteq C_r(G)$. Moreover, a simple calculation shows that the $L^2$-extension of $\Phi$, denoted by $\hat{\Phi}$, is given by

$$\hat{\Phi} \Lambda_h(a) = \left( id_{L^2(G)} - \frac{1}{2|3|_q} T^*T \right) \Lambda_h(a) \quad (a \in C_u(G)), \quad (5.8)$$

where $T : L^2(G) \to H_1 \otimes L^2(G) \otimes H_1$ is the commutator map introduced in Section 5.1. The key result of this section is the following proposition.

Proposition 5.11. Let $G = \mathbb{G}_{out}(B, \psi)$, where $B$ is a finite dimensional C*-algebra with $\delta$-trace $\psi$. If $\dim B \geq 8$, then

$$\lim_{n \to \infty} \Phi^n(x) = h(x)1_{C_r(G)} \quad (x \in C_r(G)).$$

Proof. Note that it is sufficient to prove this result for $x_0$ belonging to the dense Hopf $\ast$-subalgebra $\lambda(C_{alg}(G)) \subseteq C_r(G)$. Take $a \in C_{alg}(G)$ and consider $x = \lambda(a)$. Since $a \in C_{alg}(G)$, there is an $r \geq 0$ such that $x \in \bigoplus_{k=1}^{r+2n} \lambda(C_k(G))$, where $C_k(G)$ is the subspace of $C_{alg}(G)$ spanned by the matrix elements of the irreducible representation $U^k$. Now let $n \in \mathbb{N}$ and consider $\Phi^n(x) - h(x)1 = \Phi^n(x - h(x)1)$, which evidently belongs to $\bigoplus_{k=1}^{r+2n} \lambda(C_k(G))$.

Writing $\Phi^n(x - h(x)1) = \sum_{k=1}^{r+2n} \lambda(z_k)$ with $z_k \in C_k(G)$ and using property RD (Theorem 4.10), we obtain

$$\|\Phi^n(x) - h(x)1\|_{C_r(G)} \leq \sum_{k=1}^{r+2n} \|\lambda(z_k)\|_{C_r(G)} \leq \sum_{k=1}^{r+2n} (2k + 1) \|\Lambda_h(z_k)\|_{L^2(G)}$$

$$\leq (2r + 4n + 1) \sqrt{r + 2n} \left( \sum_{k=1}^{r+2n} \|\Lambda_h(z_k)\|^2_{L^2(G)} \right)^{1/2}$$

$$= (2r + 4n + 1) \sqrt{r + 2n} \|\Phi^n(a - h(a)1)\|_{L^2(G)}$$

$$\leq (2r + 4n + 1) \sqrt{r + 2n} \|\hat{\Phi}|_{C_{alg}(G)} \|^n \|\Lambda_h(a - h(a)1)\|_{L^2(G)}.$$
Now assume that $\delta = \sqrt{\dim B} \geq \sqrt{8}$. From (5.7) and Lemma 5.9, it follows that $T^*T|_{C\xi^0} \geq 2|q| f(\delta)^2 \text{id}_{C\xi^0}$ and therefore

$$\hat{\Phi}|_{C\xi^0} = \text{id}_{C\xi^0} - \frac{1}{2|q|} T^*T|_{C\xi^0} \leq \left(1 - \frac{1}{2} f(\delta)^2\right) \text{id}_{C\xi^0} \leq \left(1 - \frac{1}{2} f(\sqrt{8})^2\right) \text{id}_{C\xi^0}.$$ 

Since $f(\sqrt{8}) \approx 0.1111$, $\|\hat{\Phi}|_{C\xi^0}\| \leq \left(1 - \frac{1}{2} f(\sqrt{8})^2\right) < 1$, which gives

$$\limsup_{n \to \infty} \|\Phi^n(x) - h(x)1\|_{C_r(\mathbb{G})} \leq \lim_{n \to \infty} (2r + 4n + 1)\sqrt{\frac{1}{2} f(\sqrt{8})^2} \|\Lambda h(a - h(a)1)\|_{L^2(\mathbb{G})} = 0.$$ 

Simplicity and uniqueness of trace for $C_r(\mathbb{G})$ are now immediate consequences (as long as $\dim B \geq 8$).

**Corollary 5.12.** Let $\psi$ be the $\delta$-trace on a finite dimensional $C^*$-algebra $B$ and let $\mathbb{G} = \mathbb{G}_{\text{aut}}(B, \psi)$. If $\dim B \geq 8$, then $C_r(\mathbb{G})$ is simple with unique trace.

**Proof.** Let $I \subseteq C_r(\mathbb{G})$ be a non-zero closed two-sided ideal and let $a \in I$ be such that $h(a) \neq 0$. Then by Proposition 5.11, $h(a)1 = \lim_{n \to \infty} \Phi^n(a) \in I$. Therefore $I = C_r(\mathbb{G})$ and $C_r(\mathbb{G})$ is simple.

Now let $\tau$ be any tracial state on $C_r(\mathbb{G})$. Since $\tau \circ \Phi^n = \tau$ for all $n$, we have $\tau(x) = \lim_{n \to \infty} \tau(\Phi^n(x)) = \tau(h(x)1) = h(x)$ for all $x \in C_r(\mathbb{G})$.

### 5.4 Solidity

In this final section we show that the von Neumann algebras $L^\infty(\mathbb{G})$ considered in this paper always have the Akemann-Ostrand property. Recall that a von Neumann $M \subseteq B(\mathcal{H})$ has the Akemann-Ostrand property (or property (AO)) if there exist $\sigma$-weakly dense unital $C^*$-subalgebras $A \subseteq M$ and $B \subseteq M'$ such that $A$ is locally reflexive and the $\star$-homomorphism $\pi m : A \otimes_{\text{alg}} B \to B(\mathcal{H})/K(\mathcal{H})$ is continuous with respect to the minimal tensor norm. Here $m : A \otimes_{\text{alg}} B \to B(\mathcal{H})$ is the multiplication map and $\pi : B(\mathcal{H}) \to B(\mathcal{H})/K(\mathcal{H})$ is the canonical quotient map. Note that an exact $C^*$-algebra is always locally reflexive. We now state the main result of this section. Below we denote by $J : L^2(\mathbb{G}) \to L^2(\mathbb{G})$ the modular conjugation operator for $L^\infty(\mathbb{G}) = \lambda(C_u(\mathbb{G}))''$, so that $JL^\infty(\mathbb{G})J = L^\infty(\mathbb{G})'$.

**Theorem 5.13.** Let $\mathbb{G} = \mathbb{G}_{\text{aut}}(B, \psi)$, where $B$ is a finite dimensional $C^*$-algebra equipped with a $\delta$-form $\psi$. Then there exists a unital completely positive map $\theta : C_r(\mathbb{G}) \otimes JC_r(\mathbb{G})J \to B(L^2(\mathbb{G}))$ such that

$$\theta(a \otimes JbJ) - aJbJ \in K(L^2(\mathbb{G})) \quad (a, b \in C_r(\mathbb{G})).$$

In particular, $L^\infty(\mathbb{G})$ has property (AO).
In [31], Ozawa proved that a finite von Neumann algebra $M$ with property (AO) is solid. That is, for any diffuse von Neumann subalgebra $A \subset M$, the relative commutant $A' \cap M$ is injective. Recall that a II$_1$-factor $M$ is said to be prime if for every tensor product decomposition $M \cong M_1 \otimes M_2$, either $M_1$ or $M_2$ is finite dimensional. It is easy to see that a solid II$_1$-factor is prime unless it is injective. As a consequence of Corollary 4.17, Theorem 5.1 and the above discussion, the following result is immediate.

**Corollary 5.14.** Let $\psi$ be the $\delta$-trace on $B$ and let $G = G_{\text{aut}}(B, \psi)$. Then $L^\infty(G)$ is solid and it is a prime II$_1$-factor if $\dim B \geq 8$.

**Remark 5.15.** It is perhaps worthwhile pointing out that the local reflexivity of $C_r(G)$ is not actually required to establish the solidity of $L^\infty(G)$. To see this, note that an examination of the proof that property (AO) implies solidity in [31] shows that local reflexivity is only needed to establish, for each finite dimensional operator system $E \subset L^\infty(G)$, the existence of a net of unital completely positive maps $\{\Psi^E_t : E \to C_r(G)\}_t$ converging pointwise $\sigma$-weakly to $\text{id}_E$. In our case, we can just take $\Psi^E_t = \Phi_t|_E$, where $\{\Phi_t\}_{t<\dim B}$ is the net constructed in the proof of Theorem 4.2. Indeed, a standard estimate using property (RD) (see for example [14], Proposition 5.3) shows that $\Phi_t$ is ultracottractive, that is $\Phi_t(L^2(G)) \subseteq \Lambda_k(C_r(G))$ for all $t$. In particular, $\Phi_t(L^\infty(G)) \subseteq C_r(G)$ for all $t$.

**Remark 5.16.** The existence of a completely positive map $\theta$ in Theorem 5.13 such that $\pi \circ \theta = \pi \circ m$ imposes a structural property on $L^\infty(G)$ which is stronger than property (AO). Using the terminology of Isono [25], Theorem 5.13 says that $L^\infty(G)$ has property (AO)$^\tau$. Property (AO)$^\tau$ can be regarded as a generalization to finite von Neumann algebras of the description of bi-exactness for discrete groups in terms of their reduced C$^*$-algebras (see [15], Chapter 15). In [33], Popa and Vaes proved that the II$_1$-factors $M = L(G)$ associated to i.e.c., weakly amenable, bi-exact groups $\Gamma$ are strongly solid. That is, for any diffuse injective von Neumann subalgebra $A \subset M$, the von Neumann algebra generated by the normalizer of $A$, $N_M(A) = \{u \in U(M) : uAu^* = A\}$, is injective. Recently, Isono [25] extended the results of [33] and proved that a finite von Neumann algebra with property (AO)$^\tau$ and the weak$^*$-completely bounded approximation property (W$^*$-CBAP) is strongly solid. As a result, if one could prove that the von Neumann algebras $L^\infty(G)$ considered here have the W$^*$-CBAP, their strong solidity would be an immediate consequence. Based on recent work of Freslon [22] on the weak amenability free orthogonal quantum groups, we expect that $L^\infty(G)$ always has this approximation property.

**Proof of Theorem 5.13.** Our arguments follow very closely those of Vergnioux [37], where property (AO)$^\tau$ was established for the duals of universal discrete quantum groups. We note that the proof in [37] makes extensive use of the notion of a Cayley graph associated to a discrete quantum group $\hat{G}$, and assumes that this graph is a tree. In the case of $G = G_{\text{aut}}(B, \psi)$ of interest here, the Cayley graph of $\hat{G}$ is no longer a tree. In this case, since we have the fusion
rules $U^1 \otimes U^k \cong U^k \otimes U^1 \cong U^{k+1} \oplus U^k \oplus U^{k-1}$ ($k \geq 1$), the Cayley graph is obtained by taking the infinite tree $A_\infty$ and adding a loop starting and ending at each vertex $k \neq 0$ (resulting from the inclusions $U^k \subset U^1 \otimes U^k, U^k \otimes U^1$).

Since such a graph is not “too far” from being a tree, we can readily adapt the proof from [37] to our context. Below we freely use the notation from Sections 3.2, 4.2 and 5.1. We also assume that $\delta^2 = q + q^{-1} \geq 5$, so that $C_r(\mathbb{G})$ is not nuclear.

For each $l \in \mathbb{N}$, put $m_l = [2l + 1]_q$, $\alpha_l = \left( \sum_{n,k \in \mathbb{N}} n+k=l \frac{m_n m_k}{m_l} \right)^{-\frac{1}{2}}$ and define a linear map

$$V : C[\hat{G}] \to C[\hat{G}] \otimes_{alg} C[\hat{G}]; \quad V(x) = \sum_{l \in \mathbb{N}} \alpha_l \sum_{n,k \in \mathbb{N}} n+k=l \hat{\Delta}(\hat{P}_l)x(\hat{P}_n \otimes \hat{P}_k).$$

Using Lemma 4.12 and its proof, it follows that $V$ extends to an isometry from $\ell^2(\hat{G}) \to \ell^2(\hat{G}) \otimes \ell^2(\hat{G})$, and $V^* (\hat{P}_n \otimes \hat{P}_k)y = \alpha_{n+k} P_{n+k} \text{Conv}((\hat{P}_n \otimes \hat{P}_k)y)$ ($y \in \ell^2(\hat{G}) \otimes \ell^2(\hat{G})$, $n, k \in \mathbb{N}$). Identifying $\ell^2(\hat{G}) \cong L^2(\mathbb{G})$ via the Fourier transform $\mathcal{F}$, we can form a unital completely positive map

$$\theta : C_r(\mathbb{G}) \otimes JC_r(\mathbb{G})J \to B(L^2(\mathbb{G})); \quad \theta(x) = V^* x V.$$

Our goal is to prove that

$$a J b J V^* - V^* (a \otimes J b J) \in K(L^2(\mathbb{G}) \otimes L^2(\mathbb{G}), L^2(\mathbb{G})) \quad (5.9)$$

for all $a, b \in C_r(\mathbb{G})$, which will show that $\theta$ satisfies the hypotheses of the theorem. Note that since $a J b J V^* - V^* (a \otimes J b J) = a (J b J V^* - V^* (1 \otimes J b J)) + (a^* - V^* (a \otimes 1))(1 \otimes J b J)$, it suffices to prove (5.9) when one of $a$ or $b$ equals 1. Let us start with $b = 1$. Note furthermore that we only need to prove (5.9) for $a = \lambda(v) \in C_r(\mathbb{G})$, where $v \in C_u(\mathbb{G})$ is a matrix element of the irreducible representation $U^1 = U \oplus U^0$. This follows because the norm-closed subalgebra of $C_r(\mathbb{G})$ generated by all such $a = \lambda(v)$ equals $C_r(\mathbb{G})$, and (5.9) is preserved by sums, products and norm-limits.

Fix for the remainder $a = \lambda(v)$ as above, and put $K = \sum_{l \in \mathbb{N}} \alpha_l \hat{P}_l \in B(L^2(\mathbb{G}))$.

Since $1 \leq \frac{m_n m_k}{m_l} = \frac{(1-q^n+2)(1-q^{n+k+2})}{(1-q^l)(1-q^{n+k+2})} \leq \frac{1}{(1-q^l)}$ for $k + n = l$, it follows that $(1-q^l)(l+1)^{-1/2} \leq \alpha_l \leq (l+1)^{-1/2}$. Thus $K$ is compact with (unbounded) inverse $K^{-1} = \sum_{l \in \mathbb{N}} \alpha_l^{-1} \hat{P}_l$. Using $K$, we can write

$$a V^* - V^* (a \otimes 1) = K (K^{-1}a - aK^{-1})V^* - K (aK^{-1}V^* - K^{-1}V^* (a \otimes 1)),$$

so it suffices to show that $A_1 = K^{-1}a - aK^{-1}$ and $A_2 = aK^{-1}V^* - K^{-1}V^* (a \otimes 1)$ are bounded linear maps.

First consider $A_1$. Since $a = \lambda(v) = \hat{P}_0 a \hat{P}_1 + \sum_{l \geq 1} \sum_{|l| \leq l} \hat{P}_l a \hat{P}_l$, we have

$$A_1 = (1 - \alpha_0^{-1}) \hat{P}_0 a \hat{P}_1 + \sum_{l \geq 1} \sum_{|l| \leq l} (\alpha_{l}^{-1} - \alpha_{l+1}^{-1}) \hat{P}_l a \hat{P}_l.$$
It therefore suffices to show that \( S := \sup_{\xi \geq 1} \{ \max_{\xi' - 0 \leq \xi} |\alpha_\xi^{-1} - \alpha_{\xi'}^{-1}| \} < \infty. \)

Since \((l + 1)^{1/2} \leq \alpha_\xi^{-1} \leq \frac{(l + 1)^{1/2}}{q^2}, \) the fact that \( S < \infty \) is clear.

Now consider \( A_2. \) Let \( \hat{a} = F(v) \in B(H_1) \subset C[\hat{G}], \) let \( y \in C[\hat{G}] \otimes_{\text{alg}} C[\hat{G}], \) and put \( y_{k,n} = (\hat{P}_k \otimes \hat{P}_n)y \) for each \( k, n \in \mathbb{N}. \) For each inclusion \( U^l \subset U^k \boxtimes U^n (k, l, n \geq 0), \) let \( \phi^k_{\otimes \hat{G}^n} = C_{(k,n,l)}^r \rho^k_{\otimes \hat{G}^n} : H_1 \to H_k \otimes H_n \) be the (unique, up to multiplication by \( T \)) isometric morphism defined by (3.2)-(3.3). Since \( V^* (y_{k,n}) = K\hat{P}_{k+n} \text{Conv}(y_{k,n}) \) and (by Lemma 4.12 and its proof)

\[
\hat{P}_k \text{Conv}(y_{k,n}) = \frac{m_k m_n}{m_{k+n}} (\phi^k_{\otimes \hat{G}^n})^* y_{k,n} \phi^k_{\otimes \hat{G}^n} = \frac{m_k m_n}{m_{k+n}} \text{Ad}((\phi^k_{\otimes \hat{G}^n})^*) y_{k,n},
\]

it follows that

\[
A_2(y_{k,n}) = \text{Conv}(\hat{a} \otimes \hat{P}_{k+n} \text{Conv}(y_{k,n})) - K^{-1} V^*((\text{Conv} \otimes \text{id})(\hat{a} \otimes y_{k,n}))
\]

\[
= \sum_{r=0,\pm 1} \left( \hat{P}_{n+k+r} \right)
\]

\[
\times \left( \text{Conv}(\hat{a} \otimes \hat{P}_{k+n} \text{Conv}(y_{k,n})) - \text{Conv}((\hat{P}_{k+n} \text{Conv} \otimes \text{id})(\hat{a} \otimes y_{k,n}))) \right)
\]

\[
= \sum_{r=0,\pm 1} \left( \frac{m_k m_n}{m_{k+n}} (\text{Ad}((\text{id} \otimes \phi^k_{\otimes \hat{G}^n}) \phi^k_{\otimes \hat{G}^n})^* y_{k+n+r}) \right)
\]

\[
- \text{Ad}((\phi^k_{\otimes \hat{G}^n} \otimes \text{id}) y_{k+n+r})^*[\hat{a} \otimes y_{k,n}]),
\]

where \( \hat{P}_{-1} \) is zero by convention. Let \( \phi^k_{\otimes \hat{G}^n,L} = (\text{id} \otimes \phi^k_{\otimes \hat{G}^n}) \phi^k_{\otimes \hat{G}^n,r} \) and \( \phi^k_{\otimes \hat{G}^n,R} = (\phi^k_{\otimes \hat{G}^n} \otimes \text{id}) \phi^k_{\otimes \hat{G}^n}. \) By appealing to [36], Lemma A.1, note that we can find a constant \( C = C(q) > 0 \) (independent of \( k \) and \( n \)) such that

\[
\inf_{\lambda \in T} \| \phi^k_{\otimes \hat{G}^n,L} - \lambda \phi^k_{\otimes \hat{G}^n,R} \|_{\text{TL}_{2(k+n+1),2(2k+n+1)}(\delta)} \leq C q^{2k+r-1}, \tag{5.10}
\]

for each \( n, k \in \mathbb{N} \) and \( r = 0, \pm 1. \) Since we also have \( \| T^* (\hat{a} \otimes y_{k,n}) S \|_{\ell^2} \leq \left( \frac{m_{n+k+r}}{m_{n} m_{k+r}} \right)^{1/2} \| T \| \| S \| \| \hat{a} \otimes y_{k,n} \|_{\ell^2} \) \) for any \( S, T \in \text{Mor}(H_{n+k+r}, H_k \otimes H_n), \) an application of the triangle inequality and (5.10) yields

\[
\| \text{Ad}((\phi^k_{\otimes \hat{G}^n,L})^*) - \text{Ad}((\phi^k_{\otimes \hat{G}^n,R})^*) [\hat{a} \otimes y_{k,n}] \|_{\ell^2(\hat{G})} \leq 2C q^{2k+r-1} \left( \frac{m_{n+k+r}}{m_{n} m_{k+r}} \right)^{1/2} \| \hat{a} \otimes y_{k,n} \|_{\ell^2(\hat{G})}. \]

Using this last inequality and the Cauchy-Schwarz inequality, we obtain the
estimate
\[
\|A_2y\|_{\ell^2(\widehat{G})}^2 = \sum_{l \geq 0} \left\| \sum_{k,n:|n+k-l| \leq 1} \hat{P}_l A_2(y_{k,n}) \right\|_{\ell^2(\widehat{G})}^2
\]
\[
= \sum_{l \geq 0} \left\| \sum_{k,n:|n+k-l| \leq 1} \frac{m_1m_km_n}{m_l} (\text{Ad}([\phi^*_L \otimes \phi^*_R]^{\tau})(\phi^*_L)) [\hat{a} \otimes y_{k,n}] \right\|_{\ell^2}^2
\]
\[
\leq \sum_{l \geq 0} \left( \sum_{k,n:|n+k-l| \leq 1} 2Cq^{l+n-1} \left( \frac{m_1m_km_n}{m_l} \right)^{1/2} \|\hat{a}\|_{\ell^2(\widehat{G})}\|y_{k,n}\|_{\ell^2(\widehat{G})}\|y_{k,n}\|_{\ell^2(\widehat{G})} \right)^2
\]
\[
\leq 4C^2m_1\|\hat{a}\|_{\ell^2(\widehat{G})}^2
\times \sum_{l \geq 0} \left( \sum_{k,n:|n+k-l| \leq 1} \frac{m_km_n}{m_l} q^{2k+2l-2n-2} \right) \sum_{k,n:|k+n-l| \leq 1} \|y_{k,n}\|_{\ell^2(\widehat{G})}\|y_{k,n}\|_{\ell^2(\widehat{G})} \sum_{k,n:|k+n-l| \leq 1} \|y_{k,n}\|_{\ell^2(\widehat{G})}\|y_{k,n}\|_{\ell^2(\widehat{G})}.
\]

Since the above sum over \( l \) is dominated by \( 3\|y\|_{\ell^2(\widehat{G})}=\sum_{l \geq 0} \frac{m_1m_n}{m_l} \leq \frac{q^{-2}}{(1-q^2)} \) for \( |k+n-l| \leq 1 \), and
\[
\sum_{k,n:|n+k-l| \leq 1} \frac{m_km_n}{m_l} q^{2k+2l-2n-2} \leq \sum_{k,n:|k+n-l| \leq 1} \frac{q^{-2}}{(1-q^2)} q^{2k+2l-2n-2}
\]
\[
= \frac{q^{-4}}{(1-q^2)^2} \sum_{r=0}^{l+r} \sum_{k=0}^{l+r} q^{4k-2r} \leq \frac{3q^{-6}}{(1-q^2)^2(1-q^4)} \quad (l \geq 0),
\]
we conclude that \( A_2 \) is bounded.

The proof of (5.9) when \( a = 1 \) and \( b \neq 1 \) is essentially the same as the above argument. The only significant difference is that one must use a right-sided version of (5.10). The required inequality is given by Lemma A.2 from [36]. We leave the details to the reader. \( \square \)

### 5.5 Concluding Remark

Since this work was submitted for publication, the author has obtained a description of the quantum groups \( \widehat{G}_{\text{aut}}(B, \psi) \) for arbitrary invariant states \( \psi \) in terms of a free product \( \ast_{\text{aut}} \mathcal{G}_{\text{aut}}(B_i, \psi_i) \), where each \( B_i \) is a certain distinguished subalgebra of \( B \) and \( \psi_i = \psi|_{B_i} \). See Proposition 18 in [19].

### 6 Appendix – Proof of Theorem 5.4

In this section we prove Theorem 5.4. We will freely use the notation and assumptions of Section 5.1.
Proof of Theorem 5.4. We can write the commutator map $T : L^2(G) \to H_1 \otimes L^2(G) \otimes H_1$ as the sum $T = T_L - T_R$, where $T_L = \sum_{i,j=1}^{d_1} F_i e_j \otimes \lambda(u_{ij}^1) \otimes e_i$ is the part of $T$ corresponding to the left action of $\{\lambda(u_{ij}^1)\}_{i,j=1}^{d_1}$ on $L^2(G)$ and $-T_R$ is the remaining part of $T$ corresponding to the right action of $\{\lambda(u_{ij}^1)\}_{i,j=1}^{d_1}$ on $L^2(G)$. Observe that $T\xi_0 = T_L\xi_0 - T_R\xi_0 = 0$.

In terms of matrix elements of irreducible representations of $G$, the map $T_L$ (respectively $T_R$) corresponds to the operation of tensoring on the left (respectively on the right) by $U^1$. Since $U^1 \otimes U^k \cong U^{k+1} \oplus U^k \oplus U^{k-1} \cong U^k \otimes U^1$ for all $k \geq 1$, we can uniquely decompose $T_L$ and $T_R$ as the sums $T_{L,R} = \sum_{\alpha=0,\pm 1} T_{L,R}^{(\alpha)}$, where

$$T_{L,R}^{(0)}\xi_0 = T_{L,R}^{(-1)}\xi_0 = 0, \quad T_{L,R}^{(+1)}\xi_0 = T_{L,R}^{(0)}\xi_0 = \sum_{i,j=1}^{d_1} F_i e_j \otimes \lambda(u_{ij}^1)\xi_0 \otimes e_i,$$

and for $k \geq 1$, $T_{L,R}^{(\alpha)} : L^2_k(G) \to H_1 \otimes L^2_{k+\alpha}(G) \otimes H_1$ is defined by

$$T_{L,R}^{(\alpha)}|_{L^2_k(G)} = (\text{id}_{H_1} \otimes P_{k+\alpha} \otimes \text{id}_{H_1})T_{L,R}|_{L^2_k(G)}.$$

Setting $T^{(\alpha)} := T^{(\alpha)}_L - T^{(\alpha)}_R$, we obtain the decomposition $T = \sum_{\alpha=0,\pm 1} T^{(\alpha)}$ stated in Theorem 5.4.

To verify the expressions for $T^{(\alpha)}$ given by Theorem 5.4, fix $k \geq 1$, $\xi, \eta \in H_k$ and consider the matrix element $\lambda((\omega_{\eta,\xi} \otimes \text{id})U^k)\xi_0 \mapsto \xi \otimes (1 \otimes \eta^*)t_{2k}$. $(\xi, \eta \in H_k, k \geq 0)$, given by (2.4), as well as the identities

$$\sum_{i,j=1}^{d_1} F_i e_j \otimes e_j = \sum_{j=1}^{d_1} F_i F_i^* F_i \otimes e_j = \sum_{j=1}^{d_1} F_i \otimes e_j = \sum_{i,j=1}^{d_1} F_i e_j \otimes e_j = [3]^1/2 t_2,$$

$$(p_x \otimes p_y) p_{x+y} = p_{x+y} \quad (x, y \in \mathbb{N}).$$

Since $\bigoplus_{\alpha=0,\pm 1} \varphi_{k+\alpha,L}^{(-\alpha)} : \bigoplus_{\alpha=0,\pm 1} H_{k+\alpha} \to H_1 \otimes H_k$ is a unitary intertwiner, it follows that

$$T_L(\lambda((\omega_{\eta,\xi} \otimes \text{id})U^k)\xi_0) = \sum_{i,j=1}^{d_1} F_i e_j \otimes \lambda(u_{ij}^1(\omega_{\eta,\xi} \otimes \text{id})U^k)\xi_0 \otimes e_i$$

$$= \sum_{i,j=1}^{d_1} F_i e_j \otimes \lambda((\omega_{\eta,\xi} \otimes \text{id})U^{k+\alpha}\otimes \text{id})(U^1 \otimes U^k))\xi_0 \otimes e_i$$

$$= \sum_{\alpha=0,\pm 1} \sum_{i,j=1}^{d_1} F_i e_j \otimes \lambda((\omega_{\eta,\xi} \otimes \text{id})(U^{k+\alpha}\otimes \text{id})(U^1 \otimes U^k))\xi_0 \otimes e_i.$$
Let us first consider

\[ T^{(\alpha)}_L(\xi \otimes (1 \otimes \eta^*)t_{2k}) \]

\[ = \sum_{i,j=1}^{d_1} F_1 e_j \otimes \lambda(\omega(\phi_{k+\alpha,L}^{(-\alpha)})(e_i \otimes \eta) \otimes \xi \otimes \xi) \otimes \lambda)U^{k+\alpha} \xi_0 \otimes e_i \]

\[ = \sum_{j=1}^{d_1} F_1 e_j \otimes (\phi_{k+\alpha,L}^{(-\alpha)})(e_j \otimes \xi) \otimes \sum_{i=1}^{d_1} (1 \otimes (e_i^* \otimes \eta^*))\phi_{k+\alpha,L}^{(-\alpha)} t_{2k+2\alpha} \otimes e_i \]

\[ = \left[ (1_2 \otimes (\phi_{k+\alpha,L}^{(-\alpha)}))^*(\sum_j F_1 e_j \otimes e_j \otimes \xi) \right] \]

\[ \otimes \left[ (1_{2k+2\alpha} \otimes 1_2 \otimes \eta^*)(1_{2k+2\alpha} \otimes \phi_{k+\alpha,L}^{(-\alpha)})t_{2k+2\alpha} \right], \]

\[ =: A_3^{(\alpha)}(1 \otimes \eta^*)t_{2k} \]

and similarly

\[ T^{(\alpha)}_R(\xi \otimes (1 \otimes \eta^*)t_{2k}) = \left[ (1_2 \otimes (\phi_{k+\alpha,R}^{(-\alpha)}))^*(\sum_{j=1}^{d_1} F_1 e_j \otimes e_j \otimes \xi) \right] \]

\[ \otimes \left[ (1_{2k+2\alpha} \otimes \eta^*)t_{2k+2\alpha} \right] \]

\[ =: B_3^{(\alpha)}(1 \otimes \eta^*)t_{2k} \]

Let us first consider \( T^{(\alpha)}_L|_{H_k \otimes H_k} = A_1^{(\alpha)} \otimes A_2^{(\alpha)} \). When \( \alpha = 1 \), we have

\[ A_1^{(1)} = [3]_{q}^{1/2}(1_2 \otimes (\phi_{k+1,L}^{(-1)}))(t_2 \otimes \xi) = [3]_{q}^{1/2}(1_2 \otimes p_{2k+2}(p_2 \otimes p_{2k}))(t_2 \otimes \xi) \]

\[ = [3]_{q}^{1/2}(p_2 \otimes p_{2k+2})(t_2 \otimes 1_{2k})p_{2k} \xi = \left[ \frac{[2k+3]_{q}}{[2k+1]_{q}} \right]^{1/2} \phi_{k,L}^{(+1)} \xi, \]

and

\[ A_2^{(1)}(1 \otimes \eta^*)t_{2k} = (1_{2k+2} \otimes 1_2 \otimes \eta^*)(1_{2k+2} \otimes p_2 \otimes p_{2k})t_{2k+2} \]

\[ = \left[ \frac{[2k+1]_{q}[3]_{q}}{[2k+3]_{q}} \right]^{1/2} (1_{2k+2} \otimes 1_2 \otimes \eta^*)(p_{2k+2} \otimes p_2 \otimes p_{2k})(1_{2k} \otimes t_2 \otimes 1_{2k})t_{2k} \]

\[ = (1_{2k+2} \otimes 1_2 \otimes \eta^*)(1_{2k+2} \otimes 1_{2k+2})(\phi_{k,R}^{(+1)} \otimes 1_{2k})t_{2k} = \phi_{k,R}^{(+1)}(1_{2k} \otimes \eta^*)t_{2k}. \]
When $\alpha = 0$, we have

$$A_1^{(0)} \xi = [3]^{1/2} (p_2 \otimes (\phi_{k,L}^{(0)})^*)(t_2 \otimes \xi)$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} (p_2 \otimes p_{2k}(m \otimes 1_{2k-2})(p_2 \otimes p_{2k}))(t_2 \otimes \xi)$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} (p_2 \otimes p_{2k}(m \otimes 1_{2k-2}))(t_2 \otimes \xi)$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} (p_2 \otimes p_{2k}(m \otimes 1_{2k-2}))(m^* \nu \otimes 1_{2k})\xi$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} (p_2 \otimes p_{2k}(1_2 \otimes m \otimes 1_{2k-2})(m^* \otimes 1_{2k})(\nu \otimes 1_{2k})\xi$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} (p_2 \otimes p_{2k})(m^* m \otimes 1_{2k-2})(\nu \otimes 1_{2k})\xi$$

(cf. Remark 3.6)

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} (p_2 \otimes p_{2k})(m^* \otimes 1_{2k-2})\xi = \phi_{k,L}^{(0)} \xi,$$

and

$$A_2^{(0)} (1 \otimes \eta^*)(t_{2k} = (1_{2k} \otimes 12 \otimes \eta^*)(1_{2k} \otimes \phi_{k,L}^{(0)})(t_{2k} = \phi_{k,R}^{(0)}(1_{2k} \otimes \eta^*)t_{2k}.$$  

The last equation follows because both of the isometries $(\phi_{k,R}^{(0)} \otimes 1_{2k})t_{2k}$ and $(1_{2k} \otimes \phi_{k,L}^{(0)})t_{2k}$ belong to the one-dimensional space $\text{Mor}(1,U^k \otimes U^1 \otimes U^k)$ and therefore $(1_{2k} \otimes \phi_{k,L}^{(0)})t_{2k} = z(\phi_{k,R}^{(0)} \otimes 1_{2k})t_{2k}$ for some $z \in T$. But then

$$z = t_{2k}^* ((\phi_{k,R}^{(0)})^* \otimes 1_{2k})(1_{2k} \otimes \phi_{k,L}^{(0)})t_{2k}$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} t_{2k}^* (1_{2k-2} \otimes m \otimes 1_{2k})(1_{2k} \otimes p_2 \otimes 1_{2k})(1_{2k} \otimes m^* \otimes 1_{2k-2})t_{2k}$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} t_{2k}^* (1_{2k-2} \otimes m \otimes 1_{2k})(1_{2k} \otimes (1_2 - \nu \nu^*) \otimes 1_{2k})$$

$$\times (1_{2k} \otimes m^* \otimes 1_{2k-2})t_{2k}$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} t_{2k}^* (1_{2k-2} \otimes m^* m \otimes 1_{2k-2})t_{2k} - t_{2k}^* t_{2k}$$

$$= \left[ \frac{[2k]_q}{[2k+2]_q} \right]^{1/2} \left( \frac{[2k+1]_q}{[2k]_q} - 1 \right) = 1,$$

where in the last line we have used

$$(1_{2k-2} \otimes m \otimes 1_{2k-2})t_{2k} = \left( [2k]_q [2k+1]_q [2k]_q^{-1} \right)^{1/2} t_{2k-1}.$$
When \( \alpha = -1 \), we have

\[
A_{1}^{(-1)} \xi = [3]_{q}^{1/2} (1_2 \otimes (\phi_{k-1,L}^{(1)})^*)(t_2 \otimes \xi)
\]

\[
= \left[ \frac{[3]_{q}[2k-1]_{q}}{[2k+1]_{q}} \right]^{1/2} (1_2 \otimes p_{2k-2} (t_2^* \otimes 1_{2k-2}) (p_2 \otimes p_{2k})) (t_2 \otimes \xi)
\]

\[
= \left[ \frac{[3]_{q}[2k-1]_{q}}{[2k+1]_{q}} \right]^{1/2} (p_2 \otimes p_{2k-2}) (1_2 \otimes t_2^* \otimes 1_{2k-2}) (t_2 \otimes 1_{2k}) \xi
\]

\[
= \left[ \frac{[2k-1]_{q}}{[2k+1]_{q}} \right]^{1/2} (p_2 \otimes p_{2k-2}) \xi = \left[ \frac{[2k-1]_{q}}{[2k+1]_{q}} \right]^{1/2} \phi_{k,L}^{(-1)} \xi
\]

(using \( (1_2 \otimes t_2^*) (t_2 \otimes 1_2) = [3]_{q}^{-1} p_2 \)),

and

\[
A_{2}^{(-1)} (1 \otimes \eta^*) t_{2k}
\]

\[
= \left[ \frac{[3]_{q}[2k-1]_{q}}{[2k+1]_{q}} \right]^{1/2} \times (1_{2k-2} \otimes 1_2 \otimes \eta^*) (p_{2k-2} \otimes p_2 \otimes p_{2k}) (1_{2k-2} \otimes 1_{2k-2}) t_{2k-2}
\]

\[
= (1_{2k-2} \otimes 1_2 \otimes \eta^*) (p_{2k-2} \otimes p_2 \otimes 1_{2k})
\]

\[
\times \left[ \frac{[3]_{q}[2k-1]_{q}}{[2k+1]_{q}} \right]^{1/2} (1_{2k} \otimes p_{2k}) (1_{2k-2} \otimes 1_{2k-2}) t_{2k-2}
\]

\[
= (1_{2k} \otimes \eta^*) (\phi_{k,R}^{(-1)} \otimes 1_{2k}) t_{2k} = \phi_{k,R}^{(-1)} (1_{2k} \otimes \eta^*) t_{2k}.
\]

Let us now consider \( T_{R}^{(\alpha)}|_{H_{k} \otimes H_{k}} = B_{1}^{(\alpha)} \otimes B_{2}^{(\alpha)} \). When \( \alpha = 1 \), we have

\[
B_{1}^{(1)} \xi = [3]_{q}^{1/2} \sigma ((\phi_{k+1,R}^{(-1)})^* \otimes 1_2) (\xi \otimes t_2)
\]

\[
= [3]_{q}^{1/2} \sigma (p_{2k+2} (p_2 \otimes p_{2k}) (1_2 \otimes t_2)) \xi
\]

\[
= [3]_{q}^{1/2} \sigma (p_{2k+2} \otimes p_2) (1_2 \otimes t_2) \xi
\]

\[
= \left[ \frac{[2k+3]_{q}}{[2k+1]_{q}} \right]^{1/2} \sigma \phi_{k,R}^{(+1)} \xi,
\]
and

\[
B_2^{(1)}(1 \otimes \eta^*)(t_{2k}) = (1_{2k+2} \otimes \eta^* \otimes 1_2)(1_{2k+2} \otimes \phi_{k+1,R}^{-1})t_{2k+2}
\]

\[
= \left( \frac{[2k+1]_q}{[2k+3]_q} \right)^{1/2} (1_{2k+2} \otimes \eta^* \otimes 1_2)(p_{2k+2} \otimes 1_{2k+2}) \\
\times \left( \sum_{j=1}^{d_1} e_j \otimes 1_{2k} \otimes 1_{2k} \otimes F_1 e_j \right)t_{2k}
\]

\[
= \left( \frac{[2k+1]_q}{[2k+3]_q} \right)^{1/2} \sigma^*(1_2 \otimes 1_{2k+2} \otimes \eta^*)(1_2 \otimes p_{2k+2} \otimes 1_{2k}) \\
\times \left( \sum_{j=1}^{d_1} F_1 e_j \otimes e_j \otimes 1_{2k} \otimes 1_{2k} \right)t_{2k}
\]

\[
= \left( \frac{[2k+1]_q}{[2k+3]_q} \right)^{1/2} \sigma^*(1_2 \otimes p_{2k+2}) \left( \sum_{j=1}^{d_1} F_1 e_j \otimes e_j \otimes 1_{2k} \otimes \eta^* \right)t_{2k}
\]

\[
= \left( \frac{[3]_q [2k+1]_q}{[2k+3]_q} \right)^{1/2} \sigma^*(1_2 \otimes p_{2k+2})(t_2 \otimes 1_{2k} \otimes \eta^*)t_{2k}
\]

\[
= \left( \frac{[3]_q [2k+1]_q}{[2k+3]_q} \right)^{1/2} \sigma^*(1_2 \otimes p_{2k+2})(t_2 \otimes 1_{2k})(1_{2k} \otimes \eta^*)t_{2k}
\]

\[
= \sigma^* \phi_{k,L}^{(+1)}((1_{2k} \otimes \eta^*)t_{2k}).
\]

When \( \alpha = 0 \), we have

\[
B_1^{(0)} \xi = [3]_q^{1/2} \sigma((\phi_{k,R}^{(0)})^* \otimes 1_2)(\xi \otimes t_2)
\]

\[
= [3]_q^{1/2} \left( \frac{[2k]_q}{[2k+2]_q} \right)^{1/2} \sigma(p_{2k}(1_{2k-2} \otimes m)(p_{2k} \otimes p_2) \otimes 1_2)(\xi \otimes t_2)
\]

\[
= [3]_q^{1/2} \left( \frac{[2k]_q}{[2k+2]_q} \right)^{1/2} \sigma(p_{2k} \otimes p_2)(1_{2k-2} \otimes m \otimes 1_2)(1_{2k} \otimes t_2)\xi
\]

\[
= \left( \frac{[2k]_q}{[2k+2]_q} \right)^{1/2} \sigma(p_{2k} \otimes p_2)(1_{2k-2} \otimes m \otimes 1_2)(1_{2k} \otimes 1_2 \otimes p_2) \\
\times (1_{2k} \otimes m^*)(1_{2k} \otimes \nu)\xi
\]

\[
= \left( \frac{[2k]_q}{[2k+2]_q} \right)^{1/2} \sigma(p_{2k} \otimes p_2)(1_{2k-2} \otimes m \otimes 1_2)(1_{2k} \otimes m^*)(1_{2k} \otimes \nu)\xi
\]

\[
= \left( \frac{[2k]_q}{[2k+2]_q} \right)^{1/2} \sigma(p_{2k} \otimes p_2)(1_{2k-2} \otimes m \otimes 1_2)(1_{2k} \otimes m^*)\xi = \sigma \phi_{k,R}^{(0)} \xi.
\]

Using the fact that

\[
p_2^{\otimes 3}(1_2 \otimes m^*)t_2 = [3]_q^{1/2} p_2^{\otimes 3}(1_2 \otimes m^*)m^* \nu
\]

\[
= [3]_q^{1/2} p_2^{\otimes 3}(m^* \otimes 1_2)m^* \nu = p_2^{\otimes 3}(m^* \otimes 1_2)t_2,
\]
Finally, when $\alpha = -1$, we have

$$
\left( \frac{[2k]_q}{[2k+2]_q} \right)^{-1/2} B_2^{(0)} (1 \otimes \eta^*) \xi \eta
\begin{align*}
&= (1_{2k} \otimes \eta^* \otimes 1_k)(p_{2k} \otimes (p_{2k} \otimes p_2)(1_{2k-2} \otimes m^*)) t_{2k} \\
&= \left( \frac{[3]_q[2k-1]_q}{[2k+1]_q} \right)^{1/2} (p_{2k} \otimes \eta^* \otimes p_2)(1_{2k} \otimes 1_{2k-2} \otimes m^*) t_{2k} \\
&= \left( \frac{[3]_q[2k-1]_q}{[2k+1]_q} \right)^{1/2} \sigma (p_{2k} \otimes p_2 \otimes \eta^*) t_{2k} \\
&= \left( \frac{[3]_q[2k-1]_q}{[2k+1]_q} \right)^{1/2} \sigma^* (p_2 \otimes p_{2k} \otimes \eta^*) (m^* \otimes 1_{2k-2} \otimes t_{2k-2} \otimes 1_{2k-2}) t_{2k} \\
&= \sigma^* (p_2 \otimes p_{2k})(1_{2k-2} \otimes m^*) t_{2k} \\
&= \left( \frac{[2k]_q}{[2k+2]_q} \right)^{-1/2} \sigma^* \phi_{k,L}^{(0)} (1_{2k} \otimes \eta^*) t_{2k}.
\end{align*}
$$

Finally, when $\alpha = -1$, we have

$$
B_1^{(-1)} \xi = [3]_q^{1/2} \sigma ((\phi_{k-1,R})^* \otimes 1_2)(\xi \otimes t_2)
\begin{align*}
&= [3]_q^{1/2} \left( \frac{[3]_q[2k-1]_q}{[2k+1]_q} \right)^{1/2} \sigma (p_{2k-2} (1_{2k-2} \otimes t_2^* \otimes 1_2)(p_{2k} \otimes p_2) \otimes 1_2)(\xi \otimes t_2) \\
&= [3]_q^{1/2} \left( \frac{[3]_q[2k-1]_q}{[2k+1]_q} \right)^{1/2} \sigma (p_{2k-2} (1_{2k-2} \otimes t_2^* \otimes 1_2)(1_{2k} \otimes t_2)(\xi \otimes t_2) \\
&= [3]_q^{1/2} \left( \frac{[3]_q[2k-1]_q}{[2k+1]_q} \right)^{1/2} \sigma (p_{2k-2} \otimes 1_2)(1_{2k-2} \otimes t_2^* \otimes 1_2)(1_{2k} \otimes t_2)(\xi \otimes t_2) \\
&= \left( \frac{[2k-1]_q}{[2k+1]_q} \right)^{1/2} \sigma (p_{2k-2} \otimes p_2) \xi \sigma^* (\phi_{k,R}^{(-1)} \xi) \\
&= \left( \frac{[2k-1]_q}{[2k+1]_q} \right)^{1/2} \sigma (1_2 \otimes 1_2)(1_{2k} \otimes t_2),
\end{align*}
$$

(using $[3]_q^{-1} = (t_2^* \otimes 1_2)(1_2 \otimes t_2)$).
Quantum Automorphism Groups

and

\[
\left( \frac{[3]^q[2k-1]^q}{[2k+1]^q} \right)^{-1/2} B_2^{(-1)} (1 \otimes \eta^*) t_{2k} = (1_{2k-2} \otimes \eta^* \otimes 1_2)(1_{2k-2} \otimes (p_{2k} \otimes p_2)(1_{2k-2} \otimes t_2)p_{2k-2})t_{2k-2}
\]

\[
= (p_{2k-2} \otimes \eta^* \otimes p_2)(1_{2k-2} \otimes 1_{2k-2} \otimes t_2)t_{2k-2}
\]

\[
= \sigma^*(p_2 \otimes p_{2k-2} \otimes \eta^*)(1_2 \otimes t_{2k-2} \otimes 1_2)t_2
\]

\[
= \left( \frac{[3]^q[2k-1]^q}{[2k+1]^q} \right)^{-1/2} \sigma^*(p_2 \otimes p_{2k-2} \otimes \eta^*) t_{2k}
\]

\[
= \left( \frac{[3]^q[2k-1]^q}{[2k+1]^q} \right)^{-1/2} \sigma^* \phi_{k,L}^{(-1)} (1_{2k} \otimes \eta^*) t_{2k}.
\]

The theorem now follows from these identities. \qed

References


Documenta Mathematica 18 (2013) 1349–1402


Quantum Automorphism Groups


Michael Brannan  
Department of Mathematics  
University of Illinois  
Urbana, IL 61801  
USA  
mbrannan@illinois.edu
UNITARY CYCLES ON SHIMURA CURVES
AND THE SHIMURA LIFT I

SIDDARTH SANKARAN

Received: April 4, 2013
Communicated by Peter Schneider

Abstract. This paper concerns two families of divisors, which we call the ‘orthogonal’ and ‘unitary’ special cycles, defined on integral models of Shimura curves. The orthogonal family was studied extensively by Kudla-Rapoport-Yang, who showed that they are closely related to the Fourier coefficients of modular forms of weight $3/2$, while the unitary divisors are analogues of cycles appearing in more recent work of Kudla-Rapoport on unitary Shimura varieties. Our main result relates these two families by (a formal version of) the Shimura lift.


Contents

1 Introduction 1404
2 Shimura curves and global special cycles 1409
3 Local cycles on the Drinfeld upper half-plane 1415
  3.1 Drinfeld space 1416
  3.2 The structure of a local unitary cycle 1420
  3.3 Relation to local orthogonal special cycles 1436
  3.4 $p$-adic uniformizations 1440
4 Proof of the main theorem 1444
  4.1 The (formal) Shimura lift 1444
  4.2 The Shimura lift formula in fibres of bad reduction 1445
  4.3 The main theorem and applications 1459

1403
1 Introduction

In a series of works leading up to the monograph [KRY], Kudla, Rapoport and Yang study a family of arithmetic divisors that lie in the first arithmetic Chow groups of integral models of Shimura curves. One of their main results states that if one assembles these divisors into a formal generating series, the result is modular of weight 3/2: in effect, pairing this generating series with a suitable linear functional yields the $q$-expansion of a weight 3/2 modular form.

In this paper, together with its forthcoming sequel, we take up the study of the Shimura lift of this generating series. Our main result relates the Shimura lift to a generating series comprised of ‘unitary’ divisors, that are analogues of the cycles constructed by Kudla-Rapoport in their more recent work [KR3] on unitary Shimura varieties.

In the present work, we focus on the geometry of the two families of divisors, and prove that their members satisfy relations that mirror the relations between the Fourier coefficients of a holomorphic modular form and those of its Shimura lift; as we discuss in more detail below, the key step is determining the local structure of the cycles in a formal neighbourhood of a prime of bad reduction.

Let $B$ be a rational indefinite quaternion algebra with discriminant $D_B$, and fix a maximal order $\mathcal{O}_B$. The Shimura curve $C_B$ is the moduli stack over $\text{Spec}(\mathbb{Z})$ that parametrizes pairs $A = (A, \iota)$; here $A$ is an abelian scheme over some base scheme $S$, equipped with an $\mathcal{O}_B$-action $\iota: \mathcal{O}_B \to \text{End}(A)$ that satisfies a determinant condition, cf. Definition 2.1. This stack is then an integral model for the classical Shimura curve associated to $B$.

For a positive integer $n$, Kudla, Rapoport and Yang define (more or less, see Definition 2.3) an ‘orthogonal’ special cycle $Z^o(n)$ as the moduli space of diagrams

$$\xi: A \to A,$$

where $A$ is a $\mathcal{O}_B$-abelian surface (i.e. a point of $C_B$), and $\xi$ is a traceless $\mathcal{O}_B$-linear endomorphism of $A$ such that $\xi^2 = -n$. We may view $Z^o(n)$ as a cycle on $C_B$ via the natural forgetful map, and form the orthogonal generating series (which, to emphasize the connection to modular forms, we have suggestively written as a $q$-expansion in terms of the variable $\tau \in \mathfrak{s}$):

$$\Phi^o(\tau) := Z^o(0) + \sum_{n>0} Z^o(n) q^n \in \text{Div}(C_B)[[q_\tau]].$$

Here $\text{Div}(C_B)$ is the group of divisors on $C_B$, i.e., the free abelian group generated by the closed irreducible substacks of $C_B$ that are étale-locally Cartier, and $Z^o(0)$ is an appropriate constant term, cf. (4.23).

The ‘unitary’ cycles are also defined by a moduli problem. Let $k = \mathbb{Q}(\sqrt{\Delta})$ be an imaginary quadratic field, where $\Delta < 0$ is squarefree, and denote its ring of integers by $\mathcal{O}_k$. Throughout this paper, we assume that (i) $\Delta$ is even and (ii)
every prime $p | D_B$ is inert in $k$. The second condition implies the existence of an embedding $\phi: o_k \to \mathcal{O}_B$, and we fix one such embedding for the moment. Consider the moduli stack $\mathcal{E}^+$ over $\text{Spec}(o_k)$ that parametrizes tuples

$E = (E, i_E, \lambda_E)$.

Here $E$ is an elliptic curve over some base scheme $S$ over $\text{Spec}(o_k)$, endowed with an $o_k$-action $i_E: o_k \to \text{End}(E)$ and a compatible principal polarization $\lambda_E$. We also assume that on the Lie algebra $\text{Lie}(E)$, the action induced by $i_E$ agrees with the structural morphism $o_k \to \mathcal{O}_S$. Let $\mathcal{E}^-$ denote the moduli space of tuples $E = (E, i_E, \lambda_E)$ as before, except now we insist that on $\text{Lie}(E)$, the action of $i_E$ is given by the conjugate of the structural morphism. Finally we take

$\mathcal{E} = \mathcal{E}^+ \coprod \mathcal{E}^-$

to be the disjoint union of these two stacks over $\text{Spec}(o_k)$.

Fix a base scheme $S$ over $o_k$, and suppose we are given a pair of $S$-points $E \in E(S)$ and $A = (A, \iota) \in \mathcal{C}_B(S)$. Following [KR3], we form the space of special homomorphisms

$\text{Hom}_{o_k, \phi}(E, A) := \{ y \in \text{Hom}_S(E, A) \mid y \circ i_E(a) = \iota(\phi(a)) \circ y \text{ for all } a \in o_k \}$,

which is equipped with an $o_k$-hermitian form $h^\phi_{E,A}$, cf. (2.2).

For an integer $m > 0$, the unitary special cycle $Z(m, \phi)$ is the moduli stack over $\text{Spec}(o_k)$ that parametrizes tuples $(E, A, y)$, where $E$ and $A$ are as above and

$y \in \text{Hom}_{o_k, \phi}(E, A)$

such that $h^\phi_{E,A}(y, y) = m$. (1.1)

Again, we may view $Z(m, \phi)$ as a cycle on $\mathcal{C}_B/o_k := \mathcal{C}_B \times \text{Spec}(o_k)$ via the natural forgetful map, and in fact it turns out that $Z(m, \phi)$ is a divisor.

We also define the following rescaled version:

$Z^*(m, \phi) := Z\left(\frac{m}{\text{gcd}(m, D_B)}, \phi\right)$

Finally, we form the unitary generating series:

$\Phi^u(\tau) := Z(0) + \frac{1}{2h(k)} \sum_{m > 0} \left( \sum_{[\phi] \in \text{Opt}/\mathcal{O}_g^o} Z(m, \phi) + Z^*(m, \phi) \right) q^m$

$\in \text{Div}(\mathcal{C}_B/o_k) \otimes \mathbb{Z} \mathbb{Q}$

for an appropriate constant term $Z(0)$, cf. (4.24), and the sum on $[\phi]$ runs over an equivalence class of optimal embeddings (cf. the notation section below).
Our main theorem describes the relationship between $\Phi^{o}$ and $\Phi^{u}$ in terms of the Shimura lift, which is a classical operation on modular forms. It takes as its input a modular form $F$ of half-integral weight together with a squarefree integer parameter $t$, and yields a modular form $Sh(F)$ of even integral weight. Moreover, when $F = \sum a(n)q^n$ is holomorphic, there are explicit formulas (depending on $t$ together with the weight, level, and character) for the Fourier coefficients of $Sh(F)$ in terms of those Fourier coefficients of $F$ that are of the form $a(tm^2)$. We define the formal Shimura lift to be the operator on formal power series determined by these formulas, cf. §4.1.

**Main Theorem** (see Theorem 4.10). Suppose $k = \mathbb{Q}(\sqrt{\Delta})$, where $\Delta < 0$ is a squarefree even integer and assume further that every prime dividing $D_B$ is inert in $k$. Let $\Phi^{o}_{/o_k}$ denote the generating series

$$\Phi^{o}_{/o_k}(\tau) = Z^o(0)/o_k + \sum_{n>0} Z^o(n)/o_k q^{n}_e \in Div(C_B/o_k)[[q_e]]$$

obtained by taking the base change to $o_k$ of each coefficient. Then we have an equality of formal generating series

$$Sh(\Phi^{o}_{/o_k})(\tau) = \Phi^{u}(\tau),$$

in $Div(C_B/o_k)[[q_e]] \otimes \mathbb{Q}$.

We now give an outline of the proof. It turns out that both the orthogonal and unitary divisors have vertical components only at primes $p | D_B$. This leads us to study the $p$-adic uniformization of the Shimura curve $C_B$, which relates the formal completion along its fibre at such a prime $p$ to (a formal model of) the Drinfeld $p$-adic upper half-plane $\mathcal{D}$. We also have $p$-adic uniformizations for the orthogonal and unitary special cycles, which are expressed in terms of linear combinations of analogous ‘local’ cycles defined on $\mathcal{D}$.

Let $\mathbb{F} = \mathbb{F}_p^{alg}$ be an algebraic closure of $\mathbb{F}_p$, and let $W = W(\mathbb{F})$ the ring of Witt vectors. Denote by $\textbf{Nilp}$ the category of $W$-schemes such that $p$ is locally nilpotent, and for a scheme $S \in \textbf{Nilp}$, let $\mathcal{S} := S \times_W \mathbb{F}$. Then the Drinfeld upper half-plane $\mathcal{D}$ parametrizes tuples $X = (X, \iota_X, \rho_X)$, where $X$ is a $p$-divisible group of height 4 and dimension 2, over a base scheme $S \in \textbf{Nilp}$, together with a ‘special’ action $\iota_X : O_{B,p} \to \text{End}(X)$ of the maximal order $O_{B,p}$ (cf. §3.1); finally $\rho_X$ is an $O_{B,p}$-linear quasi-isogeny

$$\rho_X : X \times_S \mathcal{S} \to \mathbb{X} \times_{\mathbb{F}} \mathcal{S}$$

of height 0, where $\mathbb{X}$ is some fixed $p$-divisible group over $\mathbb{F}$ endowed with a special $O_{B,p}$-action. In this picture, the local analogues of the orthogonal and unitary cycles are described as the deformation loci of homomorphisms of $p$-divisible groups. A complete description of the orthogonal cycles can be found in [KR1], and so our aim in Section 3 is provide the same for the unitary cycles. Recent work of Kudla and Rapoport [KR4] describes the special fibre of $\mathcal{D}$ in terms of the Bruhat-Tits tree for $SU(C)$, where $C$ is the split 2-dimensional...
hermitian space over \( k_p \) (recall that by assumption, \( k_p \) is an unramified quadratic extension of \( \mathbb{Q}_p \)). Combining this description with a healthy dose of Grothendieck-Messing theory, we are able to write down explicit equations for a local unitary cycle as a closed formal subscheme of \( \mathcal{D} \), and we consequently obtain a precise description of its irreducible components. This description parallels the one found in [KR1] for the orthogonal cycles, and by comparing the two formulas, we obtain the following key result (Theorems 3.17 and 3.19): any local orthogonal cycle that appears in the \( p \)-adic uniformization of an orthogonal cycle \( \mathcal{Z}^\circ(\Delta|n^2) \) can be expressed as sum of two (explicitly determined) local unitary cycles.

We now describe the \( p \)-adic uniformizations. If we fix an \( \mathbb{F} \)-valued point \( \mathbf{A} = (\mathbf{A}, \iota) \in \mathcal{C}_B(\mathbb{F}) \), then the space of \( \mathcal{O}_B \)-linear quasi-endomorphisms

\[
\text{End}_{\mathcal{O}_B}(\mathbf{A})_{\mathbb{Q}} = B'
\]

is a definite quaternion algebra over \( \mathbb{Q} \) with discriminant \( D_B/p \). The \( p \)-adic uniformization of the Shimura curve \( \mathcal{C}_B \) can be expressed in the following way: there is a finite subgroup \( \Gamma' \subset (B')^\times \) acting on \( \mathcal{D} \), such that if we let \( \tilde{\mathcal{C}}_B \) denote the base change to \( W \) of the formal completion of \( \mathcal{C}_B \) along its fibre at \( p \), then there is an isomorphism

\[
\tilde{\mathcal{C}}_B \simeq \left[ \Gamma' \setminus \mathcal{D} \right]
\]

of formal stacks\(^1\) over \( W \).

Similarly, let \( \tilde{\mathcal{Z}}^\circ(n) \) and \( \tilde{\mathcal{Z}}(m, \phi) \) denote the base change to \( W \) of the formal completions of the special cycles \( \mathcal{Z}^\circ(n) \) and \( \mathcal{Z}(m, \phi) \) along their fibres at \( p \). Viewed as a cycle on \( \tilde{\mathcal{C}}_B \), we may express an orthogonal cycle as a sum

\[
\tilde{\mathcal{Z}}^\circ(n) = \sum_{\xi \in \Omega^\circ(n)_{mod \Gamma'}} \left[ Z^\circ(\xi[p^\infty]) \right] + \sum_{\xi' \in \Omega^\circ(n)_{mod \Gamma'}} \left[ Z^\circ(\xi'[p^\infty]) \right]
\]

where \( [Z^\circ(\xi[p^\infty])] \) is the projection to \( [\Gamma' \setminus \mathcal{D}] \) of a local orthogonal cycle on \( \mathcal{D} \), and

\[
\Omega^\circ(n) = \{ b' \in (B')^\times \mid \text{Tr}(b') = 0 \}
\]

is a \( \Gamma' \)-invariant set of vectors of reduced norm \( n \) satisfying a certain integrality property, cf. Theorem 3.21.

For the unitary cycles, we fix a triple \( \mathbf{E} \in \mathcal{E}(\mathbb{F}) \). A unitary cycle then decomposes as

\[
\tilde{\mathcal{Z}}(m, \phi) = \frac{1}{|k|} \sum_{[\alpha] \in \mathcal{C}(k)} \left( \sum_{\beta \in \Omega^+(m, \alpha, \phi)_{mod \Gamma'}} [Z(\beta[p^\infty])] + \sum_{\beta' \in \Omega^-(m, \alpha, \phi)_{mod \Gamma'}} [Z(\beta'[p^\infty])] \right),
\]

\(^1\)Here and throughout this paper, we use the term ‘isomorphism of formal stacks’ in a rather cavalier fashion. What we mean in all cases is the following: upon fixing sufficiently deep level structure away from \( p \), one obtains a formal scheme on each side, together with a covering map from the corresponding stack. Our assertion is that there is an isomorphism of formal schemes compatible with the automorphisms of the covering maps, and the morphisms induced by varying the level structure, in the natural way.
where $\text{Cl}(k)$ is the class group of $k$, and as before $[\mathbb{Z}(\beta|p^{\infty})]$ and $[\mathbb{Z}(\beta'|p^{\infty})]$ are local unitary cycles. The sets $\Omega^\pm(m, a, \phi)$ appearing above are subsets

$$\Omega^\pm(m, a, \phi) \subset \text{Hom}(E, A) \otimes_{\mathbb{Z}} \mathbb{Q},$$

consisting of quasi-morphisms of a specified norm, a linearity (or anti-linearity, in the case of $\Omega^-$) condition with respect to the action of $\mathfrak{o}_k$, and again satisfying an integrality condition, cf. Theorem 3.22.

Thus, in order to compare the unitary and orthogonal cycles, we need to compare the indexing sets $\Omega^o(n)$ and $\Omega^\pm(m, a, \phi)$, at least in the case that the squarefree part of $n$ is equal to $|\Delta|$, and as $\phi$ varies among classes of optimal embeddings. This task, which amounts to a study in the arithmetic of quaternion algebras, is carried out in §4.2. Together with the description of the local cycles discussed previously, we arrive at a relationship between the two families of cycles, in a formal neighbourhood at $p|D_B$, that matches exactly the formula for the Fourier coefficients of the Shimura lift. Since the vertical components of the cycles only occur at such primes, the main theorem follows immediately, cf. Theorem 4.10.

I would like to conclude the introduction by placing this result in the context of the sequel [San2] to the present work, whose aim is to establish the same Shimura lift formula in the first arithmetic Chow group of $\mathcal{C}_B$ (in the sense of Gillet-Soulé). Recall that in [KRY], the authors prove that the generating series

$$\hat{\Phi}^o(\tau) := \sum_{n \in \mathbb{Z}} \mathbb{Z}^o(n, v) q^n, \quad \Im(\tau) = v$$

is a non-holomorphic modular form of weight 3/2; here the coefficients are arithmetic classes

$$\mathbb{Z}^o(n, v) = (\mathbb{Z}^o(n), Gr^o(n, v)) \in \hat{\mathcal{H}}^1(\mathcal{C}_B)$$

for appropriate choices of Green functions $Gr^o(n, v)$, and constant term $\mathbb{Z}^o(0, v)$. The modularity of $\hat{\Phi}^o(\tau)$ means that in particular, applying a linear functional to its coefficients yields the $q$-expansion of a modular form in the usual sense (in fact, their proof of modularity amounts to showing that this is true for a well-chosen set of linear functionals).

In [San2], we augment the unitary cycles with Green functions $Gr(m, v, \phi)$ and obtain classes

$$\mathbb{Z}(m, v, \phi) = (\mathbb{Z}(m, \phi), Gr(m, v, \phi)) \in \hat{\mathcal{H}}^1(\mathcal{C}_B)$$

and, together with an appropriate constant term $\hat{\mathbb{Z}}(0, v)$, define

$$\hat{\Phi}^u(\tau) := \hat{\mathbb{Z}}(0, v) + \frac{1}{2h(k)} \sum_{m \in \mathbb{Z}} \sum_{[\phi]} \left[ \hat{\mathbb{Z}}(m, v, \phi) + \hat{\mathbb{Z}}^*(m, v, \phi) \right] q^m.$$

The main result of [San2] identifies this generating series with the Shimura lift of $\hat{\Phi}^o$, by combining the geometric results of the present work with calculations.
that relate the Green functions. In particular, applying a linear functional to both series yields an identity involving the classical Shimura lift of modular forms.

Some linear functionals are ‘geometric’ in nature, i.e., they only depend on the cycles and not the Green functions. When this is the case, the main result of the present work already implies that the Shimura lift relation holds for the corresponding modular forms, and we give a few examples at the end of this paper that will play a role in the sequel.

Acknowledgments: This paper is an extension of the work carried out in my Ph. D. thesis; I am grateful to my advisor S. Kudla for introducing me to the subject, and his consistent encouragement. I would also like to thank B. Howard, M. Rapoport, and the anonymous referee for helpful comments. This work was supported by the SFB/TR 45 ‘Periods, Moduli Spaces, and Arithmetic of Algebraic Varieties’ of the DFG (German Research Foundation).

Notation:

- $B$ is an indefinite quaternion algebra over $\mathbb{Q}$, with discriminant $D_B$, and with reduced trace and norm denoted by $Trd$ and $Nrd$ respectively. We fix a maximal order $O_B \subset B$.
- $k = \mathbb{Q}(\sqrt{\Delta})$ is an imaginary quadratic field, with ring of integers $\mathcal{O}_k$. We denote the non-trivial Galois automorphism by $a \mapsto a'$. We also assume throughout this paper that
  
  (i) $\Delta < 0$ is a squarefree even integer;
  
  (ii) and every prime $p | D_B$ is inert in $k$.

In particular, $D_B$ is odd.

- By definition, an embedding $\phi : \mathcal{O}_k \hookrightarrow \mathcal{O}_B$ is optimal if

\[ \phi(\mathcal{O}_k) = \phi(k) \cap \mathcal{O}_B. \]  

(1.3)

Let $Opt$ denote the set of optimal embeddings. Note that $\mathcal{O}_B^\times$ acts on $Opt$ by conjugation: for $\xi \in \mathcal{O}_B^\times$, set

\[ (\xi \cdot \phi)(a) := (\text{Ad}_\xi \circ \phi)(a) = \xi \cdot \phi(a) \cdot \xi^{-1}, \quad a \in \mathcal{O}_k. \]

Denote by $Opt/\mathcal{O}_B^\times$ the set of equivalence classes of optimal embeddings under this action.

- $\chi_k$ is the quadratic character associated to $k$, so that for a prime $p$,

\[ \chi_k(p) = \begin{cases} 
1, & \text{if } p \text{ splits}, \\
0, & \text{if } p \text{ is ramified}, \\
-1, & \text{if } p \text{ is inert}.
\end{cases} \]  

(1.4)

2 Shimura curves and global special cycles

In this section, we recall the construction of Shimura curves, and the global orthogonal and unitary special cycles on them.
**Definition 2.1 (Shimura curve).** Let $\mathcal{C}_B$ denote the moduli problem which associates to a scheme $S$ over $\text{Spec}(\mathbb{Z})$ the category whose objects are pairs

$$\mathcal{C}_B(S) = \{(A,\iota)\},$$

where (i) $A$ is an abelian surface over $S$, and (ii) $\iota: O_B \to \text{End}_S(A)$ is an action of $O_B$ on $A$. We also require that for every $b \in O_B$,

$$\det(T - \iota(b)|_{\text{Lie}(A)}) = T^2 - \text{Trd}(b)T + \text{Nrd}(b) \in O_S[T].$$

**Proposition 2.2 (cf. [KRY], Proposition 3.1.1).** The moduli problem $\mathcal{C}_B$ is representable by a Deligne-Mumford (DM) stack, which we also denote by $\mathcal{C}_B$. It is regular, proper and flat over $\text{Spec}(\mathbb{Z})$ of relative dimension 1, and smooth over $\text{Spec}(\mathbb{Z}[D_B^{-1}]).$

The orthogonal special cycles, as constructed in e.g. [KRY], are also defined by a moduli problem:

**Definition 2.3.** For $n \in \mathbb{Z}_{>0}$, let $\mathcal{Z}_\circ(n)^\sharp$ denote the DM stack which represents the following moduli problem over $\text{Spec}(\mathbb{Z})$: to a scheme $S/\mathbb{Z}$, we let $\mathcal{Z}_\circ(n)^\sharp(S)$ be the category of tuples $(A,\iota,\xi)$, where

(i) $(A,\iota) \in \mathcal{C}_B(S)$, and

(ii) $\xi \in \text{End}_{O_B}(A)$ is an $O_B$-linear endomorphism of $A$ with $\text{Tr}(\xi) = 0$ and $\xi^2 = -n$.

We view $\mathcal{Z}_\circ(n)^\sharp$ as a cycle on $C_B$ via the natural forgetful map $\mathcal{Z}_\circ(n)^\sharp \to C_B$, which is finite and unramified by [KRY, §3.4]. Let $\mathcal{Z}_\circ(n)$ denote the Cohen-Macaulayfication of $\mathcal{Z}_\circ(n)^\sharp$, as in p. 55 of loc. cit., so that $\mathcal{Z}_\circ(n)$ is of pure codimension 1 in $C_B$.

We now turn to the unitary special cycles, following [KR3]. Fix once and for all an element $\theta \in O_B$ such that $\theta^2 = -D_B$. Given a point $(A,\iota) \in \mathcal{C}_B(S)$, there exists a unique principal polarization $\lambda_A^0$ on $A$ such that the Rosati involution $\varphi \mapsto \varphi^\dagger$ satisfies

$$\iota(b)^\dagger = (\lambda_A^0)^{-1} \circ \iota(b)^\vee \circ \lambda_A^0 = \iota(\theta^{-1} b^\dagger \theta), \quad \text{for all } b \in O_B,$$

cf. [How, §3.1].

**Lemma 2.4.** Let $\phi: o_k \to O_B$ be an embedding. Then

(i) $\text{Trd}(\theta\phi(\sqrt{\Delta})) \neq 0$;

(ii) Assume, without loss of generality, that $\text{Trd}(\theta\phi(\sqrt{\Delta})) > 0$, by replacing $\theta$ by $-\theta$ if necessary. Then the isogeny

$$\lambda_{A,\phi} := \lambda_A^0 \circ \iota(\theta \phi(\sqrt{\Delta})). \tag{2.1}$$

is a (non-principal) polarization.
Proof. (i) Since $\theta^2$ and $\Delta$ are both negative, the statement follows immediately from the assumption that $B$ is indefinite.

(ii) We need to check that on geometric points, the map $\lambda_{A,\phi}$ is induced by an ample line bundle; therefore we assume $A$ is an abelian variety over an algebraically closed field. The endomorphism $\iota(\theta\phi(\sqrt{\Delta}))$ is symmetric with respect to $\dagger$, and so lies in the image of the map

$$NS(A) \to \text{End}(A)$$

where $NS(A)$ is the Néron-Severi group. An element in this image comes from an ample bundle if and only if it is totally positive, i.e. its characteristic polynomial has positive real roots, cf. [Mum, §21]. The roots of the characteristic polynomial are the same as those of its minimal polynomial, cf. loc. cit. p. 203, and $\iota(\theta\phi(\sqrt{\Delta}))$ has minimal polynomial

$$P(x) = x^2 - \text{Trd}(\theta\phi(\sqrt{\Delta})) \cdot x + \text{Nrd}(\theta\phi(\sqrt{\Delta})).$$

We may fix an isomorphism $B \otimes \mathbb{R} \simeq M_2(\mathbb{R})$ such that $\text{Trd}$ and $\text{Nrd}$ are identified with the trace and determinant respectively, and the positive involution $b \mapsto b^\dagger = \theta^{-1} \cdot b' \cdot \theta$ is identified with the transpose operator. Then $\theta\phi(\sqrt{\Delta})$ is identified with a symmetric matrix $A = A^t$; the conditions $\text{tr}(A) > 0$ and $\det(A) = |\Delta|D_B > 0$ then imply that the roots of $P(x)$ are real and positive.

Note that the Rosati involution $\ast$ associated to the polarization $\lambda_{A,\phi}$ satisfies

$$\iota_A(\phi(a))^\ast = \iota_A(\phi(a')),$$

for all $a \in k$.

Let $\mathcal{E}^+$ denote the moduli stack over $\text{Spec}(k)$ such that for a base scheme $S/k$, the $S$-points parametrize tuples

$$\mathcal{E}^+(S) = \{E = (E, i_E, \lambda_E)\};$$

here $E$ is an elliptic curve over $S$ endowed with an action $i_E: k \to \text{End}(E)$, and a principal polarization $\lambda_E$ such that the induced Rosati involution $\ast$ satisfies

$$i_E(a)^\ast = i_E(a').$$

We further impose the condition that the action of $k$ on $\text{Lie}(E)$ induced by $i_E$ agrees with the action given via the structural morphism $k \to O_S$.

Similarly, we define $\mathcal{E}^-$ to be the moduli space of tuples $E = (E, i_E, \lambda_E)$ as above, except we require that the action on $\text{Lie}(E)$ induced by $i_E$ is equal to the conjugate of the structural morphism. Finally, we take

$$\mathcal{E} = \mathcal{E}^+ \coprod \mathcal{E}^-$$
to be the disjoint union of these stacks.

Suppose $S$ is a scheme over $\mathcal{O}_k$, and we are given two points $\mathbf{A} \in \mathcal{C}_B(S)$ and $E \in \mathcal{E}(S)$. We form the space of special homomorphisms:

$$\text{Hom}_\phi(E, \mathbf{A}) := \{ y \in \text{Hom}(E, A) \mid y \circ i_E(a) = \iota_A(\phi(a)) \circ y, \text{ for all } a \in \mathcal{O}_k \}$$

This space comes equipped with an $\mathcal{O}_k$-hermitian form $h^\phi_{E, \mathbf{A}}$ defined by

$$h^\phi_{E, \mathbf{A}}(s, t) := (\lambda_E)^{-1} \circ t \circ \lambda_{A, \phi} \circ s \in \text{End}(E, i_E) \simeq \mathcal{O}_k. \quad (2.2)$$

**Definition 2.5 (Unitary special cycles.)** Suppose $m \in \mathbb{Z}_{>0}$ and $\phi: \mathcal{O}_k \to \mathcal{O}_B$ is an optimal embedding. Let $Z(m, \phi)$ denote the DM stack over $\text{Spec}(\mathcal{O}_k)$ representing the following moduli problem: for a scheme $S/\mathcal{O}_k$, we define $Z(m, \phi)(S)$ to be the category of tuples

$$Z(m, \phi)(S) = \{ (E, \mathbf{A}, y) \}$$

where (i) $E \in \mathcal{E}(S)$, (ii) $\mathbf{A} \in \mathcal{C}_B(S)$, and (iii) $y \in \text{Hom}_{\mathcal{O}_k, \phi}(E, \mathbf{A})$ such that $h^\phi_{E, \mathbf{A}}(y, y) = m$.

By the proof of [KR3, Proposition 2.10], which applies verbatim to the present setting, the forgetful map

$$Z(m, \phi) \to \mathcal{C}_B/\mathcal{O}_k := \mathcal{C}_B \times_{\mathcal{Z}} \text{Spec}(\mathcal{O}_k)$$

is finite and unramified, and so we may view $Z(m, \phi)$ as a cycle on $\mathcal{C}_B/\mathcal{O}_k$; abusing notation, we shall refer to the cycle by the same symbol, and hope that context will suffice to clarify which instance of the notation is intended.

**Proposition 2.6.** The cycle $Z(m, \phi)$ is a divisor on $\mathcal{C}_B/\mathcal{O}_k$, i.e., each irreducible component is of codimension 1. Moreover, it has vertical components in characteristic $p$ if and only if (i) $p$ divides $D_B$ and (ii) $\text{ord}_p m > 0$.

**Proof.** By the complex uniformization [San, Theorem 3.3.3], the irreducible components of the generic fibre $Z(m, \phi)_k$ are 0-dimensional, and so their closures in $\mathcal{C}_B/\mathcal{O}_k$ (i.e. the horizontal components) are 1-dimensional.

Suppose $p \mid D_B$. In Section 3.2 below, we show that $Z(m, \phi)$ is a divisor in a formal neighbourhood of the fibre at $p$ by writing down explicit equations. In particular, Theorem 3.14 asserts that $Z(m, \phi)$ contains vertical irreducible components in this fibre if and only if $\text{ord}_p (m) > 0$.

It remains to consider the fibre at primes $p \subset \mathcal{O}_k$ with $(p, D_B) = 1$. Suppose $z \in Z(m, \phi)(\mathbb{F})$ is a geometric point corresponding to a tuple $(E, \mathbf{A}, y)$ over $\mathbb{F}$, for an algebraic closure $\mathbb{F}$ of $\mathcal{O}_k/p$. Let $x \in \mathcal{C}_B/\mathcal{O}_k(\mathbb{F})$ denote the point below $z$. We need to prove that (i) the cycle $Z(m, \phi)$ is a divisor at $x$ and (ii) there are no vertical components, i.e., that the data $(E, \mathbf{A}, y)$ can be lifted to characteristic zero and $Z(m, \phi)$ does not contain the entire fibre at $p$. 

Documenta Mathematica 18 (2013) 1403–1464
Let $W$ be the completion of the maximal unramified extension of the local ring $o_{k,p}$, and denote by $\mathbf{ART}$ the category of local Artinian $W$-algebras with residue field $F$. The natural morphism

$$o_{k,p} \rightarrow W$$

endows any $W$-algebra with an $o_{k,p}$ structure, and so in particular every object of $\mathbf{ART}$ can be viewed as an $o_k$-algebra in this way.

Let $R_z$ and $R_x$ denote the étale local rings of $z$ and $x$ respectively. Then $R_x$ pro-represents the functor of deformations of the data $A = (A, \iota_A)$ to objects of $\mathbf{ART}$. By the Serre-Tate theorem, giving a deformation of $A$ is equivalent to giving a deformation of the underlying $p$-divisible group (with induced $O_B \otimes \mathbb{Z}_p$-action):

$$X = (X, \iota_X) := (A[p^\infty], \iota_A \otimes \mathbb{Z}_p).$$

Similarly, the local ring $R_z$ pro-represents deformations of the data $(E, A, y)$, which in turn are equivalent to deformations of $(Y, X, b) := (E[p^\infty], A[p^\infty], y[p^\infty])$.

Recalling that $p \nmid D_B$, we may fix an isomorphism $O_B \otimes \mathbb{Z}_p \cong M_2(\mathbb{Z}_p)$ such that $\phi(\sqrt{\Delta})$ is identified with the matrix $(\Delta^{-1})$; note that when $p = 2$, this is possible on account of the assumption that $|\Delta|$ is even. The two idempotents

$$e_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad e_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

determine a splitting $X = X^o \times X^o$, where $X^o$ is a $p$-divisible group of dimension 1 and height 2; keep in mind that $X^o$ itself does not inherit any additional endomorphism structure. In a similar manner, any deformation of $X = (X, \iota_X)$ decomposes into a product of two copies of a deformation of $X^o$. Hence, $\text{Def}(X, \iota_X) = \text{Def}(X^o)$, and it is a well-known fact, cf. e.g. [MZ, Theorem 3.8], that the latter deformation problem is pro-represented by $\text{Spf}(W[t])$.

Our aim is to analyze the deformation locus $\text{Def}(Y, X, b)$ as a closed formal subscheme of $\text{Def}(X) = \text{Def}(X^o) = \text{Spf}(W[t])$. Under the splitting $X = X^o \times X^o$, we have a decomposition $b = (b_1, b_2)$, where $b_i \in \text{Hom}(Y, X^o)$ are non-zero morphisms. Moreover, because of (2.3),

$$b_2 = b_1 \circ iv(\sqrt{\Delta}).$$
Thus, the problem of deforming \((Y, X, b)\) is equivalent to deforming \((Y, X^o, b_1)\).

The key tool is the theory of \textit{canonical lifts}: given a \(p\)-divisible group \(\mathfrak{g}\) over \(\mathbb{F}\) of height 2 and dimension 1 together with an action

\[
i_{\mathfrak{g}} : o_{k,p} \to \text{End}(\mathfrak{g}),
\]

there exists a \textit{canonical lift} \(\mathfrak{G} = (\mathfrak{G}, i_{\mathfrak{g}})\) over \(W\). This fact is a consequence of the existence of Serre-Tate canonical coordinates in the case when \(\mathfrak{g}\) is ordinary, cf. [Mes, Appendix], and of Gross’ theory of (quasi-)canonical liftings when \(\mathfrak{g}\) is supersingular, cf. [Gro]. Moreover, for any object \(R \in \text{ART}\), there exists a unique lift of \(\mathfrak{g}\) to \(R\): namely, the base change \(\mathfrak{G}_R\) of \(\mathfrak{G}\) to \(R\).

In particular, if we let \(\mathfrak{G}^0\) be the canonical lift of \(Y\) as above, then the problem of deforming \((Y, X^o, b_1)\) to \(R \in \text{ART}\) is equivalent to finding deformations \(X^o\) of \(X^o\) such that \(b_1\) lifts to a morphism \(\mathfrak{G}^0_R \to X^o\). It is not hard to show that the locus in \(\text{Def}(X^o)\) where \(b_1\) lifts is cut out by a single equation, cf. the proof of [Wew, Prop. 5.1]. Note that in our case, this equation cannot be 0. If it were, then \(R_x = R_z\), and so \(Z(m, \phi)\) would contain a positive-dimensional component in the generic fibre; this would contradict the description of the complex points above.

Moreover, we claim \(Z(m, \phi)\) cannot contain the entire fibre \((C_B)_p\). Note that if \(X^o\) and \(Y\) are isogenous, they are either both super-singular or both ordinary. But \(Y\) is necessarily ordinary when \(p\) is a split prime, and supersingular otherwise, as it admits an action of \(o_k \otimes \mathbb{Z}_p\); thus taking \(X^o\) to be supersingular in the first case, or ordinary in the second, yields a point in the fibre \((C_B)_p\) that cannot lie on \(Z(m, \phi)\).

It remains to show that there always exists a lift to characteristic 0. Suppose first that \(X^o\) and \(Y\) are ordinary. Then \(p\) splits in \(k\) and \(\text{End}(X^o) = \text{End}(Y) = o_{k,p}\). Fixing an isomorphism \(X^o \simeq Y\), we have

\[
b_1 \in \text{Hom}(Y, X^o) \simeq \text{End}(X^o) \simeq o_{k,p},
\]

which lifts to an endomorphism of the canonical lift \(X^o\) of \(X^o\).

Next, we consider the case where \(X^o\) and \(Y\) are supersingular; this necessarily implies that \(p\) is non-split in \(k\). Let \(D\) be the division quaternion algebra over \(\mathbb{Q}_p\) with maximal order denoted by \(O_D\), and fix a uniformizer \(\Pi \in O_D\).

Suppose first that \(p = p^2\) is ramified in \(k\), and let \(\varpi \in o_{k,p}\) be a uniformizer. Then we may identify

\[
\text{End}(Y) \simeq O_D
\]

such that \(i_Y(\varpi)\) is identified with \(\Pi\). We may also fix an isomorphism

\[
X^o \simeq Y, \quad (2.4)
\]

which induces isomorphisms

\[
\text{End}(X^o) \simeq \text{Hom}(Y, X^o) \simeq \text{End}(Y) \simeq O_D
\]
in the natural way. Without loss of generality, we may normalize (2.4) such that $b_1$ is identified with $\Pi \in O_D$ for some $t$. We let $i_{X^o}: o_{k,p} \to \text{End}(X^o)$ denote the action given by composing $i_Y$ with the isomorphism (2.4). Appealing again to Gross’ results, this action determines a canonical lifting $X^o = (X^o, i_{X^o})$ of $(X, i_X)$ to $W$, and it is clear by construction that $b_1$ also lifts.

Finally, we suppose that $p = p$ is inert in $k$. We may fix an identification $\text{End}(Y) \cong O_D$ such that $\Pi \cdot i_Y(a) = i_Y(a') \cdot \Pi$, for all $a \in o_{k,p}$.

As before, we may also fix an isomorphism $\alpha: X_o \cong Y$ such that $b_1$ is identified with $\Pi \in O_D$ for some $t$. We then define an action $i_{X^o}: o_{k,p} \to \text{End}(X^o)$ by the formula:

$$i_{X^o}(a) := \begin{cases} \alpha^{-1} \circ i_Y(a) \circ \alpha, & \text{if } t \text{ is even,} \\ \alpha^{-1} \circ i_Y(a') \circ \alpha, & \text{if } t \text{ is odd.} \end{cases}$$

By construction, the morphism $b_1$ will lift to a morphism $Y \to X^o$, where $Y$ and $X^o$ are the canonical lifts of $Y$ and $X^o$ to $W$ determined by their respective $o_{k,p}$-actions as above.

3 Local cycles on the Drinfeld upper half-plane

In this section, we discuss the local analogues of the unitary and orthogonal special cycles. Suppose $p|D_B$, so that in particular $p$ is inert in $k$ and $p \neq 2$.

Let $k_p$ denote the completion at $p$, and let $\delta \in k_p$ denote the image of $\sqrt{\Delta}$. Throughout this section, we fix an embedding $\phi: o_{k,p} \hookrightarrow O_{B,p}$, (3.1)

where $o_{k,p}$ and $O_{B,p}$ are the maximal orders in $k_p$ and $B_p$ respectively. We also fix a uniformizer $\Pi \in O_{B,p}$ such that $\Pi \phi(a) = \phi(a') \Pi$, for all $a \in o_{k,p}$.

Let $F = F^\text{alg}_p$ denote a fixed algebraic closure of $F_p$, and $W = W(F)$ the ring of Witt vectors, with Frobenius endomorphism $\sigma: W \to W$. We fix an embedding $\tau_0: o_{k,p}/(p) \to F$, which lifts uniquely to an embedding $\tau_0: o_{k,p} \to W$. In addition, we let $\tau_1$ denote the conjugate embedding (or its lift to characteristic 0). Note that these maps allow us to view $F$ and $W$ as $o_k$-algebras.

Let $\text{Nil}$ denote the category of $W$-schemes for which the ideal sheaf generated by $p$ is locally nilpotent, and for a scheme $\mathcal{S} \in \text{Nil}$, we set

$$\mathcal{S} := S \times_W \text{Spec}(F).$$

Additionally, we fix a trivialization of the prime-to-$p$ roots of unity in $\mathbb{F}$:

$$\mathbb{A}_f^p(1) := \left( \prod_{\ell \neq p} \lim_{\longleftarrow} \mu_{p^n}(\mathbb{F}) \right) \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathbb{A}_f^p.$$
Finally, we set $\hat{\mathbb{Z}}^p := \prod_{\ell \neq p} \mathbb{Z}_\ell$, and if $M$ is any $\mathbb{Z}$-module, we define $\hat{M}^p := M \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}^p$.

### 3.1 Drinfeld space

In this section, we recall the definition of Drinfeld space as a moduli space of $p$-divisible groups, as well as an ‘alternative’ description of its special fibre as a union of projective lines indexed by hermitian lattices, as in [KR4].

Following [BC], we define a special formal $\mathcal{O}_{B,p}$-module over a scheme $S$ to be a pair $(X, \iota_X)$ consisting of a $p$-divisible group $X$ of height 4 and dimension 2 over $S$, and a map

$$\iota_X : \mathcal{O}_{B,p} \to \text{End}(X),$$

which satisfies the following special condition: the Lie algebra $\text{Lie}(X)$ is (locally on $S$) a free $\mathcal{O}_S \otimes W_{\tau_0} \mathbb{O}_{k,p}$ module of rank 1. Here $\text{Lie}(X)$ is viewed as an $\mathcal{O}_S \otimes \mathbb{O}_{k,p}$-module via the action of $\mathbb{O}_{k,p}$ on $X$ induced by the composition $\iota_X \circ \phi$.

Fix once and for all a special formal $\mathcal{O}_{B,p}$-module $(X, \iota_X)$ over $S$; note such a pair is unique up to isogeny, by the classification of Dieudonné isocrystals over $F$ [BC, Proposition II.5.2]. The pair $(X, \iota_X)$ serves as a “base point” for the following moduli problem:

**Definition 3.1.** Let $\mathcal{D}$ denote the moduli problem on $\text{Nilp}$ that associates to a scheme $S$ the category of isomorphism classes of tuples

$$\mathcal{D}(S) = \{(X, \iota_X, \rho_X)\} / \simeq,$$

consisting of

- a special formal $\mathcal{O}_B$-module $(X, \iota_X)$ over $S$,
- and an $\mathcal{O}_{B,p}$-linear quasi-isogeny

$$\rho_X : X \times_S \overline{\mathbb{F}} \to X \times_{\text{Spec}(\overline{\mathbb{F}})} \overline{\mathbb{F}}$$

of height 0.

An isomorphism between two tuples $(X, \iota, \rho)$ and $(X', \iota', \rho')$ is an isomorphism $\alpha : X \to X'$ which is $\mathcal{O}_{B,p}$-equivariant, and such that $\rho = \rho' \circ (\alpha \times_S S_0)$.

The functor $\mathcal{D}$ is then represented by (a formal model of) the Drinfeld upper-half plane, and in particular is a formal scheme over $\text{Spf}(W)$; see [BC] for a discussion of this result.

Next, we recall the ‘alternative’ description of the reduced locus of $\mathcal{D}$ as given in [KR4]. Crucial to this description is the following theorem:

**Theorem 3.2** (Drinfeld, cf. §III.4 of [BC]). Suppose $(X, \iota_X)$ is a special formal $\mathcal{O}_{B,p}$-module, over any base $S \in \text{Nilp}$. Then there exists a principal polarization $\lambda_X^0$ on $X$ such that

$$\lambda_X^0 \circ \iota_X(b) \circ \lambda_X^0 = \iota_X(\Pi b' \Pi^{-1}) \quad \text{for all } b \in \mathcal{O}_{B,p}; \quad (3.2)$$
here the map $b \mapsto b^\dagger$ is the involution of $B$. Moreover, $\lambda_X^0$ is unique up to multiplication by $\mathbb{Z}_p^\times$.

For the base point $X$, we shall fix once and for all a polarization $\lambda_X^0$ as in Drinfeld’s theorem. Then, for any point $(X, \iota_X, \rho_X) \in \mathcal{D}(S)$, there is a unique principal polarization $\lambda_X^0$ satisfying (3.2), and such that the diagram

\[
\begin{array}{c}
X \times_S S \xrightarrow{\lambda_X^0 \times_S S} X^\vee \times_S S \\
\rho_X \downarrow \quad \quad \quad \downarrow \rho_X^\vee \\
X \times_{\mathbb{F}} S \xrightarrow{\lambda_X^0 \times_{\mathbb{F}} S} X^\vee \times_{\mathbb{F}} S
\end{array}
\]

commutes. Thus, for any point $(X, \iota_X, \rho) \in \mathcal{D}(S)$, we may associate another polarization $\lambda_{X,\phi}$ by the formula

\[
\lambda_{X,\phi} := \lambda_X^0 \circ \iota_X (\Pi \phi(\delta)).
\]  

(3.3)

Note that the Rosati involution $\ast$ induced by $\lambda_{X,\phi}$ has the property

\[
\iota_X (\phi(a))^\ast = \iota_X (\phi(\alpha')),
\]  

for all $a \in \mathcal{O}_{k,p}$.

Let $M(X)$ be the Dieudonné module of $X$ over $W = W(\mathbb{F})$, with Frobenius and Verschiebung operators denoted by $F$ and $V$ respectively. We have a decomposition

\[
M(X) = M(X)_0 \oplus M(X)_1,
\]

where

\[
M(X)_i := \{m \in M(X) \mid (\iota_X \circ \phi)(a) \cdot m = \tau_i(a)m, \text{ for all } a \in \mathcal{O}_{k,p}\}.
\]  

(3.4)

Let $N(X) := M(X) \otimes_{\mathbb{Z}} \mathbb{Q}$ be the rational Dieudonné module, with induced decomposition

\[
N(X) = N(X)_0 \oplus N(X)_1.
\]

Note that the operator $pV^{-2} = V^{-1}F: N(X)_0 \to N(X)_0$ is a $\sigma^2$-linear operator, and hence the space of invariants

\[
C := (N(X)_0)^{V^{-1}F=1}
\]  

(3.5)

is a two-dimensional vector space over $k_p$, where $k_p$ acts on $C$ via the embedding $\tau_0: k_p \to W$.

The polarization $\lambda_{X,\phi}$ as defined in (3.3) induces an alternating pairing

\[
\{\cdot, \cdot\}_X : N(X) \times N(X) \to W_{\mathbb{Q}}
\]

such that for all $x, y \in N(X)$, we have

\[
\{Fx, y\}_X = \sigma (\{x, Vy\}_X).
\]
Thus, if we define
\[ h(x, y) := \frac{1}{p^d} \{x, Fy\}_X, \]
(3.6)
it is straightforward to verify that the restriction of \( h \) to \( C \) defines a hermitian form; we denote this restricted form again by \( h \). We shall shortly see that \((C, h)\) is in fact split, cf. Remark 3.4. Let \( q(x) = h(x, x) \) denote the corresponding quadratic form.

**Definition 3.3.** (i) If \( L \) is a \( W \)-lattice in \( N(X)_0 \), we let
\[ L^\sharp := \{ n \in N(X)_0 \mid h(n, L) \subset W \}. \]
Note that \((L^\sharp)^\sharp = \Lambda\). Similarly, if \( \Lambda \subset C \) is an \( \omega_{k,p} \)-lattice, we set
\[ \Lambda^\sharp := \{ v \in C \mid h(v, \Lambda) \subset \omega_{k,p} \}. \]
(ii) Suppose \( \Lambda \subset C \). We say \( \Lambda \) is a vertex lattice of type 0 (resp. type 2) if \( \Lambda^\sharp = \Lambda \) (resp. \( \Lambda^\sharp = \LambdaA \)). We shall use the term “vertex lattice” to mean a vertex lattice of type 0 or 2.
(iii) Let \( \mathcal{B} \) denote the Bruhat-Tits tree for \( SU(C) \), which is a graph with the following description. The vertices are vertex lattices, and edges can only occur between vertex lattices of differing type. Two vertex lattices \( \Lambda \) and \( \Lambda' \) of type 0 and 2 respectively are joined by an edge if and only if
\[ p\LambdaA \subset \Lambda \subset \Lambda', \]
where the successive quotients are \( F_p^2 \) vector spaces of dimension 1. In particular, this graph is a \( p + 1 \)-regular tree.

Suppose \( x = (X, \iota_X, pX) \in \mathcal{D}(F) \). We may use the quasi-isogeny \( \rho_X \) to identify the Dieudonné module \( M(X) \) as a \( W \)-lattice inside of \( N(X) \). Furthermore, as \( \rho_X \) is \( \omega_{k,p} \)-linear, we have \( M(X)_i = M(X) \cap N(X)_i \) for \( i = 0, 1 \), where \( M(X)_i \) is defined in the same way as (3.4). Hence, to any point \( x \), we may associate a chain of \( W \)-lattices \( B \subset A \subset \Lambda \), where
\[ B = M(X)_0, \quad A = (VM(X)_1)^\sharp. \]
By [KR4, Corollary 2.3], we have either \( B^\sharp = B \) or \( A^\sharp = pA \), or both. If both conditions are satisfied, then we say the point \( x \) is superspecial; otherwise, \( x \) is ordinary. We say a point is special if both \( A \) and \( B \) are \( pV^{-2} \)-invariant, so in particular superspecial points are special.

This construction yields a bijection between \( \mathcal{D}(F) \) and pairs of \( W \)-lattices \( B \subset A \subset \Lambda \) such that either \( B^\sharp = B \) or \( A^\sharp = pA \). Moreover, if \( B^\sharp = B \), then \( B = \Lambda \otimes_{\omega_{k,p}} W \) for some vertex lattice \( \Lambda \) of type 0; on the other hand, if \( A^\sharp = pA \), then \( A = \Lambda' \otimes W \) for a vertex lattice \( \Lambda' \) of type 2, cf. [KR4, Corollary 2.3]. Suppose \( \Lambda \) is a vertex lattice of type 0. We may define a map
\[ \mathbb{P}_\Lambda(F) := \mathbb{P}(p^{-1}\Lambda/\Lambda)(F) \to \mathcal{D}(F), \]
by sending a line $\ell \subset (p^{-1}\Lambda/\Lambda) \otimes_{\mathbb{F}_p} \mathbb{F}$ to the pair of lattices $B \subset A$, where $B = \Lambda_W = \Lambda \otimes W$, and $A$ is the inverse image of $\ell$ in $p^{-1}\Lambda_W$.

If $\Lambda'$ is a vertex lattice of type 2, we obtain a map

$$P_{\Lambda'}(\mathbb{F}) := P(\Lambda'/p\Lambda')(\mathbb{F}) \to D(\mathbb{F}),$$

defined by sending a line $\ell' \subset (\Lambda'/p\Lambda') \otimes_{\mathbb{F}_p} \mathbb{F}$ to the pair of lattices $B \subset A$, where $A = \Lambda'_W$, and $B$ is the inverse image of $\ell'$ in $\Lambda'_W$.

Note that if $\Lambda$ and $\Lambda'$ are neighbours in $B$, i.e. if $p\Lambda' \subset \Lambda \subset \Lambda'$, then the lines

$$\ell = \Lambda' \otimes_{\mathbb{O}_k,p}\varpi \mathbb{F} \in P(p^{-1}\Lambda/\Lambda)(\mathbb{F}), \quad \ell' = \Lambda \otimes \mathbb{F} \in P(\Lambda'/p\Lambda')(\mathbb{F})$$

define the same point of $D(\mathbb{F})$; this point is superspecial, and all superspecial points arise in this way.

By [KR4, Proposition 2.4], the above maps are induced by embeddings of schemes over $\mathbb{F}$:

$$P_\Lambda \to D_{\text{red}}, \quad \Lambda \text{ a vertex lattice},$$

where $D_{\text{red}}$ is the underlying reduced subscheme of the formal scheme $D$, and the collection of such maps, as $\Lambda$ varies among the vertex lattices, yield a cover of $D_{\text{red}}$ by projective lines.

Remark 3.4. (i) In [KR1, §1], there is a similar description of the special fibre of $D$, but it is given in terms of homothety classes of $\mathbb{Z}_p$-lattices. These two descriptions are essentially the same. To start, note that the operator

$$\epsilon := \Pi^{-1}_X \circ V : N(X) \to N(X), \quad \Pi_X := t_{\mathbb{F}}(\Pi)$$

is 0-graded and commutes with $F^{-1}V$, and hence restricts to a Galois-semilinear operator $\epsilon : C \to C$.

Without loss of generality, we may assume $X = \mathbb{Y} \times \mathbb{Y}$, where $\mathbb{Y}$ is a supersingular $p$-divisible group of height 2 and dimension 1 over $\mathbb{F}$. Then the Dieudonné module $M(X)$ has a basis $\{e_0, e_1, f_0, f_1\}$ consisting of vectors that are both $F^{-1}V$- and $\epsilon$-invariant, and such that

$$M(X)_0 = W \cdot e_0 \oplus W \cdot f_1, \quad M(X)_1 = W \cdot e_1 \oplus W \cdot f_0,$$

cf. [BC, §III.4.5]. The set $\{e_0, f_1\}$ is an $\epsilon$-invariant $k_p$-basis for $C$ such that $h(e_0, f_1) = \delta$ and $h(e_0, e_0) = h(f_1, f_1) = 0$, and in particular, we see that $C$ is split. Let $\Lambda_0 := \text{span}_{\mathbb{O}_k,p}(e_0, f_1)$, which is a vertex lattice of type 0, i.e. $\Lambda_0 = \Lambda_0$.

Now suppose $\gamma \in \text{End}_{\mathbb{O}_k,p}(X) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$. Then $\gamma$ also acts on $C$, and is described by a matrix of the form

$$[\gamma] = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad a, b, c, d \in \mathbb{Q}_p.$$
with respect to the basis \{e_0, f_1\}. In particular (at least, when \(\gamma\) is invertible), the matrix \([\gamma]\) lies in \(GU(C)\) with \([\gamma]^* = \det(\gamma) \cdot [\gamma]^{-1}\). One can then verify directly that the map \(\gamma \mapsto [\gamma]\) induces an isomorphism

\[
\{ \gamma \in \text{End}_{\mathbb{O}_p}(X) \otimes \mathbb{Q}_p, \ \det(\gamma) = 1 \} \overset{\sim}{\longrightarrow} SU(C).
\]  

Suppose \([L]\) is a homothety class of \(\mathbb{Z}_p\)-lattices in \(M(X)^{0,1}_0 = \mathcal{C}^{e=1} \simeq (\mathbb{Q}_p)^2\), and \(L \in [L]\). Then there exists \(\gamma \in \text{End}_{\mathbb{O}_p}(X)\) \(\mathbb{Q}_p\) such that

\[
\gamma(L) = L_0 := \text{span}_{\mathbb{Q}_p}\{ e_0, f_1 \}.
\]

Set \(\Lambda := L \otimes_{\mathbb{Z}_p} \mathbb{O}_k, p\), so that

\[
\Lambda^\sharp = ([\gamma]^*)^{-1} \cdot \Lambda_0^\sharp = \det(\gamma)^{-1} [\gamma] \cdot \Lambda_0 = \det(\gamma)^{-1} \cdot \Lambda.
\]

Hence, by scaling by a power of \(p\), there is a unique representative \(L \in [L]\) such that \(L \otimes_{\mathbb{O}_k, p} \) is a vertex lattice, whose type depends on the parity of \(\text{ord}_p \det(\gamma)\).

Conversely, suppose \(\Lambda \subset C\) is a vertex lattice. Since \(SU(C)\) acts transitively on the set of lattices of a given type, there exists an element \([\gamma] \in SU(C)\) such that \([\gamma] \cdot \Lambda\) is equal to one of \(\Lambda_0\) or \(\Lambda_0^\sharp := \text{span}_{\mathbb{O}_k, p}(p^{-1} e_0, f_1)\), depending on the type of \(\Lambda\). In either case, by \((3.7)\), the transformation \([\gamma]\) commutes with \(e\), and so \(\Lambda\) admits an \(e\)-invariant basis. Thus the \(\mathbb{Z}_p\)-lattice \(\Lambda^{e=1}\) determines a homothety class of \(\mathbb{Z}_p\)-lattices \([L] := [\Lambda^{e=1}]\).

(ii) Suppose \(x = (X, \iota_X, \rho_X) \in D(\mathbb{F})\), and let \(M = M_0 \oplus M_1\) denote its \(\text{Die}\)donné module, where each \(M_i\) is viewed as a \(W\)-lattice in \(N(X)\). Each lattice \(M_i\) is self-dual with respect to the pairing induced by the polarization \(\Lambda^\sharp_X\) described in \(\text{Theorem 3.2}\), and it is then a straightforward calculation to show

\[
M_0^\sharp = \Pi^{-1} VM_0, \quad \text{and} \quad (VM_1)^\sharp = \Pi^{-1} M_1,
\]

where the \(^\sharp\) denotes the dual with respect to the hermitian form \(h\).

Recall that \(x\) is said to be “0-critical” in the sense of [BC] if and only if \(\Pi M_0 = VM_0\), which is equivalent to the relation \(M_0^\sharp = M_0\), i.e. \(x \in \mathbb{P}_A(\mathbb{F})\) for a vertex lattice \(\Lambda\) of type 0.

Similarly \(x\) is “1-critical” if and only if \(\Pi M_1 = VM_1\), which is equivalent to the relation \((VM_1)^\sharp = p(VM_1)^\sharp\). This last condition is then equivalent to the condition \(x \in \mathbb{P}_A(\mathbb{F})\) for a vertex lattice \(\Lambda'\) of type 2.

In particular, this discussion implies that our use of the terms “ordinary”, “special” and “superspecial” coincides with their use in [KR1, §1].

3.2 The Structure of a Local Unitary Cycle

We begin this section by defining the \((\text{local})\) unitary special cycles, whose construction is due to Kudla and Rapoport [KR2].

To start, fix a triple \(\mathcal{Y} = (Y, \iota_Y, \lambda_Y)\) over \(\mathbb{F}\) consisting of

(i) a (supersingular) \(p\)-divisible group \(\mathcal{Y}\) of dimension 1 and height 2 over \(\mathbb{F}\);
(ii) an action $i_Y: o_{k,p} \to \text{End}(\mathcal{Y})$ such that on $\text{Lie}(\mathcal{Y})$, this action coincides with the action of $o_{k,p}$ via the embedding $\tau_0: o_{k,p}/(p) \to \mathbb{F}$;

(iii) and finally, a principal polarization $\lambda_Y$ such that the induced Rosati involution $^\ast$ satisfies

$$i_Y(a)^\ast = i_Y(a'), \quad \text{for all } a \in o_{k,p}.$$

We define two spaces of special homomorphisms:

$$\mathcal{V}_\phi^+ := \{ b \in \text{Hom}(\mathcal{Y}, \mathcal{X}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \mid b \circ i_Y(a) = i_X(\phi(a)) \circ b \text{ for all } a \in o_{k,p} \}$$

(3.8)

and

$$\mathcal{V}_\phi^- := \{ b \in \text{Hom}(\mathcal{Y}, \mathcal{X}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \mid b \circ i_Y(a) = i_X(\phi(a)) \circ b \text{ for all } a \in o_{k,p} \}.$$  

(3.9)

Using the polarization $\lambda_Y$, and the polarization $\lambda_{X,\phi}$ as defined in (3.3), we may construct natural hermitian forms $h^+$ and $h^-$ on $\mathcal{V}_\phi^+$ and $\mathcal{V}_\phi^-$ respectively; these are defined by the formulas

$$h^+(b_1, b_2) := \lambda_Y^{-1} \circ b_2^\vee \circ \lambda_{X,\phi} \circ b_1 \in \text{End}_{o_k}(\mathcal{Y}^+) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \simeq k_p$$

(3.10)

$$h^-(b_1, b_2) := \lambda_Y^{-1} \circ b_2^\vee \circ \lambda_{X,\phi} \circ b_1 \in \text{End}_{o_k}(\mathcal{Y}^-) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \simeq k_p. $$

(3.11)

Let $q^\pm(b) := h^\pm(b, b)$ denote the corresponding quadratic forms.

Our next step is to relate the spaces $\mathcal{V}_\phi^\pm$ to the hermitian space $(C, h)$. Let $M(\mathcal{Y})$ denote the Dieudonné module over $W = W(\mathbb{F})$ attached to $\mathcal{Y}$; this is a free $W$-module of rank 2. As before, we have a grading $M(\mathcal{Y}) = M(\mathcal{Y})_0 \oplus M(\mathcal{Y})_1$, where

$$M(\mathcal{Y})_i := \{ m \in M(\mathcal{Y}) \mid i_Y(a) \cdot m = \tau_i(a)m, \text{ for all } a \in o_{k,p} \}.$$  

Moreover, we may choose generators $f_0$ and $f_1$ for $M(\mathcal{Y})_0$ and $M(\mathcal{Y})_1$ respectively, such that $Vf_0 = f_1, Vf_1 = pf_0$, and that the alternating form $\{\cdot, \cdot\}_Y$ defined by the polarization $\lambda_Y$ satisfies

$$\{f_0, f_1\}_Y = \delta;$$

(3.12)

see [KR2, Remark 2.5].

Suppose that $b \in \mathcal{V}_\phi^+$. Abusing notation, we denote the corresponding map on (rational) Dieudonné modules again by

$$b: N(\mathcal{Y}) = M(\mathcal{Y}) \otimes \mathbb{Q} \to N(\mathcal{X}).$$

Let $b := b(f_0)$. Then since $b$ is $o_{k,p}$-linear, we have that $b \in N(\mathcal{X})_0$, and furthermore,

$$V^{-1}Fb = pV^{-2}b = pb(V^{-2}f_0) = b,$$

which implies $b \in C$. Finally, we note that

$$b(f_1) = b(Vf_0) = Vb(f_0) = Vb,$$

Documenta Mathematica 18 (2013) 1403–1464
and so \( \mathbf{b} \) is determined by \( b \). We therefore obtain an isomorphism

\[
\varphi^+ : \mathcal{V}_S^+ \to C, \quad b \mapsto b := \mathbf{b}(f_0).
\]

(3.13)

One readily checks that \( q(\varphi^+ b) = p^{-1} q^+(b) \), where \( q \) and \( q^+ \) are the quadratic forms on \( C \) and \( \mathcal{V}_S^+ \) defined by (3.6) and (3.10) respectively. In a similar manner, if \( \mathbf{b} \in \mathcal{V}_S^- \), then \( b := \mathbf{b}(f_1) \in C \), and \( b \) is again determined by \( b \). Hence we obtain an isomorphism

\[
\varphi^- : \mathcal{V}_S^- \to C, \quad b \mapsto b := \mathbf{b}(f_1)
\]

(3.14)

such that \( q(\varphi^- b) = q^-(b) \).

Finally, we come to the definition of the unitary special cycles, as in [KR2]. Note that the \( p \)-divisible group \( \mathcal{V} = (\mathbb{Y}, i_Y) \), together with its \( o_{k,p} \)-action, admits a canonical lift \( \mathcal{V}_W = (\mathbb{Y}_W, i_{Y_W}) \) to \( W \). For any \( S \in \text{Nilp} \), we let \( \mathcal{V}_S = (\mathbb{Y}_S, i_{Y_S}) \) denote the base change \( \mathcal{V}_S = \mathbb{Y}_W \times_W S \), with induced \( o_{k,p} \)-action.

**Definition 3.5.** For \( \mathbf{b} \in \mathcal{V}_S^+ \), we define the local unitary special cycle \( Z(b) \) by the following moduli problem: for \( S \in \text{Nilp} \), let \( Z(b)(S) \) denote the set of points \((X, i_X, \rho) \in \mathcal{D}(S) \) such that quasi-isogeny

\[
\rho^{-1} \circ \mathbf{b} : Y \times F \mathcal{S} \to X \times_S \mathcal{S}.
\]

lifts to a morphism of \( p \)-divisible groups \( \mathcal{V}_S \to X \).

**Remark 3.6.** (i) If such a lift exists, then it is unique, by rigidity for \( p \)-divisible groups. In particular, if the lift exists, then it is \( o_{k,p} \)-linear (resp. \( o_{k,p} \)-antilinear) whenever \( \mathbf{b} \in \mathcal{V}_S^+ \) (resp. \( \mathbf{b} \in \mathcal{V}_S^- \)).

(ii) By [RZ, Proposition 2.9], the moduli problem \( Z(b) \) is represented by a closed formal subscheme of \( \mathcal{D} \).

Our first result is the following description of the \( F \)-points of the special cycles.

**Proposition 3.7.** (i) Suppose \( \mathbf{b} \in \mathcal{V}_S^- \), and \( b = \varphi^- \mathbf{b} \in C \) is the corresponding vector, cf. (3.14). Let \( \Lambda \) be any vertex lattice. Then

\[
Z(b)(F) \cap \mathcal{P}_A(F) = \begin{cases}
\emptyset, & \text{if } b \notin \Lambda, \\
\text{a single point } \{x\}, & \text{if } b \in \Lambda - p\Lambda, \\
\mathcal{P}_A(F), & \text{if } b \in p\Lambda.
\end{cases}
\]

If the second case above occurs, then \( \Lambda \) is necessarily of type 0, and the point \( x \) is special. Moreover \( x \) is superspecial if and only if \( \text{ord}_p q^-(b) > 0 \).

(ii) Similarly, for \( \mathbf{b} \in \mathcal{V}_S^+ \) with \( b = \varphi^+ \mathbf{b} \), cf. (3.13), we have

\[
Z(b)(F) \cap \mathcal{P}_A(F) = \begin{cases}
\emptyset, & \text{if } b \notin \Lambda, \\
\text{a single point } \{x\}, & \text{if } b \in \Lambda - p\Lambda, \text{ with } \Lambda^q = p\Lambda \\
\mathcal{P}_A(F), & \text{if } b \in \Lambda \text{ with } \Lambda^q = p\Lambda \\
\mathcal{P}_A(F), & \text{if } b \in p\Lambda, \Lambda \text{ arbitrary.}
\end{cases}
\]

In the second case, the unique point \( x \) is special, and is superspecial if and only if \( \text{ord}_p q^+(b) > 0 \).
Proof. (i) Suppose $b \in V_2^\circ$. We observe that a point $x = (X, \rho_X) \in \mathcal{D}(\mathbb{F})$ is in $Z(b)(\mathbb{F})$ if and only if, upon identifying $M(X)$ with a $W$-lattice in $N(X)$, we have

$$b(M(Y)) \subset M(X) \iff b(f_0) \in M(X)_1, \text{ and } b(f_1) \in M(X)_0$$

$$\iff b = b(f_1) \in VM(X)_1;$$

the last equivalence follows from the relation $Vb(f_0) = b(f_1)$. Recall also that $b(f_0) \in C = N(X)^{FV^{-1}}$, so

$$x = (X, \rho_X) \in Z(b)(\mathbb{F}) \iff b \in VM(X)_1 \cap C.$$

Now suppose $x \in \mathbb{P}_\Lambda(\mathbb{F}) \cap Z(b)(\mathbb{F})$, with $A^2 = pA$. By construction, $x$ corresponds to a lattice pair $B \subset A$, where

$$A = VM(X)^1 = \Lambda \otimes W.$$

On the other hand, note

$$pA = A^2 = (VM(X)^1)^2 = FM_1(X).$$

Thus, as $x \in Z(b)(\mathbb{F})$ as well, then

$$b \in VM_1(X) \cap C = FM_1(X) \cap C = pA \cap C.$$

However, as $pA = p\Lambda W$, the above line is true for all $x \in \mathbb{P}_\Lambda(\mathbb{F})$, as soon as it is true for a single point. Hence we have $\mathbb{P}_\Lambda(\mathbb{F}) \subset Z(b)(\mathbb{F})$.

Now suppose $x \in \mathbb{P}_\Lambda(\mathbb{F})$ with $A^2 = \Lambda$. By construction, this means $M(X)_0$ is identified with $\Lambda \otimes W$, and so if $\mathbb{P}_\Lambda(\mathbb{F}) \cap Z(b) \neq \emptyset$, then we must have $b \in \Lambda$. Furthermore, any $x \in \mathbb{P}_\Lambda(\mathbb{F})$ is determined by the sequence of inclusions of $\mathbb{F}$-codimension 1

$$pM(X)_0 \subset VM(X)_1 \subset M(X)_0$$

$$\parallel \quad \parallel$$

$$p\Lambda \otimes W \quad \Lambda \otimes W.$$

Hence, if $b \in p\Lambda$, then $b \in VM(X)_1$ for all $(X, \rho_X) \in \mathbb{P}_\Lambda(\mathbb{F})$, and so

$$Z(b)(\mathbb{F}) \cap \mathbb{P}_\Lambda(\mathbb{F}) = \mathbb{P}_\Lambda(\mathbb{F}).$$

If on the other hand $b \in \Lambda \setminus p\Lambda$, and $\mathbb{P}_\Lambda(\mathbb{F}) \cap Z(b)(\mathbb{F}) \neq \emptyset$, then this intersection necessarily consists of a single point $x = (X, \rho_X)$: namely, the unique point with

$$VM(X)_1 = W \cdot b + p\Lambda_W \subset \Lambda_W.$$

Note in this case, $(pV^{-2})VM(X)_1 = VM(X)_1$ as both $\Lambda_W$ and $b$ are $pV^{-2}$-invariant.
Now by construction, we have
$$\Lambda W \subset VM(X)^{\dagger}_1 \subset p^{-1}\Lambda W.$$ If $\text{ord}_{p,q}(b) = 0$, then $p^{-1}b \not\in VM(X)^{\dagger}_1$, and so the point is ordinary. On the other hand, if $\text{ord}_{p,q}(b) > 0$, then
$$VM(X)^{\dagger}_1 = W \cdot p^{-1}b + \Lambda W = p^{-1}VM(X)_1 = p^{-1}\left(VM(X)^{\dagger}_1\right)^\dagger,$$ and so this point is superspecial.

(ii) Suppose $b \in V_{p,+}^{\dagger}$, and let $b = b(f_0) = \varphi^\dagger(b)$. A point $x = (\sum \rho_X)$ lies in $Z(b)(F)$ if and only if $b \in M(X)_0$. The lemma follows via a similar argument to the previous case.

**Lemma 3.8.** Fix an $\epsilon$-invariant basis $\{v_0, v_1\}$ of $C$ (cf. Remark 3.4) such that $h(v_0, v_0) = h(v_1, v_1) = 0$, $h(v_0, v_1) = -h(v_1, v_0) = \delta$.

Suppose that $b = a_0v_0 + a_1v_1$, with $\text{ord}_{p,q}(b)$ either 0 or $-1$. Set $b' := \epsilon(b) = a'_0v_0 + a'_1v_1$ and
$$\Lambda_b := \text{span}_{\text{ok},p}\{b, b'\}.$$ If $\text{ord}_{p,q}(b) = 0$, then $\Lambda_b$ is of type 0 (i.e. $\Lambda_b^\dagger = \Lambda_b$), and is the unique lattice of type 0 such that $b \in \Lambda_b - p\Lambda_b$. Similarly, if $\text{ord}_{p,q}(b) = -1$, then $\Lambda_b$ is of type 2, and is the unique lattice of type 2 such that $b \in \Lambda_b - p\Lambda_b$.

**Proof.** Suppose $\text{ord}_{p,q}(b) = 0$ and $\Lambda$ is a type 0 lattice with $b \in \Lambda - p\Lambda$. Then there exists an element $\gamma \in SU(C)$ such that
$$\gamma \cdot \Lambda = \Lambda_b,$$ as $SU(C)$ acts transitively on the set of type 0 lattices. We may assume without loss of generality that $q(b) = 1$. Note the vectors $b$ and $b'$ form an orthogonal basis for $C$, with $h(b, b) = -h(b', b') = 1$. Let
$$[\gamma] = \begin{pmatrix} x & \gamma \\ w & z \end{pmatrix}$$ denote the matrix representation of $\gamma$ with respect to the basis $\{b, b'\}$. Since $b = (\frac{1}{\lambda}) \in \Lambda$, we have $x, w \in o_{k,p}$. On the other hand, the equation $\gamma \cdot \gamma^\dagger = 1$ implies
$$xx' - yy' = -ww' + zz' = 1, \quad yy' = xz',$$ which implies that $y, w \in o_{k,p}$ as well. Hence $\gamma$ and $\gamma^\dagger$ stabilize $\Lambda_b$, which yields the result $\Lambda = \Lambda_b$.

The proof in the case $\text{ord}_{p,q}(b) = -1$ is similar.  \(\square\)
Remark 3.9. Suppose \( b \in V_\pm^\phi \), with \( \text{ord}_{p}\varphi b = 0 \). Let \( b = \varphi^\pm b \in C \) denote the corresponding vector. Then by Lemma 3.8, there is a unique lattice \( \Lambda \) such that \( b \in \Lambda - p\Lambda \), and \( \Lambda \) is of type 0 (resp. of type 2) if \( b \in V_\pm^\phi \) (resp. \( b \in V_\phi^\phi \)). Combining this observation with Proposition 3.7, we find that

\[
Z(b)(F) = \{x\}
\]

is a single ordinary point lying in the component \( \mathbb{P}_{\Lambda}(F) \).

Our goal for the remainder of this section is to give a complete description of the special cycles \( Z(b) \) as cycles on \( D \), as in Theorem 3.14 below. We do this by writing down equations using the (formal) affine open cover described in [KR1, §1], which consist of affine schemes of two types:

1. \( \hat{\Omega}_\Lambda^{ord} \simeq \text{Spf} W[T, (T^p - T)^{-1}]^\vee \), for each vertex lattice \( \Lambda \), and
2. \( \hat{\Omega}_{[\Lambda, \Lambda']} \simeq \text{Spf} W[T_1, T_2, (T_1^p - 1)^{-1}(T_2^p - 1)^{-1}]^\vee \), for each pair of neighbouring vertex lattices \( \Lambda \) and \( \Lambda' \).

In both cases above, the superscript \( ^\vee \) denotes completion along the ideal generated by \( p \), and the isomorphisms are determined by a choice of basis for \( \Lambda \).

The underlying set of \( \hat{\Omega}_\Lambda^{ord} \) is the set of ordinary points \( \mathbb{P}_{\Lambda}(\text{ord}) \) in \( \mathbb{P}_{\Lambda}(\text{ord}) \) (that is, the complement of the superspecial points), while the underlying set of \( \hat{\Omega}_{[\Lambda, \Lambda']} \) is the union

\[
\mathbb{P}_{\Lambda}^{ord}(\cup \mathbb{P}_{\Lambda'}^{ord}(\cup\{x\})
\]

where \( x \) is the superspecial point at the intersection of \( \mathbb{P}_{\Lambda} \cap \mathbb{P}_{\Lambda'} \). For \( \Lambda \) a type 0 lattice, with neighbour \( \Lambda' \) of type 2, we have open immersions (cf. [KR1, §1])

\[
\hat{\Omega}_\Lambda^{ord} \to \hat{\Omega}_{[\Lambda, \Lambda']}, \quad \text{induced by } T_0 \mapsto T, \ T_1 \mapsto pT^{-1}
\]

and

\[
\hat{\Omega}_{\Lambda'}^{ord} \to \hat{\Omega}_{[\Lambda, \Lambda']}, \quad \text{induced by } T_0 \mapsto pT^{-1}, \ T_1 \mapsto T.
\]

We first consider a type 0 lattice \( \Lambda = \Lambda^2 \). Let \( b \in V_\phi^\phi \), with \( \text{ord}_{p}\varphi^\pm \varphi b \geq 0 \) and such that \( Z(b) \cap \hat{\Omega}_\Lambda^{ord} \neq \emptyset \). Let \( b = \varphi^\pm b \in C \) the corresponding vector; by Proposition 3.7, we must have \( b \in \Lambda \). Fix an \( \epsilon \)-invariant basis \( \{v_0, v_1\} \) of \( \Lambda \) with respect to which the hermitian form \( h \) has matrix

\[
h \sim \begin{pmatrix} -\delta & \delta \\ -\delta & -\delta \end{pmatrix},
\]

and write

\[
b = a_0v_0 + a_1v_1, \quad \text{where } a_0, a_1 \in \mathbb{O}_{k,p}.
\]

Proposition 3.10. Let \( b \in V_\phi^\phi \), and suppose \( Z(b) \cap \hat{\Omega}_\Lambda^{ord} \neq \emptyset \). Then, with notation as in the previous paragraph,

\[
Z(b) \cap \hat{\Omega}_\Lambda^{ord} \simeq \text{Spf} W[T, (T^p - T)^{-1}]^\vee/(f),
\]

Documenta Mathematica 18 (2013) 1403–1464
where

\[ f := \begin{cases} a_0 T + a_1, & \text{if } b \in V^- \phi \\ p(a'_0 T + a'_1), & \text{if } b \in V^+ \phi. \end{cases} \tag{3.17} \]

**Proof.** The proof of this proposition is modelled on arguments by Terstiege found in the proofs of [Ter, Propositions 2.8 and 4.5]. We consider the following cases:

**Case 1:** \( b \in V^- \phi, \) with \( b = \phi^- b \in \Lambda. \)

Fix a point \( x \in Z(b)(F) \cap \mathcal{P}_x(F) \) corresponding to \((X, \eta, \rho) \in \mathcal{D}(F).\) At present, we do not require that \( x \) be ordinary (i.e. \( X \) may be superspecial); in fact the argument we are about to discuss applies equally well in the superspecial case, and so it will be expedient to consider this slightly more general setting here. Consider the (complete) local ring

\[ R = \mathcal{O}_{D, x}. \]

Let \( m_R \) denote the maximal ideal of \( R, \) and \( I \) denote the ideal corresponding to \( Z(b)(\mathbb{F}) \cap \mathcal{P}_x(\mathbb{F}).\) At present, we do not require that \( x \) be ordinary (i.e. \( X \) may be superspecial); in fact the argument we are about to discuss applies equally well in the superspecial case, and so it will be expedient to consider this slightly more general setting here. Consider the (complete) local ring

\[ R = \mathcal{O}_{D, x}. \]

it will suffice to prove that \( I_n \) is generated by the image of \( f \) in \( R_n, \) for each \( n. \) Let \( m_n \) denote the maximal ideal of \( R_n. \) We set

\[ A := R_n/m_n I_n, \quad A' := R_n/I_n. \]

Then the kernel \( J := I_n/m_n I_n \) of the projection \( A \to A' \) satisfies \( J^2 = 0, \) and hence is endowed with a PD structure. Moreover, to prove the proposition, it suffices (by Nakayama) to show that \( J \) is generated by the image of \( f \) in \( A \) (which, abusing notation, we shall henceforth denote as \( f \)).

Finally, we note that both \( A \) and \( A' \) can be viewed as \( \mathcal{O}_{k, p} \)-algebras via the fixed embedding \( \pi_0: \mathcal{O}_{k, p} \to W \) composed with the respective structural morphisms. Now associated to the rings \( A \) and \( A' \) are two points in the moduli space \( \mathcal{D}, \) which in turn correspond to formal \( \mathcal{O}_q \)-modules \( X \) and \( X', \) both of whose special fibres are equal to \( \overline{X}. \) Moreover, by assumption, \( X' \) is in \( Z(b)(A'). \) In other words, the map

\[ \beta_X := \rho^{-1} \circ b: \mathbb{Y} \to \overline{X} \]

is a morphism of \( p \)-divisible groups, which lifts to a morphism

\[ \beta_{X'}: \mathbb{Y}_{A'} \to X'. \tag{3.18} \]

Here \( \mathbb{Y}_{A'} \) is the base change of the canonical lift \( \mathbb{Y}_W \) to \( A'. \)
Let $D(Y_{A'}/\cdot)$ denote the Grothendieck-Messing crystal of $Y_{A'}$, which carries a $\mathbb{Z}/2\mathbb{Z}$ grading induced by the action of $o_{k,p}$. In particular, for a PD extension $B \to A'$, the canonical lift of $Y_{A'}$ (together with its $o_{k,p}$-action) to $B$ is determined by the Hodge filtration:

$$D(Y_{A'}/B)_{1} = F_{Y_{B}} \subset D(Y_{A'}/B). \quad (3.19)$$

This is a consequence of the requirement that $o_{k,p}$ act on $\text{Lie}(Y_{B})$ with signature $(1,0)$.

Turning to the $p$-divisible groups $X$ and $X'$, we let $D(X/\cdot)$ and $D(X'/\cdot)$ denote their respective Grothendieck-Messing crystals. We then have a diagram

$$D(X/A) \cong D(X'/A) \downarrow_{\text{mod } J}$$

Let $B \to A'$ be a PD-extension as above. Since $X' \in \mathbb{Z}(b)(A')$, there is a map of $B$-modules

$$D(\beta X') : D(Y_{A'}/B) \to D(X'/B)$$

induced by (3.18). By Grothendieck-Messing theory, if $\tilde{X}$ is a lift of $X'$ to $B$ which corresponds to an $O_{B,p}$-stable direct summand $\mathcal{F} \subset D(X'/B)$, then

$$\tilde{X} \in \mathbb{Z}(b)(B) \iff D(\beta X')(\mathcal{F}_{Y_{B}}) \subset \mathcal{F}. \quad (3.21)$$

Fixing any basis vector $f_{B} \in \mathcal{F}_{Y_{B}} = D(Y_{A'}/B)_{1}$, and recalling that $b$ is $o_{k,p}$-antilinear, the above condition can be rewritten as

$$\tilde{X} \in \mathbb{Z}(b)(B) \iff D(\beta X')(f_{B}) \in \mathcal{F}_{0}. \quad (3.21)$$

For convenience, we fix a basis vector $f'_{1} \in D(Y_{A'}/A'_{1})$, and a lift to a basis vector $f_{1} \in D(Y_{A}/A)_{1}$.

We want to express the condition (3.21) in terms of the affine chart

$$\hat{\Omega}_{[\Lambda,\Lambda']} \simeq \text{Spf } W[T_{0}, T_{1}, (T_{0}^{p-1} - 1)^{-1}, (T_{1}^{p-1} - 1)^{-1}]/(T_{0} T_{1} - p)^{\nu}. \quad (3.22)$$

As $p$ is nilpotent in $A$, the point $X \in \hat{\Omega}_{[\Lambda,\Lambda']}(A)$ is simply determined by a map of $W$-algebras

$$W[T_{0}, T_{1}, (T_{0}^{p-1} - 1)^{-1}, (T_{1}^{p-1} - 1)^{-1}]/(T_{0} T_{1} - p) \to A. \quad (3.23)$$

Let $t \in A$ denote the image of $T_{0}$, so that $t^{p-1} - 1 \in A^{\times}$.

If $X$ is ordinary, as in the hypotheses of the proposition, then we may equally well restrict to the smaller open affine neighbourhood

$$\hat{\Omega}_{[\Lambda]}^{\nu p} \simeq \text{Spf } W[T_{1}, (T^{p} - T)^{-1}]^{\nu}, \quad (3.24)$$
via the open immersion
\[ \hat{\Omega}_A^{\text{ord}} \to \hat{\Omega}_{[A,A']}, \]
induced by \( T_0 \mapsto T, \ T_1 \mapsto pT^{-1} \).

The point \( X \in \hat{\Omega}_A^{\text{ord}}(A) \) then corresponds to the same element \( t \in A \) as before, which in this case now satisfies \( t \in A^\times \).

In any case, by carefully tracing through the construction of the charts (3.22) and (3.24) as they are constructed in [BC], we find that there exists isomorphisms
\[ D(X/A)_0 \simeq \Lambda \otimes_{ok,p} A, \quad D(X'/A')_0 \simeq \Lambda \otimes_{ok,p} A' \] (3.25)
such that

(i) the 0-th graded piece of the Hodge filtration for \( X \) is identified with
\[ (\mathcal{F}_X)_0 = \text{span}_A\{v_0 \otimes 1 - v_1 \otimes t\} \subset D(X/A)_0 \simeq \Lambda \otimes A, \]
where \( \{v_0, v_1\} \) is the fixed basis of \( \Lambda \) as in (3.15);

(ii) the element \( D(\beta_{X'})(f_1) \) is identified with \( b \otimes 1 \in \Lambda \otimes A; \)

(iii) the vertical map in (3.20) is the natural map
\[ \Lambda \otimes A \to \Lambda \otimes A', \quad l \otimes a \mapsto l \otimes a', \]
where \( a \mapsto a' \) is the projection \( A \to A' \).

Now suppose \( b \) is any ideal of \( A \) contained in \( J \), and let \( B := A/b \). Let \( t_B \) denote the image of \( t \) in \( B \) and \( X_B \) the corresponding \( B \)-valued point of \( \mathcal{D} \). We may view \( X_B \) as a deformation of \( X' \), which by Grothendieck-Messing theory corresponds to the direct summand
\[ \mathcal{F}_{X_B} = \mathcal{F}_X \otimes_A B \subset D(X'/B), \]
and in particular we have
\[ (\mathcal{F}_{X_B})_0 = \text{span}_B\{v_0 \otimes 1 - v_1 \otimes t_B\} \subset D(X'/B)_0 = \Lambda \otimes B. \]

Then in light of (3.21), and noting that \( b \) is \( ok,p \)-antilinear, we have
\[ X_B \in Z(b)(B) \iff D(\beta_{X'}(f_1))(f_1) \otimes 1 \in (\mathcal{F}_{X_B})_0 \iff b \otimes 1 \in \text{span}\{v_0 \otimes 1 - v_1 \otimes t_B\} \text{ in } \Lambda \otimes B. \]

Recall that we had written \( b = a_0v_0 + a_1v_1 \). Hence the last condition is equivalent to
\[ a_0t_B + a_1 = 0 \text{ in } B. \]

Applying the necessity of this condition to the case \( b = J \), so \( B = A/J \), we find that
\[ f = a_0t + a_1 \in J. \]

Suppose now that \( X \) is ordinary. By the sufficiency of the above condition, the map \( \text{Spf}(A/(f)) \to \hat{\Omega}_A^{\text{ord}} \) factors through \( Z(b) \). But by definition \( J \) is the
smallest ideal of $A$ such that $\text{Spf}(A/J) \to \hat{\Omega}^{\text{ord}}_\Lambda$ factors through $Z(b)$, and so $J = (f)$; this concludes the proof in this case.

When $X$ is superspecial, an identical argument implies that $\mathbb{Z}(b)$ is again cut out by $f$ in $\mathcal{O}_{D,x}$.

**Case 2:** $b \in \mathbb{V}_0^+$ with $b = \varphi^* b \in \Lambda$.

Here it is necessary to restrict our attention to the ordinary setting. Let $r = r(b, \Lambda) := \max\{ n \mid p - n b \in \Lambda \}$, and write $b = p^r b_0 = p^r (\alpha_0 v_0 + \alpha_1 v_1)$, where $b_0 = \alpha_0 v_0 + \alpha_1 v_1 \in \Lambda - p \Lambda$, and so in particular at least one of $\alpha_0, \alpha_1$ is a unit in $\mathcal{O}_{B, p}$.

As a first step, we shall prove that every $W_{r+1} := W/(p^{r+1})$ valued point of $\hat{\Omega}^{\text{ord}}_{\Lambda}$ belongs to the special cycle $Z(b)$. To this end, suppose $(X, \rho_X) \in \hat{\Omega}^{\text{ord}}_{\Lambda}(W_{r+1})$, and let $\underline{\mathbf{X}} \in \Omega^{\text{ord}}_{\Lambda}(\mathbb{F})$ denote its reduction modulo $p$. Let

$$\underline{\mathbf{M}} = \underline{\mathbf{M}(X)} = \underline{\mathbf{M}}_0 \oplus \underline{\mathbf{M}}_1$$

denote the Dieudonné module of $\underline{\mathbf{X}}$, endowed with the grading induced by the action of $\mathcal{O}_{B, p}$. As the projection $W_{r+1} \to \mathbb{F}$ has kernel generated by $p$, it is equipped with a PD structure. Hence, via Grothendieck-Messing theory, the point $X$ corresponds to an $\mathcal{O}_{B, p}$-stable summand

$$\mathcal{F} \subset \mathcal{D}(\underline{\mathbf{X}}/W_{r+1}) = \underline{\mathbf{M}} \otimes \mathbb{W} W_{r+1} = \underline{\mathbf{M}}/p^{r+1} \underline{\mathbf{M}},$$

such that

$$\mathcal{F} \otimes \mathbb{W}_{r+1} \mathbb{F} = V \underline{\mathbf{M}}/p \underline{\mathbf{M}}.$$

By Proposition 3.7, we have $\underline{\mathbf{X}} \in Z(b)(\mathbb{F})$, and so the map $\beta := \rho_X^{-1} \circ b : \mathbb{Y} \to \underline{\mathbf{X}}$ induces a morphism of Dieudonné modules

$$\beta : M(\mathbb{Y}) = W : f_0 \oplus W : f_1 \to \underline{\mathbf{M}}.$$

By definition,

$$\beta(f_0) = b \in \Lambda \otimes W = \underline{\mathbf{M}}_0, \quad \beta(f_1) = V \beta(f_0) = V b \in \underline{\mathbf{M}}_1.$$

The morphism $\beta$ also induces a map on crystals

$$\mathcal{D}(\beta/W_{r+1}) : M(\mathbb{Y}) \otimes \mathbb{W} W_{r+1} = \mathcal{D}(\mathbb{Y} / W_{r+1}) \to \mathcal{D}(\underline{\mathbf{X}}/W_{r+1}) = \underline{\mathbf{M}} \otimes \mathbb{W} W_{r+1}.$$
corresponding to the lift $Y_{W_{r+1}}$ of $Y$ is simply the rank-1 module $\mathcal{F}_Y = \text{span}_{W_{r+1}} \{f_1 \otimes 1\}$. Hence, by Grothendieck-Messing, we have

\[
X \in Z(b)(W_{r+1}) \iff D(\beta/W_{r+1})(f_1) \in \mathcal{F}_1 \iff (Vb) \otimes 1 \in \mathcal{F}_1 \text{ in } M_1 \otimes W_{r+1}.
\]

We shall show that this latter condition always holds for any $X \in \hat{\Omega}_\Lambda^{ord}(W_{r+1})$.

Consider the element $b_0 = p^{-r}b \in \Lambda \setminus p\Lambda$. If $Vb_0 \in pM_1$, then

\[
Vb \otimes 1 = p^r \cdot Vb_0 \otimes 1 \in p^{r+1}\cdot (M_1 \otimes W_{r+1}) = \{0\},
\]

and so (3.26) holds trivially. If on the other hand $Vb_0 \in M_1 - pM_1$, then

\[
\text{span}_p(Vb_0 + pM_1) = V\mathcal{M}_0/pM_1 = \mathcal{F}_1 \otimes_{W_{r+1}} \mathbb{F}.
\]

In other words, the image of $Vb_0$ in $M_1/pM_1$ is a basis vector for $\mathcal{F}_1 \otimes \mathbb{F}$. Hence, there exists $\alpha \in M_1 \otimes W_{r+1}$ such that

\[
Vb_0 \otimes 1 + p\alpha \in \mathcal{F}_1,
\]

and so

\[
p^r(Vb_0 \otimes 1) + p^{r+1}\alpha = Vb \otimes 1 \in \mathcal{F}_1,
\]

as required. Thus, we have proven

\[
Z(b) \cap \hat{\Omega}_\Lambda^{ord}(W_{r+1}) = \hat{\Omega}_\Lambda^{ord}(W_{r+1}).
\]

We now determine the local equation of $Z(b) \cap \hat{\Omega}_\Lambda^{ord}$ in $\hat{\Omega}_\Lambda^{ord} \simeq \text{Spf} W[T, (T^p - T)^{-1}]^\nu$, assuming the intersection is non-empty. Note that as $\Lambda = \Lambda^2$ is a lattice of type 0, then by Remark 3.9, we must have $ord_p q^+(b) > 0$; otherwise, $Z(b)$ would meet the special fibre only at a type-2 lattice.

Set

\[
b' := \Pi_{\mathcal{X}}^{-1} \circ b \in \mathcal{V}_\mathcal{X}^-,
\]

so that $ord_p q^-(b') = ord_p q^+(b) - 1 \geq 0$, and note

\[
\varphi^{-} b' = b'(f_1) = \Pi_{\mathcal{X}}^{-1} b(f_1) = V\Pi_{\mathcal{X}}^{-1} b(f_0) = \epsilon(b) = p^r(\alpha_0'v_0 + \alpha_1'v_1),
\]

since by assumption the basis $\{v_0, v_1\}$ was taken to be $\epsilon$-invariant. Furthermore, we have evident inclusions

\[
Z(b') \subset Z(b) \subset Z(p \cdot b').
\]

If $I \subset W[T, (T^p - T)^{-1}]^\nu$ denotes the ideal defining $Z(b) \cap \hat{\Omega}_\Lambda^{ord}$, then

\[
(p^{r+1}(\alpha_0'v_0 + \alpha_1')) \subset I \subset (p^r(\alpha_0'T + \alpha_1')),
\]

Documenta Mathematica 18 (2013) 1403–1464
which follows by applying the result from Case 1 to the two antilinear homomorphisms $b'$ and $p \cdot b'$.

Recall that we want to prove that $I$ is generated by $f := p^{r+1}(\alpha'_0 T + \alpha'_1)$. Thus, it suffices to prove that $I \subset (p^{r+1}(\alpha'_0 T + \alpha'_1))$.

To this end, suppose $g \in I$. We write

$$g = p'(\alpha'_0 T + \alpha'_1)g_0,$$

and we need to show that $p$ divides $g_0$.

Suppose not. Then the reduction modulo $p$ of $(\alpha'_0 T + \alpha'_1)g_0$ is a non-zero rational function over $F$ and so there exists $t \neq 0 \in F$ such that $(\alpha'_0 t + \alpha'_1)g_0(t) \neq 0 \in F$.

Choose any preimage $\tilde{t} \in W_{r+1}^\times$ of $t$. Then the map

$$W[T, (T^p - T)^{-1}]^\vee \to W_{r+1}, \quad T \mapsto \tilde{t}$$

does not factor through $W[T, (T^p - T)^{-1}]^\vee / I$, which contradicts the assertion (3.27). Hence, we have that $p$ divides $g_0$, and so

$$I \subset (p^{r+1}(\alpha'_0 T + \alpha'_1)) = (p(\alpha'_0 T + \alpha'_1))$$
as required.

For a type 2 lattice $\Lambda'$, we describe the analogous result: choose an $\epsilon$-invariant basis $\{w_0, w_1\}$ for $\Lambda'$ such that $q(w_0) = q(w_1) = 0$ and $h(w_0, w_1) = p^{-1}\delta$.

Suppose $b \in V_{\phi}^\times$, and let $b = \varphi^b \alpha$ denote the corresponding vector. If $Z(b) \cap \widehat{\Omega}^{\text{ord}}_{\Lambda'} \neq \emptyset$, then by Proposition 3.7, we must have $b \in \Lambda'$, and so we may write

$$b = a_0w_0 + a_1w_1, \quad a_0, a_1 \in o_{k,p}.$$

The proof of the following proposition is completely analogous to Proposition 3.10, and is therefore omitted.

**Proposition 3.11.** Let the notation be as in the previous paragraph. Then if $Z(b) \cap \widehat{\Omega}^{\text{ord}}_{\Lambda'} \neq \emptyset$, we have

$$Z(b) \cap \widehat{\Omega}^{\text{ord}}_{\Lambda'} \simeq \text{Spf} \ W[T, (T^p - T)^{-1}]^\vee / (f),$$

where

$$f = \begin{cases} a_0 + a_1 T & \text{if } b \in V_{\phi}^+ \\ a'_0 + a'_1 T & \text{if } b \in V_{\phi}^- \end{cases}.$$
Lemma 3.12. For $b \in C$, write
\[ \text{ord}_p(q(b)) = \begin{cases} 2t, & \text{if } \text{ord}_p(q(b)) \text{ is even} \\ 2t - 1, & \text{otherwise.} \end{cases} \]

For any vertex lattice $\Lambda$, let
\[ r(b, \Lambda) := \max\{ r \in \mathbb{Z} \mid p^{-r}b \in \Lambda \}; \tag{3.28} \]

note that here we do not assume $b \in \Lambda$, and so $r(b, \Lambda)$ may be negative. Finally, let $\Lambda_\bigcirc$ denote the unique vertex lattice containing $p^{-1}b$, as in Lemma 3.8. Recall that $\Lambda_\bigcirc$ is type 0 (resp. type 2) if $\text{ord}_p(q(b))$ is even (resp. odd).

Then we have the formula
\[ r(b, \Lambda) = \begin{cases} t - \frac{d(\Lambda, \bigcirc)}{2}, & \text{ord}_p(q(b)) \text{ even} \\ t - \frac{d(\Lambda, \bigcirc) + 1}{2}, & \text{ord}_p(q(b)) \text{ odd}. \end{cases} \]

Here $d(\Lambda, \bigcirc)$ is the distance function on $B$, the Bruhat-Tits tree.

Proof. By scaling by a power of $p$, it suffices to prove this lemma in the case $t = 0$; that is, we may assume that either $\text{ord}_p(q(b)) = 0$ or $\text{ord}_p(q(b)) = -1$.

We proceed by induction on $d = d(\Lambda, \bigcirc)$. If $d(\Lambda, \bigcirc) = 0$, i.e. $\Lambda = \Lambda_\bigcirc$, then $r(b, \Lambda_\bigcirc) = 0$ by the definition of $\Lambda_\bigcirc$.

Next, suppose we have proven the claim for all lattices $L$ with $0 \leq d(L, \bigcirc) \leq d$, and let $\Lambda$ be a lattice with $d(\Lambda, \bigcirc) = d$. We shall prove the desired formula holds for all the neighbours of $\Lambda$. There are four cases here to check, as $\Lambda$ can be either type 0 or 2, and $\text{ord}_p(q(b))$ can be even or odd.

For example, suppose that $\Lambda = \Lambda^2$ and $\text{ord}_p(q(b)) = 0$, so $d = d(\Lambda, \bigcirc)$ is even. There exists an $a_k, p$ basis $\{v_0, v_1\}$ for $\Lambda$ with $(v_0, v_0) = (v_1, v_1) = 0$, and $(v_0, v_1) = -1; v_0) = \delta$. With respect to this basis, a complete list of the $p + 1$ neighbours of $\Lambda$ in the Bruhat-Tits tree are described as follows:

\[ \Lambda_\bigcirc' := \text{span}_{a_k, p} \{ p^{-1}v_0, v_1 \} \]
\[ \Lambda_\bigcirc' := \text{span}_{a_k, p} \{ v_0, p^{-1}av_0 + p^{-1}v_1 \} \]

as $\alpha \in \mathbb{Z}_p$ ranges over a complete set of representatives for $\mathbb{F}_p = \mathbb{Z}_p / p \mathbb{Z}_p$.

If $d = 0$, the claim holds for all the neighbours of $\Lambda = \Lambda_\bigcirc$ by inspection. Otherwise, without loss of generality we may assume that $d(\Lambda_\bigcirc', \bigcirc) = d - 1$, and so for all the other neighbours, $d(\Lambda_\bigcirc', \bigcirc) = d + 1$. We may write
\[ b = p^r(a_0v_0 + a_1v_1) = p^r \left( pa_0 \cdot (p^{-1}v_0) + a_1 \cdot v_1 \right) = p^{r+1} \left( a_0 \cdot (p^{-1}v_0) + p^{-1}a_1 \cdot v_1 \right) \]

where $r = r(b, \Lambda) = -[d/2]$.

Since $d$ is even, the induction hypothesis applied to $\Lambda$ and $\Lambda_\bigcirc'$ yields
\[ r(b, \Lambda_\bigcirc') = -[\frac{(d - 1)}{2}] = -[\frac{d}{2}] + 1 = r(b, \Lambda) + 1 = r + 1, \]

Documenta Mathematica 18 (2013) 1403–1464
and so we must have $a_0 \in o_{k,p}^\times$ and $p|a_1$. By inspecting the remaining neighbours $\Lambda'_\alpha$ of $\Lambda$, we immediately see

$$r(b, \Lambda'_\alpha) = r = -\left\lfloor \frac{d+1}{2} \right\rfloor,$$

as required. The remaining cases all follow in the same manner.

We now determine the equations of the special cycles $Z(b)$ in the local ring at a superspecial point $x \in \mathbb{P}_\Lambda(\mathbb{F}) \cap \mathbb{P}_{\Lambda'}(\mathbb{F})$, for a pair of neighbouring vertex lattices $\Lambda$ and $\Lambda'$ of type 0 and 2 respectively. Recall that we have a (formal) affine open neighbourhood of $x$

$$\hat{\Omega}_{[\Lambda, \Lambda']} \simeq \text{Spf} \left( W[T_0, T_1, (T_0^{p-1} - 1)^{-1}, (T_1^{p-1} - 1)^{-1}] / (T_0 T_1 - p) \right)^\vee,$$

where the point $x$ corresponds to the maximal ideal $m_x = (T_0, T_1)$.

Without loss of generality, we suppose that there is a basis $\{v_0, v_1\}$ for $\Lambda$ such that $(v_0, v_1) = -(v_1, v_0) = \delta$, $(v_0, v_0) = (v_1, v_1) = 0$, and $\Lambda = \text{span}\{v_0, v_1\}$, $\Lambda' = \text{span}\{p^{-1}v_0, v_1\}$. (3.29)

**Proposition 3.13.** Suppose $b \in \mathbb{V}_\phi^\pm$, with $b = \varphi^\pm b \in C$ the corresponding vector. Suppose $x \in Z(b)(\mathbb{F})$ is a superspecial point, with $x \in \mathbb{P}_\Lambda(\mathbb{F}) \cap \mathbb{P}_{\Lambda'}(\mathbb{F})$ as above. If we set

$$r = r(b, \Lambda), \quad r' = r(b, \Lambda'),$$

as in (3.28), then the equation for $Z(b)$ in the local ring $\mathcal{O}_{\mathcal{D}, x}$ is given by

$$\begin{cases} (T_0)^r' (T_1)^r = 0, & \text{if } b \in \mathbb{V}_\phi^- \\ (T_0)^r' (T_1)^{r+1} = 0, & \text{if } b \in \mathbb{V}_\phi^+. \end{cases}$$

**Proof.** We begin with the case $b \in \mathbb{V}_\phi^-$. With respect to a basis $\{v_0, v_1\}$ of $\Lambda$ as in (3.29), we may write

$$b = p^r (a_0 v_0 + a_1 v_1) = p^r (pa_0 \cdot (p^{-1}v_0) + a_1 v_1)$$

where, as usual, $r = r(b, \Lambda)$, and at least one of $a_0$ or $a_1$ is a unit in $o_{k,p}$. Then in light of the choice of basis (3.29), we have

$$r' = \begin{cases} r, & \text{if } a_1 \in o_{k,p}^\times \\ r + 1, & \text{if } p \mid a_1. \end{cases}$$

By the proof of **Proposition 3.10**, Case (i), the special cycle $Z(b)$ is defined at $x$ by the vanishing of the element

$$f := p^r (a_0 T_0 + a_1) \in \mathcal{O}_{\mathcal{D}, x}.$$
Now consider the case \( r' = r \) (i.e. \( \alpha_1 \in \mathfrak{o}_{k,p}^\times \)). Then the factor \( \alpha_0 T_0 + \alpha_1 \) is a unit in \( \mathcal{O}_{D,x} \) and so \( \mathcal{Z}(\mathbf{b}) \) is given by
\[
p^r = (T_0)^r \cdot (T_1)^r = 0.
\]
If on the other hand \( r' = r + 1 \), then \( p \) divides \( \alpha_1 \) and \( \alpha_0 \in \mathfrak{o}_{k,p}^\times \). Thus
\[
p^r(\alpha_0 T_0 + \alpha_1) = p^r T_0 (\alpha_0 + (\alpha_1/p) T_1) = p^r T_0 \cdot u
\]
with \( u = (\alpha_0 + (\alpha_1/p) T_1) \in \mathcal{O}_{D,x}^\times \), and so the cycle \( \mathcal{Z}(\mathbf{b}) \) is given by
\[
(T_0)^{r+1} (T_1)^r = (T_0)^r (T_1)^r = 0,
\]
as required.

The proof for a homomorphism \( \mathbf{b} \in \mathcal{V}^+_{\phi} \) is completely analogous.\[\square\]

**Theorem 3.14.** Suppose \( \mathbf{b} \in \mathcal{V}^\pm_{\phi} \), with \( \mathcal{O}_{D,x} \) is given by
\[
p \mathbf{r} = \binom{T_0 \mathbf{r}}{T_1 \mathbf{r}} = 0.
\]
If on the other hand \( r' = r + 1 \), then \( p \) divides \( \alpha_1 \) and \( \alpha_0 \in \mathfrak{o}_{k,p}^\times \). Thus
\[
p^r(\alpha_0 T_0 + \alpha_1) = p^r T_0 (\alpha_0 + (\alpha_1/p) T_1) = p^r T_0 \cdot u
\]
with \( u = (\alpha_0 + (\alpha_1/p) T_1) \in \mathcal{O}_{D,x}^\times \), and so the cycle \( \mathcal{Z}(\mathbf{b}) \) is given by
\[
(T_0)^{r+1} (T_1)^r = (T_0)^r (T_1)^r = 0,
\]
as required.

The proof for a homomorphism \( \mathbf{b} \in \mathcal{V}^+_{\phi} \) is completely analogous.\[\square\]

**Theorem 3.14.** Suppose \( \mathbf{b} \in \mathcal{V}^\pm_{\phi} \), with \( \mathcal{O}_{D,x} \) is given by
\[
\text{ord}_p q^\pm(\mathbf{b}) = \begin{cases} 2t, & \text{if } \text{ord}_p q^\pm(\mathbf{b}) \text{ is even} \\ 2t - 1, & \text{if } \text{ord}_p q^\pm(\mathbf{b}) \text{ is odd} \end{cases}
\]

Let \( \mathbf{b} = \varphi^\pm \mathbf{b} \in \mathcal{C} \) be the corresponding vector, and write \( \mathbf{b} = p^k \mathbf{b}^\circ \), where \( \mathcal{O}_{D,x} \) is either 0 or -1. Then by Lemma 3.8, there is a unique vertex lattice \( \Lambda^\circ \) (the “central lattice”) such that \( \mathbf{b}^\circ \in \Lambda^\circ - p \Lambda^\circ \).

Finally, for any vertex lattice \( \Lambda \) we define
\[
m(\mathbf{b}, \Lambda) := \begin{cases} 0, & \text{if } \mathbf{b} \not\in \Lambda \\ t - \lfloor d(\Lambda, \Lambda^\circ) / 2 \rfloor, & \text{if } \mathbf{b} \in \Lambda \text{ and } \mathcal{O}_{D,x} \end{cases}
\]

Then we have the equality of cycles on \( \mathcal{D} \):
\[
\mathcal{Z}(\mathbf{b}) = \mathcal{Z}(\mathbf{b})^\text{hor} + \sum_{\Lambda} m(\mathbf{b}, \Lambda) \mathbb{P}_\Lambda,
\]
where \( \mathcal{Z}(\mathbf{b})^\text{hor} \simeq \text{Spf} \mathcal{W} \) is a horizontal divisor meeting the special fibre of \( \mathcal{D} \) at a single ordinary special point in the component \( \mathbb{P}_{\Lambda^\circ} \).

**Proof.** To start, we have the following relations:
\[
m(\mathbf{b}, \Lambda) := \begin{cases} 0, & \text{if } \mathbf{b} \not\in \Lambda \\ r(\mathbf{b}, \Lambda) + 1, & \text{if } \mathbf{b} \in \Lambda, \Lambda^z = \Lambda \quad \text{for } \mathbf{b} \in \mathcal{V}^+_{\phi}, \mathbf{b} = \varphi^+ \mathbf{b} \\ r(\mathbf{b}, \Lambda), & \text{if } \mathbf{b} \in \Lambda, \Lambda^z = p \Lambda \end{cases}
\]
and
\[
m(\mathbf{b}, \Lambda) := \begin{cases} 0, & \text{if } \mathbf{b} \not\in \Lambda \\ r(\mathbf{b}, \Lambda), & \text{if } \mathbf{b} \in \Lambda \quad \text{for } \mathbf{b} \in \mathcal{V}^-_{\phi}, \mathbf{b} = \varphi^- \mathbf{b} \end{cases}
\]
which are easily verified by comparing the definition of \( m(b, \Lambda) \) above with Lemma 3.12.

The proof of this theorem amounts to collating the information contained in the local equations given by Propositions 3.10, 3.11, and 3.13. We shall illustrate this argument in the case \( b \in V_\varphi \) with \( ord_b q^{-1}(b) = 2t \) even; the remaining cases follow in an identical manner.

Suppose \( b = \varphi^{-1} b \in C \) is the vector corresponding to \( b \in V_\varphi \), so by assumption \( ord_b q(b) = ord_b q^{-1}(b) = 2t \) is also even. In particular, we have \( b_\circ = p^{-1} b \).

Let \( \Lambda \) denote a type 0 lattice, and choose a basis \( \{v_0, v_1\} \) such that \( h(v_0, v_1) = \delta \) and \( q(v_0) = q(v_1) = 0 \). If we write

\[
b = p^r (\alpha_0 v_0 + \alpha_1 v_1) = p^r b_0,
\]

where \( r = r(b, \Lambda) \) and \( b_0 \in \Lambda - p\Lambda \) as usual, then by Proposition 3.10, the intersection \( Z(b) \cap \hat{\Omega}_\Lambda^{\text{ord}} \) with the ordinary locus \( \hat{\Omega}_\Lambda^{\text{ord}} \) of \( \hat{\Omega}_\Lambda \) is defined by the vanishing of the element

\[
p^r (\alpha_0 T + \alpha_1) \in W[T, (T^p - T)^{-1}]^\vee.
\]

(3.32)

Note that by Lemma 3.12,

\[
ord_b q(b_0) = 0 \iff b_0 = b_\circ \iff r = r(b, \Lambda) = t \iff \Lambda = \Lambda_\circ.
\]

By the choice of basis \( \{v_0, v_1\} \), we also have

\[
q(b_0) = \delta (\alpha_0 \alpha_1^* - \alpha_0^* \alpha_1).
\]

If \( \Lambda \neq \Lambda_\circ \), so that \( ord_b q(b_0) > 0 \), then \( \alpha_0 \alpha_1^* \equiv \alpha_0^* \alpha_1 \mod p \). A short calculation implies that the term \( (\alpha_0 T + \alpha_1) \) is a unit in \( W[T, (T^p - T)^{-1}]^\vee \). Hence, for \( \Lambda \neq \Lambda_\circ \), we have that \( Z(b) \cap \hat{\Omega}_\Lambda^{\text{ord}} \) is determined by the equation

\[
p^r = 0.
\]

As the component \( \mathbb{P}_A^{\text{ord}} \) is given by the equation \( p = 0 \), we then obtain the equality of cycles

\[
Z(b) \cap \hat{\Omega}_\Lambda^{\text{ord}} = m(b, \Lambda) \mathbb{P}_A^{\text{ord}}, \quad \text{for } \Lambda \text{ of type 0, } \Lambda \neq \Lambda_\circ.
\]

(3.33)

On the other hand, if \( \Lambda = \Lambda_\circ \), we have \( \alpha_0, \alpha_1 \in o_k^x \), and so the cycle defined by (3.32) is

\[
Z(b) \cap \hat{\Omega}_\Lambda^{\text{ord}} = Z(b)^{\text{hor}} + m(b, \Lambda_\circ) \mathbb{P}_A^{\text{ord}}
\]

(3.34)

where

\[
Z(b)^{\text{hor}} := \text{Spf } W[T, (T^p - T)^{-1}]^\vee / (\alpha_0 T + \alpha_1) \cong \text{Spf } W.
\]

Now suppose \( \Lambda' \) is a type 2 lattice. We choose a basis \( \{w_0, w_1\} \) of \( \Lambda' \) such that \( h(w_0, w_1) = p^{-1} \delta \) and \( q(w_0) = q(w_1) = 0 \). Writing

\[
b = p^r (\alpha_0 w_0 + \alpha_1 w_1),
\]
with \( r' = r(b, \Lambda') \) and \( b_0 := \alpha_0 w_0 + \alpha_1 w_1 \), we have

\[
q(b_0) = \delta (\alpha_0 a'_1 - \alpha'_0 a_1) .
\]

By Proposition 3.11, the cycle \( Z(b) \cap \hat{\Omega}^{ord}_{\Lambda'} \) is given by the vanishing of the element

\[
p^{r'}(\alpha'_0 + \alpha'_1 T) \in W[T, (T^p - T)^{-1}].
\]

Since we have assumed at the outset that \( ord_p q(b) \) is even, it follows that \( ord_p q(b_0) > -1 \). As before, this implies that the factor \( (\alpha'_0 + \alpha'_1 T) \) is a unit, and so the cycle \( Z(b) \cap \hat{\Omega}^{ord}_{\Lambda'} \) is given by the equation \( p^{r'} = 0 \). Hence, we have the equality of cycles

\[
Z(b) \cap \hat{\Omega}^{ord}_{\Lambda'} = m(b, \Lambda') \mathbb{P}^{ord}_{\Lambda'}, \quad \text{for } \Lambda' \text{ type 2}.
\]

Combining (3.33), (3.34) and (3.36), we have

\[
Z(b)^{ord} = Z(b)^{hor} + \sum_{\Lambda} m(b, \Lambda) \mathbb{P}^{ord}_{\Lambda},
\]

where \( Z(b)^{ord} \) denotes the restriction of \( Z(b) \) to the ordinary locus of \( D \) (i.e. the complement of the superspecial points).

Now suppose \( x \) is a superspecial point lying in the intersection \( \mathbb{P}_A \cap \mathbb{P}_{A'} \), for a type 0 lattice \( A \) and its type 2 neighbour \( A' \). Then Proposition 3.13 tells us that in a neighbourhood of \( x \), the special cycle \( Z(b) \) is determined by the equation

\[
(T_0)^{r'} \cdot (T_1)^{r} = 0.
\]

Recall that in such a neighbourhood, the components \( \mathbb{P}_A \) and \( \mathbb{P}_{A'} \) are given by the equations \( T_1 = 0 \) and \( T_0 = 0 \) respectively, and so meet \( Z(b) \) with multiplicities \( r = m(b, \Lambda) \) and \( r' = m(b, \Lambda') \) respectively. Therefore, we have

\[
Z(b) = Z(b)^{hor} + \sum_{\Lambda} m(b, \Lambda) \mathbb{P}_{\Lambda}
\]

as required.

We have proven the theorem in the case of an anti-linear special homomorphism \( b \in \mathcal{ Cru} \) with \( ord_p q^{-1}(b) \) even; the other cases all follow in a completely analogous manner.

\[\square\]

### 3.3 Relation to local orthogonal special cycles

In this section, we briefly recall definition of the local orthogonal special cycles constructed by Kudla and Rapoport [KR1], and relate them to the unitary special cycles of the previous section.
Definition 3.15. Suppose
\[ j \in \text{End}_{O_{B,p}}(X)_{\mathbb{Q}_p} := \text{End}_{O_{B,p}}(X) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \]
is an \( O_{B,p} \)-linear quasi-endomorphism of \( X \). Define \( Z^o(j)^\sharp \) to be the closed formal subscheme of \( \mathcal{D} \) which represents the following moduli problem: if \( S \in \text{Nilp} \), we take \( Z^o(j)^\sharp(S) \) to be the locus of points \( (X, t_X, \rho_X) \in \mathcal{D}(S) \) such that the quasi-morphism \( \rho_X^{-1} \circ j \circ \rho_X : X \times S \to X \times S \) lifts to an endomorphism of \( X \).

We also define \( Z^o(j) \) to be the Cohen-Macaulayfication of \( Z^o(j)^\sharp \), namely the closed subscheme of \( Z^o(j)^\sharp \) defined by the ideal sheaf of sections with finite support, cf. [KR1, §4].

These cycles (and their global counterparts, discussed in the next section) have been studied extensively by Kudla-Rapoport [KR1] and Kudla-Rapoport-Yang [KRY]. Of immediate interest to us is the following special case of a result of Kudla and Rapoport, which is the counterpart to Theorem 3.14.

Proposition 3.16 ([KR1] Proposition 4.5). Suppose \( j \in \text{End}_{O_{B,p}}(X)_{\mathbb{Q}_p} \) such that \( j^2 = u^2 p^2 \alpha^2 \delta^2 \) for some \( u \in \mathbb{Z}_{p}^\times \) (so in particular, \( Tr(j) = 0 \), and \( \mathbb{Q}_p(j) \) is isomorphic to \( k_p \)). Then \( Z^o(j) \) is a divisor on \( \mathcal{D} \).

Moreover, let \( \Lambda_0 \) be the unique vertex lattice such that \( j(\Lambda_0) = p^\alpha \Lambda_0 \). For any vertex lattice \( \Lambda \), define
\[ \mu^o(j, \Lambda) := \max \{ \alpha - d(\Lambda, \Lambda_0), \ 0 \} . \]

Then there is an equality of cycles on \( \mathcal{D} \):
\[ Z^o(j) = Z^o(j)^{hor} + \sum_{\Lambda} \mu^o(j, \Lambda) \mathbb{P}_\Lambda, \]
where \( Z^o(j)^{hor} \) is a disjoint sum of two divisors, each of which is isomorphic to \( \text{Spf} W \) and meets the special fibre of \( \mathcal{D} \) at an ordinary special point in the component \( \mathbb{P}_{\Lambda_0} \).

Suppose \( \gamma \in \text{End}_{O_{B,p}}(X)_{\mathbb{Q}_p} \). Then \( \gamma \) induces an endomorphism of \( N(X)_0 \), the 0-graded component of the rational Dieudonné module of \( X \), and commutes with the operators \( F \) and \( V \). Therefore \( \gamma \) defines an \( O_{k,p} \)-linear endomorphism on \( C := N(X)_0^{\langle 0 \rangle -1} F \), which, abusing notation, we also denote by \( \gamma \).

Now we observe that if \( j \) is as in Proposition 3.16, then as an endomorphism of \( C \), its characteristic polynomial splits. In particular, we may find an eigenvector \( b_0 \in C \) with eigenvalue \( p^\alpha \delta \). By scaling, we may assume \( \text{ord}_p(b_0) \) is either 0 or \(-1 \). Note that \( j \) also commutes with the Galois-semilinear operator \( \epsilon := V^{-1} \circ \Pi_X \), and so
\[ (j \circ \epsilon)(b_0) = (\epsilon \circ j)(b_0) = -p^\alpha \delta \cdot \epsilon(b_0) . \]
In other words, $b_0$ and $\epsilon(b_0)$ are eigenvectors for $j$, with eigenvalues $p^\alpha \delta$ and $-p^\alpha \delta$ respectively. Moreover, the “central” lattice $\Lambda_0$ attached to $j$, as in Proposition 3.16 (ii), is $\Lambda_0 = \text{span}(b_0, \epsilon(b_0))$.

In anticipation of the next section, we make the following definition: for $j$ as above, set

$$\nu_p(j) = \nu_p(j, \phi) := \begin{cases} 1, & \text{if } \exists \text{ an eigenvector } b \in C \text{ with } \text{ord}_p q(b) \text{ odd} \\ p, & \text{otherwise.} \end{cases} \quad (3.37)$$

In other words, we have $\nu_p(j) = 1$ if and only if there exists a homomorphism $b \in V^+_{\phi}$ such that

$$j \circ b = p^\alpha \cdot (b \circ i_Y(\delta)), \quad \text{with } \text{ord}_p q(b) \text{ even.}$$

The next two theorems relate the orthogonal and unitary special cycles on $D$.

**Theorem 3.17.** Suppose $j \in \text{End}_{\mathbb{B}_{\mathbb{R}}(X)} \mathbb{Q}_p$, with $j^2 = u^2 p^{2\alpha} \delta^2$ for some $\alpha > 0$ and $u \in \mathbb{Z}_{p^\infty}^\times$. Abbreviate $\nu = \nu_p(j)$, and fix $b_0 \in C$ an eigenvector with eigenvalue $p^\alpha \delta$ such that $q(b_0) = p^{-1}\nu$.

(i) If $\alpha$ is even, define $b^+ \in V^+_{\phi}$ and $b^- \in V^-_{\phi}$ by the relations:

$$\varphi^+(b^+) = \nu^{-1} \cdot p^{\alpha/2} \cdot b_0, \quad \varphi^-(b^-) = p^{\alpha/2} \cdot b_0.$$

(ii) If $\alpha$ is odd, define $b^+$ and $b^-$ by the relations:

$$\varphi^+(b^+) = p^{\alpha-1/2} \cdot b_0, \quad \varphi^-(b^-) = \nu^{-1} \cdot p^{\alpha+1/2} \cdot b_0$$

Then we have an equality of cycles on $D$

$$Z^\alpha(j) = Z(b^+) + Z(b^-). \quad (3.38)$$

**Remark 3.18.** The key feature of this formula is that in every case, exactly one of the special homomorphism $b^+$ appearing on the right hand side has norm $p^{\alpha}$, while the other has norm $p^{\alpha-1}$ (up to units in $\mathbb{Z}_{p^\infty}^\times$).

**Proof.** Note that the central lattices for the cycles $Z^\alpha(j)$, $Z(b^+)$, and $Z(b^-)$ are all the same lattice, namely $\Lambda_\circ := \text{span}_{\mathbb{B}_{\mathbb{R}, p}} \{b, \epsilon(b)\}$. One can easily verify that in every case in the statement of the theorem, we have

$$m(b^+, \Lambda) + m(b^-, \Lambda) = \alpha - \lfloor d(\Lambda, \Lambda_\circ) / 2 \rfloor - \lfloor (d(\Lambda, \Lambda_\circ) + 1) / 2 \rfloor$$

$$= \alpha - d(\Lambda, \Lambda_\circ)$$

$$= \mu^\circ(j, \Lambda)$$

for any vertex lattice $\Lambda$ with $d(\Lambda, \Lambda_\circ) \leq \alpha$. Indeed, the vectors $b^+$ and $b^-$ were scaled precisely so that this relation holds. On the other hand, if $d(\Lambda, \Lambda_\circ) > \alpha$, 

...
then neither side of (3.38) meets the component $\mathbb{P}_\Lambda$. Hence, any component $\mathbb{P}_\Lambda$ always occurs with the same multiplicity on both sides of (3.38).

It therefore suffices to show the horizontal components of both sides, which live in the open affine $\tilde{\Omega}_\Lambda^{\text{ord}}$, are equal. Suppose $\Lambda_{\hat{\circ}}$ is type 0, i.e.

$$\text{ord}_p q(b_0) = 0 = \text{ord}_p v_p(j).$$

Let $\{v_0, v_1\}$ denote an $e$-invariant basis for $\Lambda_{\hat{\circ}}$, with $q(v_0) = q(v_1) = 0$, and $h(v_0, v_1) = \delta$, and write

$$b_0 = a_0 v_0 + a_1 v_1, \quad \epsilon(b_0) = a'_0 v_0 + a'_1 v_1.$$

Note that $q(b_0) = -q(\epsilon(b_0)) = \delta(a_0 a'_1 - a'_0 a_1)$ is a unit in $\mathbb{Z}_p^\times$, by assumption.

Let $j_0 := p^{-\alpha} j$. Then $(j_0)^2 = \delta^2$, and by [KR1, Proposition 3.2], we have

$$Z^\alpha(j)^{\text{hor}} = Z^\alpha(j_0).$$

As an endomorphism of $C$, the element $j_0$ is determined by the fact that $b_0$ and $\epsilon(b_0)$ are eigenvectors with eigenvalues $\delta$ and $-\delta$ respectively. It is therefore defined by the matrix

$$[j] = \frac{\delta}{a_0 a'_1 - a'_0 a_1} \begin{pmatrix} a_0 a'_1 + a'_0 a_1 & -2n(a_0) \\ 2n(a_1) & -a_0 a'_1 - a'_0 a_1 \end{pmatrix} \in M_2(\mathbb{Z}_p)$$

with respect to the basis $\{v_0, v_1\}$. Now in the affine neighbourhood

$$\tilde{\Omega}_\Lambda^{\text{ord}} \simeq \text{Spf} W[T, (T^p - T)^{-1}]^\varphi,$$

the proof of Proposition 3.10 says that the cycle $Z(b^+)^{\text{hor}} + Z(b^-)^{\text{hor}}$ is given by the vanishing of

$$(a_0 T + a_1) \cdot (a'_0 T + a'_1) = n(a_0) T^2 + (a_0 a'_1 + a'_0 a_1) T + n(a_1). \quad (3.39)$$

On the other hand, by [KR1, Equation (3.5)], the equation for $Z^\alpha(j_0)$ in $\tilde{\Omega}_\Lambda^{\text{ord}}$ is given by

$$\frac{-2\delta}{a_0 a'_1 - a'_0 a_1} (n(a_0) T^2 + (a_0 a'_1 + a'_0 a_1) T + n(a_1)), $$

which differs from (3.39) by a scalar in $\mathbb{Z}_p^\times$, and hence defines the same divisor. If $\Lambda_{\hat{\circ}} = \text{span}\{b_0, \epsilon(b_0)\}$ is a type 2 lattice, fix an $e$-invariant basis $\{w_0, w_1\}$ for $\Lambda_{\hat{\circ}}$, such that $h(w_0, w_1) = p^{-1} \delta$ and $q(w_0) = q(w_1) = 0$. As before, we write

$$b_0 = a_0 w_0 + a_1 w_1, \quad \epsilon(b_0) = a'_0 w_0 + a'_1 w_1,$$

and note that $a_0 a'_1 - a'_0 a_1 \in \mathbb{Z}_p^\times$.

The cycle $Z(b^+)^{\text{hor}} + Z(b^-)^{\text{hor}}$ in $\tilde{\Omega}_\Lambda^{\text{ord}}$ is then given by the vanishing of the element

$$(a_0 + a_1 T) \cdot (a'_0 + a'_1 T) = n(a_0) + (a_0 a'_1 + a'_0 a_1) T + n(a_1) T^2.$$
Turning to the special endomorphism $j$, we note that it acts on $C$ via the matrix

$$
[j] = \begin{pmatrix} \delta & a_0a_1' + a_0'a_1 & -2n(a_0) \\ 2n(a_1) & -a_0a_1' - a_0'a_1 \end{pmatrix}
$$

with respect to the basis $\{w_0, w_1\}$. One can then check (e.g. by exploiting the action of $GL_2(\mathbb{Q}_p)$ on $D$ cf. [KR1, §1]), that the equation for $Z^\alpha(j_0)$ in this case is given by

$$\frac{-2\delta}{a_0a_1' - a_0'a_1} (n(a_0) + (a_0a_1' + a_0'a_1)T + n(a_1)T^2),$$

and hence defines the divisor $Z(b^+)\text{hor} + Z(b^-)\text{hor}$, as required.

We also have the corresponding theorem in the case $\alpha = 0$; note that all the cycles involved are horizontal.

**Theorem 3.19.** Suppose $j \in \text{End}_{\mathcal{O}_k}(X)\mathbb{Q}_p$ with $j^2 = u^2\delta^2$ with $u \in \mathbb{Z}_p^\times$. Let $b_0 \in C$ denote an eigenvector with eigenvalue $\delta$, and suppose that $q(b_0) = p^{-1}u \cdot (j, \phi)$, so $\text{ord}_p q(b_0)$ is 0 or $-1$.

Then if $\nu_p(j, \phi) = 1$, and we define $b_1, b_2 \in V_\phi^+$ by the formulas

$$b_1 = \phi^+(b_0), \quad b_2 = \phi^+(\epsilon(b_0)),$$

then

$$Z^\alpha(j) = Z(b_1) + Z(b_2). \quad (3.40)$$

Similarly, if $\nu_p(j, \phi) = p$ and we define $b_1, b_2 \in V_\phi^-$ by

$$b_1 = \phi^-(b_0), \quad b_2 = \phi^-(\epsilon(b_0)),$$

then

$$Z^\alpha(j) = Z(b_1) + Z(b_2). \quad (3.41)$$

**Proof.** The proof is completely analogous to the proof of Theorem 3.17, and is omitted.

3.4 $p$-adic uniformizations

Recall that we have fixed maps

$$\tau_0: \mathcal{O}_k/p \to \mathbb{F} \quad \text{and} \quad \tau_0: \mathcal{O}_{k,p} \to W.$$

In this section, when we take the $F$- or $W$-valued points of a stack defined over $\mathcal{O}_k$, we shall implicitly do so via the $\mathcal{O}_k$-structure induced by $\tau_0$.

If $\mathcal{X}$ denotes any of the stacks we have defined so far, we let $\tilde{\mathcal{X}}$ denote the base change to $W$ of the formal completion of its fibre at $p$. The $p$-adic uniformizations we are about to describe relate these completions to the moduli spaces of $p$-divisible groups discussed in the previous section.
We begin with the Shimura curve $C_B$. Fix a pair $\underline{A} = (A, \iota_A) \in C_B(\mathbb{F})$, such that the $p$-divisible group

$$\underline{A}[p^\infty] = (A[p^\infty], \iota_A \otimes \mathbb{Z}_p) \in \mathcal{D}(\mathbb{F}),$$

together with the induced $\mathcal{O}_{B,p}$-action, is equal to the “base-point” $(X, \iota_X)$ that we had fixed in defining the Drinfeld moduli space $\mathcal{D}$, cf. Definition 3.1. Let $B' = \text{End}_{\mathcal{O}_B}(A) \otimes \mathbb{Q}$. Then $B'$ is a quaternion algebra over $\mathbb{Q}$ whose invariants differ from those of $B$ at exactly $p$ and $\infty$. We let $H' = (B')^\times$, viewed as an algebraic group over $\mathbb{Q}$, and define

$$H'(\mathbb{Q})^* := \{ h' \in H'(\mathbb{Q}) \mid \text{ord}_p Nrd'(h') = 0 \},$$

where $Nrd'$ is the reduced norm on $B'$. Let $K_p \subset H'(A_p)$ denote the image of $(\hat{\mathcal{O}}_B)^\times$ under the natural identification $B'(A_p) \simeq B^{op}(A_p)^\times$, and set

$$\Gamma' := K_p \cap H'(\mathbb{Q})^*.$$

Then $\Gamma'$ acts on the Drinfeld upper-half plane $\mathcal{D}$ as follows: if $X = (X, \iota_X, \rho_X) \in \mathcal{D}(S)$ for some $S \in \text{Nilp}$, and $\gamma \in \Gamma'$, then we set

$$\gamma \cdot X := (X, \iota_X, \gamma_p \circ \rho_X),$$

where $\gamma_p$ denotes the image of $\gamma$ in $B' \otimes \mathbb{Q}_p$. We then obtain the following $p$-adic uniformization of $C_B$:

**Theorem 3.20.** There is an isomorphism of formal stacks over $\text{Nilp}$:

$$\tilde{C}_B \simeq [\Gamma' \backslash \mathcal{D}].$$

**Proof.** This is a restatement of [BC, Theorem III.5.3], see also [KR1, §8]. Primarily to set up notation for the sequel, we indicate the idea of the proof. We begin by fixing an $\mathcal{O}_B$-linear isomorphism

$$\eta_A : Ta^p(\mathcal{A}) \xrightarrow{\sim} \hat{\mathcal{O}}_B^p,$$

where $Ta^p(\mathcal{A}) := \prod_{\ell \neq p} Ta(\ell)$ is the prime-to-$p$ Tate module, and $\hat{\mathcal{O}}_B^p := \mathcal{O}_B \otimes \hat{\mathbb{Z}}_p$ viewed as an $\mathcal{O}_B$-module via left-multiplication.

Let $(A, \iota) \in C_B(S)$, for some $S \in \text{Nilp}$. Then there exists an $\mathcal{O}_B$-linear quasi-isogeny

$$\psi_A : A \rightarrow A$$

such that (i) the induced map on $p$-divisible groups (over $\mathcal{S} = S \times_{W} \mathbb{F}$)

$$\psi_A[p^\infty]_{\mathcal{S}} : A[p^\infty] \times \mathcal{S} \rightarrow A[p^\infty] \times \mathcal{S}$$

is a quasi-isogeny of height 0, and (ii) the composition

$$\eta_A \circ \psi_A^p : Ta^p(A) \otimes \mathbb{Q}_p \rightarrow B(\mathcal{A}_p)$$
maps $Ta^p(A)$ isomorphically onto $\hat{O}_k^p$. Noting that any two quasi-isogenies satisfying both these properties necessarily differ by an element of $\Gamma'$, it follows that the map

$$\widetilde{C}_B \to [\Gamma' \setminus D], \quad (A, t) \mapsto \left[ A[p^\infty], \ t \otimes \mathbb{Z}_p, \ \psi_A[p^\infty] \right]$$

is an isomorphism. \qed

We have a similar statement for the orthogonal special cycles (Definition 2.3).

**Theorem 3.21.** For $n \in \mathbb{Z}_{>0}$, let

$$\Omega^0(n) := \left\{ \xi \in B'(\mathbb{Q}) \mid Trd(\xi) = 0, \xi^2 = -n, \ and \ \eta_A \circ Ta^p(\xi) \circ \eta_A^{-1} \in \mbox{End}(\hat{O}_B^p) \right\},$$

(3.44)

and note that $\Gamma'$ acts on $\Omega^p(n)$ by conjugation. Then there is an isomorphism

$$\tilde{Z}^0(n) \simeq \left[ \Gamma' \ \setminus \ \bigoplus_{\xi \in \Omega^p(n)} Z^0(\xi[p^\infty]) \right]$$

as formal stacks over $W$. Here $\xi[p^\infty] \in \mbox{End}_{O_B}(X)$ is the endomorphism of $X = A[p^\infty]$ induced by $\xi$, and $Z^0(\xi[p^\infty])$ is a local orthogonal special cycle as in Definition 3.15.

**Proof.** Rewrite [KR1, (8.17)] in terms of the uniformization Theorem 3.20. \qed

We now turn to the $p$-adic uniformization of the unitary special cycles, following [KR3, §6]. Recall that we had fixed an embedding $\tau_0 : o_k/(p) \to F = \mathbb{F}_p^{alg}$, which lifts uniquely to an embedding $\tau_0 : o_k \to W = W(F)$; via these maps, we may view both $F$ and $W$ as $o_k$-algebras.

Fix a triple $E = (E, i_E, \lambda_E) \in \delta^+(F)$. We may also identify the $p$-divisible group $E[p^\infty]$, with its extra data, with the triple $Y = (Y, i_Y, \lambda_Y)$ of the previous section. We may further assume, by replacing $E$ with an isogenous elliptic curve if necessary, that there is an $o_k$-linear isomorphism

$$\eta_E : Ta^p(E) \sim \hat{o}_k^p$$

(3.45)

such that the pullback of the symplectic form

$$(a, b) \mapsto \mbox{tr}(ab'(\sqrt{\Delta})^{-1})$$

on $\hat{o}_k^p$ is equal to the Weil pairing

$$e_E : Ta^p(E)_{\mathbb{Q}} \times Ta^p(E)_{\mathbb{Q}} \to \mathbb{A}_f^p(1) \simeq \mathbb{A}_f^p.$$

*Documenta Mathematica* 18 (2013) 1403–1464
defined by $\lambda_E$. The two base points $E = (E, i_E, \lambda_E)$ and $A = (A, \iota_A)$ allow us
to define the global analogues of the spaces of special homomorphisms of the
previous section, as follows. For any optimal embedding $\phi : \mathcal{O}_k \to \mathcal{O}_B$, let

$$V^+_\phi := \{ \beta \in \text{Hom}(E, A)_\mathbb{Q} \mid \beta \circ i_E(a) = \iota_A(\phi(a)) \circ \beta \text{ for all } a \in \mathcal{O}_{k,p} \}$$
and $$V^-_\phi := \{ \beta \in \text{Hom}(E, A)_\mathbb{Q} \mid \beta \circ i_E(a) = \iota_A(\phi(a')) \circ \beta \text{ for all } a \in \mathcal{O}_{k,p} \}.$$ We view these spaces as $k$-vector spaces via the action

$$a \cdot \beta := \beta \circ i_E(a), \quad \text{for all } a \in k, \beta \in V^\pm_\phi.$$ Recall that for any optimal embedding $\phi$, we had defined a (non-principal)
polarization $\lambda_{A,\phi}$ on $A$, as in (2.1). The spaces $V^\pm_\phi$ can then be equipped with
Hermitian forms $h^\pm$, defined by the formulas:

$$h^+(\beta_1, \beta_2) := \lambda_E^{-1} \circ \beta_2^\vee \circ \lambda_{A,\phi} \circ \beta_1 \in \text{End}_{\mathcal{O}_k}(E)_\mathbb{Q} \simeq k, \quad \beta_1, \beta_2 \in V^+_\phi \quad (3.46)$$
and

$$h^-(\beta_1, \beta_2) := \lambda_E^{-1} \circ \beta_2^\vee \circ \lambda_{A,\phi} \circ \beta_1 \in \text{End}_{\mathcal{O}_k}(E)_\mathbb{Q} \simeq k, \quad \beta_1, \beta_2 \in V^-_\phi \quad (3.47)$$ respectively. Let $q^\pm(\beta) = h^\pm(\beta, \beta)$ denote the associated quadratic forms.
Finally, we recall the decomposition $\mathcal{D} = \mathcal{D}^+ \coprod \mathcal{D}^-$, which induces a decompo-
sition of the (global) cycles

$$Z(m, \phi) = Z^+(m, \phi) \coprod Z^-(m, \phi).$$ We may now state their $p$-adic uniformization:

**Theorem 3.22.** Suppose $m \in \mathbb{Z}_{>0}$. For any fractional ideal $a$ in $k$, and any
optimal embedding $\phi : \mathcal{O}_k \to \mathcal{O}_B$, let

$$\Omega^\pm(m, a, \phi) := \left\{ \beta \in V^\pm_\phi \mid q^\pm(\beta) = m \cdot \frac{\nu_p \cdot N(a)}{N(a)}, \quad \iota_A \circ \beta^p \circ \iota_{E^p}^{-1}(\hat{\alpha}) \subset \hat{O}_B^p \right\}. \quad (3.48)$$
where, for any $\beta \in V^\pm_\phi$, we denote by $\beta^p : \text{Tate}(E)_\mathbb{Q} \to \text{Tate}(A)_\mathbb{Q}$ the induced map
on rational prime-to-$p$ Tate modules. Note $\Gamma'$ acts on these sets by composition,
and we let $\mathcal{O}^\pm_k$ act via the $k$-vector space structure on $V^\pm_\phi$. Then

$$\tilde{Z}^\pm(m, \phi) \simeq \left[ \Gamma' \times \mathcal{O}_k^\pm \right] \left\{ \prod_{[\beta] \in \mathcal{C}(k)} \prod_{\beta \in \Omega^\pm(m, a, \phi)} Z(\beta[p\infty]) \right\}. \quad (3.49)$$
where $\beta[p\infty] \in V^\pm_\phi$ is the quasi-morphism of $p$-divisible groups induced by $\beta$. Here $a$ ranges any set of representatives of the class group $\mathcal{C}(k)$ of $k$. 

Documenta Mathematica 18 (2013) 1403–1464
Proof. This is a reformulation of [KR3, Proposition 6.3] in the present context.

Remark 3.23. Note that the right hand side of (3.49) is independent of the choices of representatives \( a \in Cl(k) \). Indeed, for any \( a \in k^\times \), we have a bijection

\[
\Omega^\pm(m, a, \phi) \xrightarrow{\sim} \Omega^\pm(m, a, \phi), \quad \beta \mapsto p^{-ord_p(n(a)/2)}a \cdot \beta.
\]

(3.50)

However, the cycles \( Z(\beta[p^\infty]) \) on \( D \) depend only on (i) the image of \( \beta[p^\infty] \) in \( \mathbb{P}(V_{\phi}^+) \) (which determines the horizontal part), and (ii) the \( p \)-adic valuation of \( q^\pm(\beta) \), (which determines the vertical part). By construction, both of these are the same for elements \( \beta \) appearing on either side of (3.50).

4 Proof of the main theorem

4.1 The (formal) Shimura lift

The Shimura lift is a classical operation that takes modular forms of half-integral weight to modular forms of even weight. In its original formulation [Shim], the lift of a modular form \( F \) is realized by writing down a formal \( q \)-expansion involving the Fourier coefficients of \( F \), and then proving that it is indeed the \( q \)-expansion of a modular form with the desired properties. This recipe inspires the following definition:

**Definition 4.1** (The formal Shimura lift). Let \( M \) be any \( \mathbb{Z} \)-module, and

\[
F = \sum_{n \geq 0} a(n) \ q^n \in M[[q]]
\]

a formal power series with coefficients in \( M \). Suppose that

(i) \( \kappa \geq 3 \) is an odd integer and \( \lambda := (\kappa - 1)/2 \);
(ii) \( N \) and \( t \) are positive integers with \( t \) squarefree;
(iii) and \( \chi \) is a Dirichlet character modulo \( 4N \).

We define a new character

\[
\chi_t(a) := \chi(a) \left( \frac{-1}{a} \right)^{(\kappa-1)/2} \left( \frac{t}{a} \right),
\]

where \( (\cdot/\cdot) \) denotes Shimura’s modification of the Kronecker symbol, cf. [Cip, Appendix A]. Then the Shimura lift of \( F \), with respect to the parameters \((\kappa, N, t, \chi)\), is by definition the formal power series

\[
Sh(F) := \sum_{m \geq 0} b(m) \ q^m \in M[q] \otimes \mathbb{C}
\]

(4.1)

whose coefficients are given as follows. For \( m > 0 \), we set

\[
b(m) := \sum_{n|m} \chi_t(n) \ n^{(\kappa-3)/2} \ a \left( \frac{tm^2}{n^2} \right).
\]

(4.2)
The constant term $b(0)$ is given as follows: let

$$\chi_t(a) = \frac{4}{4Nt - 1} \sum_{h=0}^{4Nt-1} \chi_t(h) \exp(2\pi iah/4Nt)$$

denote the Gauss sum, and set

$$b(0) = a(0) \left( -\frac{1}{4} \right)^{\lambda} (2\pi i)^{\lambda} (\pi)^{-2\lambda} (Nt)^{\lambda-1} \Gamma(\lambda) \sum_{m>0} m^{-\lambda} \chi_t(m).$$  \hspace{1cm} \text{(4.3)}$$

Note that if $\chi^2 = 1$ then $b(m) \in M$ for all $m > 0$.

As alluded to above, if $F$ is the $q$-expansion of a holomorphic modular form of weight $\kappa/2$, level $\Gamma_0(4Nt)$ and nebenncharacter $\chi$, then for each squarefree integer $t$, the Shimura lift $Sh(F)$ is the $q$-expansion of a modular form of weight $\kappa - 1$ for $\Gamma_0(4Nt)$ with character $\chi^2$, cf. [Cip, Proposition 2.17].

4.2 The Shimura lift formula in fibres of bad reduction

In this section, which is the technical heart of the paper, we show that a relation in the spirit of (4.2) holds for the orthogonal and unitary special cycles in a formal neighbourhood of the fibre at $p$ for $p|D_B$ by comparing their $p$-adic uniformizations. Let $F = F_p$, and fix a base point $\overline{A} \in \mathcal{C}_B(\mathbb{F})$ as in Section 3.4.

Our first aim is to relate the sets $\Omega_o(\sqrt{\Delta}a^2)$ and $\Omega^\pm(m, a, \varphi)$ in the case of interest, when the squarefree part of $n$ is equal to $|\Delta|$. Suppose $\xi \in \Omega^o(\sqrt{\Delta}|a^2|)$, for some $a \in \mathbb{Z}_{>0}$. Then, by assumption, the endomorphism

$$\eta^{-1} \circ \xi^p \circ \eta_A \in \text{End}_{\mathcal{O}_B}(\hat{\mathcal{O}_B})$$

is given by right-multiplication by some finite adele $(x_\ell)_{\ell \neq p} \in \hat{\mathcal{O}_B}$, such that $(x_\ell)^2 = a^2 \Delta$. Therefore, for every $\ell \neq p$, we obtain an embedding

$$\varphi_\ell : k_\ell = k \otimes \mathbb{Q}_\ell \to B_\ell, \quad a\sqrt{\Delta} \mapsto x_\ell.$$

We define the conductor $c = c(\xi)$ of $\xi$ to be the smallest (rational) positive integer such that for all $\ell \neq p$, we have

$$\varphi_\ell \left( \mathbb{Z}_\ell[c\sqrt{\Delta}] \right) \subset O_{B, \ell}.$$

In other words, $c(\xi)$ is the smallest integer such that $\varphi_\ell$ maps the unique $\mathcal{O}_k$-order of conductor $c(\xi)$ into $\mathcal{O}_{B, \ell}$, for all $\ell \neq p$. We note that since

$$\varphi_\ell \left( \mathbb{Z}_\ell[a\sqrt{\Delta}] \right) = \mathbb{Z}_\ell + \mathbb{Z}_\ell x_\ell \subset O_{B, \ell},$$

we have that $c(\xi)$ divides $a$, and by valuation considerations it is easy to see $(c(\xi), D_B) = 1$. In particular, we obtain a disjoint decomposition

$$\Omega^o(\sqrt{|\Delta}|a^2) = \prod_{(a, D_B) = 1} \Omega^o(\sqrt{|\Delta}|a^2, c),$$  \hspace{1cm} \text{(4.4)}$$

Documenta Mathematica 18 (2013) 1403–1464
where $\Omega^p(|\Delta|a^2, c)$ denotes the subset of elements $\xi \in \Omega^p(|\Delta|a^2)$ with $c(\xi) = c$. Note also that if $\xi \in \Omega^p(|\Delta|a^2, c)$ and $t$ is any integer, then the conductor of $t \cdot \xi$ is again $c$, and one checks easily that we have a bijection

$$\Omega^p(|\Delta|a^2, c) \cong \Omega^p(|\Delta|a^2t^2, c), \quad \xi \mapsto t \cdot \xi. \quad (4.5)$$

Now suppose $\phi: \mathcal{O}_k \to \mathcal{O}_L$ is an optimal embedding, and $\ell \neq p$ is a prime dividing $D_B$; recall our standing assumption that such a prime is inert in $k$.

Reducing modulo $\ell$, we obtain two maps

$$\overline{\phi}_\ell, \overline{\phi}_\ell: o_{k,\ell}/(\ell) \to \mathcal{O}_{B,\ell}/(\ell),$$

where $\theta \in \mathcal{O}_B$ is a fixed element such that $\theta^2 = -D_B$. As both the source and target of the maps are isomorphic to the field $\mathbb{F}_{\ell^2}$, the two maps $\overline{\phi}_\ell$ are either equal, or they differ by the Frobenius automorphism on $\mathbb{F}_{\ell^2}$. This observation allows to define the Frobenius type away from $p$ of $\xi$ as follows:

$$\nu_p(\xi, \phi) := \prod_{\ell \mid D_B, \ell \neq p} \nu(\xi, \phi), \quad \text{where } \nu(\xi, \phi) := \begin{cases} 1, & \text{if } \overline{\phi}_\ell = \overline{\phi}_\ell, \\ \ell, & \text{otherwise.} \end{cases} \quad (4.6)$$

Recall that we also had the notion of a Frobenius type

$$\nu_p(\xi, \phi) := \nu_p(\xi[p^{\infty}], \phi_0) \quad (4.7)$$

at $p$, which was defined in (3.37) in terms of the parity of the $p$-valuation of the norm of any eigenvector of the induced map $\xi[p^{\infty}]$ on Dieudonné modules.

For an optimal embedding $\phi$, and an element $\beta \in \mathcal{V}_G^+$, we may also define the notion of a conductor, as follows. Let

$$(h_\ell)_{\ell \neq p} := \eta_\Lambda \circ \beta^p \circ \eta_{E}^{-1}(1) \in B(A_f^p),$$

and for each $\ell \neq p$, define an embedding $\varphi'_\ell: k_\ell \to B_\ell$ by the formula

$$\varphi'_\ell(\sqrt{\Delta}) = (h_\ell)^{-1} \cdot \phi(\sqrt{\Delta}) \cdot h_\ell.$$ 

As before, we define the conductor $c(\beta)$ of $\beta$ to be the smallest integer $c$ such that $\varphi'_\ell(o_c) \subset \mathcal{O}_{B,\ell}$ for all $\ell \neq p$, where $o_c = \mathbb{Z}[c\sqrt{\Delta}]$ is the unique order of conductor $c$. Note that this quantity depends on the choice of embedding $\phi$.

**Lemma 4.2.** Suppose $\beta \in \mathcal{V}_G^+$, and let $h = (h_\ell)_{\ell \neq p} := \eta_\Lambda \circ \beta^p \circ \eta_{E}^{-1}(1) \in B(A_f^p)$. Then

(i) $q^+(\beta) = \Delta \cdot \text{Nrd}_{B(A_f^p)}(h) \in A_f^p$.

(ii) If $\beta \in \Omega^+(m, \alpha, \phi)$, then the conductor $c(\beta)$ divides $m$.

**Proof.**

(i) Recalling our fixed trivialization $A_f^p(1) \simeq A_f^p$, consider the diagram

$$\begin{array}{ccc}
\text{Ta}^p(E)_Q \times \text{Ta}^p(E)_Q & \xrightarrow{\beta^p \times \beta^p} & \text{Ta}^p(A)_Q \times \text{Ta}^p(A)_Q \\
\downarrow \eta_E \times \eta_E & & \downarrow \eta_A \times \eta_A \\
A_{k,f}^p \times A_{k,f}^p & \xrightarrow{(\phi(x), \phi(y))} & B(A_f^p) \times B(A_f^p) \\
(x,y) \mapsto (\phi(x), \phi(y)) & \mapsto & (x,y) \mapsto \text{Trd}(x^{\psi} \phi(\sqrt{\Delta})),
\end{array}$$

where $\text{Trd}(x^{\psi} \phi(\sqrt{\Delta}))$ is again $c$, and one checks easily that we have a bijection $\Omega^p(|\Delta|a^2, c) \cong \Omega^p(|\Delta|a^2t^2, c), \quad \xi \mapsto t \cdot \xi.$
where $e_A$ is the Weil pairing on $A$ defined by the polarization $\lambda_A, \phi$. It is an immediate consequence of the definitions that both squares commute. Let

$$e_E : \text{Tor}^p(E)_\mathbb{Q} \times \text{Tor}^p(E)_\mathbb{Q} \to \mathbb{A}^p_f$$

denote the Weil pairing on $E$ defined by $\lambda_E$. Then on the one hand, by taking adjoints, we have

$$e_A(\beta_p(x), \beta_p(y)) = q^+(\beta) \cdot e_E(x, y), \quad x, y \in \text{Tor}^p(E)_\mathbb{Q},$$

while on the other hand, if we write $s = \eta_E(x)$ and $t = \eta_E(y)$, then the commutative diagram above tells us that

$$e_A(\beta_p(x), \beta_p(y)) = \text{Trd} \left[ \phi(s) h \cdot (\phi(t) h)^+ \cdot \phi(\sqrt{\Delta}) \right]$$

$$= Nrd(h) \cdot \text{Trd} \left( \phi(s \cdot t' \cdot \sqrt{\Delta}) \right)$$

$$= Nrd(h) \cdot \Delta \cdot e_E(x, y),$$

where the last line follows by the choice of $\eta_E$, as in (3.45). This proves (i).

(ii) Suppose $\beta \in \Omega^+(m, a, \phi)$. It suffices to check that

$$m \cdot (h_\ell)^{-1} \phi(\sqrt{\Delta}) h_\ell \in \mathcal{O}_{B, \ell}, \quad \text{for all } \ell \neq p.$$ 

Choose $a_\ell \in k^\times$ such that $a \otimes \mathbb{Z}_\ell = a_\ell \cdot a_{k, \ell}$. Then by the definition of $\Omega^+(m, a, \phi)$, we have $h_\ell \in \phi(a_\ell^{-1}) \cdot \mathcal{O}_{B, \ell}$, and in particular,

$$(h_\ell)^+ \phi(\sqrt{\Delta}) h_\ell \in \frac{1}{N(a)} \mathcal{O}_{B, \ell}.$$ 

Hence, by combining part (i) of the lemma with the assumption

$$q^+(\beta) = m \cdot \frac{p^{\text{ord}_\ell(N(a))}}{N(a)} h_\ell,$$

we have

$$m \cdot (h_\ell)^{-1} \phi(\sqrt{\Delta}) h_\ell = p^{-\text{ord}_\ell(N(a))} \cdot \Delta \cdot N(a) \cdot \left( (h_\ell)^+ \phi(\sqrt{\Delta}) h_\ell \right) \in \mathcal{O}_{B, \ell}$$

as required.

Lemma 4.3. Let $\beta \in \Omega^+(m, a, \phi)$. Then there is a unique element $\xi = \xi(\beta) \in \Omega^p(|\Delta|m^2)$ such that $\xi \circ \beta = m \sqrt{\Delta} \cdot \beta$. Moreover, the conductor $c(\xi)$ of $\xi$ is equal to $c(\beta)$.

Proof. Note that each $\xi \in B'(\mathbb{Q})$ defines a $k$-linear endomorphism $[\xi]$ of $\mathcal{V}_\phi^+$ by composition:

$$[\xi] \cdot \beta := \xi \circ \beta.$$

Documenta Mathematica 18 (2013) 1403–1464
Moreover, it follows immediately from definitions that
\[ q^+(\xi \cdot \beta_1) = N\text{rd}'(\xi) \cdot q^+(\beta_1, \beta) \quad \text{for all } \beta \in V^+_{\Delta}. \]

Hence, if \( \beta \in \Omega^+(m, a, \phi) \) and \( \xi, \xi_2 \) both satisfy \( [\xi_1] \cdot \beta = m\sqrt{\Delta} \cdot \beta \), then \( N\text{rd}'(\xi_1 - \xi_2) = 0 \), so \( \xi_1 = \xi_2 \), as \( B' \) is division. This proves uniqueness.

To show existence, we choose an element \( \vartheta \in O_B \) with \( \vartheta^2 = -\vartheta \), and such that for all \( a \in k \), we have \( \vartheta \phi(a) = \phi(a') \vartheta \). Then we may define a quasi-isogeny
\[ \psi_\beta : E \times E \to A, \quad (x, y) \mapsto \beta(x) + \iota_A(\vartheta) \cdot \beta(y). \]

Note that for all \( a \in o_k \), we have
\[ \iota_A(\phi(a)) \cdot \vartheta(x, y) = \psi_\beta(i_E(a) \cdot x, i_E(a') \cdot y) \quad (4.8) \]
and
\[ \iota_A(\vartheta) \cdot \vartheta(x, y) = \psi_\beta(\vartheta^2 y, x). \quad (4.9) \]

We may then define an element \( \xi \in \End(A)_\mathbb{Q} \) by the formula
\[ \xi := m \cdot \psi_\beta \circ (i_E(\sqrt{\Delta}), i_E(\sqrt{\Delta})) \circ \psi_\beta^{-1}. \]

It follows from (4.8) and (4.9) that \( \xi \) commutes with \( \iota_A(\vartheta) \), and \( \iota_A(\phi(a)) \) for all \( a \in o_k \), so in fact \( \xi \in B'(\mathbb{Q}) = \End_{o_k}(A)_\mathbb{Q} \), and it is easily seen that
\[ \xi \circ \beta = m \cdot \beta \circ i_E(\sqrt{\Delta}) = m\sqrt{\Delta} \cdot \beta. \quad (4.10) \]

It is also straightforward to verify that \( \xi \) satisfies the conditions (3.44) defining \( \Omega^m(m^2|\Delta|) \), as well as the claim regarding conductors.

**Proposition 4.4.** Fix a set of representatives \( \{a_1, \ldots, a_h\} \) for the class group \( \Cl(k) \) of \( k \), such that each \( a_i \) is relatively prime to \( (p) \). Consider the map
\[ f : \prod_{i=1}^h \Omega^+(m, a_i, \phi) \to \Omega^m(|\Delta|m^2), \quad \beta \mapsto \xi(\beta) \]
where \( \xi(\beta) \) is the unique element which has \( \beta \) as an eigenvector with eigenvalue \( mn\sqrt{\Delta} \), as in Lemma 4.3. For any \( \xi \in \Omega^m(|\Delta|m^2) \), let \( \nu_p = \nu^\beta(\xi, \phi) \) (resp. \( \nu_p = \nu_p(\xi|p^\infty, \phi) \)) denotes its Frobenius type away from (resp. at) \( p \), and \( c = c(\xi) \) its conductor. Then we have
\[ \#(f^{-1}(\xi)) = |a_k^m| \cdot \rho \left( \frac{m}{c|\Delta|\nu_p\nu^p} \right), \]
where for any rational number \( N \), \( \rho(N) \) denotes the number of integral ideals of \( k \) of norm \( N \).

In particular, if \( m/c|\Delta|\nu_p\nu^p \) is not an integer, then the fibre \( f^{-1}(\xi) \) is empty.
Proof. Let $\xi \in \Omega^+(|\Delta|m^2)$. Viewing $\xi$ as a $k$-linear endomorphism of $\mathcal{V}_\phi^+$, it has two distinct eigenvalues $\pm m\sqrt{\Delta}$ and so there certainly exists some $\beta_0 \in \mathcal{V}_\phi^+$ such that

$$\xi \cdot \beta_0 = m\sqrt{\Delta} \cdot \beta_0.$$ 

Moreover, any other eigenvector with the same eigenvalue differs from $\beta_0$ by a scalar in $k^\times$. Hence, we have

$$\# f^{-1}(\xi) = \# \prod_{\ell=1}^h \{ \beta \in \mathcal{V}_\phi^+(m, a, \phi) \mid \xi(\beta) = \xi \}$$

$$= \# \prod_{\ell=1}^h \left\{ a \in k^\times \mid q^+(a\beta_0) = \frac{m}{N(a)}, \eta_A \circ \beta_0^p \circ \eta^{-1}_E(a\hat{a}^p) \subset \hat{O}_B^p \right\}$$

$$= |\mathcal{O}_k^\times| \cdot \# \left\{ a \in k \text{ a fractional ideal} \mid N(a) = \frac{m}{q^+(\beta_0)}, \eta_A \circ \beta_0^p \circ \eta^{-1}_E(1) \subset \hat{O}_B^p \right\}. \quad (4.11)$$

As before, set

$$h = (h\ell)_{\ell \neq p} := \eta_A \circ \beta_0^p \circ \eta^{-1}_E(1), \quad \text{and} \quad x = (x\ell)_{\ell \neq p} := \eta_A \circ \xi^p \circ \eta^{-1}_E(1).$$

Then the condition $\xi \cdot \beta_0 = m\sqrt{\Delta} \cdot \beta_0$ implies

$$x_{\ell} = m \cdot h_{\ell}^{-1} \phi(\sqrt{\Delta}) h_{\ell}, \quad \text{for all } \ell \neq p.$$ 

For each $\ell \neq p$, let $\varphi_\ell: k_\ell \to B_\ell$ be the embedding determined by the relation

$$\varphi_\ell(\sqrt{\Delta}) = m^{-1}x_{\ell} = h_{\ell}^{-1} \phi(\sqrt{\Delta}) h_{\ell}. \quad (4.12)$$

We shall translate the conditions on the right hand side of (4.11) to a collection of local ones. First, suppose that $\ell$ is a prime not dividing $D_B$. We may fix an isomorphism $\mathcal{O}_{B,\ell} \simeq M_2(Z\ell)$ that identifies

$$\phi(\sqrt{\Delta}) \quad \text{with} \quad \begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix}; \quad (4.13)$$

note that this is possible when $\ell = 2$ on account of the assumption that $|\Delta|$ is even. Let $c = c(\xi)$ be the conductor of $\xi$, and define

$$w(c)_{\ell} := \begin{pmatrix} c \\ 1 \end{pmatrix} \in \mathcal{O}_{B,\ell}.$$ 

Then the map

$$\varphi_c: k_\ell \to B_\ell, \quad \varphi_c(\sqrt{\Delta}) = w(c)_{\ell}^{-1} \phi(\sqrt{\Delta}) w(c)_{\ell}$$

is a local embedding of conductor $c$; that is, we have

$$\varphi_c(k_\ell) \cap \mathcal{O}_{B,\ell} = \varphi_c \left( Z\ell[\sqrt{\Delta}] \right).$$
By the definition of the conductor of $\xi$, the same is true for the embedding $\varphi_\ell : k_\ell \to B_\ell$ as in (4.12). However, by [Vig, Théorème II.3.2], any two embeddings of conductor $c$ necessarily differ by an inner automorphism determined by an element of $(O_{B_\ell})^\times$. Furthermore, for $x, y \in B_\ell^\times$,

$$Ad_{x^{-1}} \circ \phi = Ad_{y^{-1}} \circ \phi \iff x = \phi(a)y$$

for some $a \in k_\ell^\times$.

Hence, for each $\ell$ not dividing $D_B$, we may write

$$h_\ell = \phi(a_\ell) \cdot w(c)_\ell \cdot u_\ell$$

(4.14)

for some $a_\ell \in k_\ell^\times$ and $u_\ell \in (O_{B_\ell})^\times$. Note moreover that $h_\ell \in (O_{B_\ell})^\times$ for almost all $\ell$.

Now consider a prime $\ell | D_B$, $\ell \neq p$. Fix a uniformizer $\Pi_\ell \in O_{B_\ell}$ such that $\Pi_\ell \phi(a) = \phi(a') \Pi_\ell$ for all $a \in k_\ell$. Recalling our assumption that $\ell$ is inert in $k$, we may write

$$h_\ell = \phi(a_\ell) \cdot (\Pi_\ell)^{\epsilon_\ell},$$

(4.15)

for some $a_\ell \in k_\ell^\times$ and $\epsilon_\ell \in \{0,1\}$. One checks that for the reductions

$$\overline{\varphi_\ell}, \overline{\phi} \in \text{Hom}( (o_{k_\ell}/(\ell)), (O_{B_\ell}/(\Pi_\ell)) ),$$

we have

$$\overline{\varphi_\ell} = \overline{\phi} \iff \epsilon = 0,$$

where $\varphi_\ell$ is the embedding determined by $h_\ell$ as in (4.12). Hence, by definition of the Frobenius type (4.6),

$$\epsilon_\ell = \text{ord}_\ell(\nu_\ell(\xi, \phi)).$$

At this point, we have amassed a list of elements $(a_\ell) \in (A_{p,k,f})^\times$, as they appear in (4.14) and (4.15). We supplement this list with an element $a_p \in k_p^\times$ defined as follows: note that by the definition of the Frobenius type at $p$ (cf. (4.7) and (3.37)), we have

$$\text{ord}_p(\nu_p(\xi, \phi)) \equiv \text{ord}_p q^+(\beta_0) \pmod{2}.$$

We then set

$$a_p := p^\frac{1}{2}(\text{ord}_p q^+(\beta_0) - \text{ord}_p(\xi, \phi)).$$

Let $a_0$ denote the fractional ideal defined by $(a_\ell) \in A_{p,k,f}$. 

*Documenta Mathematica 18 (2013) 1403–1464*
Abbreviating $\nu_\ell = \nu_\ell(\xi, \phi)$, the product formula for $\mathbb{Q}$ implies

\[
q^+(\beta_0) = \left( \prod_{\ell \neq p} |q^+(\beta_0)_\ell|^{-1} \right) |q^+(\beta_0)_p|^{-1} = \left( \prod_{\ell \neq p} |\Delta|^{-1}_\ell \cdot |Nrd(h_\ell)|^{-1}_\ell \right) |q^+(\beta_0)_p|^{-1} = \prod_{\ell \neq p} |\Delta|^{-1}_\ell \cdot |Nrd(w(c)_\ell)|^{-1}_\ell
\]

\[
\times \left( \prod_{\ell \in D_B} |n(a_\ell)|^{-1}_\ell \cdot \nu_\ell \right) |n(a_p)|^{-1}_p \cdot \nu_p
\]

\[
= |\Delta| \cdot N(a_0) \left( \prod_{\ell \neq p} |Nrd(w(c)_\ell)|^{-1}_\ell \cdot \nu_\ell \right) \cdot \nu_p
\]

\[
= |\Delta| \cdot N(a_0) \cdot c \cdot \nu^p(\xi, \phi) \cdot \nu_p(\xi, \phi),
\]

where we have used the facts: (i) $Nrd(w(c)_\ell) = c$ for all $\ell$ and (ii) $|\Delta|$ is relatively prime to $D_B$.

At long last, we return to the quantity we wish to compute. Note that for a fractional ideal $a$ appearing in the right hand side of (4.11), we have the equivalence

\[
\eta_A \circ \beta^p \circ \eta_E^1(\widehat{\alpha}^p) \subset \widehat{\mathcal{O}}_B^p \iff \phi(\widehat{\alpha}^p) \subset \widehat{\mathcal{O}}_B^p \cdot h^{-1}.
\]

This in turn is equivalent to the collection of local statements, for all $\ell \neq p$:

\[
\phi(a_\ell) \subset \mathcal{O}_{B,\ell} \cdot h^{-1}_\ell = \begin{cases} \mathcal{O}_{B,\ell} \cdot \Pi^{-\nu_\ell}_{\ell} \phi(a_\ell)^{-1}, & \text{if } \ell \nmid D_B \\ \mathcal{O}_{B,\ell} \cdot w(c)_\ell^{-1} \cdot \phi(a_\ell)^{-1}, & \text{if } \ell \nmid D_B. \end{cases}
\]

Hence, replacing the ideals $a$ appearing in (4.11) by $a_0 \cdot a$, we obtain

\[
\#f^{-1}(\xi) = |a_0^*| \cdot \# \left\{ a \subset k \text{ a fractional ideal} \mid N(a) = \frac{m}{c|\Delta|^{p^\nu_p}}, \phi(a_{\ell}) \subset \mathcal{O}_{B,\ell} \cdot w(c)_{\ell}^{-1} \text{ for } \ell \nmid D_B, \text{ and } \phi(a_{\ell}) \subset \mathcal{O}_{B,\ell} \Pi_{\ell}^{-\nu_\ell} \text{ for } \ell \nmid D_B \right\}
\]

(4.16)

To conclude the proof, we show that an ideal $a$ appearing on the right hand side above is necessarily integral. Note that for $\ell \mid D_B$, including the case $\ell = p$, the condition on the norm of $a$ implies $\phi(a_{\ell}) \subset \mathcal{O}_{B,\ell}$. If $\ell \mid D_B$, then with respect to the isomorphism $\mathcal{O}_{B,\ell} \simeq M_2(\mathbb{Z}_\ell)$ as in (4.13), we compute

\[
\mathcal{O}_{B,\ell} \cdot w(c)_{\ell}^{-1} = \left\{ \begin{pmatrix} c^{-1}_x & y \\ c^{-1}_w & z \end{pmatrix}, x, y, w, z \in \mathbb{Z}_\ell \right\}.
\]
Recall that for $a, b \in \mathbb{Q}_\ell$,
\[
\phi(a + b\sqrt{\Delta}) = \begin{pmatrix} a & b \\ b\Delta & a \end{pmatrix}.
\]
Therefore,
\[
\left( \phi(k_\ell) \cap \mathcal{O}_{B, \ell} w(\varpi)^{-1} \right) \subset \phi \left( \mathbb{Z}_\ell[\sqrt{\Delta}] \right) \subset \mathcal{O}_{B, \ell}
\]
and so for all fractional ideals $\mathfrak{a}$ appearing in (4.16), and all finite primes $\ell$, we have $\phi(\mathfrak{a}_\ell) \subset \mathcal{O}_{B, \ell}$. As $\phi : \mathfrak{o}_k \to \mathcal{O}_B$ is an optimal embedding, it follows that $\mathfrak{a} \subset \mathfrak{o}_k$ is integral.

Before stating the main result of this section, we need a few more lemmas.

**Lemma 4.5.** There exists a prime $q$ that is split in $k$, and an element
\[
\varpi \in \text{End}(E)_\mathbb{Q}, \quad -Nm(\varpi) = \varpi^2 = -pq,
\]
such that $\varpi \circ \iota_E(a) = \iota_E(a') \circ \varpi$ for all $a \in \mathfrak{o}_k$.

**Proof.** See [Mann, p. 144]. The idea is that since $\text{End}(E)_\mathbb{Q}$ is the quaternion algebra ramified at exactly $p$ and $\infty$, the existence of such an element $\varpi$ is equivalent to the existence of a prime $q$ such that
\[
\left( -pq, \Delta \right)_\ell = \begin{cases} -1, & \text{if } \ell \in \{ \infty, p \} \\ 1, & \text{otherwise}, \end{cases}
\]
where $(\cdot, \cdot)_\ell$ is the Hilbert symbol. This imposes a finite set of congruence conditions on $q$, for which (infinitely many) solutions exists by Dirichlet’s theorem, and furthermore such a solution is necessarily split in $k$.

**Lemma 4.6.** Suppose $\phi : \mathfrak{o}_k \to \mathcal{O}_B$ is optimal, and let $\phi'$ denote the conjugate embedding. Then $\phi$ and $\phi'$ are not $\mathcal{O}_B^\times$-equivalent.

**Proof.** Write $\phi' = Ad_t \circ \phi$ for some $t \in B^\times$, which is always possible by Noether-Skolem. Let $\ell$ be a prime dividing $D_B$; then there is a uniformizer $\Pi_\ell$ such that
\[
\phi' = Ad_{\Pi_\ell^{-1}} \circ \phi \in \text{Hom}(\mathfrak{o}_k, \mathcal{O}_B, \ell).
\]
Hence $\Pi_\ell^{-1} \cdot t \in \phi(k_\ell^\times)$, and so $ord_\ell \text{Nrd}(t)$ is necessarily odd. In particular, $t \notin \mathcal{O}_B^\times$.

**Lemma 4.7.** Suppose $m \in \mathbb{Z}_{>0}$ is a positive integer, $\phi_1 : \mathfrak{o}_k \to \mathcal{O}_B$ is an optimal embedding, and $\mathfrak{a}$ is a fractional ideal of $k$. Suppose further that $t \in \mathcal{O}_B$ such that $\text{Nrd}(t)$ divides $\gcd(D_B, m)$. Note that $t$ normalizes $\mathcal{O}_B$, and in particular
\[
\phi_2 := Ad_{t^{-1}} \circ \phi_1
\]
Lemma 4.2

As $B$ holds. Write $\hat{\beta}$ we only need to check that the inclusion $\beta \in \Omega^\pm(m,Nrd(t),a,\phi_2)$. Thus valuation considerations.

Proof. Suppose $\beta \in \Omega^\pm(m,a,\phi_1)$, and set $\beta' := \imath_A(t^{-1}) \circ \beta$. We first verify $\beta' \in \Omega^\pm(m,Nrd(t),a,\phi_2)$.

The condition on the norm of $\beta'$ follows immediately from definitions, and so we only need to check that the inclusion

$$\eta_A \circ (\beta')^p \circ \eta_{E^{-1}}^{-1} (\widehat{a}) \in \widehat{\mathcal{O}}_B^p$$

holds. Write $\widehat{a} = (a_\ell) \cdot \widehat{o}_k^p$ for some prime-to-$p$ idele $(a_\ell)$, and set

$$h = (h_\ell)_{\ell \neq p} := \eta_A \circ (\beta')^p \circ \eta_{E^{-1}}^{-1} ((a_\ell)) \in \widehat{\mathcal{O}}_B^p,$$

and

$$h' = (h'_\ell) := \eta_A \circ (\beta')^p \circ \eta_{E^{-1}}^{-1} ((a_\ell)).$$

Note that to prove (4.17), it suffices to show that $h' \in \widehat{\mathcal{O}}_B^p$.

Recall that we had chosen $\eta_A : T a^p(A) \to \widehat{\mathcal{O}}_B^p$ to be an $\mathcal{O}_B$-linear isomorphism, where the action of $\mathcal{O}_B$ on $\widehat{\mathcal{O}}_B^p$ is given by left-multiplication. Hence we have

$$h' = t^{-1} \cdot h \in t^{-1} \cdot \widehat{\mathcal{O}}_B^p.$$

For primes $\ell$ not dividing $Nrd(t)$, note that $t^{-1} \cdot \mathcal{O}_{B,\ell} = \mathcal{O}_{B,\ell}$. On the other hand, suppose that $\ell$ divides $Nrd(t)$; in particular, $\ell$ divides $gcd(D_B,m)$. Then, by Lemma 4.2,

$$ord_\ell Nrd(h_\ell) = ord_\ell(m) \geq 1.$$

As $B_\ell$ is division, and $ord_\ell Nrd(t) = 1$, it follow that $h'_\ell = t^{-1} \cdot h_\ell \in \mathcal{O}_{B,\ell}$ by valuation considerations.

Thus $h'_\ell \in \mathcal{O}_{B,\ell}$ for all $\ell \neq p$, as required, and so we have shown that the assignment $\beta \mapsto \imath_A(t^{-1}) \circ \beta$ indeed defines a map $\Omega^\pm(m,a,\phi_1) \to \Omega^\pm(m,Nrd(t),a,\phi_2)$. By a similar argument, there is an inverse map

$$\Omega^\pm(m,Nrd(t),a,\phi_2) \to \Omega^\pm(m,a,\phi_1), \quad \beta' \mapsto \beta := \imath_A(t) \circ \beta',$$

which concludes the proof of the lemma.

Recall that $\widehat{Z}^\pm(m,\phi)$ denotes the formal completion of $Z^\pm(m,\phi)$ along its fibre at $p$, and that we have $p$-adic uniformizations

$$\widehat{Z}^\pm(m,\phi) \simeq \left[ \alpha_k^x \times \Gamma' \backslash \bigotimes_{[\alpha]} \prod_{\beta \in \Omega^\pm(m,a,\phi)} Z(\beta[p^\infty]) \right]$$

(4.18)

cf. Theorem 3.22.
Lemma 4.8. Let \( m \in \mathbb{Z}_{>0} \) be a positive integer, \( \phi_1 : \mathcal{O}_B \rightarrow \mathcal{O}_B \) an optimal embedding, and \( a \) a fractional ideal of \( k \). Suppose \( t \in \mathcal{O}_B \) such that \( \text{Nrd}(t) \) divides \( \gcd(DB/p, m) \), and let \( \phi_2 := \text{Ad}_{t^{-1}} \circ \phi_1 \). Then we have an equality of cycles on \( \mathcal{O}_B^a \):

\[
\tilde{Z}^{\pm}(m, \phi_1) = \tilde{Z}^{\pm}(m/\text{Nrd}(t), \phi_2).
\]

In particular, the cycle \( \tilde{Z}^{\pm}(m, \phi) \) only depends on the equivalence class \([\phi] \in \text{Opt}/\mathcal{O}_B^a\) of \( \phi \).

Proof. Let \( \beta \in \Omega^{\pm}(m,a,\phi_1) \), set \( \beta' = \iota_X(t^{-1}) \circ \beta \), and let \( b = \beta[p^\infty] \) and \( b' = \beta'[p^\infty] \) denote the corresponding maps of \( p \)-divisible groups. Recall that for a scheme \( S \in \text{Nilp} \), with special fibre \( \overline{S} = S \times \mathbb{F} \), we had defined the \( S \)-points of the cycle \( Z(b) \) on \( D \) to be the locus of points \( (X,\iota_X,\rho_X) \in D(S) \) such that the map

\[
\rho_X^{-1} \circ b : \overline{Y}_S \rightarrow X_{\overline{S}}
\]

lifts to a map \( \overline{Y}_S \rightarrow X \).

Similarly, the cycle \( Z(b') \) parametrizes tuples \( (X,\iota_X,\rho_X) \) such that \( \rho_X^{-1} \circ b' : \overline{Y}_S \rightarrow X_{\overline{S}} \) lifts to a map \( \overline{Y}_S \rightarrow X \). Note that the image of \( t \) in \( B_p \) in fact lies in \( (\mathcal{O}_{B,p})^\times \) by valuation considerations, and the endomorphism

\[
\iota_X(t^{-1})_{\overline{S}} : X \times \overline{S} \rightarrow X \times \overline{S}
\]

evidently admits a lift to \( S \), namely \( \iota_X(t^{-1}) \in \text{End}_S(X)^\times \). Combining these observations with the fact that the quasi-isogeny \( \rho_X \) is assumed to be \( \mathcal{O}_{B,p} \)-linear, we have that

\[
\rho_X^{-1} \circ b' = \rho_X^{-1} \circ \iota_X(t^{-1})_{\overline{S}} \circ b \text{ lifts } \iff \iota_X(t^{-1})_{\overline{S}} \circ \rho_X^{-1} \circ b \text{ lifts } \iff \rho_X^{-1} \circ b \text{ lifts}.
\]

Hence, we find that as cycles on \( D \), we have \( Z(b) = Z(b') \). The lemma then follows from Lemma 4.7 and the \( p \)-adic uniformization (4.18).

We now state and prove our key result:

Theorem 4.9. Let \( m \in \mathbb{Z}_{>0} \). Then we have the following equalities of cycles:

(i) If \( |\Delta| \) does not divide \( m \), then \( \tilde{Z}(m, \phi) = 0 \).

(ii) If \( m = |\Delta|m' \), then we have

\[
\sum_{[\phi] \in \text{Opt}/\mathcal{O}_B^a} \tilde{Z}(m, \phi) + \tilde{Z} \left( \frac{m}{\gcd(DB, m)}, \phi \right) = 2h(k) \sum_{\substack{\chi(m') \chi(\alpha) \tilde{Z} \left( \frac{\Delta |(m')^2}{\alpha^2} \right) \right)}
\]

where \( \chi_k \) is the quadratic character associated to \( k \), cf. (1.4).
Proof. (i) Suppose $|\Delta|$ does not divide $m$. Then for all possible values of $c$, $\nu_p$, and $\nu_p$, the quantity $m/c|\Delta|\nu_p\nu_p$ is not an integer, and so by Proposition 4.4,

$$\Omega^+(m, a, \phi) = \emptyset$$

for any fractional ideal $a$ and any embedding $\phi$. Hence we also have

$$\Omega^-(m, a, \phi) = \Omega^+(m, a, \phi') = \emptyset,$$

where $\phi'$ is the conjugate embedding. By the $p$-adic uniformization (4.18), it follows that $\tilde{Z}^+(m, \phi) = \emptyset$, and so as a cycle, $\tilde{Z}(m, \phi) = 0$.

(ii) Suppose $|\Delta|$ divides $m$, and $\phi: \mathcal{O}_k \to \mathcal{O}_B$ is an embedding with conjugate $\phi'$. By Lemma 4.6, the embeddings $\phi$ and $\phi'$ are $\mathcal{O}_B$-inequivalent. Moreover, for any $n$ we have $\tilde{Z}^+(n, \phi) = \tilde{Z}^-(n, \phi')$ essentially by definition. Summing over classes of optimal embeddings, it follows that

$$\sum_{[\phi] \in \mathcal{O}_k/\mathcal{O}_B^\times} \tilde{Z}(n, \phi) = 2 \sum_{[\phi] \in \mathcal{O}_k/\mathcal{O}_B^\times} \tilde{Z}^+(n, \phi)$$

$$= 2 \sum_{[\phi] \in \mathcal{O}_k/\mathcal{O}_B^\times} \tilde{Z}^-(n, \phi). \quad (4.20)$$

We proceed by cases:

**Case 1: $ord_p(m) > 0$**. For any cycle $Z$ on $\mathcal{D}$, let $[Z]$ denote the corresponding cycle on $\mathcal{C}_B = [\Gamma' \backslash \mathcal{D}]$, and for convenience, let

$$A := \sum_{[\phi]} \tilde{Z}(m, \phi) + \tilde{Z}(m/gcd(m, d), \phi)$$

denote the cycle we wish to compute (the left hand side of (4.19)).

Write $gcd(m, D_B) = p \cdot \mu$, and fix an element $t \in \mathcal{O}_B$ such that $Nrd(t) = \mu$.

Suppose $\phi_1: \mathcal{O}_k \to \mathcal{O}_B$ is an optimal embedding. By Lemma 4.8,

$$\tilde{Z}^-(m/p, \phi_1) = \tilde{Z}^{-}\left(\frac{m}{gcd(m, D_B)}, Ad_{t^{-1}} \circ \phi_1\right),$$

and so, upon taking the sum over all optimal embeddings and using (4.20),

$$A = 2 \sum_{[\phi] \in \mathcal{O}_k/\mathcal{O}_B^\times} \tilde{Z}^+(m, \phi) + \tilde{Z}^\left(\frac{m}{gcd(m, D_B)}, \phi\right)$$

$$= 2 \sum_{[\phi]} \tilde{Z}^+(m, \phi) + \tilde{Z}^-(m/p, \phi)$$

$$= 2 \sum_{[\phi]} \tilde{Z}^+(m, \phi) + \tilde{Z}^-(m/p, \phi).$$
Now suppose $\beta \in \Omega^+(m, a, \phi)$ and set

$$\beta' := \beta \circ \varpi,$$

where $\varpi \in \text{End}(E)$ satisfies $\varpi^2 = -pq$ for some split prime $q$, and $\varpi \circ i_E(a) = i_E(a') \circ \varpi$ for all $a \in \mathcal{O}_k$, cf. Lemma 4.5. Then it is easily verified that

$$\beta' \in \Omega^-(m/p, a', q, \phi),$$

where $a'$ is the conjugate of $a$, and $q$ is one of the ideals above $q$.

Furthermore, let $\xi = \xi(\beta) \in \Omega^+(m^2|\Delta|)$ denote the unique element such that $\beta$ is an eigenvector with eigenvalue $m\sqrt{\Delta}$, cf. Lemma 4.3. Then $\beta'$ is also an eigenvector, with eigenvalue $-m\sqrt{\Delta}$.

Moreover, the corresponding maps $\beta[p^\infty]$ and $\beta'[p^\infty]$ on $p$-divisible groups are (up to scalars in $\mathbb{Z}^\times$) precisely the special homomorphisms $b^+$ and $b^-$ appearing in Theorem 3.17, with $\xi[p^\infty]$ playing the role of $j$. The conclusion of that theorem then reads

$$Z^a(\xi[p^\infty]) = Z(\beta[p^\infty]) + Z(\beta'[p^\infty]),$$

as cycles on $D$. Hence

$$A = 2 \sum_{|\phi|} \frac{1}{|q_k^\phi|} \left\{ \sum_{\beta \in \Omega^+(m, a, \phi) \mod \Gamma'} [Z(\beta[p^\infty])] + \sum_{\beta' \in \Omega^-(m/p, a, \phi)} [Z(\beta'[p^\infty])] \right\}$$

$$= 2 \sum_{|\phi|} \frac{1}{|q_k^\phi|} \left\{ \sum_{\beta \in \Omega^+(m, a, \phi) \mod \Gamma'} [Z(\beta[p^\infty])] + \sum_{\beta' \in \Omega^-(m/p, a', \phi) \mod \Gamma'} [Z(\beta'[p^\infty])] \right\}$$

$$= 2 \sum_{|\phi|} \frac{1}{|q_k^\phi|} \left\{ \sum_{\beta \in \Omega^+(m, a, \phi) \mod \Gamma'} [Z^a(\xi[p^\infty])] \right\} \text{ where } \xi = \xi(\beta). \quad (4.21)$$

Given integers $c$, $\nu$, and $\nu^p$ such that $c|m$, $\nu_p \in \{1, p\}$ and $\nu^p|(D_B/p)$, we set

$$\Omega^p(|\Delta|m^2, c, \nu, \phi) := \{ \xi \in \Omega^p(|\Delta|m^2) \mid c(\xi) = c, \nu_p(\xi, \phi) = \nu, \nu^p(\xi, \phi) = \nu^p \};$$

that is, the set of elements of $\Omega^p(|\Delta|m^2)$ whose conductor and Frobenius types relative to $\phi$ are as specified, and where, for ease of notation, we have written $\nu = \nu_p \circ \nu^p$. Note that the action of $\Gamma'$ by conjugation preserves these sets. By Proposition 4.4, we may continue:

$$A = 2 \sum_{|\phi|} \sum_{c|m} \sum_{\nu|D_B} \rho \left( \frac{m}{|\Delta|c(m^2 \nu)} \right) \left\{ \sum_{\xi \in \Omega^p(|\Delta|m^2, c, \nu, \phi) \mod \Gamma'} [Z^a(\xi[p^\infty])] \right\}.$$
Writing \( m = |\Delta|m' \), only terms with \( c \) dividing \( m' \) contribute to the above sum. Moreover, recall that all primes dividing \( D_B \) are inert in \( k \) by assumption. Let \( \nu^* \) be the unique integer dividing \( D_B \) such that \( \text{ord}_\ell(m'/\nu^*) \) is even for all \( \ell \) dividing \( D_B \): only terms involving \( \nu = \nu^* \) can contribute to the above sum, since \( o_k \) has no ideals with norm an odd power of an inert prime. Thus

\[
A = 2 \sum_{c|m'} \rho\left(\frac{m'}{c\nu^*}\right) \left\{ \sum_{[\phi]} \sum_{\xi \in \Omega^*(m'|\Delta,\nu^*,\phi) \mod \Gamma'} [Z^*(\xi[p^\infty])] \right\}.
\]

Note that by [Vig, Corollaire III.5.12],

\[
\#\text{Opt}/\mathcal{O}_B^\times = h(k) \cdot 2^{o(D_B)},
\]

where \( o(D_B) \) denotes the number of prime factors of \( D_B \). Now consider the action of the normalizer \( N_{B^\times}(\mathcal{O}_B) \) on the set \( \text{Opt}/\mathcal{O}_B^\times \), acting by conjugation. For a fixed element \( \xi \in \Omega^*(m'|\Delta) \), the various values of the Frobenius types \( \nu(\xi,\phi) \), as \( \phi \) varies in an \( N(\mathcal{O}_B) \)-orbit of optimal embeddings, will cover all \( 2^{o(D_B)} \) possibilities. Thus, for fixed \( \xi \), there are exactly \( h(k) \) classes \([\phi]\) of optimal embeddings such that

\[
\nu(\xi,\phi) = \nu_p(\xi,\phi) \cdot \nu(\xi,\phi) = \nu^*.
\]

Hence, it follows that

\[
A = 2h(k) \sum_{c|m'} \rho\left(\frac{m'}{c\nu^*}\right) \sum_{\xi \in \Omega^*(m'|\Delta) \mod \Gamma'} [Z^*(\xi[p^\infty])].
\]

Next, we claim that for any integer \( N > 0 \), we may write

\[
\rho(N) = \sum_{a|N} \chi_k(a).
\]

Indeed, both sides of the above formula are multiplicative, and for \( N = \ell^n \) a prime power, the formula can immediately be verified by considering the cases \( \ell \) split, inert, and ramified separately. Moreover, as \( \text{ord}_\ell(m'/c\nu^*) \) is even for all \( \ell|D_B \), it follows that

\[
\rho\left(\frac{m'}{c\nu^*}\right) = \sum_{a|m'/c\nu^* \atop (a,D_B)=1} \chi_k(a) = \sum_{a|m'/c \atop (a,D_B)=1} \chi_k(a).
\]
Substituting, we obtain
\[
A = 2h(k) \sum_{c|m'} \sum_{a \mid m'/c} \chi_k(a) \sum_{\xi \in \Omega^c(m^2|\Delta)} [Z^o(\xi|p^\infty)] = 2h(k) \sum_{a \mid m'} \chi_k(a) \cdot \sum_{\xi \in \Omega^c(m^2|\Delta)} [Z^o(\xi|p^\infty)].
\]

Applying (4.5), we have that for each \(a|m'\) with \((a, D_B) = 1\),
\[
\sum_{c|(m'/a)} \sum_{\xi \in \Omega^c(m^2|\Delta)a^2} [Z^o(\xi|p^\infty)] = \sum_{c|(m'/a)} \sum_{\xi \in \Omega^c(m^2|\Delta)a^2} [Z^o(a|\Delta \cdot \xi|p^\infty)].
\]
\[
= \sum_{c|(m'/a)} \sum_{\xi \in \Omega^c(m^2|\Delta)a^2} [Z^o(\xi|p^\infty)] \quad \text{(since } a|\Delta \in \mathbb{Z}_p^\times)\]
\[
= \sum_{\xi \in \Omega^c(\Delta m^2/a^2)} [Z^o(\xi|p^\infty)] \quad \text{(by (4.4))}
\]
\[
= \tilde{Z}^o \left( \Delta \frac{(m')^2}{a^2} \right).
\]

Substituting this back into (4.22) concludes the proof of the theorem in the case \(ord_p(m) > 0\).

**Case 2: \(ord_p(m) = 0\).** As the proof is along similar lines as the previous case, we shall only indicate the necessary modifications. Fix an element \(t \in \mathcal{O}_B\) with \(Nrd(t) = p \cdot \gcd(m, D_B)\), and such that the image of \(t\) in \(\mathcal{O}_{B,p}\) is the uniformizer \(\Pi\) as in Section 2. Let \(\varpi \in \text{End(}E\rangle\) be as in Lemma 4.5, so that in particular \(\varpi^2 = -pq\) for a split prime \(q\). Then if \(\beta \in \Omega^+(m, a, \phi)\), it follows that
\[
\beta' := t_A(t) \circ \beta \circ \varpi^{-1} \in \Omega^+ \left( \frac{m}{\gcd(m, D_B)}, \ q^{-1} \cdot a', \ Ad_t \circ \phi \right) = \Omega^+ \left( \frac{m}{\gcd(m, D_B)}, \ q^{-1} \cdot a', \ Ad_t \circ \phi' \right),
\]
where \(q\) is one of the prime ideals above \(q\), and \(\phi'\) is the conjugate of \(\phi\).

Let \(\xi = \xi(\beta) \in \Omega^o(\Delta m^2)\) denote the special endomorphism corresponding to \(\beta\) as in Lemma 4.3. If \(j = \xi|p^\infty\) is the corresponding map on \(p\)-divisible
groups, then $\beta[p^\infty]$ and $\beta'[p^\infty]$ are equal (up to scaling by $\mathbb{Z}^\times$) to the elements $b_1$ and $b_2$ described in Theorem 3.19, and so

$$Z^o(\xi[p^\infty]) = Z(\beta[p^\infty]) + Z(\beta'[p^\infty]).$$

Therefore, we have

$$\left(\text{left hand side of (4.19)}\right) = \sum_{[\phi]} \tilde{Z}(m, \phi) + \tilde{Z}(m/gcd(m, D_B), \phi)$$

$$= 2 \sum_{[\phi]} \tilde{Z}^+(m, \phi) + \tilde{Z}^+(m/gcd(m, D_B), \phi) \quad [\text{by (4.20)}]$$

$$= 2 \sum_{[\phi]} \sum_{[a]} \frac{1}{\mathcal{O}_k^\times} \left\{ \sum_{\beta \in \Omega^+(m,a,\phi)} [Z(\beta[p^\infty])] \right. $$

$$\left. + \sum_{\beta' \in \Omega^+(m/(m,D_B),a,\phi)} [Z(\beta'[p^\infty])] \right\}$$

$$= 2 \sum_{[\phi]} \sum_{[a]} \frac{1}{\mathcal{O}_k^\times} \left\{ \sum_{\beta \in \Omega^+(m,a,\phi)} [Z(\beta[p^\infty])] \right. $$

$$\left. + \sum_{\beta' \in \Omega^+(m/(m,D_B),a',A\phi)} [Z(\beta'[p^\infty])] \right\}$$

where in the last line $\xi = \xi(\beta)$. The proof proceeds from this point exactly as in the previous case, cf. (4.21).

4.3 The main theorem and applications

Let $\omega$ denote the relative dualizing sheaf of $\mathcal{C}_B$, which we view as a divisor class in the Chow group $CH^1(\mathcal{C}_B)$, and choose any divisor $K$ in this class. Define the orthogonal generating series

$$\Phi^o(\tau) := -K + \sum_{n>0} \mathbb{Z}^o(n) q^n \in \text{Div}(\mathcal{C}_B)[[q]], \quad (4.23)$$

and consider the base change

$$\Phi^o_{/\mathcal{O}_k}(\tau) := -K_{/\mathcal{O}_k} + \sum_{n>0} \mathbb{Z}^o(n)_{/\mathcal{O}_k} q^n \in \text{Div}(\mathcal{C}_{B/\mathcal{O}_k})[[q]].$$
Let \( \chi_k = (\cdot/\Delta) \) denote the quadratic character attached to \( k \), and \( \chi'_k \) denote its induction to level \( 4D_B|\Delta| \). Define the Gauss sum

\[
\hat{\chi}'_k(a) := \sum_{h=0}^{4D_B|\Delta|-1} \chi'_k(h) \exp(2\pi iah/4D_B|\Delta|),
\]
and the ‘L-function’

\[
L(s, \hat{\chi}'_k) := \sum_{m>0} m^{-s} \hat{\chi}'_k(m),
\]
which is analytic (as written) in the half-plane \( \Re(s) > 0 \).

We then define the unitary generating series:

\[
\Phi^u(\tau) := \frac{i}{2\pi} L(1, \hat{\chi}'_k) \cdot K/o_k + \frac{1}{2h(k)} \sum_{m>0} [\mathcal{Z}(m, \phi) + \mathcal{Z}^*(m, \phi)] q^m
\]
\[\in \text{Div}(\mathcal{C}_{B/o_k})[q] \otimes \mathbb{Q},\tag{4.24}\]
where, as we recall,

\[
\mathcal{Z}^*(m, \phi) = \mathcal{Z}(m/gcd(m, D_B), \phi),
\]
and the sum on \( [\phi] \) is over the set \( \text{Opt}/\mathcal{O}_B^n \) of optimal embeddings taken up to \( \mathcal{O}_B^n \)-conjugacy. Note that

\[
\frac{i}{2\pi} L(1, \hat{\chi}'_k) = -\frac{h(k)}{|\alpha_k^o|} \prod_{\ell|D_B} \left( \frac{2\ell^2 + \ell - 1}{\ell^2} \right) \in \mathbb{Q},
\]
which can be seen by evaluating the \( L \)-function on the left via Dirichlet’s class number formula.

**Theorem 4.10 (Main theorem).** Suppose \( \Delta < 0 \) is squarefree and even, and every prime dividing \( D_B \) is inert in \( k \). Then we have an equality

\[
\text{Sh}(\Phi^o_{o_k})(\tau) = \Phi^u(\tau) \in \text{Div}(\mathcal{C}_{B/o_k})[q] \otimes \mathbb{Q} \tag{4.25}\]
of formal generating series, where \( \text{Sh} \) is the formal Shimura lift with parameters \( \kappa = 3, N = D_B, t = |\Delta| \) and \( \chi = 1 \) in the notation of Definition 4.1.

**Proof.** We apply the formulas (4.2) and (4.3) for the formal Shimura lift. Note that the constant terms match by design, and so – keeping in mind the shift by \( t = |\Delta| \) in the exponent in (4.1) – it suffices to prove that

(i) if \( |\Delta| \) does not divide \( m \), then \( \sum_{[\phi]} \mathcal{Z}(m, \phi) = 0 \);

(ii) and if \( m = m'|\Delta| \), then

\[
\sum_{[\phi] \in \text{Opt}/\mathcal{O}_B^n} \mathcal{Z}(m, \phi) + \mathcal{Z}\left( \frac{m}{\gcd(D_B, m)}, \phi \right)
\]
\[= 2h(k) \sum_{\alpha|m'} \chi_k(\alpha) \mathcal{Z}^o\left( \frac{|\Delta(m')^2}{\alpha^2} \right)_{o_k},\]
in $\text{Div}(\mathcal{C}_{B/o_k})$.  
By [KRY, Proposition 3.4.5], each orthogonal special cycle decomposes as 

$$Z^o(n)/o_k = Z^o(n)_{\text{hor}} + \sum_{q|D_B} Z^o(n)_q^{\text{ver}},$$

where $Z^o(n)_{\text{hor}}$ is the closure of the generic fibre in $Z^o(n)/o_k$, and $Z^o(n)_q^{\text{ver}}$ is the sum of the vertical irreducible components supported in characteristic $q$.  
By Proposition 2.6, we have an identical decomposition for the unitary cycles.  
Now for each $q|D_B$, Theorem 4.9 implies that (i) and (ii) hold for the vertical components at $q$. Moreover, since $\mathcal{C}_{B/o_k}$ is proper over $\text{Spec}(o_k)$, the same proposition implies (i) and (ii) hold in the generic fibre, by Grothendieck’s existence theorem [EGA3, Thm. 5.1.4]; more precisely, if we choose a prime $p|D_B$, then the existence theorem asserts that the desired relations hold over $o_k,p$, and then we observe that $o_k,p$ is faithfully flat over the localization $(o_k)_p$.

Our next step is to recall how certain quantities involving the orthogonal special cycles $Z^o(n)$ arise as Fourier coefficients of actual modular forms of weight 3/2, as in [KRY, §4].  
First, we consider the ‘rational degree’ generating series 

$$\Phi^o_{\text{deg}}(\tau) := -\deg_k(\omega) + \sum_{n>0} \deg_k Z^o(n) q^n \tau \in \mathbb{Q}[q]\{\tau\};$$

obtained by taking the degrees of the generic fibres. Next, we may consider the generating series 

$$\Phi^o_{\text{deg}}(\tau) := -\deg_k(\omega) + \sum_{n>0} Z^o(n)/k q^n \tau \in CH^1(\mathcal{C}_{B/k})[q]\{\tau\};$$

formed by taking generic fibres and then passing to the Chow group $CH^1(\mathcal{C}_{B/k}) = CH^1(\mathcal{C}_{B/k}) \otimes \mathbb{C}$.  
Finally, let $\mathcal{Y}$ denote an irreducible component of the fibre $(\mathcal{C}_{B/o_k})_p$ of the Shimura curve, where $p|D_B$ is a prime of bad reduction. For a closed substack $\mathcal{Z}$ of $\mathcal{C}_{B/o_k}$, we define the pairing 

$$\langle \mathcal{Z}, \mathcal{Y} \rangle := 2 \log(p) \cdot \chi(\mathcal{O}_\mathcal{Y} \otimes \mathcal{O}_{\mathcal{Z}}),$$

where $\chi$ is the stack version of the Euler-Poincaré characteristic, which takes into account the automorphism groups of points, cf. [DeRa, §VI.4]. We then form the generating series 

$$\Phi^o_{\mathcal{Y}}(\tau) := -(\omega, \mathcal{Y}) + \sum_{n>0} \langle Z^o(n), \mathcal{Y} \rangle q^n$$  \hspace{1cm} (4.26)

We may form the analogues $\Phi^o_{\text{deg}}(\tau)$, $\Phi^u_{\mathcal{Y}}(\tau)$ and $\Phi^u_{\mathcal{Y}}(\tau)$ for the unitary cycles, by applying the appropriate maps to the coefficients.
Corollary 4.11. (i) Each of $\Phi_o(\tau), \Phi_{o/k}(\tau)$ and $\Phi_{o/Y}(\tau)$ is (the $q$-expansion of) a holomorphic modular form of weight $3/2$, level $4DB$ and trivial character. (ii) Under the hypotheses of Theorem 4.10, the Shimura lift of each of these modular forms is equal to its unitary counterpart. In particular, the series $\Phi_{o/deg}, \Phi_{o/k}$ and $\Phi_{o/Y}$ are the $q$-expansions of modular forms of weight $2$.

Proof. (i) The modularity of these three generating series is proved in [KRY, §4]; specifically, see Equation (4.2.12), Theorem 4.5.1 and Theorem 4.3.4 there. (ii) This follows immediately by applying the appropriate map to both sides of (4.25).

Remark 4.12. Consider the more natural generating series

$$\Phi_{o/naive} = C + \frac{1}{2h(k)} \sum_{n > 0} \mathbb{Z}(n, \phi) q^n \in \text{Div}(\mathbb{C}_B/o)[q] \otimes \mathbb{Q}$$

which omits the $\mathbb{Z}^*$ terms appearing in $\Phi^u$. On account of the general philosophy of such generating series, one may be tempted to conjecture that for an appropriate constant term $C$, the series $\Phi_{o/naive}$ is already ‘modular’ (to fix ideas, we may take this to mean that the application of a linear functional which factors through the Chow group yields the $q$-expansion of a modular form). If this is indeed the case, we may explain the modularity of $\Phi^u$ in the following way. For an integer $d$, let $U_d$ and $B_d$ denote the following operators on formal $q$-expansions: if $F = \sum_{n \geq 0} a(n)q^n$, then

$$U_d(F) = \sum a(dn)q^n, \quad B_d(F) = \sum a(n)q^{dn}.$$  

When $F$ is the $q$-expansion of a modular form, these operators are induced by the maps $U_d = U_d^1$ and $B_d$ described in [Li]; both are maps between spaces of modular forms that preserve the weight, but may change the level. Set

$$\varphi_d := B_d \circ (1 - U_d), \quad \varphi_d(F) = \sum (a(n) - a(dn)) q^{dn}$$

and note that if $(d,d') = 1$, then $\varphi_d$ and $\varphi_{d'}$ commute. Next, let $\mathcal{P}(D_B)$ denote the set of primes dividing $D_B$, and for a non-empty subset $I = \{p_1, \ldots, p_n\} \subset \mathcal{P}(D_B)$, we define

$$\varphi_I := \varphi_{p_1} \circ \varphi_{p_2} \circ \cdots \circ \varphi_{p_n}.$$  

Then a direct calculation reveals that at the level of formal generating series,

$$\Phi^u = C' + \sum_{[\phi]} \left( 2 + \sum_{I \subset \mathcal{P}(D_B)} \varphi_I \right) (\Phi_{o/naive}),$$

where

$$C' = \frac{i}{2\pi} L(1, \chi'_k) \cdot K - 2 \cdot \# |\text{Opt} / \mathcal{O}_B | \cdot C.$$
In particular, if $C$ happens to take on the value

$$C = \frac{i}{4\pi} \frac{L(1, \chi')}{\# \text{Opt} / \mathcal{O}_B^\times} \cdot K = -\frac{2^{2s(D_B)}}{2^{\omega(D_B)}} \prod_{\ell \mid D_B} \left( \frac{2\ell^2 + \ell - 1}{\ell^2} \right) \cdot K,$$

then $C' = 0$; this would imply that whenever the application of a suitable functional to the coefficients of $\Phi^{\text{even}}$ yields a modular form, the same is true for $\Phi^{\text{odd}}$. It would be interesting to know if this implication can be reversed, and also if there is a geometric interpretation to the value of $C$ given above.

References


Siddarth Sankaran
Mathematisches Institut
der Universität Bonn
Endenicher Allee 60
53115 Bonn, Germany
sankaran@math.uni-bonn.de

*Documenta Mathematica* 18 (2013) 1403–1464
Stability of the Tangent Bundle of the Wonderful Compactification of an Adjoint Group

Indranil Biswas and S. Senthamarai Kannan

Received: April 3, 2013
Communicated by Thomas Peternell

Abstract. Let $G$ be a complex linear algebraic group which is simple of adjoint type. Let $\overline{G}$ be the wonderful compactification of $G$. We prove that the tangent bundle of $\overline{G}$ is stable with respect to every polarization on $\overline{G}$.

2010 Mathematics Subject Classification: 32Q26, 14M27, 14M17
Keywords and Phrases: Wonderful compactification, adjoint group, stability, irreducibility, equivariant bundle

1. Introduction

De Concini and Procesi constructed compactifications of complex simple groups of adjoint type, which are known as wonderful compactifications. These compactifications have turned out to be very useful objects. Our aim here is to investigate equivariant vector bundles on a wonderful compactification. One of the key concepts associated to a vector bundle on a projective variety is the notion of stability introduced by Mumford.

We prove the following (see Theorem 1.1):

Theorem 1.1. Let $\overline{G}$ be the wonderful compactification of a complex simple group $G$ of adjoint type. Take any polarization $L$ on $\overline{G}$. Then the tangent bundle of $\overline{G}$ is stable with respect to $L$.

Theorem 1.1 is proved using a result proved here on equivariant vector bundles over $\overline{G}$ which we will now explain.

Take $G$ as in Theorem 1.1. Let $\tilde{G}$ be the universal cover of $G$. The action of $G \times G$ on $\overline{G}$ produces an action of $\tilde{G} \times \tilde{G}$ on $\overline{G}$. A holomorphic vector bundle on $\overline{G}$ is called equivariant if it is equipped with a lift of the action of $\tilde{G} \times \tilde{G}$; see Definition 2.1 for the details. Let $e_0 \in G$ be the identity element. The group $\tilde{G}$ is the connected component, containing the identity element, of the isotropy group of $e_0$ for the action of $\tilde{G} \times \tilde{G}$ on $\overline{G}$. If $(E, \gamma)$ is an equivariant
vector bundle on \( \tilde{G} \), then the action \( \gamma \) of \( \tilde{G} \times \tilde{G} \) on \( E \) produces an action of \( \tilde{G} \) on the fiber \( E_{e_0} \).

We prove the following (see Proposition 3.3):

**Proposition 1.2.** Let \((E, \gamma)\) be an equivariant vector bundle of rank \( r \) on \( \tilde{G} \) such that the action of \( \tilde{G} \) on \( E_{e_0} \) is irreducible. Then either \( E \) is stable or there is a holomorphic line bundle \( \xi \) on \( \tilde{G} \) such that \( E \) is holomorphically isomorphic to \( \xi \oplus r \).

We show that the tangent bundle \( T\tilde{G} \) is not isomorphic to \( \xi \oplus d \), where \( \xi \) is some holomorphic line bundle on \( \tilde{G} \), and \( d = \text{dim}_C G \). In view of this result, Theorem 1.1 follows from Proposition 1.2.

A stable vector bundle admits an irreducible Einstein–Hermitian connection. It would be very interesting to be able to describe the Einstein–Hermitian structure of the tangent bundle of \( \tilde{G} \).

In [Ka], Kato has carried out a detailed investigation of the equivariant bundles on partial compactifications of reductive groups.

## 2. Equivariant vector bundles on \( \tilde{G} \)

Let \( G \) be a connected linear algebraic group defined over \( \mathbb{C} \) such that the Lie algebra of \( G \) is simple and the center of \( G \) is trivial. In other words, \( G \) is simple of adjoint type. The group \( G \times G \) acts on \( G \): the action of any \((g_1, g_2) \in G \times G\) is the map \( y \mapsto g_1yg_2^{-1} \).

Let \( \tilde{G} \) be the wonderful compactification of \( G \) [DP]. A key property of the wonderful compactification of \( G \) is that the above action of \( G \times G \) on \( G \) extends to an action of \( \tilde{G} \times \tilde{G} \) on \( \tilde{G} \).

Let

\[
\pi : \tilde{G} \longrightarrow G
\]

be the universal cover. Using the projection \( \pi \) in (2.1), the above mentioned action of \( G \times G \) on \( \tilde{G} \) produces an action of \( \tilde{G} \times \tilde{G} \) on \( \tilde{G} \).

\[
\beta : \tilde{G} \times \tilde{G} \longrightarrow \text{Aut}^0(\tilde{G}),
\]

where \( \text{Aut}^0(\tilde{G}) \) is the connected component, containing the identity element, of the group of automorphisms of the variety \( \tilde{G} \).

**Definition 2.1.** An *equivariant* vector bundle on \( \tilde{G} \) is a pair \((E, \gamma)\), where \( E \) is a holomorphic vector bundle on \( \tilde{G} \) and

\[
\gamma : \tilde{G} \times \tilde{G} \times E \longrightarrow E
\]

is a holomorphic action of \( \tilde{G} \times \tilde{G} \) on the total space of \( E \), such that the following two conditions hold:

1. The projection of \( E \) to \( \tilde{G} \) intertwines the actions of \( \tilde{G} \times \tilde{G} \) on \( E \) and \( \tilde{G} \), and
2. The action of \( \tilde{G} \times \tilde{G} \) on \( E \) preserves the linear structure of the fibers of \( E \).
Note that the first condition in Definition 2.1 implies that the action of any $g \in \tilde{G} \times \tilde{G}$ sends a fiber $E_x$ to the fiber $E_{\beta(g)(x)}$, where $\beta$ is the homomorphism in (2.2). The second condition in Definition 2.1 implies that the self-map of $E$ defined by $v \mapsto \gamma(g, v)$ is a holomorphic isomorphism of the vector bundle $E$ with the pullback $\beta(g^{-1})^* E$. Therefore, if $(E, \gamma)$ is an equivariant vector bundle on $\tilde{G}$, then for every $g \in \tilde{G} \times \tilde{G}$, the pulled back holomorphic vector bundle $\beta(g)^* E$ is holomorphically isomorphic to $E$. The following proposition is a converse statement of it.

**Proposition 2.2.** Let $E$ be a holomorphic vector bundle on $\tilde{G}$ such that for every $g \in \tilde{G} \times \tilde{G}$, the pulled back holomorphic vector bundle $\beta(g)^* E$ is holomorphically isomorphic to $E$. Then there is a holomorphic action $\gamma$ of $\tilde{G} \times \tilde{G}$ on $E$ such that the pair $(E, \gamma)$ is an equivariant vector bundle on $\tilde{G}$.

**Proof.** Let $\text{Aut}(E)$ denote the group of holomorphic automorphisms of the vector bundle $E$ over the identity map of $\tilde{G}$. This set $\text{Aut}(E)$ is the Zariski open subset of the affine space $H^0(\tilde{G}, E \otimes E^\vee)$ defined by the locus of invertible endomorphisms of $E$. Therefore, $\text{Aut}(E)$ is a connected complex algebraic group.

Let $\tilde{\text{Aut}}(E)$ denote the set of all pairs of the form $(g, f)$, where $g \in \tilde{G} \times \tilde{G}$ and $f : \beta(g^{-1})^* E \to E$ is a holomorphic isomorphism of vector bundles. This set $\tilde{\text{Aut}}(E)$ has a tautological structure of a group

$$(g_2, f_2) \cdot (g_1, f_1) = (g_2 g_1, f_2 \circ f_1).$$

We will show that it is a connected complex algebraic group.

Let $p_1 : \tilde{G} \times \tilde{G} \times \tilde{G} \to \tilde{G}$ be the projection to the last factor. Let

$$\tilde{\beta} : \tilde{G} \times \tilde{G} \times \tilde{G} \to \tilde{G}$$

be the algebraic morphism defined by $(g, y) \mapsto \beta(g^{-1})(y)$, where $g \in \tilde{G} \times \tilde{G}$ and $y \in \tilde{G}$. Let

$$q : \tilde{G} \times \tilde{G} \times \tilde{G} \to \tilde{G} \times \tilde{G}$$

be the projection to the first two factors. Now consider the direct image

$$\mathcal{E} := q_* (p_1^* E \otimes (\beta^* E)^\vee) \to \tilde{G} \times \tilde{G}.$$

It is locally free. The set $\tilde{\text{Aut}}(E)$ is a Zariski open subset of the total space of the algebraic vector bundle $\mathcal{E}$. Therefore, $\tilde{\text{Aut}}(E)$ is a connected complex algebraic group.

The Lie algebra of $G$ will be denoted by $\mathfrak{g}$. The Lie algebra of $\tilde{\text{Aut}}(E)$ will be denoted by $\mathfrak{A}(E)$. We have a short exact sequence of groups

$$e \to \text{Aut}(E) \to \tilde{\text{Aut}}(E) \xrightarrow{\rho} \tilde{G} \times \tilde{G} \to e,$$

where $\rho$ sends any $(g, f)$ to $g$. Let

$$\rho' : \mathfrak{A}(E) \to \mathfrak{g} \oplus \mathfrak{g}$$
be the homomorphism of Lie algebras corresponding to \( \rho \) in (2.3). Since \( g \oplus g \) is semisimple, there is a homomorphism of Lie algebras

\[
\tau : g \oplus g \longrightarrow A(E)
\]
such that

\[
(2.5) \quad \rho' \circ \tau = \text{Id}_{g \oplus g}
\]

[Bo, p. 91, Corollaire 3]. Fix such a homomorphism \( \tau \) satisfying (2.5). Since the group \( \tilde{G} \times \tilde{G} \) is simply connected, there is a unique holomorphic homomorphism

\[
\tilde{\tau} : \tilde{G} \times \tilde{G} \longrightarrow \tilde{\text{Aut}}(E)
\]
such that the corresponding homomorphism of Lie algebras coincides with \( \tau \).

From (2.5) it follows immediately that

\[
\rho \circ \tilde{\tau} = \text{Id}_{\tilde{G} \times \tilde{G}}.
\]

We now note that \( \tilde{\tau} \) defines an action of \( \tilde{G} \times \tilde{G} \) on \( E \). The pair \( (E, \tilde{\tau}) \) is an equivariant vector bundle. \( \square \)

3. IRREDUCIBLE REPRESENTATIONS AND STABILITY

Fix a very ample class \( L \in \text{NS}(\overline{G}) \), where \( \text{NS}(\overline{G}) \) is the Néron–Severi group of \( \overline{G} \). The degree of a torsionfree coherent sheaf \( F \) on \( \overline{G} \) is defined to be

\[
\text{degree}(F) := (c_1(F) \cup c_1(L)^{d-1}) \cap [\overline{G}] \in \mathbb{Z},
\]

where \( d = \dim C \). If \( \text{rank}(F) \geq 1 \), then

\[
\mu(F) := \frac{\text{degree}(F)}{\text{rank}(F)} \in \mathbb{Q}
\]
is called the slope of \( F \).

A holomorphic vector bundle \( F \) over \( \overline{G} \) is called stable (respectively, semistable) if for every nonzero coherent subsheaf \( F' \subset F \) with \( \text{rank}(F') < \text{rank}(F) \), the inequality

\[
\mu(F') < \mu(F) \quad \text{(respectively,} \quad \mu(F') \leq \mu(F))
\]
holds. A holomorphic vector bundle on \( \overline{G} \) is called polystable if it is a direct sum of stable vector bundles of same slope.

Let

\[
e_0 \in G \subset \overline{G}
\]
be the identity element. Let \( \text{Iso}_{e_0} \subset \tilde{G} \times \tilde{G} \) be the isotropy subgroup of \( e_0 \) for the action of \( \tilde{G} \times \tilde{G} \) on \( \overline{G} \). The connected component of \( \text{Iso}_{e_0} \) containing the identity element is \( \tilde{G} \).

If \( (E, \gamma) \) is an equivariant vector bundle on \( \overline{G} \), then \( \gamma \) gives an action of \( \text{Iso}_{e_0} \) on the fiber \( E_{e_0} \). In particular, we get an action of \( \tilde{G} \) on \( E_{e_0} \).

**Lemma 3.1.** Let \( (E, \gamma) \) be an equivariant vector bundle on \( \overline{G} \) such that the above action of \( \tilde{G} \) on \( E_{e_0} \) is irreducible. Then the vector bundle \( E \) is polystable.
(3.1) \[ E_1 \subset \cdots \subset E_n = E \]

be the Harder–Narasimhan filtration of \( E \) [HL, p. 16, Theorem 1.3.4]. Since \( \tilde{G} \times \tilde{G} \) is connected, the action of \( \tilde{G} \times \tilde{G} \) on \( G \) preserves the Néron–Severi class \( L \). Therefore, the filtration in (3.1) is preserved by the action of \( \tilde{G} \times \tilde{G} \) on \( E \).

Note that \((E_1)_{e_0} \neq 0\) because in that case \( E_1|_{E} = 0 \) by the equivariance of \( E \), which in turn implies that \( E_1 = 0 \). Now, from the irreducibility of the action of \( \tilde{G} \) on \( E_{e_0} \) we conclude that \((E_1)_{e_0} = E_{e_0} \). In particular, \( \text{rank}(E_1) = \text{rank}(E) \).

This implies that \( E_1 = E \). Hence \( E \) is semistable.

Let \( F \subset E \) be the unique maximal polystable subsheaf of the semistable vector bundle \( E \) [HL, p. 23, Theorem 1.5.9]. From the uniqueness of \( F \) and the connectivity of \( \tilde{G} \times \tilde{G} \) we conclude that \( F \) is preserved by the action of \( \tilde{G} \times \tilde{G} \) on \( E \). Just as done above, using the irreducibility of the action of \( \tilde{G} \) on \( E_{e_0} \) we conclude that \( F_{e_0} = E_{e_0} \). Hence \( F = E \), implying that \( E \) is polystable. \( \square \)

The following lemma is well-known.

**Lemma 3.2.** Let \( V_1 \) and \( V_2 \) be two finite dimensional irreducible complex \( \tilde{G} \)-modules such that both \( V_1 \) and \( V_2 \) are nontrivial. Then the \( \tilde{G} \)-module \( V_1 \otimes V_2 \) is not irreducible.

**Proposition 3.3.** Let \((E, \gamma)\) be an equivariant vector bundle of rank \( r \) on \( G \) such that the action of \( \tilde{G} \) on \( E_{e_0} \) is irreducible. Then either \( E \) is stable or there is a holomorphic line bundle \( \xi \) on \( \overline{G} \) such that \( E \) is holomorphically isomorphic to \( \xi^{\oplus r} \).

**Proof.** From Lemma 3.1 we know that \( E \) is polystable. Therefore, there are distinct stable vector bundles \( F_1, \ldots, F_\ell \) and positive integers \( n_1, \ldots, n_\ell \), such that \( \mu(F_i) = \mu(E) \) for every \( i \) and

\[ E = \bigoplus_{i=1}^{\ell} F_i^{\oplus n_i}. \]

We emphasize that \( F_i \neq F_j \) if \( i \neq j \). The vector bundles \( F_1, \ldots, F_\ell \) are uniquely determined by \( E \) up to a permutation of \( \{1, \ldots, \ell\} \) [AU, p. 315, Theorem 2].

Fix a holomorphic isomorphism between the two vector bundles in the two sides of (3.2). Take \( F_i \) and \( F_j \) with \( i \neq j \). Since they are nonisomorphic stable
vector bundles of same slope, we have
\[ H^0(\mathcal{G}, F_j \otimes F_j^\vee) = 0 = H^0(\mathcal{G}, F_i \otimes F_i^\vee). \]
Consequently, for every \( i \in \{1, \cdots, \ell\} \), there is a unique subbundle of \( E \) which is isomorphic to \( F_i^{\oplus n_i} \). Using this it follows that for any \( g \in \tilde{G} \times \tilde{G} \), and any \( j \in \{1, \cdots, \ell\} \), there is a \( k \in \{1, \cdots, \ell\} \) such that the action of \( g \) on \( E \) takes the subbundle \( F_j^{\oplus n_j} \) to \( F_k^{\oplus n_k} \). Since \( \tilde{G} \times \tilde{G} \) is connected, this implies that the action \( \gamma \) of \( \tilde{G} \times \tilde{G} \) on \( E \) preserves the subbundle \( F_i^{\oplus n_i} \) for every \( i \in \{1, \cdots, \ell\} \).

Now from the irreducibility of the action of \( \tilde{G} \times \tilde{G} \) on \( E_{e_0} \) we conclude that \( \ell = 1 \).

We will denote \( F_1 \) and \( n_1 \) by \( F \) and \( n \) respectively. So
\[(3.3) \quad F = F^{\oplus n}. \]
Since for every \( g \in \tilde{G} \times \tilde{G} \), the pulled back holomorphic vector bundle \( \beta(g)^*(F^{\oplus n}) \) is holomorphically isomorphic to \( F^{\oplus n} \), using [At, p. 315, Theorem 2] and the fact that \( F \) is indecomposable (recall that \( F \) is stable), we conclude that the pulled back holomorphic vector bundle \( \beta(g)^*F \) is holomorphically isomorphic to \( F \) for every \( g \in \tilde{G} \times \tilde{G} \). Therefore, from Proposition 2.2 we know that there is an action \( \delta \) of \( \tilde{G} \times \tilde{G} \) on \( F \) such that \( (F, \delta) \) is an equivariant vector bundle on \( \mathcal{G} \).

The actions \( \gamma \) and \( \delta \) together define an action of \( \tilde{G} \times \tilde{G} \) on the vector bundle \( \text{Hom}(F, E) = E \otimes F^\vee \).

This action of \( \tilde{G} \times \tilde{G} \) on \( \text{Hom}(F, E) \) produces an action of \( \tilde{G} \times \tilde{G} \) on the vector space \( H^0(\mathcal{G}, \text{Hom}(F, E)) \).

In view of (3.3), we have a canonical isomorphism
\[(3.4) \quad E = F \otimes \mathbb{C} H^0(\mathcal{G}, \text{Hom}(F, E)). \]
This isomorphism sends any \( (v, \sigma) \in (F_x, H^0(\mathcal{G}, \text{Hom}(F, E))) \) to the evaluation \( \sigma_x(v) \in E_x \). The isomorphism in (3.4) is \( \tilde{G} \times \tilde{G} \)-equivariant. Since the action of \( \tilde{G} \) on \( E_{e_0} \) is irreducible, from Lemma 3.2 we conclude that either \( \text{rank}(F) = 1 \) or
\[ \dim H^0(\mathcal{G}, \text{Hom}(F, E)) = 1. \]
If \( \dim H^0(\mathcal{G}, \text{Hom}(F, E)) = 1 \), then from (3.1) and the fact that \( F \) is stable it follows immediately that \( E \) is stable. If \( \text{rank}(F) = 1 \), then from (3.4) it follows that
\[ E = F^{\oplus r}, \]
where \( r = \text{rank}(E) \).

\[ \square \]

**Remark 3.4.** Let \( V \) be any irreducible \( G \)-module. Consider the trivial right action of \( \tilde{G} \) on \( V \) as well as the left action of \( \tilde{G} \) on \( V \) given by the combination of the action of \( G \) on \( V \) and the projection in (2.1). Therefore, we get the diagonal action of \( \tilde{G} \times \tilde{G} \) on the trivial vector bundle \( \tilde{G} \times V \) over \( \mathcal{G} \). Consequently, the trivial vector bundle \( \tilde{G} \times V \) gets the structure on an equivariant vector bundle.
We note that the action of $\tilde{G} \subset \text{Iso}_{e_0}$ on the fiber of $\overline{G} \times V$ over the point $e_0$ is irreducible because the $G$–module $V$ is irreducible.

4. The tangent bundle

**Theorem 4.1.** Let $L \in \text{NS}(\overline{G})$ be any ample class. The tangent bundle of $\overline{G}$ is stable with respect to $L$.

*Proof.* Since $G$ is simple, the adjoint action of $\tilde{G}$ on the Lie algebra $\mathfrak{g}$ of $G$ is irreducible. In view of Proposition 3.3, it suffices to show that the tangent bundle $T\overline{G}$ is not of the form $\xi^\oplus d$, where $\xi$ is a holomorphic line bundle on $\overline{G}$ and $d = \dim_{\mathbb{C}} G$.

Assume that

$$T\overline{G} = \xi^\oplus d,$$

where $\xi$ is a holomorphic line bundle on $\overline{G}$.

Since the variety $G$ is unirational (cf. [Gi, p. 4, Theorem 3.1]), the compactification $\overline{G}$ is also unirational. Hence $G$ is simply connected [Sa, p. 483, Proposition 1]. As $T\overline{G}$ holomorphically splits into a direct sum of line bundles (see (4.1)) and $G$ is simply connected, it follows that $\overline{G} = (\mathbb{C}\mathbb{P}^1)^d$.

[BPT] p. 242, Theorem 1.2. But the tangent bundle of $(\mathbb{C}\mathbb{P}^1)^d$ is not of the form $\xi^\oplus d$ (see (4.1)); although the tangent bundle of $(\mathbb{C}\mathbb{P}^1)^d$ is a direct sum of line bundles, the line bundles in its decomposition are not isomorphic. Therefore, $T\overline{G}$ is not of the form $\xi^\oplus d$. This completes the proof. □

Acknowledgements

We thank the referee and D. S. Nagaraj for some comments. This work was carried out when both the authors were visiting KSOM, Calicut. We thank KSOM for hospitality. The first-named author acknowledges the support of the J. C. Bose Fellowship.

References


Holomorphic Connections on Filtered Bundles over Curves

Indranil Biswas and Viktoria Heu

Received: July 1, 2013

Communicated by Edward Frenkel

Abstract. Let $X$ be a compact connected Riemann surface and $E_P$ a holomorphic principal $P$–bundle over $X$, where $P$ is a parabolic subgroup of a complex reductive affine algebraic group $G$. If the Levi bundle associated to $E_P$ admits a holomorphic connection, and the reduction $E_P \subset E_P \times^P G$ is rigid, we prove that $E_P$ admits a holomorphic connection. As an immediate consequence, we obtain a sufficient condition for a filtered holomorphic vector bundle over $X$ to admit a filtration preserving holomorphic connection. Moreover, we state a weaker sufficient condition in the special case of a filtration of length two.

2010 Mathematics Subject Classification: 14H60, 14F05, 53C07

Keywords and Phrases: Holomorphic connection, filtration, Atiyah bundle, parabolic subgroup

1. Introduction

Let $X$ be a compact connected Riemann surface. A holomorphic vector bundle $E$ over $X$ admits a holomorphic connection if and only if every indecomposable component of $E$ is of degree zero [W6, A]. This criterion generalizes to the context of principal bundles over $X$ with a complex reductive affine algebraic group as the structure group [AB1]. Note that since there are no nonzero $(2,0)$–forms on $X$, holomorphic connections on a holomorphic bundle on $X$ are the same as flat connections compatible with the holomorphic structure of the bundle.

Our aim here is to consider flat connections on vector bundles compatible with a given filtration of the bundle. Let

$$0 = E_0 \subset E_1 \subset \cdots \subset E_{\ell-1} \subset E_\ell = E$$

be a filtration of a holomorphic vector bundle $E$ on $X$. If $E$ admits a flat connection

$$D : E \rightarrow E \otimes \Omega^1_X$$
preserving the filtration, meaning $D(E_i) \subset E_i \otimes \Omega^1_X$ for every $i$, then this connection induces a flat connection $D_i$ on each successive quotient $E_i/E_{i-1}$ with $i \in [1, \ell]$. The question is the following: which supplementary condition is needed in order to ensure the existence of a filtration preserving holomorphic connection $D$? Suppose for example that $E$ is semi-stable of degree zero such that each successive quotient in (1.1) admits a flat connection. Then it follows immediately that each subbundle $E_i$, $i \in [1, \ell]$, is also semi-stable of degree zero. According to Corollary 3.10 in [Si], p. 40, the filtered vector bundle $E$ then admits a filtration preserving holomorphic connection $D$. In this paper, we show that the rigidity of the filtration (1.1) is another sufficient supplementary condition for the existence of a filtration preserving holomorphic connection on $E$. We note that a related example is quoted in [Bi] (see [Bi], p. 119, Example 3.6).

More generally, we consider holomorphic connections on principal bundles with a parabolic group as the structure group. Let $P$ be a parabolic subgroup of a complex reductive affine algebraic group $G$, and let $E_P$ be a holomorphic principal $P$-bundle over $X$. Let $L(P) := P/R_u(P)$ be the Levi quotient of $P$, where $R_u(P)$ is the unipotent radical of $P$. Assume that the associated holomorphic principal $L(P)$-bundle $E_P/R_u(P)$ admits a holomorphic connection. We are interested in the question of finding sufficient conditions for the existence of a holomorphic connection on $E_P$. Let $E_P \times_F^P G$ be the holomorphic principal $G$-bundle obtained by extending the structure group $E_P$ using the inclusion of $P$ in $G$. We shall prove that the rigidity of the reduction of structure group $E_P \subset E_P \times_F^P G$ ensures the existence of a holomorphic connection on $E_P$ (see Theorem 2.1).

2. Connections on principal bundles with parabolic structure group

Let $G$ be a connected reductive affine algebraic group defined over $\mathbb{C}$. Let $P \subset G$ be a parabolic subgroup, i.e., $P$ is a Zariski closed connected algebraic subgroup of $G$ such that the quotient variety $G/P$ is complete. The unipotent radical of $P$ will be denoted by $R_u(P)$. The quotient $L(P) := P/R_u(P)$, which is a connected reductive complex affine algebraic group, is called the Levi quotient of $P$. The Lie algebra of $G$ (respectively, $P$) will be denoted by $\mathfrak{g}$ (respectively, $\mathfrak{p}$).

Let $X$ be a compact connected Riemann surface. Let

$$f : E_P \longrightarrow X$$

(2.1)

be a holomorphic principal $P$-bundle. The quotient

$$E_{L(P)} := E_P/R_u(P)$$

(2.2)

is a holomorphic principal $L(P)$-bundle on $X$. We note that $E_{L(P)}$ is identified with the principal $L(P)$-bundle obtained by extending the structure group of $E_P$ using the quotient map $P \longrightarrow L(P)$.

Documenta Mathematica 18 (2013) 1473–1480
Let
\[ E_G := E_P \times^P G \longrightarrow X \]
be the holomorphic principal \( G \)-bundle obtained by extending the structure group of \( E_P \) using the inclusion of \( P \) in \( G \). Let
\[
\text{ad}(E_G) := E_G \times^G \mathfrak{g} \quad \text{and} \quad \text{ad}(E_P) := E_P \times^P \mathfrak{p}
\]
be the adjoint vector bundles for \( E_G \) and \( E_P \) respectively. The reduction of structure group \( E_P \subset E_G \) is called rigid if
\[
H^0(X, \text{ad}(E_G)/\text{ad}(E_P)) = 0.
\]
Let us give a brief geometric interpretation of this property. Recall that the space of infinitesimal deformations of the principal bundle \( E_G \) (respectively, \( E_P \)) can be identified with \( H^1(X, \text{ad}(E_G)) \) (respectively, \( H^1(X, \text{ad}(E_P)) \)) [SU]. We have a short exact sequence of vector bundles
\[
0 \longrightarrow \text{ad}(E_P) \longrightarrow \text{ad}(E_G) \longrightarrow \text{ad}(E_G)/\text{ad}(E_P) \longrightarrow 0.
\]
The rigidity of the reduction of structure group \( E_P \subset E_G \) thus translates as
\[
H^1(X, \text{ad}(E_P)) \hookrightarrow H^1(X, \text{ad}(E_G)),
\]
i.e. the infinitesimal deformations of \( E_P \) are uniquely determined by the infinitesimal deformations of \( E_G \) that they induce. In other words, if we fix the principal bundle \( E_G \), then the parabolic subbundle \( E_P \) cannot be deformed.

**Theorem 2.1.** Assume that the holomorphic principal \( L(P) \)-bundle \( E_{L(P)} \) in \([2.2]\) admits a holomorphic connection, and the reduction of structure group \( E_P \subset E_G \) is rigid. Then the holomorphic principal \( P \)-bundle \( E_P \) admits a holomorphic connection.

**Proof.** Let \( \text{At}(E_P) := (f_*TE_P)^P \subset f_*TE_P \) be the Atiyah bundle for \( E_P \), where \( f \) is the projection in \([2.1]\) [AD]. It fits in a short exact sequence of holomorphic vector bundles on \( X \)
\[
0 \longrightarrow \text{ad}(E_P) \longrightarrow \text{At}(E_P) \overset{p_0}{\longrightarrow} TX \longrightarrow 0,
\]
where \( p_0 \) is given by the differential \( df : T_E_P \longrightarrow f^*TX \) of \( f \). We recall that a holomorphic connection on \( E_P \) is a holomorphic splitting of \([2.3]\) [AD].

Let \( R_n(p) \) be the Lie algebra of the unipotent radical \( R_n(P) \). We note that \( R_n(p) \) is the nilpotent radical of the Lie algebra \( p \). Let
\[
V_0 := E_P \times^P R_n(p) \longrightarrow X
\]
be the holomorphic vector bundle associated to the principal \( P \)-bundle \( E_P \) for the \( P \)-module \( R_n(p) \).

Let \( \tilde{f} : E_{L(P)} \longrightarrow X \) be the projection induced by \( f \). Let
\[
\text{At}(E_{L(P)}) := (\tilde{f}_*TE_{L(P)})^{L(P)} \subset \tilde{f}_*TE_{L(P)}
\]
be the Atiyah bundle for $E_{L(P)}$. We have a commutative diagram

\[
\begin{array}{ccccccccc}
0 & \rightarrow & \text{ad}(E_{P}) & \rightarrow & \text{At}(E_{P}) & \rightarrow & TX & \rightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \rightarrow & \text{ad}(E_{L(P)}) & \rightarrow & \text{At}(E_{L(P)}) & \rightarrow & TX & \rightarrow & 0 \\
\end{array}
\]

(2.5)

where $\mathcal{V}_0$ is defined in (2.4).

By assumption, $E_{L(P)}$ admits a holomorphic connection. Hence there is a holomorphic homomorphism

\[
\beta : TX \rightarrow \text{At}(E_{L(P)})
\]

(2.6)

such that $p_1 \circ \beta = \text{Id}_{TX}$, where $p_1$ is the projection in (2.5). Therefore, we have a short exact sequence of holomorphic vector bundles

\[
\begin{array}{ccccccccc}
0 & \rightarrow & \mathcal{V}_0 & \rightarrow & \mathcal{V} := q^{-1}(\beta(TX)) & \overset{p_0}{\rightarrow} & TX & \rightarrow & 0 \\
\end{array}
\]

(2.7)

where $q$ is the projection in (2.5).

The short exact sequence in (2.7) splits holomorphically if the the short exact sequence in (2.3) splits holomorphically. The obstruction for splitting of (2.7) is a cohomology class

\[
\psi \in H^1(X, \mathcal{V}_0 \otimes (TX)^*) = H^0(X, \mathcal{V}_0^*)
\]

(2.8)

(by Serre duality).

Since the group $G$ is reductive, its Lie algebra $\mathfrak{g}$ has a $G$-invariant symmetric non-degenerate bilinear form. For example, let $B$ be the direct sum of the Killing form on $[\mathfrak{g}, \mathfrak{g}]$ and a symmetric non-degenerate bilinear form on the center of $\mathfrak{g}$. Note that $\mathfrak{p} \subset R_n(\mathfrak{p})^\perp$ (the annihilator of $R_n(\mathfrak{p})^\perp$) and actually

\[
\mathfrak{p} = R_n(\mathfrak{p})^\perp
\]

since they have the same dimension. We thus have

\[
R_n(\mathfrak{p})^\ast = \mathfrak{g}/R_n(\mathfrak{p})^\perp = \mathfrak{g}/\mathfrak{p}.
\]

As the above isomorphism between $R_n(\mathfrak{p})^*$ and $\mathfrak{g}/\mathfrak{p}$ is $P$-equivariant, it follows that

\[
\mathcal{V}_0^* = E_{P} \times^P R_n(\mathfrak{p})^* = \text{ad}(E_G)/\text{ad}(E_{P}).
\]

Now the given condition that $E_{P} \subset E_G$ is rigid implies that that

\[
H^0(X, \mathcal{V}_0^*) = 0.
\]
Therefore, $\psi$ in (2.8) vanishes. Consequently, the short exact sequence in (2.7) splits, implying that the short exact sequence in (2.3) splits.

Some criteria for the existence of a holomorphic connection on $E_{L(P)}$ can be found in [AB1] and [AB2]. Theorem 2.1 has the following immediate corollary:

**Corollary 2.2.** Let

$$0 = E_0 \subset E_1 \subset \cdots \subset E_{\ell-1} \subset E_{\ell} = E$$

be a filtration of holomorphic vector bundles on $X$, and let $\text{End}(E_{\bullet}) \subset \text{End}(E)$ be the subbundle defined by the sheaf of filtration preserving endomorphisms. Assume that each successive quotient $E_i/E_{i-1}$, with $i \in [1, \ell]$, admits a holomorphic connection, and

$$H^0(X, \text{End}(E)/\text{End}(E_{\bullet})) = 0.$$  

Then $E$ admits a holomorphic connection $D$ such that $D$ preserves each subbundle $E_i$ with $i \in [1, \ell]$.

Note that (2.9) is not a necessary condition for the existence of a filtration preserving connection $D$, as one can see by the example of trivial bundles filtered by trivial subbundles. In the next section, we state a weaker sufficient condition when the of length $\ell$ of the filtration is two.

### 3. Holomorphic Connections on Extensions

Let $E$ and $F$ be holomorphic vector bundles on $X$ admitting holomorphic connections. A holomorphic connection on $E$ and a holomorphic connection on $F$ together define a holomorphic connection on the vector bundle $\text{Hom}(E,F) = E^* \otimes F$.

**Proposition 3.1.** Assume that $E$ and $F$ admit holomorphic connections $D_E$ and $D_F$ respectively, such that every holomorphic section of $\text{Hom}(E,F)$ is flat with respect to the connection on $\text{Hom}(E,F)$ given by $D_E$ and $D_F$. Then for any holomorphic extension

$$0 \rightarrow E \rightarrow W \rightarrow F \rightarrow 0,$

the holomorphic vector bundle $W$ admits a holomorphic connection that preserves the subbundle $E$.

**Proof.** Let $r_1$ and $r_2$ be the ranks of $E$ and $F$ respectively. Take the group

$$G = \text{GL}(r_1 + r_2, \mathbb{C})$$

let $P \subset G$ be the parabolic subgroup that preserves the subspace $\mathbb{C}^{r_1} \subset \mathbb{C}^{r_1+r_2}$ given by the first $r_1$ vectors of the standard basis. We note that $L(P) = \text{GL}(r_1) \times \text{GL}(r_2)$. Take an extension $W$ as in the proposition. Then the pair $(W, E)$ defines a holomorphic principal $P$–bundle $E_P$ over $X$ and $E \oplus F$ defines the associated $L(P)$–bundle $E_{L(P)}$. The holomorphic connection $D_{E_P} \oplus D_{E_P}$ on $E \oplus F$ gives a section $\beta$ as in (2.6).
After we fix the above set-up, the vector bundle \( V_0 \) in (2.4) is \( E \otimes F^* \). Consider \( \psi \in H^1(X, E \otimes F^* \otimes K_X) = H^0(X, E^* \otimes F)^* = H^0(X, \text{Hom}(E, F))^* \) in (2.8). Given any \( T \in H^0(X, \text{Hom}(E, F)) \), we will explicitly describe the evaluation \( \psi(T) \in \mathbb{C} \).

Fix a \( C^\infty \) splitting

\[
\begin{pmatrix}
\overline{\partial}_E & A \\
0 & \overline{\partial}_F
\end{pmatrix},
\]

where \( A \) is a smooth section of \( \text{Hom}(F, E) \otimes \Omega^{0,1}_X \).

Let \( D_{F,E} \) be the holomorphic connection on \( \text{Hom}(F, E) \) given by \( D_E \) and \( D_F \). We have

\[
D_{F,E}(A) \in C^\infty(X; \text{Hom}(F, E) \otimes \Omega^{1,1}_X).
\]

Take any \( T \in H^0(X, \text{Hom}(E, F)) \). We will show that

(3.2) \[
\psi(T) = \int_X \text{trace}(D_{F,E}(A) \circ T) \in \mathbb{C}.
\]

To prove this, consider the holomorphic connection \( D_E \oplus D_F \) on \( E \oplus F \). Using the \( C^\infty \) isomorphism in (3.1), this connection produces a \( C^\infty \) connection \( \nabla^W \) on \( W \). We should clarify that \( \nabla^W \) is holomorphic if and only if the isomorphism in (3.1) is holomorphic. Let

\[
K(\nabla^W) \in C^\infty(X; \text{End}(W) \otimes \Omega^{1,1}_X)
\]

be the curvature of the connection \( \nabla^W \). Since \( D_E \oplus D_F \) is a flat connection on \( E \oplus F \), and the inclusion of \( E \) in \( W \) is holomorphic, it follows that \( K(\nabla^W) \) lies in the subspace

\[
C^\infty(X; E \otimes F^* \otimes \Omega^{1,1}_X) \subset C^\infty(X; \text{End}(W) \otimes \Omega^{1,1}_X)
\]

constructed using the inclusion of the vector bundle \( \text{Hom}(F, E) \) in \( \text{End}(W) \). From the definition of the cohomology class \( \psi \in H^1(X, E \otimes F^* \otimes K_X) \) it follows that the Dolbeault cohomology class in \( H^1(X, E \otimes F^* \otimes K_X) \) represented by the form \( K(\nabla^W) \) represents a form \( \psi \). On the other hand, the form \( D_{F,E}(A) \in C^\infty(X; E \otimes F^* \otimes \Omega^{1,1}_X) \) coincides with \( \psi \). Therefore, the equality in (3.2) follows. We note that \( \int_X \text{trace}(D_{F,E}(A) \circ T) \) is independent of the choice of the homomorphism \( \eta \). Indeed, for a different choice of \( \eta \), the section \( A \) is replaced by
$A + \partial_{E\otimes F^*}(A')$, where $A'$ is a smooth section of $\text{Hom}(F, E)$, and $\partial_{F,E}$ is the Dolbeault operator defining the holomorphic structure of $\text{Hom}(F, E)$. Now

$$\int_X \text{trace}(D_{F,E}(\partial_{F,E}(A')) \circ T) = \int_X \text{trace}(\partial_{F,E}(D_{F,E}(A')) \circ T)$$

since the connection $D_{F,E}$ is flat and compatible with the holomorphic structure, and we also have

$$\int_X \text{trace}(\partial_{F,E}(D_{F,E}(A')) \circ T) = \int_X \partial(\text{trace}(D_{F,E}(A') \circ T)) = 0$$

because the section $T$ is holomorphic. Therefore, $\int_X \text{trace}(D_{F,E}(A) \circ T)$ is independent of the choice of $\eta$.

We also note that $\text{trace}(D_{F,E}(A) \circ T) = \text{trace}(T \circ D_{E,F}(A))$. Let $D_{E,E}$ be the holomorphic connection on $\text{End}(E)$ induced by $D_E$. Let $D_{E,F}$ be the holomorphic connection on $\text{Hom}(E, F)$ induced by $D_E$ and $D_F$. Note that

$$D_{E,E}(T) = 0$$

by the condition given in the proposition. Therefore, we have

$$D_{F,E}(A) \circ T = D_{F,E}(A) \circ T + A \circ D_{E,F}(T) = D_{E,E}(A \circ T).$$

On the other hand,

$$\int_X \text{trace}(D_{E,E}(A \circ T)) = \int_X \partial(\text{trace}(A \circ T)) = 0.$$

Combining these, from (3.2), it follows that $\psi = 0$. The principal $P$–bundle $E_P$ thus admits a holomorphic connection. In other words, the holomorphic vector bundle $W$ admits a holomorphic connection that preserves the subbundle $E$.

**Corollary 3.2.** Let $E$ be a holomorphic vector bundle on $X$ of degree zero such that

$$H^0(X, \text{End}(E)) = \mathbb{C} \cdot \text{Id}_E.$$

Then given any short exact sequence of holomorphic vector bundles

$$0 \rightarrow E \rightarrow W \rightarrow E \rightarrow 0,$$

the holomorphic vector bundle $W$ admits a holomorphic connection that preserves the subbundle $E$.

**Proof.** The holomorphic vector bundle $E$ is indecomposable because

$$H^0(X, \text{End}(E)) = \mathbb{C} \cdot \text{Id}_E.$$

Therefore, the given condition that $\text{deg}(E) = 0$ implies that $E$ admits a holomorphic connection [We, A1, p. 203, Theorem 10]. For any holomorphic connection on $E$, the corresponding connection on $\text{End}(E)$ has the property that the section $\text{Id}_E$ is flat with respect to it. Hence Proposition 3.1 completes the proof.
ACKNOWLEDGEMENTS

We are very grateful to Claude Sabbah for going through the paper carefully and making helpful suggestions. The first-named author acknowledges the support of the J. C. Bose Fellowship.

REFERENCES


Indranil Biswas
School of Mathematics
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 400005
India
indranil@math.tifr.res.in

Viktoria Heu
IRMA
UMR 7501
7 rue René-Descartes
67084 Strasbourg Cedex
France
heu@math.unistra.fr
MINIMIZATION OF THE ENERGY OF THE NONRELATIVISTIC
ONE-ELECTRON PAULI-FIERZ MODEL OVER QUASIFREE STATES

VOLKER BACH, SÉBASTIEN BRETEAUX, TIM Tzaneteas

Received: February 14, 2013
Revised: April 30, 2013
Communicated by Heinz Siedentop

ABSTRACT. In this article the existence of a minimizer for the energy for the nonrelativistic one-electron Pauli-Fierz model within the class of quasifree states is established. To this end it is shown that the minimum of the energy on quasifree states coincides with the minimum of the energy on pure quasifree states, where existence and uniqueness of a minimizer holds. Infrared and ultraviolet cutoffs are assumed, along with sufficiently small coupling constant and momentum of the dressed electron. A perturbative expression of the minimum of the energy on quasifree states for a small momentum of the dressed electron and small coupling constant is given. We also express the Lagrange equation for the minimizer in terms of the generalized one particle density matrix of the pure quasifree state.

2010 Mathematics Subject Classification: 81V10, 81L05, 46T99
Keywords and Phrases: Pauli-Fierz model, Quasifree states, Mass renormalization, Minimization techniques

I INTRODUCTION AND MAIN RESULTS

We begin by introducing the mathematical model studied in this paper and mention some well-known results before we describe the main results of the paper.
I.1 The Hamiltonian

According to the Standard Model of Nonrelativistic Quantum Electrodynamics \[4\] the unitary time evolution of a free nonrelativistic particle coupled to the quantized radiation field is generated by the Hamiltonian

\[
\tilde{H}_g := \frac{1}{2} (\frac{\mathbf{i}}{\hbar} \nabla_x - \mathbf{A}(x))^2 + H_f \tag{I.1}
\]

acting on the Hilbert space \(L^2(\mathbb{R}^3; \mathfrak{F})\) of square-integrable functions with values in the photon Fock space \(
\mathfrak{F} := \mathfrak{F}_+ := \bigoplus_{n=0}^\infty \mathfrak{F}_+^{(n)}(h).
\tag{I.2}
\)

Here \(\mathfrak{F}_+^{(0)}(h) = \mathbb{C} \cdot \Omega\) is the vacuum sector and the \(n\)-photon sector \(\mathfrak{F}_+^{(n)}(h) = S_n(h^{\otimes n})\) is the subspace of totally symmetric vectors on the \(n\)-fold tensor product of the one-photon Hilbert space

\[
h = \{ \vec{f} \in L^2(S_{\sigma,\Lambda}; \mathbb{C} \otimes \mathbb{R}^3) \mid \forall \vec{k} \in S_{\sigma,\Lambda}, \text{ a.e. : } \vec{k} \cdot \vec{f}(\vec{k}) = 0 \} \tag{I.3}
\]

of square-integrable, transversal vector fields which are supported in the momentum shell

\[
S_{\sigma,\Lambda} := \{ \vec{k} \in \mathbb{R}^3 \mid \sigma \leq |\vec{k}| \leq \Lambda \}, \tag{I.4}
\]

where \(0 \leq \sigma < \Lambda < \infty\) are infrared and ultraviolet cutoffs, respectively. The condition \(\vec{k} \cdot \vec{f}(\vec{k}) = 0\) reflects our choice of gauge, namely, the Coulomb gauge. It is convenient to fix real polarization vectors \(\vec{\varepsilon}_+^{(\pm)}(\vec{k}) \in \mathbb{R}^3\) such that \(\{\vec{\varepsilon}_+^{(\pm)}(\vec{k}), \vec{\varepsilon}_-^{(\pm)}(\vec{k}), \frac{\vec{k}}{|\vec{k}|}\} \subseteq \mathbb{R}^3\) form a right-handed orthonormal basis (Dreibein) and replace (I.3) by

\[
h = L^2(S_{\sigma,\Lambda} \times \mathbb{Z}_2), \tag{I.5}
\]

with the understanding that \(\vec{f}(\vec{k}) = \vec{\varepsilon}_+ \cdot f(\vec{k},+) + \vec{\varepsilon}_- \cdot f(\vec{k},-).\)

In (I.1) the energy of the photon field is represented by

\[
H_f = \int |k| a^*(k) a(k) \, dk, \tag{I.6}
\]

where \(\int f(k)dk := \sum_{\tau=\pm} \int_{S_{\sigma,\Lambda}} f(\vec{k},\tau) \, d^3k\) and \(\{a(k), a^*(k)\}_{k \in S_{\sigma,\Lambda} \times \mathbb{Z}_2}\) are the usual boson creation and annihilation operators constituting a Fock representation of the CCR on \(\mathfrak{F}\), i.e.,

\[
[a(k), a(k')] = [a^*(k), a^*(k')] = 0, \quad [a(k), a^*(k')] = \delta(k - k') \mathbf{1}, \quad a(k)\Omega = 0, \tag{I.7}
\]

\[
[a(k), a(k')] = \delta(k - k') \mathbf{1}, \quad a(k)\Omega = 0, \tag{I.8}
\]
for all \( k, k' \in S_{\sigma, \Lambda} \times \mathbb{Z}_2 \). The magnetic vector potential \( \vec{A}(\vec{x}) \) is given by

\[
\vec{A}(\vec{x}) = \int \vec{G}(k) \left( e^{-i\vec{k} \cdot \vec{x}} a^*(k) + e^{i\vec{k} \cdot \vec{x}} a(k) \right) dk,
\]

with \( k = (\vec{k}, \tau) \in \mathbb{R}^3 \times \mathbb{Z}_2 \).

The magnetic vector potential \( \vec{A}(\vec{x}) \) is given by

\[
\vec{A}(\vec{x}) = \hat{\vec{G}}(\vec{k}) \left( e^{-i\vec{k} \cdot \vec{x}} a^*(k) + e^{i\vec{k} \cdot \vec{x}} a(k) \right) dk,
\]

with \( k = (\vec{k}, \tau) \in \mathbb{R}^3 \times \mathbb{Z}_2 \).

\[
\hat{G}(k, \tau) := g \varepsilon_{\tau} |\vec{k}|^{-1/2},
\]

and \( g \in \mathbb{R} \) being the coupling constant. In our units, the mass of the particle and the speed of light equal one, so the coupling constant is given as \( g = \frac{1}{4\pi} \sqrt{\alpha} \), with \( \alpha \approx 1/137 \) being Sommerfeld’s fine structure constant.

The Hamiltonian \( \tilde{H}_g \) preserves (i.e., commutes with) the total momentum operator \( \vec{p} = \int \hat{\vec{A}}_x + \vec{P}_f \) of the system, where

\[
\vec{P}_f = \int \hat{k} a^*(k) a(k) dk
\]

is the photon field momentum. This fact allows us to eliminate the particle degree of freedom. More specifically, introducing the unitary operator

\[
\mathbb{U} : L^2(\mathbb{R}^3_x; \mathfrak{g}) \to L^2(\mathbb{R}^3_p; \mathfrak{g}), \quad (\mathbb{U}\Psi)(\vec{p}) := \int e^{-i\vec{x} \cdot \vec{p} - \vec{P}_f} \Psi(\vec{x}) d^3x \quad (2\pi)^{3/2}
\]

one finds that

\[
\mathbb{U} \tilde{H}_g \mathbb{U}^* = \int H_{g, \vec{p}} d^3p,
\]

where

\[
H_{g, \vec{p}} = \frac{1}{2}(\vec{P}_f + \vec{G}(0) - \vec{p})^2 + H_f
\]

is a self-adjoint operator on \( \text{dom}(H_{0, \vec{0}}) \), the natural domain of \( H_{0, \vec{0}} = \frac{1}{2} \vec{P}_f^2 + H_f \).

I.2 Ground State Energy and Bogolubov-Hartree-Fock Energy

Due to (I.13), all spectral properties of \( \tilde{H}_g \) are obtained from those of \( \{ H_{g, \vec{p}} \}_{\vec{p} \in \mathbb{R}^3} \).

Of particular physical interest is the mass shell for fixed total momentum \( \vec{p} \in \mathbb{R}^3 \), coupling constant \( g \geq 0 \), and infrared and ultraviolet cutoffs \( 0 \leq \sigma < \Lambda < \infty \), i.e., the value of the ground state energy

\[
E_{gs}(g, \vec{p}, \sigma, \Lambda) := \inf \sigma[H_{g, \vec{p}}] \geq 0
\]

and the corresponding ground states (or approximate ground states).
We express the ground state energy in terms of density matrices with finite energy expectation value and accordingly introduce

$$\tilde{\mathcal{DM}} := \{ \rho \in L^1(\mathcal{F}) \mid \rho \geq 0, \quad \text{Tr}\left[ \rho H_{0,0} \rho \right] \in L^1(\mathcal{F}) \}, \quad (I.16)$$

so that the Rayleigh-Ritz principle appears in the form

$$E_{gs}(g,\vec{p}) = \inf \left\{ \text{Tr}\left[ \rho H_{g,\vec{p}} \rho \right] \mid \rho \in \tilde{\mathcal{DM}} \right\}. \quad (I.17)$$

Note that \( \text{Tr}\left[ \rho H_{g,\vec{p}} \rho \right] = \text{Tr}\left[ \rho^{1-\beta} H_{g,\vec{p}} \rho^{\beta} \right] \), for all \( 0 \leq \beta \leq 1 \), due to our assumption \( \rho H_{0,0} \in L^1(\mathcal{F}) \).

The determination of \( E_{gs}(g,\vec{p}) \) and the corresponding ground state \( \rho_{gs}(g,\vec{p}) \in \tilde{\mathcal{DM}} \) (provided the infimum is attained) is a difficult task. In this paper we rather study approximations to \( E_{gs}(g,\vec{p}) \) and \( \rho_{gs}(g,\vec{p}) \) that we borrow from the quantum mechanics of atoms and molecules, namely, the Bogolubov-Hartree-Fock (BHF) approximation.

We define the BHF energy as

$$E_{BHF}(g,\vec{p},\sigma,\Lambda) = \inf \left\{ \text{Tr}\left[ \rho H_{g,\vec{p}}(\sigma,\Lambda) \right] \mid \rho \in \mathcal{Q}_F \right\}, \quad (I.18)$$

with corresponding BHF ground state(s) \( \rho_{BHF}(g,\vec{p},\sigma,\Lambda) \in \mathcal{Q}_F \), determined by

$$\text{Tr}\left[ \rho_{BHF}(g,\vec{p},\sigma,\Lambda) H_{g,\vec{p}}(\sigma,\Lambda) \right] = E_{BHF}(g,\vec{p},\sigma,\Lambda), \quad (I.19)$$

where

$$\mathcal{Q}_F := \{ \rho \in \mathcal{DM} \mid \rho \text{ is quasifree} \} \subseteq \mathcal{DM} \quad (I.20)$$

denotes the subset of quasifree density matrices (of finite particle number; see Sect. II.1).

### I.3 Results

Our first result in Theorem [IV.5] is that the minimal energy expectation value for all quasifree density matrices \( \mathcal{Q}_F \) is already obtained if the variation is restricted to pure quasifree density matrices \( \rho \in \mathcal{Q}_F \), i.e.,

$$E_{BHF}(g,\vec{p},\sigma,\Lambda) = \inf_{\rho \in \mathcal{Q}_F} \text{Tr}[H_{g,\vec{p}} \rho] = \inf_{\rho \in \mathcal{Q}_F} \text{Tr}[H_{g,\vec{p}} \rho], \quad (I.21)$$

see (I.20) and (II.40). The physical relevance of the minimization over (pure) quasifree density matrices is seen by the fact that it includes density matrices of the form

$$\rho_{sq} = |e^{i(a^*a)^2 + a^2}| \langle e^{i(a^*)^2 + a^2} \rangle, \quad (I.22)$$

where \( a \equiv a(f) \) for some one-photon state \( f \). These are important states in quantum optics known as *squeezed light*. 

Documenta Mathematica 18 (2013) 1481–1519
The restriction to pure quasifree states has the great advantage that the latter have a very convenient parametrization of their reduced one-particle density matrix given in Proposition IV.12. This enables us to prove the existence and uniqueness of a (pure) quasifree minimizer \((f, \gamma, \tilde{\alpha}, \gamma_0, \tilde{\alpha}_0)\) which minimizes the energy \(E_{g, \vec{p}}(f, \gamma, \tilde{\alpha})\) in Theorem VIII.6. The minimizer is characterized in Theorem VIII.8 by the Euler-Lagrange equations corresponding to \(E_{g, \vec{p}}\). We obtain expansions of the minimizer and the corresponding minimal energy for small \(g\) and \(\vec{p}\) in Theorem VIII.6.

Our proof uses a convexity and coercivity argument and the assumption that \(|g|, |\vec{p}| \leq C\) are smaller than a certain constant \(C \equiv C(\sigma, \Lambda)\) which, however, is not uniformly bounded as \(\Lambda \to \infty\) or \(\sigma \to 0\).

We also determine the minimizer in the case where the variation over pure quasifree density matrices is further restricted to coherent states introduced in (II.41). More precisely, we minimize the energy functional \(L^2(S_{\sigma, \Lambda} \times \mathbb{Z}_2) \ni f \mapsto E_{g, \vec{p}}(f, 0, 0)\) and obtain the existence of a minimizer \(f_{g, \vec{p}}\) and its uniqueness in Theorem VI.2 again by a fixed-point argument. Compared to the general case studied in Theorem VIII.3, our assumption for the minimizing coherent state is much milder, namely, that \(|\vec{p}| \leq 1/3\) and that \(g^2\) is small compared to \(1/\ln[\Lambda + 2]\). As our equations are fairly explicit in the coherent state case, we determine the leading orders in the expansions of the minimizer \(f_{g, \vec{p}}\) and of the minimal energy \(E_{g, \vec{p}}(f_{g, \vec{p}}, 0, 0)\) in powers of \(g\) and \(|\vec{p}|\). In particular, the coefficient for the term proportional to \(|\vec{p}|^2\), which gives the “renormalized electron mass for coherent states”, is computed in Proposition VII.1 and is found to agree with the first order expansion in \(\alpha\) of the renormalized mass of the electron, as computed for example in [12]. Our result holds uniformly in \(\sigma \to 0\) but not in \(\Lambda \to \infty\).

**Outline of the article** In Section II we discuss density matrices, density matrices of finite particle number, pure density matrices and quasifree density matrices in greater detail. We introduce our notation to describe the second quantization framework in Section III. Section IV introduces two parametrizations of pure quasifree states and contains the proof of Theorem IV.5. The energy functional for a fixed value of the momentum \(\vec{p}\) of the dressed electron is computed in Section V and some positivity properties of the different parts of the energy are established. From Section VI on we tacitly assume that the coupling constant \(|g| > 0\) is small. The energy is then minimized in the particular case of coherent states in Section VI, providing a first upper bound to the energy of the ground state and a proof of Theorem VI.2. We then turn in Section VIII to the problem of minimizing the energy over all pure quasifree states. The existence and uniqueness of a minimizer among the class of pure quasifree state is then proven in Section VIII.1 provided \(|\vec{p}|\) is small enough. The first terms of a perturbative expansion for small \(g\) and \(\vec{p}\) of the energy at the minimizer is computed in Section VIII.2. Finally the Lagrange equations associated with the problem of minimization in the generalized one particle density matrix variables are presented in Section VIII.3.
II Density Matrices and Quasifree Density Matrices

We now further discuss density matrices on Fock space and in particular give more details about quasifree density matrices.

II.1 Density Matrices of Finite Particle Number

Recall that the ground state energy is obtained as

$$E_{gs}(g, \vec{p}) = \inf \left\{ \text{Tr}_\mathcal{F} \left[ \rho H_{g,\vec{p}} \right] \bigg| \rho \in \hat{\mathcal{DM}} \right\}.$$  \hfill (II.23)

It is not difficult to see that $E_{gs}(g, \vec{p})$ is already obtained as an infimum over all density matrices

$$\mathcal{DM} := \left\{ \rho \in \hat{\mathcal{DM}} \bigg| \rho N_f, N_f \rho \in L^1(\mathcal{F}) \right\}.$$  \hfill (II.24)

of finite photon number expectation value, where

$$N_f = \int a^*(k) a(k) \, dk.$$  \hfill (II.25)

is the photon number operator. Indeed, if $\sigma > 0$ then

$$H_{g,\vec{p}} \geq H_f \geq \sigma N_f,$$  \hfill (II.26)

and $\mathcal{DM} = \hat{\mathcal{DM}}$ is automatic. Furthermore, if $\sigma = 0$ then it is not hard to see \[4\] that $E_{gs}(g, \vec{p}, 0, \Lambda) = \lim_{\sigma \to 0} E_{gs}(g, \vec{p}, \sigma, \Lambda)$, by using the standard relative bound

$$\left\| \tilde{A}_{<\sigma}(\vec{0}) \, \psi \right\| \leq O(\sigma) \left\| (H_{f, <\sigma} + 1)^{1/2} \, \psi \right\|,$$  \hfill (II.27)

where $\tilde{A}_{<\sigma}(\vec{0})$ and $H_{f, <\sigma}$ are the quantized magnetic vector potential and field energy, respectively, for momenta below $\sigma$. So, for all $0 \leq \sigma < \Lambda < \infty$, we have that

$$E_{gs}(g, \vec{p}, \sigma, \Lambda) = \inf \left\{ \text{Tr}_\mathcal{F} \left[ \rho H_{g,\vec{p}}(\sigma, \Lambda) \right] \bigg| \rho \in \mathcal{DM} \right\}.$$  \hfill (II.28)

Indeed, if the infimum \[II.28\] is attained at $\rho_{gs}(g, \vec{p}, \sigma, \Lambda) = \bigg( H_{g,\vec{p}}(\sigma, \Lambda) \bigg)^{1/2}$ then we call $\rho_{gs}(g, \vec{p}, \sigma, \Lambda)$ a ground state of $H_{g,\vec{p}}(\sigma, \Lambda)$.

Since $\mathcal{DM}$ is convex, we may restrict the density matrices in \[II.28\] to vary only over pure density matrices,

$$E_{gs}(g, \vec{p}, \sigma, \Lambda) = \inf \left\{ \text{Tr}_\mathcal{F} \left[ \rho H_{g,\vec{p}}(\sigma, \Lambda) \right] \bigg| \rho \in \hat{\mathcal{DM}} \right\}.$$  \hfill (II.29)

where pure density matrices are those of rank one,

$$\hat{\mathcal{DM}} := \left\{ \rho \in \hat{\mathcal{DM}} \bigg| \exists \Psi \in \mathcal{F}, \|\Psi\| = 1 : \rho = |\Psi\rangle\langle\Psi| \right\}.$$  \hfill (II.30)
and

\[ p\DM := \DM \cap \widetilde{p\DM}. \quad (\text{II.31}) \]

Another class of states that play an important role in our work is the set of \emph{centered} density matrices,

\[ c\DM := \{ \rho \in \DM \mid \forall f \in \mathfrak{h} : \Tr_{\tilde{\mathfrak{g}}} [\rho a^\ast (f)] = 0 \}. \quad (\text{II.32}) \]

\section{Quasifree Density Matrices}

A density matrix \( \rho \in \DM \) is called \emph{quasifree}, if there exist \( f_\rho \in \mathfrak{h} \), a symplectomorphism \( T_\rho \) (see Definition\[\text{III.5}\]) and a positive, self-adjoint operator \( h_\rho = h_\rho^\ast \geq 0 \) on \( \mathfrak{h} \) such that

\[ \big\langle W(\sqrt{2}f/i)\big\rangle_{\rho} := \Tr_{\tilde{\mathfrak{g}}} [\rho W(\sqrt{2}f/i)] \]

\[ = \exp \left[ 2i \operatorname{Im} \langle f_\rho | f \rangle - \frac{1}{2} \langle T_\rho f | (1 + h_\rho)T_\rho f \rangle \right], \quad (\text{II.33}) \]

for all \( f \in \mathfrak{h} \), where

\[ W(f) := \exp \left[ i\Phi(f) \right] := \exp \left[ \frac{1}{\sqrt{2}}(a^\ast (f) + a(f)) \right] \quad (\text{II.34}) \]

denotes the Weyl operator corresponding to \( f \) and we write expectation values w.r.t. the density matrix \( \rho \) as \( \langle \cdot \rangle_{\rho} \).

There are several important facts about quasifree density matrices, which do not hold true for general density matrices in \( \DM \). See, e.g., [5, 13, 7, 8]. The first such fact is that if \( \rho \in \Omega_{\mathfrak{g}} \) is a quasifree density matrix then so is \( W(-i\sqrt{2}f_\rho)\rho W^\ast (\sqrt{2}f_\rho) \) for any \( g \in \mathfrak{h} \), as follows from the Weyl commutation relations

\[ \forall f, g \in \mathfrak{h} : \quad W(f) W(g) = e^{-4 \operatorname{Im} \langle f \rangle g} W(f + g). \quad (\text{II.35}) \]

Choosing \( g := -i\sqrt{2}f_\rho \), we find that \( \tilde{\rho} := W(-i\sqrt{2}f_\rho)^\ast \rho W(-i\sqrt{2}f_\rho) \) is a centered quasifree density matrix, i.e.,

\[ \tilde{\rho} := W(\sqrt{2}f_\rho/i)^\ast \rho W(\sqrt{2}f_\rho/i) \in \mathfrak{c}\mathfrak{g} := \Omega_{\mathfrak{g}} \cap c\DM. \quad (\text{II.36}) \]

A characterization of centered quasifree density matrices is given in Appendix\[\text{VII}\].

A second important fact is that any quasifree state \( \rho \in \Omega_{\mathfrak{g}} \) is completely determined by its one-point function \( \langle a(\varphi) \rangle_\rho = \langle \varphi, f_\rho \rangle \) and its two-point function (one-particle reduced density matrix)

\[ \Gamma[\gamma_\rho, \tilde{\alpha}_\rho] := \left( \begin{array}{c} \gamma_\rho \\ \tilde{\alpha}_\rho \end{array} \right) \quad (\text{II.37}) \]

where the operators \( \gamma_\rho, \tilde{\alpha}_\rho \in B(\mathfrak{h} \oplus \mathfrak{h}) \) are defined as

\[ \langle \varphi, \gamma_\rho \psi \rangle := \langle a^\ast (\psi) a(\varphi) \rangle_{\tilde{\rho}} \quad \text{and} \quad \langle \varphi, \tilde{\alpha}_\rho \psi \rangle := \langle a(\varphi) a(\mathcal{J} \psi) \rangle_{\tilde{\rho}}, \quad (\text{II.38}) \]
and $\mathcal{J} : \hbar \to \hbar$ is a conjugation. See Definition \[\text{IV.8}\] and Remark \[\text{IV.10}\] The positivity of the density matrix $\rho$ implies that $\Gamma[\gamma_\rho, \tilde{\alpha}_\rho] \geq 0$ and, in particular, $\gamma_\rho \geq 0$, too. Moreover, the additional finiteness of the particle number expectation value, which distinguishes $\mathcal{D}\mathfrak{M}$ from $\widetilde{\mathcal{D}}\mathfrak{M}$, ensures that $\gamma_\rho \in \mathcal{L}^1(\hbar)$ is trace-class, namely,

$$\text{Tr}_\hbar[\gamma_\rho] = \langle N_f \rangle_\rho < \infty,$$

(II.39)

and that $\tilde{\alpha}_\rho \in \mathcal{L}^2(\hbar)$ is Hilbert-Schmidt.

Similar to (II.30)-(II.31), we introduce pure quasifree density matrices,

$$p\mathcal{Q}\mathfrak{F} := \mathcal{Q}\mathfrak{F} \cap \widetilde{p}\mathcal{D}\mathfrak{M}.$$  \hspace{1cm} (II.40)

A subset of $p\mathcal{Q}\mathfrak{F}$ of special interest is given by coherent states, i.e., pure quasifree states of the form $|W(−i\sqrt{2}f)\rangle\langle W(−i\sqrt{2}f)|$, which we collect in

$$\text{coh} := \{ |W(−i\sqrt{2}f)\rangle\langle W(−i\sqrt{2}f)| \mid f \in \hbar \}.$$  \hspace{1cm} (II.41)

For these, $\gamma_\rho = \tilde{\alpha}_\rho = 0$.

Conversely, if $\gamma \in \mathcal{L}^1(\hbar)$ is a positive trace-class operator and $\tilde{\alpha} \in \mathcal{L}^2(\hbar)$ is a Hilbert-Schmidt operator such that $\Gamma[\gamma, \tilde{\alpha}] \geq 0$ is positive then there exists a unique centered quasifree density matrix $\rho \in \mathfrak{c}\mathcal{Q}\mathfrak{F}$ such that $\gamma_\rho = \gamma$ and $\alpha = \alpha_\rho$ are its one-particle reduced density matrices.

Summarizing these two relations, the set $\mathcal{Q}\mathfrak{F}$ of quasifree density matrices is in one-to-one correspondence to the convex set

$$1−\text{pdm} := \{(f, \gamma, \tilde{\alpha}) \in \hbar \oplus \mathcal{L}^1_+(\hbar) \oplus \mathcal{L}^2(\hbar) \mid \Gamma[\gamma, \tilde{\alpha}] \geq 0\}. $$  \hspace{1cm} (II.42)

Note that coherent states correspond to elements of $1−\text{pdm}$ of the form $(f, 0, 0)$. Next, we observe in accordance with (II.42) that, if $\rho \in \mathcal{Q}\mathfrak{F}$ is quasifree then its energy expectation value $\langle H_{g,\vec{p}} \rangle_\rho$ is a functional of $(f_\rho, \gamma_\rho, \tilde{\alpha}_\rho)$, namely,

$$\langle H_{g,\vec{p}} \rangle_\rho = \mathcal{E}_{g,\vec{p}}(f_\rho, \gamma_\rho, \tilde{\alpha}_\rho).$$  \hspace{1cm} (II.43)

where, as shown in Section \[\text{V}\]

$$\mathcal{E}_{g,\vec{p}}(f, \gamma, \tilde{\alpha}) = \frac{1}{2} \left\{ (\text{Tr}[\gamma \vec{k}]) + \langle f, \vec{k}f \rangle + 2\text{Re}(\langle f, \vec{G} \rangle) - \vec{p} - \vec{k} \right\}^2$$

$$+ \text{Tr}[\gamma \vec{k} \cdot \vec{k}] + \text{Tr}[\tilde{\alpha} \vec{k} \cdot \vec{k}] + \text{Tr}[\vec{k}^2]$$

$$+ 2\text{Re}(\langle G + \vec{k}f, \tilde{\alpha}(\vec{G} + \vec{k}f) \rangle + \langle \vec{G} + \vec{k}f, \cdot(2\gamma + 1)(\vec{G} + \vec{k}f) \rangle)$$

$$+ \text{Tr}[\gamma |\vec{k}|] + \langle f, |\vec{k}|f \rangle,$$  \hspace{1cm} (II.44)

where $|\vec{a}| \cdot |\vec{a}| = \sum_{j=1}^3 |a_j|^2 |a_j|$. Note that in this expression $\vec{k}$ denotes the triple of multiplication operators $(k_1, k_3, k_3)$. We also use the same notation $\vec{k}$ for the momentum variable, the meaning being clear from the context.
III Second Quantization

In this section we describe in detail the second quantization framework we use, and in particular we explain the notation introduced below, which may be unfamiliar to some readers.

In what follows $\mathcal{H}$ will denote a $C$-Hilbert space with a scalar product $C$-linear in the right variable and $C$-antilinear in the left variable.

Let $\mathcal{B}(X;Y)$ be the space of bounded operators between two Banach spaces $X$ and $Y$, and $L^1(\mathcal{H})$ the space of trace class operators on $\mathcal{H}$. Given two $C$-Hilbert spaces $(\mathcal{H}_j, \langle \cdot, \cdot \rangle_j), j = 1, 2$ and a bounded linear operator $A : \mathcal{H}_1 \to \mathcal{H}_2$, set $A^* : \mathcal{H}_2 \to \mathcal{H}_1$ to be the operator such that

$$\forall z_1 \in \mathcal{H}_1, z_2 \in \mathcal{H}_2, \quad \langle z_2, Az_1 \rangle_2 = \overline{\langle z_1, A^*z_2 \rangle_1},$$

and $\text{Re} A := \frac{1}{2}(A + A^*), \text{Im} A := \frac{1}{2i}(A - (A^*)) \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2) \oplus \mathcal{B}(\mathcal{H}_2, \mathcal{H}_1)$.

**Example III.1.** For $z, z' \in \mathcal{H}$,

$$\langle z, z' \rangle = z^* z'.$$

The adjoint of a bounded operator $A$ on $\mathcal{H}$ is $A^*$.

**Remark III.2.** This notation applies in particular to one-particle vectors $f \in \mathcal{H}$ identified with linear applications from $C$ to $\mathcal{H}$ or to two-particle vectors $\alpha \in \mathcal{H}^\otimes 2$ identified with linear applications from $C$ to $\mathcal{H}^\otimes 2$. For this purpose a slight generalization of the Dirac notation with bras and kets would have been sufficient, but we would like to emphasize that in some situations, like in Equation (IV.58), it is natural to apply this operation to more general objects. For vectors and operators in a finite dimensional space $\mathcal{H}$ this notation is consistent with the usual notation on matrices.

The symmetrization operator $S_n$ on $\mathcal{H}^\otimes n$ is the orthogonal projection defined by

$$S_n(z_1 \otimes \cdots \otimes z_n) = \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} z_{\pi_1} \otimes \cdots \otimes z_{\pi_n}$$

and extension by linearity and continuity. The symmetric tensor product for vectors is $z_1 \vee z_2 = S_{n_1+n_2}(z_1 \otimes z_2)$ and more generally for operators is $A_1 \vee A_2 = S_{p_1+p_2} \circ (A_1 \otimes A_2) \circ S_{p_1+p_2}$ for $A_j \in \mathcal{B}(\mathcal{H}^{\otimes p_j}; \mathcal{H}^{\otimes q_j})$. We set

$$\mathcal{H}^\vee n := S_n \mathcal{H}^{\otimes n}, \quad \mathcal{B}^{p,q} := \mathcal{B}(\mathcal{H}^{\otimes p}; \mathcal{H}^{\otimes q}).$$

**Definition III.3.** The symmetric Fock space on a Hilbert space $\mathcal{H}$ is defined to be

$$\mathcal{F}_+(\mathcal{H}) := \bigoplus_{n=0}^{\infty} \mathcal{H}^{\vee n},$$

where $\mathcal{H}^{\vee 0} := C\Omega, \Omega$ being the normalized vacuum vector.
For a linear operator $C$ on $\mathfrak{h}$ such that $\|C\|_{B(\mathfrak{h})} \leq 1$, let $\Gamma(C)$ defined on each $\mathfrak{h}^\vee n$ by $C^\vee n$ and extended by continuity to the symmetric Fock space on $\mathfrak{h}$.

For an operator $A$ on $\mathfrak{h}$, the second quantization $d\Gamma(A)$ of $A$ is defined on each $\mathfrak{h}^\vee n$ by

$$d\Gamma(A)|_{\mathfrak{h}^\vee n} = n\mathfrak{h}^\vee n \cdot A$$

and extended by linearity to $\bigoplus_{n \geq 1} \mathfrak{h}^\vee n$. The number operator is $Nf = d\Gamma(1)$.

The self-adjoint field operator associated to $f$ is $\Phi(f) = \frac{1}{\sqrt{n}}(a^\ast(f) + a(f))$. For more details on the second quantization see the book of Berezin [6].

A dot “$\cdot$” denotes an operation analogous to the scalar product in $\mathbb{R}^3$. For every two objects $\vec{a} = (a_1, a_2, a_3)$ and $\vec{b} = (b_1, b_2, b_3)$ with three components such that the products $a_j b_j$ are well defined

$$\vec{a} \cdot \vec{b} = \sum_{j=1}^{3} a_j b_j.$$  

**Example III.4.** With $\vec{p} \in \mathbb{R}^3$, $\vec{G} \in \mathfrak{h}^3$, $\vec{k} \in (\mathcal{B}^{1,1})^3$

$$\vec{p}^2 = \sum_{j=1}^{3} p_j^2 \in \mathbb{R}, \quad \vec{k} \cdot \vec{p} = \sum_{j=1}^{3} p_j k_j \in \mathcal{B}^{1,1}, \quad \vec{p} \cdot \vec{G} = \sum_{j=1}^{3} p_j G_j \in \mathfrak{h},$$

$$\vec{k}^2 = \sum_{j=1}^{3} k_j^2 \in \mathcal{B}^{1,1}, \quad \vec{k} \cdot \vec{G} = \sum_{j=1}^{3} k_j G_j \in \mathfrak{h}, \quad \vec{G}^\ast \cdot \vec{k} = \sum_{j=1}^{3} G_j^\ast k_j \in \mathfrak{h}^\ast,$$

$$\vec{G} \cdot \vec{G}^\ast = \sum_{j=1}^{3} G_j G_j^\ast \in \mathcal{B}^{1,1}, \quad \vec{G} \cdot \vec{G} = \sum_{j=1}^{3} G_j G_j \in \mathbb{C},$$

where for an object with three components $\vec{a} = (a_1, a_2, a_3)$ such that $a_j^\ast$ is well-defined, $\vec{a}^\ast = (a_1^\ast, a_2^\ast, a_3^\ast)$. We sometimes use the notation $\vec{p}^2 = |\vec{p}|^2$, or $\vec{k}^2 = |\vec{k}|^2$.

And with another product, such as the symmetric tensor product $\vee$,

$$\vec{k} \vee^2 = \sum_{j=1}^{3} k_j^2 \in \mathcal{B}^{2,2}, \quad \vec{k} \vee \vec{G} = \sum_{j=1}^{3} k_j \vee G_j \in \mathcal{B}^{2,3}.$$
Recall that the Weyl operators are the unitary operators $W(f) = \exp(i\Phi(f))$ satisfying the relations

$$W(z_1)W(z_2) = e^{-\frac{i}{2}\text{Im}(z_1 \bar{z}_2)}W(z_1 + z_2), \quad (\text{III.46})$$

$$W(-i\sqrt{2}z)\Omega = e^{-\frac{|z|^2}{2}}\sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}}. \quad (\text{III.47})$$

We now introduce the usual parametrization of coherent states by vectors in $\mathfrak{h}$ and of Bogolubov transformations by symplectomorphisms.

**Definition III.5.** The coherent vectors are the vectors of the form

$$E_z = W(-i\sqrt{2}z)\Omega$$

for some $z \in \mathfrak{h}$ and the coherent states are the states of the form

$$|E_z\rangle\langle E_z|.$$  

**Definition III.6.** A symplectomorphism $T$ for the symplectic form $\text{Im}(\cdot, \cdot)$ on a C-Hilbert space $\mathfrak{h}$ is a continuous $\mathbb{R}$-linear automorphism on $\mathfrak{h}$ which preserves this symplectic form, i.e.,

$$\forall z_1, z_2 \in \mathfrak{h} : \quad \text{Im}(Tz_1, Tz_2) = \text{Im}(z_1, z_2).$$

A symplectomorphism $T$ is implementable if there is a unitary operator $U_T$ on $\mathfrak{h}$ such that

$$\forall z \in \mathfrak{h}, \quad U_TW(z)U_T^* = W(Tz).$$

In this case $U_T$ is a Bogolubov transformation corresponding to $T$.

We recall a well-known parametrization, in the spirit of the polar decomposition, of implementable symplectomorphisms.

**Proposition III.7.** The set of implementable symplectomorphisms is the set of operators

$$T = u \exp[\hat{r}] = u \sum_{n=0}^{\infty} \frac{1}{n!} \hat{r}^n,$$

where $u$ is an isometry and $\hat{r}$ is an antilinear operator, self-adjoint in the sense that $\forall z, z' \in \mathfrak{h}, \langle z, \hat{r}z' \rangle = \langle z', \hat{r}z \rangle$, and Hilbert-Schmidt in the sense that the positive operator $\hat{r}^2$ is trace-class. Equivalently, there exist a Hilbert basis $(\varphi_j)_{j \in \mathbb{N}}$ of $\mathfrak{h}$ and $(\hat{r}_{i,j})_{i,j} \in \ell^2(\mathbb{N}^2; \mathbb{C})$ such that

$$\hat{r} = \sum_{i,j=1}^{\infty} \hat{r}_{i,j} \langle \cdot, \varphi_j \rangle \varphi_i, \quad \forall i, j \in \mathbb{N}^2 : \quad \hat{r}_{i,j} = \hat{r}_{j,i}, \quad \text{and} \quad \sum_{i,j=1}^{\infty} |\hat{r}_{i,j}|^2 < \infty.$$
Proof. On the one hand, every operator of the form \( T = u \exp[\tilde{r}] \) with \( u \) a unitary operator and \( \tilde{r} \) a self-adjoint antilinear operator is a symplectomorphism. Since a unitary operator is a symplectomorphism, and the set of symplectomorphisms is a group for the composition, it is enough to prove that \( \exp[\tilde{r}] \) is a symplectomorphism. It is indeed the case since, for all \( z, z' \) in \( \mathfrak{h} \),

\[
\text{Im} \langle e^{\bar{r}} z, e^{\bar{r}} z' \rangle = \text{Im} \langle e^{\bar{r}} z, \cosh(\tilde{r}) z' \rangle + \text{Im} \langle e^{\bar{r}} z, \sinh(\tilde{r}) z' \rangle = \text{Im} \langle e^{\bar{r}} z, z' \rangle + \text{Im} \langle \sinh(\tilde{r}) e^{\bar{r}} z, z' \rangle = \text{Im} \langle e^{-\bar{r}} e^{\bar{r}} z, z' \rangle.
\]

The implementability condition is then satisfied if we suppose \( \tilde{r} \) to be Hilbert-Schmidt. On the other hand, to get exactly this formulation we give the step to go from the result given in Appendix A in [9] to the decomposition in Proposition III.7. In [9] an implementable symplectomorphism is decomposed as

\[
T = u e^{\bar{r}},
\]

where \( u \) is a unitary operator, \( c \) is a conjugation and \( \tilde{r} \) is a Hilbert-Schmidt, self-adjoint, non-negative operator commuting with \( c \). It is then enough to set \( \tilde{r} = c \tilde{r} \) to get the expected decomposition. To check the self-adjointness of \( \tilde{r} \), observe that, for all \( z, z' \) in \( \mathfrak{h} \),

\[
\langle z', \tilde{r} z \rangle = \langle z', \tilde{r} c z \rangle = \langle \tilde{r} z', c z \rangle = \langle z, c \tilde{r} z' \rangle = \langle z, \tilde{r} z' \rangle.
\]

For the convenience of the reader we recall the main steps to obtain the decomposition in Eq. (III.48). First decompose the reader can recall the main steps to obtain the decomposition in Eq. (III.48). First decompose \( T \) in its \( \mathbb{C} \)-linear and antilinear parts, \( T = L + A \), then write the polar decomposition \( L = u \|L\| \). It is then enough to prove that \( \|L\| + u^* A \) is of the form \( e^{\bar{r}} \). From certain properties of symplectomorphisms (also recalled in [9]) it follows that the antilinear operator \( u^* A \) is self-adjoint and \( \|L\|^2 + 1_{\mathfrak{h}} = (u^* A)^2 \). A decomposition of the positive trace class operator \( (u^* A)^2 = \sum_j \lambda_j^2 e_j e_j^* \) with \( e_j \) an orthonormal basis of \( \mathfrak{h} \) yields \( \|L\| = \sum_j (1 + \lambda_j^2)^{1/2} e_j e_j^* \). Using that \( \lambda_j \to 0 \) one can study the operator \( \|L\| \) and \( u^* A \) on the finite dimensional subspaces \( \ker(\|L\| - \mu 1_{\mathfrak{h}}) \) which are invariant under \( u^* A \). It is then enough to prove that for a \( \mathbb{C} \)-antilinear self-adjoint operator \( f \) such that \( f f^* = \lambda^2 \) on a finite dimensional space, there is an orthonormal basis \( \{ \varphi_k \}_k \) such that \( f(\varphi_k) = \lambda_k \varphi_k \). The conjugation is then defined such that \( c(\sum \beta_k \varphi_k) = \sum \bar{\beta}_k \varphi_k \) and \( \tilde{r} = \sinh^{-1}(\lambda_j) 1 \) on that subspace. \( \square \)

IV Pure Quasifree States

In this section we give a characterization of quasifree states and use this to show that the infimum of the energy over quasifree states is equal to the infimum of the energy over pure quasifree states. This result has been generalized to a wider class of Hamiltonians and also to the case of fermion Fock space in [2].
IV.1 From Quasifree States to Pure Quasifree States

Let \( \mathfrak{h} \) be the \( \mathbb{C} \)-Hilbert space \( L^2(S_{\sigma,\lambda} \times \mathbb{Z}_2) \). We make use of the following characterization of quasifree density matrices.

**Lemma IV.1.** The set of quasifree density matrices and pure quasifree density matrices, respectively, of finite photon number expectation value can be characterized by

\[
\Omega^\text{QF} = \Omega \cap \left\{ W(-i\sqrt{2}f) U^\star \frac{\Gamma(C)}{\text{Tr}[\Gamma(C)]} UW(-i\sqrt{2}f)^* \right\}
\]

where 
\[ f \in \mathfrak{h}, \quad U \text{ a Bogolubov transformation}, \]
\[ C \in L^1(\mathfrak{h}), \quad C \geq 0, \quad \|C\|_{B(\mathfrak{h})} < 1 \]

\[
p\Omega^\text{QF} = \Omega \cap \left\{ W(-i\sqrt{2}f) U^\star |\Omega\rangle \langle \Omega| UW(-i\sqrt{2}f)^* \right\}
\]

where 
\[ f \in \mathfrak{h}, \quad U \text{ a Bogolubov transformation}. \]

**Proof.** We only sketch the argument, details can be found in [6],[13]. It is not difficult to see that any density matrix of the form \( W(-i\sqrt{2}f) U^\star \frac{\Gamma(C)}{\text{Tr}[\Gamma(C)]} UW(-i\sqrt{2}f)^* \) is indeed quasifree. Conversely, if \( \rho \in \Omega^\text{QF} \) is a quasifree density matrix then it is fully characterized by its one-point function \( f_\rho \in \mathfrak{h} \) and two-point functions \( \langle \gamma_\rho, \vec{\alpha}_\rho \rangle \). Moreover, \( W(-i\sqrt{2}f_\rho)^* \rho W(-i\sqrt{2}f_\rho) \in \Omega^\text{QF} \) is a centered quasifree density matrix with the same one-particle density matrix, that is, the density matrix \( W(-i\sqrt{2}f_\rho)^* \rho W(-i\sqrt{2}f_\rho) \) corresponds to \( (0, \gamma_\rho - f_\rho f_\rho^\star, \vec{\alpha}_\rho - f_\rho f_\rho^\star) \). Obviously, \( \gamma_\rho - f_\rho f_\rho^\star \) is again trace-class and \( \vec{\alpha}_\rho - f_\rho f_\rho^\star \) is Hilbert-Schmidt. Now, we use that there exists a Bogolubov transformation \( U \) which eliminates \( \vec{\alpha}_\rho \), i.e., \( U^\star W(\sqrt{2}f_\rho/i)^* \rho W(\sqrt{2}f_\rho/i)U \) corresponds to \( (0, \vec{\gamma}_\rho, 0) \). While this is the only non-trivial step of the proof, we note that if \( U \) is characterized by \( u \) and \( v \) as in Lemma IV.2 then there is an involved, but explicit formula that determines \( u \) and \( v \). Again \( \vec{\gamma}_\rho \) is trace-class because the photon number operator \( N_f \) transforms under \( U^\star \) to itself plus lower order corrections, \( U^\star N_f U = N_f + O(N_f^{1/2} + 1) \). Finally, it is easy to see that \( (0, \vec{\gamma}_\rho, 0) \) corresponds to the quasifree density matrix \( \Gamma(C_\rho)/\text{Tr}[\Gamma(C_\rho)] \) with 
\[
C_\rho := \vec{\gamma}_\rho (1 + \vec{\gamma}_\rho)^{-1}
\]

Following these steps we finally obtain

\[
\rho = W(f_\rho) U^\star \frac{\Gamma(C_\rho)}{\text{Tr}[\Gamma(C_\rho)]} U^\star W(f_\rho)^* ,
\]

as asserted. The additional characterization of pure quasifree density matrices is obvious.

**Lemma IV.2.** Let \( U \in B(\mathfrak{h}) \) be a unitary operator. The following statements are
1494 Volker Bach, Sébastien Breteaux, Tim Tzaneteas

equivalent:
\[ U \in B(\mathfrak{g}) \text{ is a Bogolubov transformation; } \quad (\text{IV.49}) \]
⇔ \exists T \text{ implementable symplectomorphism, } \quad (\text{IV.50})
\[ U = \tilde{U}_T, \quad \tilde{U}_T W(f) \tilde{U}^* = W(T f). \]
⇔ \exists u \in B(\mathfrak{h}), v \in L^2(\mathfrak{h}) \forall f \in \mathfrak{h} :
\[ U a^* (f) U^* = a^* (uf) + a(J v f); \quad (\text{IV.51}) \]
⇔ U = \exp(iH), where H = H^* is a semibounded operator, quadratic in a^* and a and without linear term. \quad (\text{IV.52})

Proof. Again, we only sketch the argument. First note that (IV.49)⇔(IV.50) is the definition of a Bogolubov transformation. Secondly, \( \tilde{U}_T \Phi(f) \tilde{U}^* = \Phi(T f) \) is equivalent to \( \tilde{U}_T \Phi(f) \tilde{U}^* = \Phi(T f) \). Hence, using that \( a^*(f) = \frac{1}{\sqrt{2}} [\Phi(f) - i \Phi(if)] \) and \( a(f) = \frac{1}{\sqrt{2}} [\Phi(f) + i \Phi(if)] \) we obtain the equivalence (IV.50)⇔(IV.51). Thirdly, setting \( U_\lambda = \exp(i\lambda H) \) and \( a^*_\lambda(f) := U_\lambda a^*(f) U_\lambda^* \), we observe that \( \partial_\lambda a^*_\lambda(f) = i[H,a^*_\lambda(f)] \). Furthermore, \([H,a^*_\lambda(f)]\) is linear in \( a^* \) and \( a \) if, and only if, \( H \) is quadratic in \( a^* \) and \( a \). Solving this linear differential equation, we finally obtain (IV.51)⇔(IV.52).

As a consequence, the class of quasifree states (resp. centered quasifree states) is invariant under conjugation by Weyl transformations and Bogolubov transformations (resp. Bogolubov transformations):

Lemma IV.3. For all Bogolubov transformations \( U \) and all \( g \in \mathfrak{h} \):
\[ W(g) \Omega \tilde{U} \Omega \tilde{U}^* W(g)^* = \Omega \tilde{U} \Omega ; \quad (\text{IV.53}) \]
\[ U c \Omega \tilde{U} \Omega = c \Omega \tilde{U} \Omega . \quad (\text{IV.54}) \]

Remark IV.4. A pure quasifree state is a particular case of quasifree state with \( C = 0 \), that is \( \Gamma(C) = |\Omega \rangle \langle \Omega | \).

We come to the main result of this section.

Theorem IV.5. Let \( 0 \leq \sigma < \Lambda < \infty, g \in \mathbb{R} \) and \( \vec{p} \in \mathbb{R}^3 \). Minimizing the energy over quasifree states is the same as minimizing the energy over pure quasifree states, i.e.,
\[ E_{BHF}(g,\vec{p},\sigma,\Lambda) := \inf_{\rho \in \Omega \tilde{U}} \text{Tr}[H_g,\vec{p},\rho] = \inf_{\rho \in \Omega \tilde{U}} \text{Tr}[H_g,\vec{p},\rho]. \]

For the proof of Theorem IV.5 we derive a couple of preparatory lemmata.

Proposition IV.6. Let \( C \) be a non-negative operator on \( \mathfrak{h} \), then
\[ \{ \text{Tr}[\Gamma(C)] < \infty \} \quad \Leftrightarrow \quad \{ C \in L^1(\mathfrak{h}) \text{ and } \|C\|_{B(\mathfrak{h})} < 1 \}. \]

In this case \( \text{Tr}[\Gamma(C)] = \det(1 - C)^{-1} \). (We refrain from defining the determinant.) For the direction \( \Leftarrow \) the non-negativity assumption is not necessary.
Minimization of the Energy of the Nonrelativistic . . . 1495

Proof. Let us decompose \( \mathfrak{h} = \bigoplus_{j \geq 0} \mathfrak{c} e_j \) where \( \mathfrak{c} = \sum c_j e_j e_j^* \) with \( (e_j)_{j \geq 0} \) an orthonormal basis of \( \mathfrak{h} \). Then \( \mathfrak{h}^+ = \bigotimes_{j \geq 0} \mathfrak{h}^+(\mathfrak{c} e_j) \) and

\[
\text{Tr}[\Gamma(C)] = \prod_{j \geq 0} \text{Tr}[\Gamma(c_j)] = \prod_{j \geq 0} \frac{1}{1 - c_j}
\]

and the infinite product converges exactly when \( C \in \mathcal{L}^1(\mathfrak{h}) \) and \( \|C\|_{\mathcal{B}(\mathfrak{h})} < 1. \)

Lemma IV.7. Suppose \( \mathfrak{h}_d \) is of dimension \( d < \infty \). Then, for any non-negative operator \( C_d \neq 0 \) such that \( C_d \in \mathcal{L}^1(\mathfrak{h}_d) \) and \( \|C_d\|_{\mathcal{B}(\mathfrak{h}_d)} < 1 \), there exist a non-negative measure \( \mu_d \) (depending on \( C \)) of mass one on \( \mathfrak{h}_d \) and a family \( \{\rho_d(z)\}_{z \in \mathfrak{h}_d} \) of pure quasifree states such that

\[
\frac{\Gamma(C)}{\text{Tr}[\Gamma(C)]]} = \int_{\mathfrak{h}_d} \rho_d(z) \, d\mu_d(z).
\]

Proof. In finite dimension \( d \) we can use a resolution of the identity with coherent states (see, e.g., [6])

\[ 1_{\mathfrak{h}_d} = \int_{\mathfrak{h}_d} |E_{z_d}\rangle \langle E_{z_d}| \frac{dz_d}{\pi^d}, \]

where \( \mathfrak{h}_d \) is identified with \( \mathbb{C}^d \) and \( dz_d = dx_d \, dy_d \), \( z_d = x_d + iy_d \). Using Equation (III.47) we get

\[
\Gamma(C) = \int_{\mathfrak{h}_d} \Gamma(C^{1/2}) |E_{z_d}\rangle \langle E_{z_d}| \Gamma(C^{1/2}) \frac{dz_d}{\pi^d} = \int_{\mathfrak{h}_d} |E_{C^{1/2} z_d}\rangle \langle E_{C^{1/2} z_d}| \exp(||C^{1/2} z_d||^2 - |z_d|^2) dz_d.
\]

The measure \( d\mu_d(z) = \pi^{-d} \exp(||C^{1/2} z_d||^2 - |z_d|^2) dz_d / \text{Tr}[\Gamma(C)] \) has mass one. Indeed

\[
\int_{\mathfrak{h}_d} \exp(-z_d^* (1_{\mathfrak{h}_d} - C) z_d) \frac{dz_d}{\pi^d} = \prod_{j=1}^d \int_{\mathbb{R}^2} \exp(-(1 - c_j)(x^2 + y^2)) \frac{dx \, dy}{\pi} = \prod_{j=1}^d \frac{1}{1 - c_j} = \text{Tr}[\Gamma(C)],
\]

where \( C = \sum_{j=1}^d c_j e_j e_j^* \) with \( (e_j)_{j=1}^d \) an orthonormal basis of \( \mathfrak{h}_d \).

Proof of Theorem IV.5. The inclusion \( p \Omega \preceq \Omega \) implies that

\[
\inf_{\rho \in \Omega \preceq} \text{Tr}[H_{g, f} \rho] \leq \inf_{\rho \in p \Omega \preceq} \text{Tr}[H_{g, f} \rho],
\]

and it is hence enough to prove for any quasifree state

\[
\rho_{af} = W(-i \sqrt{2} f) U_T^* \frac{\Gamma(C)}{\text{Tr}[\Gamma(C)]} U_T W(-i \sqrt{2} f)^*,
\]

Documenta Mathematica 18 (2013) 1481–1519
that the inequality
\[ \text{Tr}[H_{g,\vec{p}} \rho_{qf}] \geq \inf_{\rho \in \mathcal{P} \Omega} \text{Tr}[H_{g,\vec{p}} \rho] \]
holds true. The operator \( C \) is decomposed as
\[ C = \sum_{j \geq 0} c_j e_j e_j^\ast \]
where \( (e_j) \) is an orthonormal basis of the Hilbert space \( \mathcal{H} \) and \( c_j \geq 0 \). Let \( C_d = \sum_{j \leq d} c_j e_j e_j^\ast \) where
\[ \rho_{qf,d} = W(-i\sqrt{2}f) U_T^* \Gamma(C_d) U_T W(-i\sqrt{2}f)^* \]
then using Lemma IV.7 with \( h_d = \bigoplus_{j \leq d} C e_j, \mathbf{K} = \mathbf{K}_+ h_d \oplus \mathbf{K}_+ h_d^\perp \) and the extension of the operator \( \Gamma(C_d) \) on \( \mathbf{K}_+ h_d \) to \( \mathbf{K}_+ h_d \oplus \mathbf{K}_+ h_d^\perp \) by \( \Gamma(C_d) \otimes (|\Omega_{h_d^\perp}\rangle \langle \Omega_{h_d^\perp}|) \) (which we still denote by \( \Gamma(C_d) \)), we obtain
\[ \rho_{qf,d} = \int_{h_d} \rho_d(z_d) \, d\mu_d(z_d) , \]
where \( \rho_d(z_d) \) are pure quasifree states and the \( \mu_d \) are non-negative measures with mass one. Note that
\[ \nu_d := \frac{\text{Tr}[\Gamma(C_d)]}{\text{Tr}[\Gamma(C)]} = \prod_{j > d} (1 - c_j) \nearrow 1 , \]
as \( d \to \infty \). Further note that \( \rho_{qf} \geq \nu_d \rho_{qf,d} \), for any \( d \in \mathbb{N} \), since \( \Gamma(C) \geq \Gamma(C_d) \).
Thus
\[ \text{Tr}[H_{g,\vec{p}} \rho_{qf}] \geq \text{Tr}[H_{g,\vec{p}} \nu_d \rho_{qf,d}] = \nu_d \int_{h_d} \text{Tr}[H_{g,\vec{p}} \rho_d(z_d)] \, d\mu_d(z_d) \geq \nu_d \inf_{z_d \in h_d} \text{Tr}[H_{g,\vec{p}} \rho_d(z_d)] \geq \nu_d \inf_{\rho \in \mathcal{P} \Omega} \text{Tr}[H_{g,\vec{p}} \rho] , \]
for all \( d \in \mathbb{N} \), and in the limit \( d \to \infty \), we obtain
\[ \text{Tr}[H_{g,\vec{p}} \rho_{qf}] \geq \lim_{d \to \infty} \{ \nu_d \} \inf_{\rho \in \mathcal{P} \Omega} \text{Tr}[H_{g,\vec{p}} \rho] = \inf_{\rho \in \mathcal{P} \Omega} \text{Tr}[H_{g,\vec{p}} \rho] . \]

IV.2 Pure Quasifree States and Their One-Particle Density Matrices

We now define reduced density matrices \( \rho^{p,q} \) resulting as marginals from a given density matrix \( \rho \) on Fock space and derive a convenient parametrization for them in case \( \rho \) is pure and quasifree. We also recall a characterization of the admissible one-particle density matrices for pure quasifree states.
Let \( \mathfrak{h} \) be a \( \mathbb{C} \)-Hilbert space.
Minimization of the Energy of the Nonrelativistic . . . 1497

**Definition IV.8.** Let \( \rho \in \mathcal{D} \mathfrak{M} \) be a density matrix on the bosonic Fock space \( \mathfrak{F}_+(\mathfrak{h}) \) over \( \mathfrak{h} \). If \( \text{Tr}[\rho N_{f}^{\frac{p+q}{2}}] < \infty \), we define \( \rho^{p,q} \in B^{p,q}(\mathfrak{h}) \) through

\[
\forall \varphi, \psi \in \mathfrak{h}, \quad \psi^* \varphi \rho^{p,q} \varphi^* \psi = \text{Tr}[a^*(\varphi)\gamma a(\psi)\rho].
\]

We single out

\[
f = \rho^{0,1} \in B^{0,1} \cong \mathfrak{h},
\]

i.e., \( f_{\rho} \in \mathfrak{h} \) is the unique vector such that \( \text{Tr}[a(\psi)\rho] = \psi^* f_{\rho} \) for all \( \psi \in \mathfrak{h} \). Furthermore, with \( \tilde{\rho} = W(\sqrt{2}f_{\rho}/i)^* \rho W(\sqrt{2}f_{\rho}/i) \), the matrix elements of the (generalized) one-particle density matrix are defined by

\[
\gamma_{\rho} = \tilde{\rho}^{1,1} \in B^{1,1} \quad \text{and} \quad \alpha_{\rho} = \rho^{0,2} \in B^{0,2} \cong \mathfrak{h}^{\sqrt{2}},
\]

in other words

\[
\forall \varphi, \psi \in \mathfrak{h} : \quad \langle \psi, \gamma_{\rho} \varphi \rangle = \text{Tr}[\tilde{\rho} a^*(\varphi) a(\psi)],
\]

\[
\langle \psi \otimes \varphi, \alpha_{\rho} \rangle = \text{Tr}[\tilde{\rho} a(\psi) a(\varphi)].
\]

Note that \( f_{\rho}, \gamma_{\rho}, \) and \( \alpha_{\rho} \) exist for any \( \rho \in \mathcal{D} \mathfrak{M} \) since \( N_{f}\rho, \rho N_{f} \in L^{1}(\mathfrak{F}_+) \).

**Remark IV.9.** For a centered pure quasifree state \( \tilde{\rho}, \tilde{\rho}^{p,q} \) vanishes when \( p + q \) is odd.

**Remark IV.10.** Another definition of the one-particle density matrix \( \gamma_{\rho} \) would be through the relation \( \langle \psi, \gamma_{\rho} \varphi \rangle = \text{Tr}[a^*(\varphi) a(\psi)\rho] \). We prefer here a definition with a “centered” version \( \tilde{\rho} \) of the state \( \rho \), because this centered quasifree state \( \tilde{\rho} \) then satisfies the usual Wick theorem. The same considerations hold for \( \alpha_{\rho} \).

Hence, any quasifree density matrix is characterized by \( (f_{\rho}, \gamma_{\rho}, \alpha_{\rho}) \), since \( \rho^{p,q} \) can be expressed in terms of \( (f_{\rho}, \gamma_{\rho}, \alpha_{\rho}) \).

When \( f_{\rho} = 0 \), the definition of \( \gamma_{\rho} \) is consistent with the usual one, for \( z_{1}, z_{2} \in \mathfrak{h} \),

\[
\langle z_{1}, \gamma_{\rho} z_{2} \rangle = \text{Tr}[a^*(z_{2}) a(z_{1})\rho].
\]

The definition of \( \alpha_{\rho} \) is related with the definition of the operator \( \tilde{\alpha}_{\rho} \) (here denoted with a hat for clarity) used in the article of Bach, Lieb and Solovej [5], through the relation \( \langle z_{1} \otimes z_{2}, \alpha_{\rho} \rangle_{\mathfrak{h} \otimes \mathfrak{h}} = \langle z_{1}, \tilde{\alpha}_{\rho} z_{2} \rangle_{\mathfrak{h}} \) with \( \tilde{\epsilon} \) a conjugation on \( \mathfrak{h} \).

**Example IV.11.** A centered pure quasifree state satisfies the relation,

\[
\tilde{\rho}^{2,2} = \gamma \otimes \gamma + \gamma \otimes \gamma \quad \text{Ex} + \alpha \alpha^* \in B^{2,2}, \quad (IV.55)
\]

where the exchange operator is the linear operator on \( \mathfrak{h}^{\otimes 2} \) such that

\[
\forall z_{1}, z_{2} \in \mathfrak{h}, \quad \text{Ex}(z_{1} \otimes z_{2}) = z_{2} \otimes z_{1}
\]

and where for any \( b \in \mathfrak{h}^{\otimes 2} \), \( \alpha \alpha^* b = \langle \alpha, b \rangle_{\mathfrak{h}^{\otimes 2}} \alpha \).

We now turn to another parametrization of quasifree states, by vectors in a real Hilbert space. This parametrization enables us to use convexity arguments.
Proposition IV.12. Let $T = u e^t$ be an implementable symplectomorphism and $\rho$ a quasifree state of the form $\rho = U_T^* |\Omega\rangle \langle \Omega | U_T$. Then

$$\gamma_{\rho} = \frac{1}{4} (\cosh(2\tilde{r}) - 1), \quad (IV.56)$$

$$\forall z_1, z_2 \in \mathfrak{h} : \quad \langle z_1 \otimes z_2, \rho_{\mathfrak{h} \otimes \mathfrak{h}} \rangle = \langle z_1, \frac{1}{2} \sinh(2\tilde{r}) z_2 \rangle. \quad (IV.57)$$

Proof of Proposition IV.12. We have $T i = u e^t i = u i e^{-t} = i u e^{-t}$ and for all $z \in \mathfrak{h}$

$$\text{Tr}[\rho W(-i\sqrt{2} z)] = \text{Tr}[U_T^* |\Omega\rangle \langle \Omega | U_T W(-i\sqrt{2} z)]$$

$$= \langle \Omega | W(u e^t(-i\sqrt{2} z)) |\Omega\rangle$$

$$= \langle \Omega | W(-i\sqrt{2} u e^{-t} z) |\Omega\rangle$$

$$= \exp \left( -\frac{1}{2} |u e^{-t} z|^2 \right)$$

$$= \exp \left( -\frac{1}{2} |e^{-t} z|^2 \right).$$

From this formula we can easily compute the function

$$h(t, s) := \text{Tr}[\rho W(-t i \sqrt{2} z) W(-s i \sqrt{2} z)] = \exp \left( -\frac{1}{2} |e^{-t} (t + s) z|^2 \right),$$

whose derivative $\partial_i \partial_s$ at $(t, s) = (0, 0)$ involves $\alpha$ and $\gamma$:

$$\partial_i \partial_s h(0, 0) = \text{Tr}[\rho (\alpha^* (z) - a(z))^2]$$

$$= -2z^* \gamma z + 2 \text{Re}(\alpha^* z^\vee^2) - z^* z.$$

But we also have

$$\partial_i \partial_s \exp \left( -\frac{1}{2} |e^{-t} (t + s) z|^2 \right) \bigg|_{t=s=0}$$

$$= - (e^{-t} z)^* (e^{-t} z)$$

$$= -(\cosh(\tilde{r}) z - \sinh(\tilde{r}) z)^* (\cosh(\tilde{r}) z - \sinh(\tilde{r}) z)$$

$$= -(\cosh(\tilde{r}) z)^* (\cosh(\tilde{r}) z)$$

$$+ 2 \text{Re}(\sinh(\tilde{r}) z)^* (\cosh(\tilde{r}) z) - (\sinh(\tilde{r}) z)^* (\sinh(\tilde{r}) z)$$

$$= -z^* (\cosh^2 \tilde{r} + \sinh^2 \tilde{r}) z + 2 \text{Re}(z^* (\sinh \tilde{r} \cosh \tilde{r}) z)$$

$$= -z^* \cosh(2\tilde{r}) z + 2 \text{Re}(z^* \frac{1}{2} \sinh(2\tilde{r}) z)$$

and hence, using the polarization identity

$$4z \vee z' = (z + z')^\otimes^2 - (z - z')^\otimes^2$$

to recover every vector from $h|\mathfrak{h}|^2$ from linear combinations of vectors of the form $z^\vee^2$, we arrive at (IV.56) - (IV.57).

Proposition IV.13. The admissible $\gamma$, $\alpha$ for a pure quasifree state are exactly those satisfying the relation

$$\gamma + \gamma^2 = (\alpha \otimes 1)^* (1 \otimes \alpha), \quad (IV.58)$$

with $\gamma \geq 0$.  

Documenta Mathematica 18 (2013) 1481–1519
Minimization of the Energy of the Nonrelativistic . . .

This is the constraint when we minimize the energy as a function of \((f, \gamma, \alpha)\) with the method of Lagrange multipliers in Section VIII.3.

**Proof.** If \(\gamma, \alpha\) are associated with a quasifree state, then there is an \(\hat{\ell}\) such that \(\gamma, \alpha\) and \(\hat{\ell}\) satisfy Equations (IV.56) and (IV.57), then

\[
\langle z_1, (\alpha^* \otimes \mathbf{1})(1 \otimes \alpha)z_2 \rangle = (\alpha^* \otimes z_1^*)(z_2 \otimes \alpha)
\]

\[
= (\alpha^*(z_2 \otimes \mathbf{1})) \otimes z_1^* \alpha
\]

\[
= \langle \alpha^*(z_2 \otimes \mathbf{1}), \frac{1}{2} \sinh(2\hat{\ell})z_1 \rangle_h
\]

\[
= \langle \alpha^*, z_2 \otimes \frac{1}{2} \sinh(2\hat{\ell})z_1 \rangle_{h \otimes \mathbf{1}}
\]

\[
= \langle \frac{1}{4} \sinh^2(2\hat{\ell})z_1, z_2 \rangle_h
\]

\[
= \langle \left(\frac{1}{2}(\cosh(2\hat{\ell}) - 1) + \frac{1}{4}(\cosh(2\hat{\ell}) - 1)^2\right)z_1, z_2 \rangle_h.
\]

Conversely, if \(\gamma\) and \(\alpha\) satisfy Eq. (IV.58) then we define the \(\hat{C}\)-antilinear operator \(\hat{\alpha}\) such that \(\langle z_1, \hat{\alpha}z_2 \rangle = (z_1 \otimes z_2)^* \alpha\), and set \(\hat{\varrho} = \frac{1}{4} \sinh^{-1}(2\hat{\ell})\), then

\[
\forall z_1, z_2 \in \mathfrak{h} : \quad \langle z_1 \otimes z_2, \alpha\rho \rangle_{h^2} = \langle z_1, \hat{\alpha}z_2 \rangle = \left\langle z_1, \frac{1}{2} \sinh(2\hat{\ell})z_2 \right\rangle,
\]

which, in turn, implies that \((\alpha^* \otimes \mathbf{1})(1 \otimes \alpha) = \frac{1}{2} \sinh^2(2\hat{\ell})\). Hence, we have

\[
\gamma + \gamma^2 = \frac{1}{4} \sinh^2(2\hat{\ell})
\]

and as \(\gamma \geq 0\), it follows that \(\gamma = \frac{1}{2}(\cosh(2\hat{\ell}) - 1)\). Then \(\gamma, \alpha\) is associated with the centered pure quasifree state whose symplectic transformation is \(\exp[\hat{\varrho}]\).

\(\Box\)

V Energy Functional

In this section we calculate the energy of a quasifree state in terms of its characterizing parameters, i.e., in terms of \((f, \gamma, \alpha)\) and \((f, \varrho)\).

**Notation:** We first recall that, as before, we denote by \(\tilde{k}\) and \(|\tilde{k}|\) the multiplication operators \(\tilde{k} \otimes 1_{\mathbb{C}^2}\) and \(|\tilde{k}| \otimes 1_{\mathbb{C}^2}\) on \(\mathfrak{h} = L^2(S_{\sigma, A} \times \mathbb{Z}_2)\), with three components in the case of \(\tilde{k}\).

We now work at fixed values of total momentum \(\tilde{p} \in \mathbb{R}^3\). The operator \(H_{\tilde{g}, \tilde{p}}\) is given by

\[
H_{\tilde{g}, \tilde{p}} = \frac{1}{2} (d\Gamma(\tilde{k}) + 2Re \alpha^*(\tilde{G} - \tilde{p})^2 + d\Gamma(|\tilde{k}|)),
\]

where \(\tilde{G}(k) = \tilde{G}(|k|, \pm) := g(|k|^{-1/2}e_{\tilde{p}})(k)\). The energy of a quasifree state \(\rho\) associated with \(f \in \mathfrak{h}, \gamma \in L^1(\mathfrak{h}), \alpha \in \mathfrak{h}^\vee\) is

\[
E_{\tilde{g}, \tilde{p}}(f, \gamma, \alpha) := \text{Tr}[H_{\tilde{g}, \tilde{p}} \rho], \quad (V.59)
\]

where \(\mathfrak{h}\) is the \(\mathbb{C}\)-Hilbert space \(\mathfrak{h} = L^2(S_{\sigma, A} \times \mathbb{Z}_2)\) and \(L^1(\mathfrak{h})\) is the space of trace class operators on \(\mathfrak{h}\).
Proposition V.1. The energy functional (V.59) is

\[ E_{\rho,\tilde{\rho}}(f, \gamma, \alpha) = \frac{1}{2} \left\{ (\text{Tr}[\gamma \tilde{k}]) + f^* \tilde{k} f + 2\text{Re}(f^* \tilde{G}) - \tilde{p} \right\}^2 \\
+ \text{Tr}[\gamma \tilde{k} \cdot \gamma \tilde{k}] + \alpha^*(\tilde{k} \cdot \tilde{k}) + \text{Tr}[|\tilde{k}|^2 \gamma] \\
+ 2\text{Re}(\alpha^*[(\tilde{G} + \tilde{k} f)^{\gamma^2}]) + \text{Tr}[(2\gamma + 1)(\tilde{G} + \tilde{k} f) \cdot (\tilde{G} + \tilde{k} f)^*] \\
+ \text{Tr}[\gamma|\tilde{k}] + f^*|\tilde{k}| f \]  

(V.60)

where the following positivity properties hold

1. \((\text{Tr}[\gamma \tilde{k}]) + f^* \tilde{k} f + 2\text{Re}(f^* \tilde{G}) - \tilde{p})^2 \geq 0,
2. \text{Tr}[\gamma \tilde{k} \cdot \gamma \tilde{k}] + \text{Tr}[|\tilde{k}|^2 \gamma] \geq 0,
3. \text{Tr}[\gamma \tilde{k} + f^* \tilde{k} f + 2\text{Re}(f^* \tilde{G}) - \tilde{p})^2 \\
+ \text{Tr}[\gamma \tilde{k} \cdot \gamma \tilde{k}] + \alpha^*(\tilde{k} \cdot \tilde{k}) + \text{Tr}[|\tilde{k}|^2 \gamma] \geq 0,
4. 2\text{Re}(\alpha^*[(\tilde{G} + \tilde{k} f)^{\gamma^2}]) + \text{Tr}[(2\gamma + 1)(\tilde{G} + \tilde{k} f) \cdot (\tilde{G} + \tilde{k} f)^*] \geq 0.

The energy of a pure quasifree state in the variables \( f \) and \( \tilde{\rho} \) is

\[ \hat{E}_{\rho,\tilde{\rho}}(f, \tilde{\rho}) = \frac{1}{2} \left\{ (\text{Tr}[\frac{1}{2} \cosh(2\tilde{\rho} - 1) \tilde{k}]) + f^* \tilde{k} f + 2\text{Re}(f^* \tilde{G}) - \tilde{p} \right\}^2 \\
+ \text{Tr}[\frac{1}{2} \cosh(2\tilde{\rho} - 1) \tilde{k}] \cdot \frac{1}{2} (\cosh(2\tilde{\rho} - 1) \tilde{k}] \\
+ \text{Tr}[\frac{1}{2} \sinh(2\tilde{\rho}) \tilde{k}] \cdot \frac{1}{2} \sinh(2\tilde{\rho}) \tilde{k}] + \text{Tr}[|\tilde{k}|^2 \frac{1}{2} (\cosh(2\tilde{\rho} - 1)] \\
+ 2\text{Re}(\frac{1}{2} \sinh(2\tilde{\rho}) (\tilde{G} + \tilde{k} f) \cdot (\tilde{G} + \tilde{k} f)^*) \\
+ \text{Tr}[(2\frac{1}{2} (\cosh(2\tilde{\rho} - 1) + 1)(\tilde{G} + \tilde{k} f) \cdot (\tilde{G} + \tilde{k} f)^*)] \\
+ \text{Tr}[\frac{1}{2} (\cosh(2\tilde{\rho} - 1)) |\tilde{k}] + f^*|\tilde{k}| f \]  

(V.61)

Proof. Using the Weyl operators,

\[ E_{\rho,\tilde{\rho}}(f, \gamma, \alpha) := \text{Tr}[H_{\rho,\tilde{\rho}}] = \text{Tr}[H_{\rho,\tilde{\rho}}(f) \tilde{\rho}] \]

where \( H_{\rho,\tilde{\rho}}(f) = W(\sqrt{2} f / i)^* H_{\rho,\tilde{\rho}} W(\sqrt{2} f / i) \) and \( \tilde{\rho} = W(\sqrt{2} f / i)^* \rho W(\sqrt{2} f / i). \) so that \( \tilde{\rho} \) is centered. Modulo terms of odd order, which vanish when we take the trace against a centered quasifree state, \( H_{\rho,\tilde{\rho}}(f) \) equals

\[ H_{\rho,\tilde{\rho}}(f) = \frac{1}{2} (d\Gamma(\tilde{k}) + f^* \tilde{k} f + 2\text{Re}(\alpha^* \tilde{k} f + \tilde{G})) + 2\text{Re}(f^* \tilde{G}) - \tilde{p})^2 \\
+ d\Gamma(|\tilde{k}|) + f^*|\tilde{k}| f + \text{odd} \]

\[ = \frac{1}{2} (d\Gamma(\tilde{k}) + f^* \tilde{k} f + 2\text{Re}(f^* \tilde{G}) - \tilde{p})^2 \\
+ \frac{1}{2} (2\text{Re}(\alpha^* (\tilde{k} f + \tilde{G})))^2 + d\Gamma(|\tilde{k}|) + f^*|\tilde{k}| f + \text{odd}. \]
To compute \( \mathcal{E}(f, \gamma, \alpha) \) we are thus lead to compute, for \( \tilde{\varphi} \in \mathfrak{h}^3 \) and \( \tilde{u} \in \mathbb{R}^3 \),
\[
\text{Tr} \left[ \tilde{\rho} \left( d\Gamma(\tilde{k}) + \tilde{u} \right)^2 \right] \quad \text{and} \quad \text{Tr} \left[ \tilde{\rho} \left( 2\text{Re}(a(\tilde{\varphi})) \right)^2 \right].
\]
The expression of the energy as a function of \((f, \gamma, \alpha)\) then follows from Propositions \ref{prop:v2} and \ref{prop:v4}. The expression of the energy as a function of \((f, r)\) follows from Proposition \ref{prop:iv12}.

**Proposition V.2.** Let \( \tilde{u} \in \mathbb{R}^3 \), then
\[
0 \leq \text{Tr}[\tilde{\rho}(d\Gamma(\tilde{k}) + \tilde{u})^2] = (\text{Tr}[\gamma \tilde{k}] + \tilde{u})^2 - \text{Tr}[\gamma \tilde{k}]^2 + \text{Tr}[\gamma \tilde{k} \cdot \gamma \tilde{k}] + \text{Tr}[\gamma \tilde{k}]^2 + \alpha^*(\tilde{k} \cdot \tilde{k})\alpha + \text{Tr}[\tilde{k}^2 \gamma].
\]
This condition is used with \( \tilde{u} = \tilde{p} - f^* \tilde{k} f - 2\text{Re}(f^* \tilde{G}). \)

**Proof.** Indeed,
\[
(d\Gamma(\tilde{k}) + \tilde{u})^2 = d\Gamma(\tilde{k})^2 + 2d\Gamma(\tilde{k}) \cdot \tilde{u} + \tilde{u}^2.
\]
Then we use that \( \text{Tr}[\tilde{\rho} d\Gamma(\tilde{k})] = \text{Tr}[\gamma \tilde{k}] \), add and subtract \( \text{Tr}[\gamma \tilde{k}]^2 \) to complete the square and compute \( \text{Tr}[\tilde{\rho} d\Gamma(\tilde{k})^2] \) using Lemma \ref{lemma:v3}.

**Lemma V.3.** Let \( X \in \mathcal{B}^{1,1} \), then
\[
0 \leq \text{Tr}[\tilde{\rho}d\Gamma(X)d\Gamma(X)^*] = \text{Tr}[\gamma X \gamma X^*] + \text{Tr}[\gamma X]^2 + \alpha^*(X \otimes X^*)\alpha + \text{Tr}[XX^* \gamma].
\]

**Proof.** Indeed, using Equation \ref{eq:v55},
\[
\text{Tr}[\tilde{\rho}d\Gamma(X)d\Gamma(X)^*] = \text{Tr}[\tilde{\rho} \int X(k_1, k'_1)X(k_2, k'_2)\alpha^*(k_1)\alpha^*(k_2)\alpha(k'_2)\alpha(k'_1)dk_1dk_2dk'_1dk'_2 + d\Gamma(X^*)] = \text{Tr}[\gamma X \gamma X^*] + \text{Tr}[\gamma X^* \gamma] + \text{Tr}[\gamma X \gamma X^*] + \alpha^*(X \otimes X^*)\alpha + \text{Tr}[XX^* \gamma].
\]

**Proposition V.4.** Let \( \varphi \in \mathfrak{h} \), then
\[
0 \leq \text{Tr}[\tilde{\rho}(\alpha^*(\varphi) + a(\varphi))^2] = 2\text{Re}(\alpha^*(\varphi^2)) + \text{Tr}[(2\gamma + 1)\varphi^2] \quad \text{(V.62)}
\]
and \( |2\text{Re}(\alpha^*(\varphi^2))| \leq \text{Tr}[(2\gamma + 1)\varphi^2] \).

This condition is used with the three components of \( \varphi = \tilde{G} + \tilde{k} f \).

**Proof.** A computation using the canonical commutation relations yields
\[
\text{Tr}[\tilde{\rho}(\alpha^*(\varphi) + a(\varphi))^2] = \text{Tr}[\tilde{\rho} (\alpha^*(\varphi))^2] + \text{Tr}[(\alpha^*(\varphi) + \alpha(\varphi)\alpha + a(\varphi)\alpha^*(\varphi))^2] = \alpha^* \varphi^2 + \varphi \varphi^2 \alpha + \text{Tr}[\gamma \varphi^2 + (\gamma + 1)\psi^2].
\]
VI Minimization over Coherent States

In this section we consider the problem of minimizing the energy over coherent states and show that there is a unique minimizer. We also calculate the lower orders of the energy at the minimizer seen as a function of the total momentum $\vec{p}$.

For this section we can take $\sigma = 0$ if we consider the parameter $f$ in the energy to be in $\tilde{h} := L^2(S_{\sigma, \Lambda} \times \mathbb{Z}_2, (\frac{1}{2} |\vec{k}|^2 + |\vec{k}|)dk)$. Recall that $S_{\sigma, \Lambda} = \{ \vec{k} \in \mathbb{R}^3 | \sigma \leq |\vec{k}| \leq \Lambda \}$. We also recall that for $z$ and $z'$ in some Hilbert space, $z^* z' = \langle z, z' \rangle$ (see Section III).

**Remark VI.1.** For a coherent state (see Definition III.5) the energy reduces to

$$E_{g, \vec{p}}(f) = \frac{1}{2} ||\vec{G}||^2 + \frac{1}{2} (f^* \vec{k} f + 2 \text{Re}(f^* \vec{G}) - \vec{p})^2 + f^* (\frac{1}{2} |\vec{k}|^2 + |\vec{k}|) f. \quad (VI.63)$$

Note that, for $\sigma > 0$, $h = L^2(S_{\sigma, \Lambda} \times \mathbb{Z}_2, dk) = \tilde{h}$, while for $\sigma = 0$, $h \subset \tilde{h}$, and $E_{g, \vec{p}}(f)$ extends to $\tilde{h}$ by using Equation (VI.63).

**Theorem VI.2.** There exists a universal constant $C < \infty$ such that, for $0 \leq \sigma < \Lambda < \infty$, $g^2 \ln(\Lambda + 2) \leq C$ and $|\vec{p}| \leq 1/3$, there exists a unique $f_{\vec{p}}$ which minimizes $E_{g, \vec{p}}$ in $\tilde{h}$. The minimizer $f_{\vec{p}}$ solves the system of equations

$$f_{\vec{p}} = \frac{\vec{u}_{\vec{p}} \cdot \vec{G}}{\frac{1}{2} |\vec{k}|^2 + \frac{1}{2} |\vec{k}| - \vec{k} \cdot \vec{u}_{\vec{p}}}, \quad (VI.64)$$

$$\vec{u}_{\vec{p}} = \vec{p} - 2 \text{Re}(f_{\vec{p}} \vec{G}) - f_{\vec{p}} \vec{k} f_{\vec{p}}, \quad (VI.65)$$

with $|\vec{u}_{\vec{p}}| \leq |\vec{p}|$.

**Remark VI.3.** Our hypotheses are similar those of Chen, Fröhlich, and Pizzo [10], where their vector $\nabla E_{g, \vec{p}}$ is analogous to $\vec{u}_{\vec{p}}$ in our notations. The construction of $\vec{u}_{\vec{p}}$ as the solution of a fixed point problem and the dependency in the parameter $\vec{p}$ imply that the map $\vec{p} \mapsto \vec{u}_{\vec{p}}$ is of class $C^\infty$.

**Remark VI.4.** We note that we also expect to have $\vec{u}_{\vec{p}}$ in the neighborhood of $\vec{p}$.

**Remark VI.5.** The minimizer is constructed as the solution of a fixed point problem. As a result the application

$$(\sigma, \Lambda, g, \vec{p}) \mapsto \inf_{\rho \in \text{coh}} \text{Tr}[H_{g, \vec{p}} \rho]$$

is continuous on the domain defined by Theorem VI.2 and at $\sigma, \Lambda$ fixed,

$$(g, \vec{p}) \mapsto \inf_{\rho \in \text{coh}} \text{Tr}[H_{g, \vec{p}} \rho]$$

is analytic for $g^2 < C/ \ln(\Lambda + 2)$ and $|\vec{p}| < 1/3$. 
The assumption on $|\vec{p}| \leq 1/3$ is much weaker than in the quasifree state case where we need to have $\vec{p}$ smaller than a constant $C$ which may be small. In fact the $1/3$ is arbitrary and one may suppose only $|\vec{p}| \leq R$ for some constant $R < 1$, but since this would result in a heavier exposition without providing additional relevant information, we restrict to $|\vec{p}| \leq 1/3$.

To prove Theorem VI.2 we first show that the Equations (VI.64) and (VI.65) are necessarily verified by a minimizer. We then show the existence and uniqueness of a solution to these equations by a fixed point argument.

Proof of Theorem VI.2. Assume there is a point $f_{\vec{p}}$ where the minimum is attained. The partial derivative of the energy at the point $f_{\vec{p}}$

$$\partial_{f_{\vec{p}}} \mathcal{E}(f_{\vec{p}}) = ((f_{\vec{p}}^* \vec{f} \cdot \vec{k} + |\vec{k}|^2 + |\vec{k}|) f_{\vec{p}} - (\vec{p} - f_{\vec{p}}^* \vec{k} f_{\vec{p}} - 2\Re(\vec{p} \cdot \vec{G})) \cdot \vec{G}$$

then vanishes, where the derivative $\partial_{f_{\vec{p}}} \mathcal{E}(f)$ at a point $f$ is the unique vector in $\tilde{\mathfrak{h}}^* \approx L^2(S_{\sigma, \Lambda}, (1/2 |\vec{k}|^2 + |\vec{k}|)^{-1} dk)$ defined by

$$\mathcal{E}(f + \delta f) - \mathcal{E}(f) = 2\Re(\delta f^* \partial_{f_{\vec{p}}} \mathcal{E}(f)) + o(\|\delta f\|_{\tilde{\mathfrak{h}}})$$

with $f, \delta f \in \tilde{\mathfrak{h}}$. Observe that

$$0 \leq \mathcal{E}_{\vec{p}, \vec{p}}(0) - \mathcal{E}_{\vec{p}, \vec{p}}(f_{\vec{p}}) = \frac{1}{2} |\vec{p}|^2 - \frac{1}{2} (f_{\vec{p}}^* \vec{k} f_{\vec{p}} + 2\Re(f_{\vec{p}}^* \vec{G}) - \vec{p})^2 - f_{\vec{p}}^* (\frac{1}{2} |\vec{k}|^2 + |\vec{k}|) f_{\vec{p}}$$

and hence $|\vec{p}| \geq |\bar{u}_{\vec{p}}|$ with $\bar{u}_{\vec{p}} := \vec{p} - f_{\vec{p}}^* \vec{k} f_{\vec{p}} - 2\Re(f_{\vec{p}}^* \vec{G})$. Since $|\bar{u}_{\vec{p}}| \leq |\vec{p}| < 1$, it makes sense to write

$$f_{\vec{p}} = \frac{\bar{u}_{\vec{p}} \cdot \vec{G}}{1/2 |\vec{k}|^2 + |\vec{k}| - \bar{u}_{\vec{p}} \cdot \vec{k}}.$$ 

Hence the minimum point $f_{\vec{p}}$ satisfies Equations (VI.64) and (VI.65). It is in particular sufficient to prove that there exist a unique $\bar{u}_{\vec{p}}$ in a ball $\tilde{B}(0, r)$ with $r \geq |\vec{p}|$ such that the function in Equation (VI.64) satisfies also Equation (VI.65) to prove the existence and uniqueness of a minimizer.

Proof of the existence and uniqueness of a solution. Let $1/3 < r < 1, \bar{u} \in \mathbb{R}^3, |\bar{u}| \leq r < 1$ and

$$\Phi_{\bar{u}}(\vec{k}) = \frac{\bar{u} \cdot \vec{G}(\vec{k})}{1/2 |\vec{k}|^2 + |\vec{k}| - \bar{u} \cdot \vec{k}}.$$ 

Observe that $\Phi_{\bar{u}} \in \tilde{\mathfrak{Z}}$, indeed, if $|\bar{u}| < 1$ then $1/2 |\vec{k}|^2 + |\vec{k}| - \bar{u} \cdot \vec{k} \geq (1 - r)(1/2 |\vec{k}|^2 + |\vec{k}|)$. 


and with $\xi(\vec{k}) = \xi(\vec{k}, +) + \xi(\vec{k}, -)$,

$$\int_{|\vec{k}| \in |\sigma, \Lambda|} \frac{|\vec{k}|}{\Pi(\vec{k})} d\vec{k} < g^2 \int_{|\vec{k}| \in |\sigma, \Lambda|} \frac{1}{|\vec{k}|} \frac{|\vec{u} \cdot \xi(\vec{k})|^2}{2|\vec{k}|^2 + |\vec{k}|} d\vec{k} < +\infty,$$

for some universal constant $C_0 > 0$. Observe then that

$$\int_{|\vec{k}| \in |\sigma, \Lambda|} \frac{|\vec{k}|^2}{2|\vec{k}|^2 + |\vec{k}|} d\vec{k} \leq C_0 g^2 \ln(\Lambda + 2)$$

for some universal constant $C_0 > 0$. It follows that $\Phi_{\vec{u}}^* \vec{G} \in L^1(S_{\sigma, \Lambda} \times \mathbb{Z}_2)$. Note that if $\sigma = 0$ then $\Phi_{\vec{u}} \notin L^2(S_{\sigma, \Lambda} \times \mathbb{Z}_2)$ (for $\vec{u} \neq 0$).

We can thus define the application

$$\vec{B}(0, r) \ni \vec{u} : = \vec{p} - \Phi_{\vec{u}}^* \vec{k} \Phi_{\vec{u}} - 2 \text{Re}(\Phi_{\vec{u}}^* \vec{G}) \in \mathbb{R}^3.$$

We check that the hypotheses of the Banach-Picard fixed point theorem are verified on the ball $B(0, r)$, which will prove the result.

**Stability:** If $g^2 \ln(\Lambda + 2)$ is sufficiently small, we get from

$$|\vec{\Psi}(\vec{u})| \leq |\Phi_{\vec{u}}^* \vec{k} \Phi_{\vec{u}}| + |2 \text{Re}(\Phi_{\vec{u}}^* \vec{G})| + |\vec{p}|$$

and the estimates above that the sum of the two first terms is smaller than $r - 1/3$ and since $|\vec{p}| \leq 1/3$ the map $\vec{\Psi}$ sends $B(0, r)$ into itself,

$$\vec{\Psi}(B(0, r)) \subseteq B(0, r).$$

**Contraction:** For $\vec{u}$ and $\vec{v}$ in $\vec{B}(0, r)$, we have that

$$|\Phi_{\vec{u}}(\vec{k}) - \Phi_{\vec{v}}(\vec{k})| = |\vec{u} \cdot \vec{G}(\vec{k}) - \vec{v} \cdot \vec{G}(\vec{k})| = \frac{1}{2} |\vec{k}|^2 + |\vec{k}| - \vec{k} \cdot \vec{u} - \frac{1}{2} |\vec{k}|^2 + |\vec{k}| - \vec{k} \cdot \vec{u} - \frac{1}{2} |\vec{k}|^2 + |\vec{k}| - \vec{k} \cdot \vec{u} + |\vec{u} - \vec{v}| |\vec{G}(\vec{k})| |\vec{k} - \vec{k} \cdot \vec{u}|(\frac{1}{2} |\vec{k}|^2 + |\vec{k}||)

\leq |\vec{u} - \vec{v}| |\vec{G}(\vec{k})| - \frac{1}{(1 - r)^2} + \frac{1}{(1 - r)}$$

$$\leq |\vec{u} - \vec{v}| |\vec{G}(\vec{k})| \frac{1}{(1 - r)^2}.$$
For the term $2\text{Re}(\Phi^*_u \vec{G})$, we observe that

\[
|2\text{Re}(\Phi^*_u \vec{G}) - 2\text{Re}(\Phi^*_v \vec{G})| \\
\leq g^2 2|\vec{u} - \vec{v}| \frac{1}{(1 - r)^2} \int_{[k] \in [\sigma, \Lambda]} \frac{1}{|k|} \frac{1}{\frac{1}{2}|k|^2 + |k|} d^3 k \\
\leq C_1 g^2 \ln(2 + \Lambda) 2|\vec{u} - \vec{v}| \frac{1}{(1 - r)^2}.
\]

Note that, for $g^2 \ln(2 + \Lambda) < (1 - r)^2 / (3C_1)$,

\[
|2\text{Re}(\Phi^*_u \vec{G}) - 2\text{Re}(\Phi^*_v \vec{G})| < \frac{1}{3} |\vec{u} - \vec{v}|.
\]

Finally, for the term $\Phi^*_u \vec{k} \Phi^*_u - \Phi^*_v \vec{k} \Phi^*_v$, we obtain the estimate

\[
|\Phi^*_u \vec{k} \Phi^*_u - \Phi^*_v \vec{k} \Phi^*_v| \\
\leq \int_{[k] \in [\sigma, \Lambda]} \left( \frac{1}{|k|^2 + |\vec{k}|} |\Phi^*_u(\vec{k}) - \Phi^*_v(\vec{k})| |(\Phi^*_u(\vec{k}) + |\Phi^*_v(\vec{k})|)\right) d^3 k \\
\leq \frac{|\vec{u} - \vec{v}|}{(1 - r)^2} \int_{[k] \in [\sigma, \Lambda]} |\vec{G}(\vec{k})||\Phi^*_u(\vec{k})| + |\Phi^*_v(\vec{k})|\right) d\vec{k} \\
\leq \frac{|\vec{u} - \vec{v}|}{(1 - r)^2} \left( \frac{1}{2} |\vec{k}|^2 + |\vec{k}| \right)^{1/2} \left( \frac{1}{2} |\vec{k}|^2 + |\vec{k}| + 2 |\vec{G} \cdot \vec{k}| \right) \\
\leq C_2 |\vec{u} - \vec{v}| \left( |\vec{u}| + |\vec{v}|\right) g^2 \ln(\Lambda + 2),
\]

and thus this term can be controlled for $|g \ln(\Lambda + 2)|$ sufficiently small by $\frac{1}{3} |\vec{u} - \vec{v}|$.

We thus get a contraction

\[
|\vec{\Psi}(\vec{u}) - \vec{\Psi}(\vec{u}')| \leq \frac{2}{3} |\vec{u} - \vec{u}'|
\]

and with $f_\vec{p} = \Phi_{\vec{p}}$ Equation [VI.64] is solved.

We then obtain several explicit results from the expression of this minimizer.

**Corollary VI.7.** With the same hypotheses as in Theorem [VI.2]

1. For $0 \leq \sigma < \Lambda < \infty$,

\[
\inf_{f \in \mathfrak{h}} E_{g, \vec{p}}(f) = \inf_{f \in \mathfrak{h}} E_{g, \vec{p}}(f) = E_{g, \vec{p}}(f_\vec{p}),
\]

and for $0 < \sigma < \Lambda < \infty$, we have that $f_\vec{p} \in \mathfrak{h}$.

2. For fixed $g, \sigma, \Lambda$, as a function of $\vec{p}$,

\[
E_{g, \vec{p}}(f_\vec{p}) = E_{g, \vec{p}}(0) - \vec{p} \cdot \vec{G}^\star \frac{1}{\frac{1}{2}|\vec{k}|^2 + |\vec{k}| + 2 \vec{G} \cdot \vec{k} + \vec{G}^\star \vec{G} \cdot \vec{p} + O(|\vec{p}|^3)).
\]
3. For all $f$ in $\tilde{\mathcal{H}}$, 

\[
E_{g, \tilde{p}}(f_{\tilde{p}} + f) = E_{g, \tilde{p}}(f_{\tilde{p}}) + f^*(\frac{1}{2}|\tilde{k}|^2 + |\tilde{k}| - \tilde{u}_{\tilde{p}} \cdot \tilde{k}) f + \frac{1}{4}(f^* \tilde{k} f + 2\text{Re}(f_{\tilde{p}}^* \tilde{k} f) + 2\text{Re}(f^* \tilde{G}))^2 . \quad (VI.66)
\]

4. The energy $E_{g, \tilde{p}}(f_{\tilde{p}})$ of the minimizer compared to the energy of the vacuum state $E_{g, \tilde{p}}(0)$ is 

\[
E_{g, \tilde{p}}(f_{\tilde{p}}) = E_{g, \tilde{p}}(0) - \frac{1}{2} 2\text{Re}(f_{\tilde{p}}^* \tilde{u}_{\tilde{p}} \cdot \tilde{G}) - \frac{1}{2}|\tilde{u}_{\tilde{p}} - \tilde{p}|^2 .
\]

Note that the term $2\text{Re}(f_{\tilde{p}}^* \tilde{u}_{\tilde{p}} \cdot \tilde{G})$ is non-negative.

Proof of 1. is straightforward for $\sigma > 0$. For $\sigma = 0$ one can approximate the minimizer in $\tilde{\mathcal{H}}$ by functions in $\mathcal{H}$.

Proof of 2. The expression of the energy $E_{g, \tilde{p}}(f)$ given in Equation (VI.63) implies that $E_{g, \tilde{p}}(f) \geq \frac{1}{2}||\tilde{G}||^2$, and for $\tilde{p} = 0$ this minimum is only attained at the point $f_{\tilde{0}} = 0$. It follows that $f_{\tilde{p}} = \partial_{\tilde{p}}f_{\tilde{0}} \cdot \tilde{p} + O(|\tilde{p}|^2)$.

From Equation (VI.65) we deduce 

\[
\tilde{u}_{\tilde{p}} = \tilde{p} - 2\text{Re}((\partial_{\tilde{p}}f_{\tilde{0}} \cdot \tilde{p}^*) \tilde{G}) + O(|\tilde{p}|^2)
\]

and thus 

\[
f_{\tilde{p}} = (\tilde{p} - 2\text{Re}((\partial_{\tilde{p}}f_{\tilde{0}} \cdot \tilde{p}^*) \tilde{G})) \cdot \tilde{G} + O(|\tilde{p}|^2)
\]

Expanding the left hand side of this equality in $\tilde{0}$ brings 

\[
\partial_{\tilde{p}}f_{\tilde{0}} \cdot \tilde{p} = (\frac{1}{2} |\tilde{k}|^2 + |\tilde{k}|)^{-1}(\tilde{p} - 2\text{Re}((\partial_{\tilde{p}}f_{\tilde{0}} \cdot \tilde{p}^*) \tilde{G})) \cdot \tilde{G}
\]

and hence $\partial_{\tilde{p}}f_{\tilde{0}} = (\frac{1}{2} |\tilde{k}|^2 + |\tilde{k}| + 2\tilde{G} \cdot \tilde{G}^*)^{-1}\tilde{G}$. The expansion of $f_{\tilde{p}}$ to the second order is then 

\[
f_{\tilde{p}} = (\frac{1}{2} |\tilde{k}|^2 + |\tilde{k}| + 2\tilde{G} \cdot \tilde{G}^*)^{-1}\tilde{G} \cdot \tilde{p} + O(|\tilde{p}|^2) .
\]

We can compute the energy modulo error terms in $O(|\tilde{p}|^3)$. To have less heavy com-
putations we set $A = \frac{1}{2}|\vec{k}|^2 + |\vec{k}| + 2\vec{G} \cdot \vec{G^*}$ and get
\[
\mathcal{E}_{g,\bar{p}}(f_{\bar{p}}) - \frac{1}{2} ||\vec{G}||^2 - \frac{1}{2} |\bar{p}|^2
\equiv - \frac{1}{2} |\bar{p}|^2 + \frac{1}{2} (2 \text{Re}(\bar{p} \cdot \partial_{\bar{p}} f_{\bar{p}} \cdot \vec{G} - \bar{p})^2 + \bar{p} \cdot \partial_{\bar{p}} f_{\bar{p}} (\frac{1}{2} |\vec{k}|^2 + |\vec{k}|)) \partial_{\bar{p}} f_{\bar{p}} \cdot \bar{p} \\
\equiv \frac{1}{2} (2 \text{Re}(\bar{p} \cdot \vec{G^*} A^{-1} \vec{G}))^2 - 2 \bar{p} \cdot \vec{G^*} A^{-1} \vec{G} \cdot \bar{p} \\
+ \bar{p} \cdot \vec{G^*} A^{-1} (\frac{1}{2} |\vec{k}|^2 + |\vec{k}|) A^{-1} \vec{G} \cdot \bar{p} \\
\equiv 2(\bar{p} \cdot \vec{G^*} A^{-1} \vec{G})^2 + \bar{p} \cdot \vec{G^*} A^{-1} ((\frac{1}{2} |\vec{k}|^2 + |\vec{k}|) - 2A) A^{-1} \vec{G} \cdot \bar{p} \\
\equiv \bar{p} \cdot \vec{G^*} A^{-1} 2\vec{G} \cdot \vec{G^*} A^{-1} \vec{G} \cdot \bar{p} - \bar{p} \cdot \vec{G^*} A^{-1} (\frac{1}{2} |\vec{k}|^2 + |\vec{k}| + 4\vec{G} \cdot \vec{G^*}) A^{-1} \vec{G} \cdot \bar{p} \\
\equiv -\bar{p} \cdot \vec{G^*} (\frac{1}{2} |\vec{k}|^2 + |\vec{k}| + 2\vec{G} \cdot \vec{G^*})^{-1} \vec{G} \cdot \bar{p}
\]
which yields the result.

Proof of 3] The Taylor expansion of the energy around $f_{\bar{p}}$ is
\[
\mathcal{E}_{g,\bar{p}}(f_{\bar{p}} + f) = \mathcal{E}_{g,\bar{p}}(f_{\bar{p}}) + f^* \partial_f \mathcal{E}(f_{\bar{p}}) + \partial_{f} \mathcal{E}(f_{\bar{p}}) f \\
+ \frac{1}{2} \left( (f^* \vec{k} \vec{f} + 2 \text{Re}(f^* \vec{G})) + 2 \text{Re}(f_{\bar{p}} \vec{k} \vec{f}))^2 \right. \\
\left. + 2(f_{\bar{p}} \vec{k} \vec{f} + 2 \text{Re}(f_{\bar{p}} \vec{G}) - \bar{p} \cdot f^* \vec{k} \vec{f} + f^* |\vec{k}|^2 \vec{f}) + f^* |\vec{k}| \vec{f} \right).
\]
Since $\partial_f \mathcal{E}(f_{\bar{p}})$ vanishes this gives Equation (VI.66).

Proof of 4] It is sufficient to replace $f$ by $-f_{\bar{p}}$ in Equation (VI.66). The observation
\[
f_{\bar{p}} \vec{u}_{\bar{p}} \cdot \vec{G} = \int \frac{(\vec{u}_{\bar{p}} \cdot \vec{G}(\vec{k}))^2 \vec{k}}{\frac{1}{2} |\vec{k}|^2 + |\vec{k}| - \vec{k} \cdot \vec{u}_{\bar{p}}} dk
\]
shows that $2 \text{Re}(f_{\bar{p}}^* \vec{u}_{\bar{p}} \cdot \vec{G})$ is non-negative since $|\vec{u}_{\bar{p}}| < 1$. \qed

VII Renormalized Electron Mass for Coherent States

In this section we use the coherent state minimizer found in Section VI in order to determine the renormalized electron mass.

Proposition VII.1. The renormalized electron mass for coherent states $m_{coh}(g, \Lambda)$ defined by
\[
\mathcal{E}_{g,\bar{p}}(f_{\bar{p}}, \bar{p}) - \mathcal{E}_{g,\bar{p}}(f_{\bar{p}}, \bar{p}) = \frac{1}{2m_{coh}(g, \Lambda)} |\bar{p}|^2 + \mathcal{O}(|\bar{p}|^3),
\]
is
\[
m_{coh}(g, \Lambda) = 1 + \frac{32\pi}{3} g^2 \ln(1 + \frac{\Lambda}{2}).
\]
Remark VII.2. This result agrees with $\frac{m_{\text{eff}}}{m}$ obtained in [12] to second order in $g$, taking into account that $g$ in [12] equals $\sqrt{\frac{\alpha}{\pi}}$ in the present paper, that $\omega(\vec{k}) = |\vec{k}|$, and that the mass $m$ of the electron is one in our units. See also (among others) [11] or [3].

Proof. From

$$E_{g,\vec{p}}(f_{g,\vec{p}}) - E_{g,\vec{p}}(0) = \frac{1}{2} |\vec{p}|^2 - \vec{p} \cdot \vec{G}^* + \frac{1}{2} |\vec{k}|^2 + |\vec{k}| + 2 \vec{G} \cdot \vec{G}^* \vec{p} + O(|\vec{p}|^3)$$

and the fact that for $\vec{p} = \vec{0}$ the minimizer is $f_{g,\vec{0}} = 0$, it follows that

$$E_{g,\vec{p}}(f_{g,\vec{p}}) - E_{g,\vec{0}}(f_{g,\vec{0}}) = E_{g,\vec{p}}(f_{g,\vec{p}}) - E_{g,\vec{p}}(0) + \frac{1}{2} |\vec{p}|^2$$

$$= \frac{1}{2} |\vec{p}|^2 - \vec{p} \cdot \vec{G}^* + \frac{1}{2} |\vec{k}|^2 + |\vec{k}| + 2 \vec{G} \cdot \vec{G}^* \vec{p} + O(|\vec{p}|^3)$$

where we used point 2 of Theorem VI.2 for the last equality. The power expansion of $(1 - t)^{-1}$ yields

$$\frac{1}{2} |\vec{k}|^2 + |\vec{k}| + 2 \vec{G} \cdot \vec{G}^* = \sum_{j=0}^{\infty} \frac{1}{j! |\vec{k}|^2 + |\vec{k}|} \left( -2 \vec{G} \cdot \vec{G}^* + \frac{1}{2} |\vec{k}|^2 + |\vec{k}| \right)^j$$

and hence

$$E_{g,\vec{p}}(f_{g,\vec{p}}) - E_{g,\vec{0}}(f_{g,\vec{0}}) = \frac{1}{2} |\vec{p}|^2 - \sum_{j=1}^{\infty} \vec{p} \left( -2 \vec{G} \cdot \vec{G}^* + \frac{1}{2} |\vec{k}|^2 + |\vec{k}| \right) \vec{p} + O(|\vec{p}|^3)$$

$$= \frac{1}{2} \vec{p} \cdot \left( Id_{\mathbb{R}^3} + 2 \vec{G} \cdot \vec{G}^* + \frac{1}{2} |\vec{k}|^2 + |\vec{k}| \right) \vec{p} + O(|\vec{p}|^3).$$

The coherent renormalized mass is thus

$$m_{\text{coh}} = Id_{\mathbb{R}^3} + 2 \vec{G} \cdot \vec{G}^* + \frac{1}{2} |\vec{k}|^2 + |\vec{k}| \vec{G}.$$

The proof is achieved by the computation of an integral that we give in a separate lemma since it will be useful again later.

Lemma VII.3. With $\sigma = 0$,

$$2 \vec{G} \cdot \vec{G}^* = g^2 \frac{4}{3} \ln(1 + \frac{\Lambda}{2}) Id_{\mathbb{R}^3}.$$
Proof. Using spherical coordinates

\[ 2\vec{p} \cdot \vec{G} = 2g^2 \int_{|k| \leq \Lambda} \frac{|\vec{p} \cdot \vec{e}_+(\vec{k})|^2 + |\vec{p} \cdot \vec{e}_-(\vec{k})|^2}{\frac{1}{2} |k|^3 + |k|^2} d\vec{k} \]

which yields the result.

VIII Minimization over Quasifree States

In this section we minimize the energy functional over the set of pure quasifree states. We first obtain the existence and uniqueness of a pure quasifree state in Theorem VIII.3. As shown in Section IV this proves the existence (but not the uniqueness) of a minimizer in the class of quasifree states. We then compute asymptotics for small coupling \( g \) and momentum \( p \) of the parameters \( f_g, \vec{p} \) and \( r_g, \vec{p} \) describing the pure quasifree state minimizing the energy in Theorem VIII.6. From these asymptotics we deduce an expansion of the energy at the minimizer. Finally we give in Theorem VIII.8 the Lagrange equations in terms of the parameters \( f_g, \alpha \) and \( \gamma \).

Remark VIII.1. We believe that the assumption \( \sigma > 0 \) is unnecessary, but we have no proof for this assertion and do not follow it here, because the infrared singular limit \( \sigma \to 0 \) is not our concern in this paper.

VIII.1 Existence and Uniqueness of a Minimizer of the Energy over Pure Quasifree States

In this section we minimize the energy over all pure quasifree states and show that there is a unique minimizer.

Definition VIII.2. Let \( \mathfrak{h} \) be a \( \mathbb{C} \)-Hilbert space. Let \( Y \) be the \( \mathbb{R} \)-Hilbert space of antilinear operators \( \hat{r} \) on \( \mathfrak{h} \), self-adjoint in the sense that \( \forall z, z' \in \mathfrak{h}, \langle z, \hat{r} z' \rangle = \langle z', \hat{r} z \rangle \), and Hilbert-Schmidt in the sense that the positive operator \( \hat{r}^2 \) is trace class. The space \( X = \mathfrak{h} \times Y \) with the scalar product

\[ \langle (f, \hat{r}), (f', \hat{r}') \rangle_X = f^* f' + \text{Tr}[\hat{r}\hat{r}'] \]

is an \( \mathbb{R} \)-Hilbert space.

Keeping \( \sigma > 0 \), we only need to use \( \mathfrak{h} = L^2(S_{\sigma, \Lambda} \times \mathbb{Z}_2) \) in this section.
Theorem VIII.3. Let \( 0 < \sigma < \Lambda < \infty \). There exists \( C > 0 \) such that for \( g, |\tilde{\rho}| \leq C \) there exists a unique minimizer for \( \hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) \).

Proof. This result follows from convexity and coercivity arguments. By Proposition VIII.4, \( \hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) \) is strictly \( \theta \)-convex (i.e., uniformly strictly convex) on \( \bar{B}_X(0, R) \) for some \( R > 0 \) and \( \theta > 0 \). Since \( \hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) \) is strongly continuous on the closed and convex set \( \bar{B}_X(0, R) \) of the Hilbert space \( X \) we get the existence and uniqueness of a minimizer in \( \bar{B}_X(0, R) \). (See for example [1].) The uniform strict convexity allows to prove directly that a minimizing sequence is a Cauchy sequence.) Proposition VIII.5 then proves that it is the only minimum of \( \hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) \) on the whole space.

Note that to use Propositions VIII.4 and VIII.5 we need to restrict to values of \( g \) and \( |\tilde{\rho}| \) smaller than some constant \( C > 0 \).

Proposition VIII.4 (Convexity). There exist \( 0 < C, R < \infty \) such that for \( g \leq C \) and \( |\tilde{\rho}| \leq \frac{1}{2} \), the Hessian of the energy satisfies \( H\hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) \geq \frac{4}{\mu} 1_X \) on the ball \( B_X(0, R) \).

Proof. We use that strict positivity of the Hessian implies strict convexity and thus first compute the Hessian in \((0, 0)\). The Hessian \( H\hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) \in B(X) \) is defined using the Fréchet derivative

\[
\hat{E}_{g, \tilde{\rho}}(f + \delta f, \hat{\rho} + \delta \hat{\rho}) - \hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) = D\hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) (\delta f, \delta \hat{\rho}) + \frac{1}{2} \langle (\delta f, \delta \hat{\rho}), H\hat{E}_{g, \tilde{\rho}}(f, \hat{\rho}) (\delta f, \delta \hat{\rho}) \rangle_X + o(\|\delta f, \delta \hat{\rho}\|_X^2)
\]

with \( D\hat{E}_{g, \tilde{\rho}}(0, 0) \in B(X, \mathbb{R}) \). (Note that differentiability is granted in this case because \( |\tilde{\rho}| \leq \Lambda < \infty \).) For any \( \mu > 0 \), \( \forall (f, \hat{\rho}) \in X \),

\[
\langle (f, \hat{\rho}), \frac{1}{2} H\hat{E}_{g, \tilde{\rho}}(0, 0) (f, \hat{\rho}) \rangle_X
\]

\[
= 2\text{Re}(\hat{\rho} \hat{\rho}^* f; \hat{\rho}^* \hat{\rho}) + \frac{1}{2}(2\text{Re}(f^* \hat{\rho}))^2 + \text{Tr}[\hat{\rho}^2 \hat{\rho}^* \hat{\rho}] + \frac{1}{2} \{ \text{Tr}[\hat{\rho}^2 \hat{\rho}^* \hat{\rho}] + \text{Tr}[|\hat{\rho}|^2 |\hat{\rho}|^2] \}
\]

\[
+ \text{Tr}[\hat{\rho}^2 (|\hat{\rho}| - \hat{\rho} \cdot \hat{\rho})] + f^*(\frac{1}{2}|\hat{\rho}|^2 + |\hat{\rho}| - \hat{\rho} \cdot \hat{\rho}) f
\]

\[
\geq \text{Tr}[\hat{\rho}^2 \hat{\rho}^* \hat{\rho}] - \mu \|\hat{\rho} \hat{\rho}^*\|^2 - \frac{1}{\mu} \|\hat{\rho}^2 \hat{\rho}\|^2
\]

\[
+ \frac{1}{2} \{ (2\text{Re}(f^* \hat{\rho}))^2 + \text{Tr}[\hat{\rho}^2 \hat{\rho}^* \hat{\rho}] + \text{Tr}[|\hat{\rho}|^2 |\hat{\rho}|^2] \}
\]

\[
+ \text{Tr}[\hat{\rho}^2 (|\hat{\rho}| - \hat{\rho} \cdot \hat{\rho})] + f^*(\frac{1}{2}|\hat{\rho}|^2 + |\hat{\rho}| - \hat{\rho} \cdot \hat{\rho}) f
\]

\[
\geq \text{Tr}[\hat{\rho}^2 (|\hat{\rho}| - \hat{\rho} \cdot \hat{\rho} + (1 - \mu)|\hat{\rho}^* \hat{\rho}|)] + f^*(\frac{1}{2}(1 - \mu)|\hat{\rho}|^2 + |\hat{\rho} - \hat{\rho} \cdot \hat{\rho}| f)
\]

since

\[
|2\text{Re}(\hat{\rho} \hat{\rho}^* f; \hat{\rho}^* \hat{\rho})| \leq 2 \|\hat{\rho} \hat{\rho}^*\| \|\hat{\rho}^2 \hat{\rho}\| \leq 2\sqrt{\mu} \|\hat{\rho} \hat{\rho}^*\| \frac{1}{\sqrt{\mu}} \|\hat{\rho}^2 \hat{\rho}\| \leq \mu \|\hat{\rho} \hat{\rho}^*\|^2 + \frac{1}{\mu} \|\hat{\rho}^2 \hat{\rho}\|^2.
\]
With $\mu = 2$ we obtain (with $|\vec{p}| \leq \frac{1}{2}$)
\[
\frac{1}{2} \mathcal{H}_{\mathcal{g}, \vec{p}}(0, 0)(f, \vec{r}) \geq \operatorname{Tr}[\hat{\vec{r}}^2(|\vec{k}| - \vec{k} \cdot \vec{p} - \vec{G} \cdot \vec{G}^*]) + f^*(|\vec{k}| - \vec{k} \cdot \vec{p}) f \\
\geq \operatorname{Tr}[\hat{\vec{r}}^2(|\vec{k}| \left(1 - ||\vec{k}||^{-1/2}\vec{G}||^2\right) - \vec{k} \cdot \vec{p})] + f^*(|\vec{k}| - \vec{k} \cdot \vec{p}) f \\
\geq \operatorname{Tr}[\hat{\vec{r}}^2\sigma(\frac{1}{2} - ||\vec{k}||^{-1/2}\vec{G}||^2)] + f^*\sigma f
\]
and for $g$ small enough
\[
\frac{1}{2} \mathcal{H}_{\mathcal{g}, \vec{p}}(0, 0) \geq \frac{\sigma}{4}.
\]
We then compare it with the Hessian in points near zero. Observing that the Hessian is continuous with respect to $(f, \vec{r}, \vec{p}, g)$, we deduce that there exist $R < \infty$ and $C > 0$, as asserted.

**Proposition VIII.5 (Coercivity).** Suppose $\vec{p}$ and $C > 0$ are fixed such that $\frac{1}{4}|\vec{p}|^2 + \frac{1}{4}||\vec{G}||^2 < \sigma R^2$, with the value of $R$ given by Proposition VIII.4 for any $0 < g < C$. For every $(f, \vec{r}) \in X$,
\[
\mathcal{E}_{\mathcal{g}, \vec{p}}(f, \vec{r}) \geq \operatorname{Tr}[\hat{\vec{r}}^2|\vec{k}|] + f^*|\vec{k}| f \geq \sigma \| (f, \vec{r}) \|^2_X.
\]
Since $\mathcal{E}_{\mathcal{g}, \vec{p}}(0, 0) = \frac{1}{4}|\vec{p}|^2 + \frac{1}{4}||\vec{G}||^2 < \sigma R^2$, any minimizing sequence can be assumed to take its values in $B_X(0, R)$.

**VIII.2 Asymptotics for small Coupling and Momentum**

In this section we calculate the ground state energy for small orders of the coupling constant $g$ and total momentum $\vec{p}$.

We use below an identification between self-adjoint $\mathcal{C}$-antilinear Hilbert-Schmidt operator $\hat{\vec{r}}$ and symmetric two vector generator $\hat{\vec{r}}\hat{\vec{r}}^*$ given by the relation $\langle \phi, \hat{\vec{r}}\hat{\vec{r}}^*\psi \rangle = \langle \phi \otimes \psi, r \rangle_{\mathbb{H}^2}$. Note that the self-adjointness condition for $\hat{\vec{r}}$ is equivalent to the symmetry condition $r \in \mathbb{h}^{\vee 2}$.

**Theorem VIII.6.** Let $0 < \sigma < \Lambda < \infty$. There exists $C > 0$ such that for $|g|, |\vec{p}| < C$, there exist two functions $f_{g, \vec{p}}$ and $\vec{r}_{g, \vec{p}}$ which are smooth in $(g, \vec{p})$ such that the minimum of the energy $\hat{\mathcal{E}}_{g, \vec{p}}(f, \vec{r})$ is attained at $(f_{g, \vec{p}}, \vec{r}_{g, \vec{p}})$. These functions satisfy
\[
f_{g, \vec{p}} = \frac{\vec{p} \cdot \vec{G}}{\frac{1}{2}||\vec{k}||^2 + ||\vec{k}||} + \mathcal{O}((g, \vec{p})^3)
\]
\[
\vec{r}_{g, \vec{p}} = -S^{-1}\vec{G}^{\vee 2} + \mathcal{O}((g, \vec{p})^3),
\]
with $S = \vec{k} \otimes \vec{k} + 2(\frac{1}{2}||\vec{k}||^2 + ||\vec{k}||) \vee 1_{\mathbb{h}}$ and where $\vec{G}^{\vee 2} = \sum_{j=1}^3 \vec{G}_j \vee \vec{G}_j \in \mathbb{h} \vee \mathbb{h}$ (recall that $\vee$ denotes the symmetric tensor product). As a consequence
\[
\mathcal{E}_{\text{BHF}}(g, \vec{p}, \sigma, \Lambda) = \hat{\mathcal{E}}_{g, \vec{p}}(0_X) - \vec{p} \cdot \vec{G}^* \frac{1}{\frac{1}{2}||\vec{k}||^2 + ||\vec{k}||} \vec{G} \cdot \vec{p} - \frac{1}{2} \vec{G}^{\vee 2} S^{-1} \vec{G}^{\vee 2} + \mathcal{O}((||g, \vec{p}||)^3).
\]
Recall that for \( z \) and \( z' \) in a Hilbert space, \( z^* z' = \langle z, z' \rangle \).

**Remark VIII.7.** The energy in \( 0_X \) is the energy of the vacuum state and is 
\[
\hat{E}_{g,\vec{p}}(0_X) = \frac{1}{2} |\vec{p}|^2 + \frac{1}{2} \| \vec{G} \|^2.
\]
Further note that by Lemma VII.3
\[
2(\vec{p} \cdot \vec{G}^*) \frac{1}{|k|^2 + |\vec{k}|} (\vec{G} \cdot \vec{p}) = g^2 |\vec{p}|^2 \frac{32\pi}{3} \ln \left( \frac{\Lambda + 2\sigma + 2}{\sigma + 2} \right)
\]
and in particular does not depend on the choice of the polarization vectors \( \vec{\varepsilon} \).

The quantity \( \vec{G} \cdot \vec{G}^* S^{-1} \vec{G} \cdot \vec{G}^* \) does not depend on the choice of the vectors \( \vec{\varepsilon} \) either since
\[
\sum_{\mu, \nu = \pm} |\vec{\varepsilon}(k_1, \mu) \cdot \vec{\varepsilon}(k_2, \nu)|^2 = \sum_{\mu, \nu = \pm} \text{Tr}_{\mathbb{R}^3}[P_{\vec{k}_1} P_{\vec{k}_2}]
= 1 + \text{Tr}_{\mathbb{R}^3}[P_{\vec{k}_1}^\perp P_{\vec{k}_2}^\perp]
= 1 + \left( \frac{\vec{k}_1}{|\vec{k}_1|} \cdot \frac{\vec{k}_2}{|\vec{k}_2|} \right)^2.
\]

**Proof of Theorem VIII.6.** Let
\[
F : (g, \vec{p}, f, \hat{r}) \mapsto \partial_{f, \hat{r}} \hat{E}_{g,\vec{p}}(f, \hat{r})
\]
and \( \left(f, f' \right) \) \( (g, \vec{p}) := \left( \frac{f}{\hat{r}(g, \vec{p})} \right) \) such that
\[
F(g, \vec{p}, \left( f, f' \right) (g, \vec{p})) = 0,
\] (VIII.67)
then a derivation of Equation (VIII.67) with respect to \( (f, \hat{r}) \) brings
\[
\partial_{g,\vec{p}} \left( f, f' \right) (0_g, \vec{p}) = - \left[ \partial_{f, \hat{r}} F(0_g, \vec{p}, 0_f, \hat{r}) \right]^{-1} \partial_{g,\vec{p}} F(0_g, \vec{p}, 0_f, \hat{r})
\]
The term which is independent of \( (g, \vec{p}) \) and quadratic in \( (f, f') \) in the energy is
\[
\frac{1}{2} \left\{ \text{Tr}[\hat{r}S\hat{r}] + f^* (|\vec{k}|^2 + 2|\vec{k}|) f \right\}
\]
thus, in \( (0_g, \vec{p}, 0_f, \hat{r}) \),
\[
\partial_{f, \hat{r}} F = \left( \begin{array}{cc}
|\vec{k}|^2 + 2|\vec{k}| & 0 \\
0 & S
\end{array} \right).
\]
To compute \( \partial_{g,\vec{p}} F \) in 0, observe that no part in the energy is linear in \( (g, \vec{p}) \) and linear in \( (f, \hat{r}) \). Thus \( \partial_{g,\vec{p}} F(0_g, \vec{p}, 0_f, \hat{r}) = 0 \) and we get
\[
\partial_{g,\vec{p}} F(0_g, \vec{p}) = 0.
\]
Minimization of the Energy of the Nonrelativistic . . . 1513

Differentiating a second time Equation (VIII.67) brings
\[0 = \partial_{g,\tilde{p}}^2 F + 2\partial_{f,\tilde{r}}^2 \partial_{g,\tilde{p}} F \circ \partial_{g,\tilde{p}} (f) + \partial_{f,\tilde{r}} F \circ \partial_{g,\tilde{p}}^2 (f) + \partial_{f,\tilde{r}} F (\partial_{g,\tilde{p}} (f), \partial_{g,\tilde{p}} (f)).\]

Since \(\partial_{g,\tilde{p}} (f) (0_{g,\tilde{p}}) = 0\), it follows that
\[\partial_{g,\tilde{p}}^2 (f) (0_{g,\tilde{p}}) = -[\partial_{f,\tilde{r}} F (0_{g,\tilde{p}}, 0_{f,\tilde{r}})]^{-1} \partial_{g,\tilde{p}} F (0_{g,\tilde{p}}, 0_{f,\tilde{r}}).\]

The part of the energy which is quadratic in \((g, \tilde{p})\) and linear in \((f, \tilde{r})\) is
\[-2\text{Re}(\tilde{g}^* \tilde{p}) \cdot \tilde{p} + \text{Re}(\tilde{r} \tilde{G}; \tilde{G})\], it follows that, in \((0_{g,\tilde{p}}, 0_{f,\tilde{r}}),\)
\[\partial_{g,\tilde{p}}^2 F = 2 \begin{pmatrix} 1 & 0 \\ -\partial_{g} \tilde{G} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & \partial_{g} \tilde{G} \\ 0 & 0 \end{pmatrix} + \text{Tr}[\tilde{G}]^2 f\]
which gives in \(0_{g,\tilde{p}}\)
\[\partial_{g,\tilde{p}}^2 (f) = 2 \begin{pmatrix} 1 & 0 \\ -S^{-1} \partial_{g} \tilde{G} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & \partial_{g} \tilde{G} \\ 0 & 0 \end{pmatrix} \]
Hence the expansion of \((f)\) up to order 2.

We can thus express the energy around \(0_{g,\tilde{p}}\) modulo error terms in \(O(||(g, \tilde{p})||^5)\)
\[
\min_{f,\tilde{r}} \tilde{E}_{g,\tilde{p}}(f, \tilde{r}) - \tilde{E}_{g,\tilde{p}}(0,0)
\equiv \frac{1}{2} \left\{ (\text{Tr}[\tilde{r}^2 \tilde{k}^2] + f^* \tilde{k} f + 2\text{Re}(f^* \tilde{G}) - \tilde{p} \tilde{k})^2 + \text{Tr}[\tilde{r} \tilde{k} \cdot \tilde{k}] + \text{Tr}[\tilde{k}]^2 \tilde{r}^2 \right\}
\[+ 2\text{Re}(\tilde{r} \tilde{G} + \tilde{k} f) (f + \tilde{G}) + ||\tilde{G}||^2 + f^* \tilde{k}^2 f \]
\[+ \text{Tr}[\tilde{k}]^2 \tilde{f}] + f^* \tilde{k} f - \tilde{E}_{g,\tilde{p}}(0,0)\]
\[= -2\text{Re}(f^* \tilde{G}) \cdot \tilde{p} + \frac{1}{2} \text{Tr}[\tilde{r} \tilde{S} \tilde{r}] + \text{Re}(\tilde{r} \tilde{G}; \tilde{G}) f + f^* (\frac{1}{2} ||\tilde{k}||^2 + ||\tilde{G}||) f\]
\[= -2\text{Re}(f^* \tilde{G}) \cdot \tilde{p} + \frac{1}{2} \text{Tr}[\tilde{r} \tilde{S} \tilde{r}] + \text{Re}(\tilde{r} \tilde{G}; \tilde{G}) + f^* (\frac{1}{2} ||\tilde{k}||^2 + ||\tilde{G}||) f\]
\[= -2 (\tilde{p} \cdot \tilde{G}) (\tilde{p} \cdot \tilde{G}) + \frac{1}{2} \tilde{G} \cdot \tilde{G} S^{-1} \tilde{G} \cdot \tilde{G} - \tilde{G} \cdot \tilde{G} S^{-1} \tilde{G} \cdot \tilde{G} + \frac{1}{4} \tilde{G} \cdot \tilde{G} \cdot (\tilde{p} \cdot \tilde{G})^2 (\tilde{p} \cdot \tilde{G})\]
which completes the proof.

VIII.3 Lagrange Equations

In this section we derive a system of equations that determine critical points of the energy functional. We formulate the results of Section VIII.1 in terms of \(\gamma\) and \(\alpha\) subject to the constraints \(\gamma + \gamma^2 = (\alpha^* \otimes 1_b) (1_b \otimes \alpha)\), without reference to the parametrization of \(\gamma\) and \(\alpha\) in terms of \(\tilde{r}\).

Suppose \(f \in \mathfrak{g}, \alpha \in \mathfrak{h}^{\mathbb{R}}, \gamma \in \mathcal{L}^1(h), \lambda \in \mathcal{B}(h) = B\) and \(\tilde{u} \in \mathbb{R}^3\). Let \(A(\lambda) = \frac{1}{2} \tilde{k} \cdot \nabla \tilde{k} + \lambda \tilde{v} + 1\) and \(G(\gamma) = \gamma + \gamma^2\).
Theorem VIII.8. Suppose \((f, \gamma, \alpha)\) is a minimizer of the energy functional \(E\) such that \(\|\gamma\|_{E(\mathbf{R})} < \frac{1}{2}\). Then there is a unique \((\lambda, \bar{u})\) such that \((f, \gamma, \alpha, \lambda, \bar{u})\) satisfies the following equations, equivalent to Lagrange equations

\[
M(\gamma, \bar{u})f = -(\vec{k} - \bar{u}) \cdot \vec{G} - \vec{k} \cdot \nabla (\vec{G} + \vec{k}f)^* \alpha \tag{VIII.68}
\]

\[
\mathcal{A}(\lambda)\alpha = -\frac{1}{2}(\vec{G} + \vec{k}f)^{\alpha^2} \tag{VIII.69}
\]

\[
\gamma = \mathcal{G}^{-1}((\alpha^* \otimes 1_b)(1_b \otimes \alpha)) \tag{VIII.70}
\]

\[
\lambda = \int_0^\infty e^{-t(\frac{1}{2} + \gamma)} (M(\gamma, \bar{u}) + (\vec{G} + \vec{k}f) \cdot (\vec{G} + \vec{k}f)^*) e^{-t(\frac{1}{2} + \gamma)} dt \tag{VIII.71}
\]

\[
\bar{u} = \bar{p} - \text{Tr}[\gamma \vec{k}] - f^*\vec{k}f - 2\text{Re}(f^*\vec{G}) \tag{VIII.72}
\]

with \(M(\gamma, \bar{u}) = \frac{1}{2}|\vec{k}|^2 + |\vec{k}| - \vec{k} \cdot \bar{u} + \vec{k} \cdot \gamma \vec{k}\).

Assuming \(|\bar{p}| < \frac{1}{2}\), sufficient conditions such that \(M(\gamma, \bar{u})\) and \(\mathcal{A}(\lambda)\) are invertible operators are \(|\bar{u}| < 1/2\), \(\gamma \geq 0\) and \(\|\lambda - (|\vec{k}|^2/2 + |\vec{k}| - \bar{p} \cdot \vec{k})\|_{\mathbf{R}} < \sigma/2\). Equations (VIII.68) to (VIII.72) then form a system of coupled explicit equations.

Remark VIII.9. To prove that Equations (VIII.68) to (VIII.72) admit a solution we use here the result of existence of a minimizer proved in Section VIII.1. It can also be proved directly by a fixed point argument by defining the applications

\[
\Psi_f(f, \alpha, \gamma, \bar{u}) = -M(\gamma, \bar{u})^{-1}(\vec{k} - \bar{u}) \cdot \vec{G} - \vec{k} \cdot \nabla (\vec{G} + \vec{k}f)^* \alpha
\]

\[
\Psi_\alpha(f, \lambda) = -\mathcal{A}(\lambda)^{-1/2}(\vec{G} + \vec{k}f)^{\alpha^2}
\]

\[
\Psi_\gamma(\alpha) = \mathcal{G}^{-1}((\alpha^* \otimes 1_b)(1_b \otimes \alpha))
\]

\[
\Psi_\lambda(f, \gamma, \bar{u}) = \int_0^\infty e^{-t(\frac{1}{2} + \gamma)} (M(\gamma, \bar{u}) + (\vec{G} + \vec{k}f) \cdot (\vec{G} + \vec{k}f)^*) e^{-t(\frac{1}{2} + \gamma)} dt
\]

\[
\Psi_\bar{u}(f, \gamma) = \bar{p} - \text{Tr}[\gamma \vec{k}] - f^*\vec{k}f - 2\text{Re}(f^*\vec{G})
\]

defined on balls of centers 0, 0, 0, \(\frac{1}{2}|\vec{k}|^2 + |\vec{k}| - \vec{k} \cdot \bar{u}\) and \(\bar{p}\) and proving that the application

\[
\Psi_{(f, \lambda)}(f, \lambda) = (\Psi_f(f, \Psi_\alpha(f, \lambda)), \Psi_\alpha(f, \lambda), \Psi_\lambda(f, \Psi_\alpha(f, \lambda)), \Psi_\bar{u}(f, \Psi_\gamma(\Psi_\alpha(f, \lambda))))
\]

is a contraction for a convenient choice of the radiuses and a sufficiently small coupling constant \(g\). Note that it is then convenient to consider the norm of \(L^2(S_{\gamma, \Lambda} \times Z_2, |\vec{k}|^2)\) for \(f\).

Proof of Theorem VIII.8. Indeed, set \(\bar{u} = \bar{p} - \text{Tr}[\gamma \vec{k}] - f^*\vec{k}f - 2\text{Re}(f^*\vec{G})\) and define the partial derivatives as \(\partial_f E(f, \gamma, \alpha) \in \mathbf{h}, \partial_\alpha E(f, \gamma, \alpha) \in \mathbf{h}^{\gamma^2}\) and \(\partial_\gamma E(f, \gamma, \alpha) \in \mathbf{h}\).
Recall the energy functional is given by Equation (V.60) and this yields
\[ E(f + \delta f , \gamma + \delta \gamma, \alpha + \delta \alpha) - E(f, \gamma, \alpha) = 2\text{Re}(\delta f^* \partial_f E(f, \gamma, \alpha)) + 2\text{Re}(\delta \alpha^* \partial_{\alpha^*} E(f, \gamma, \alpha)) + \text{Tr}[\delta \gamma \partial_{\gamma} E(f, \gamma, \alpha)] + \alpha(\|\delta f, \delta \gamma, \delta \alpha\|_{\mathfrak{h} \times L^1(\mathfrak{h}) \times \mathfrak{h}^{\vee \mathfrak{h}}}). \]

Recall the energy functional is given by Equation (V.60) and this yields
\[ \partial_f E(f, \gamma, \alpha) = \frac{1}{2} \{ (2\tilde{k}f + \tilde{G}) \cdot (\text{Tr}[\gamma \tilde{k}] + f^* \tilde{k}f + 2\text{Re}(f^* \tilde{G}) - \bar{p}) 
+ 2\tilde{k} \cdot \nabla(\tilde{G} + \tilde{k}f) \cdot \alpha + \tilde{k} \cdot (2\gamma + 1)(\tilde{G} + \tilde{k}f) \} + |\tilde{k}|f \n= -(\tilde{k}f + \tilde{G}) \cdot \bar{u} + \tilde{k} \cdot \nabla(\tilde{G} + \tilde{k}f) \cdot \alpha + \tilde{k} \cdot \gamma(\tilde{G} + \tilde{k}f) + |\tilde{k}|f \n= M(\gamma, \bar{u}) f + (\tilde{k} - \bar{u}) \cdot \tilde{G} + \tilde{k} \cdot \nabla(\tilde{G} + \tilde{k}f) \cdot \alpha, \]
\[ \partial_{\alpha^*} E(f, \gamma, \alpha) = \frac{1}{2} (\tilde{G} \cdot \bar{\alpha}) \alpha + 1\frac{1}{2} (\tilde{G} \cdot \tilde{k}f) \gamma^2, \]
\[ \partial_{\gamma} E(f, \gamma, \alpha) = \frac{1}{2} \{ (2\tilde{\gamma}f + \tilde{G}) \cdot (f^* \tilde{k}f + 2\text{Re}(f^* \tilde{G}) - \tilde{p}) 
+ 2\tilde{\gamma} \cdot \gamma \tilde{k} + |\tilde{k}|^2 + 2(\tilde{G} + \tilde{k}f) \cdot (\tilde{G} + \tilde{k}f)^* \} + |\tilde{k}| \n= M(\gamma, \bar{u}) + (\tilde{G} + \tilde{k}f) \cdot (\tilde{G} + \tilde{k}f)^*. \]

The constraint given by Equation (IV.58) can be expressed as
\[ \mathcal{C}(f, \gamma, \alpha) = 0 \quad \text{(VIII.73)} \]
with
\[ \mathcal{C} : \mathfrak{h} \times L^1(\mathfrak{h}) \times \mathfrak{h}^{\vee \mathfrak{h}} \rightarrow L^1(\mathfrak{h}) \]
\[ (f, \gamma, \alpha) \mapsto \gamma + \gamma^2 - (\alpha^* \otimes 1_\mathfrak{h})(1_\mathfrak{h} \otimes \alpha). \]

Equation (VIII.73) is equivalent to Equation (VIII.70). The application \( \mathcal{C} \) has a differential \( D\mathcal{C}(f, \gamma, \alpha) : \mathfrak{h} \times L^1(\mathfrak{h}) \times \mathfrak{h}^{\vee \mathfrak{h}} \rightarrow L^1(\mathfrak{h}) \) such that
\[ D\mathcal{C}(f, \gamma, \alpha)(\delta f, \delta \gamma, \delta \alpha) \n= \delta \gamma + \delta \gamma \gamma + \delta \gamma - (\delta \alpha^* \otimes 1_\mathfrak{h})(1_\mathfrak{h} \otimes \alpha) - (\alpha^* \otimes 1_\mathfrak{h})(1_\mathfrak{h} \otimes \delta \alpha). \]

For \( \|\gamma\|_{\mathcal{B}(\mathfrak{h})} < \frac{1}{4} \) the application \( D\mathcal{C}(f, \gamma, \alpha) \) is surjective. Indeed it is already surjective on \( \{0\} \times L^1(\mathfrak{h}) \times \{0\} \), since, for every \( \gamma' \in L^1(\mathfrak{h}) \) the equation \( \delta \gamma + \delta \gamma \gamma + \delta \gamma = \gamma' \) with unknown \( \delta \gamma \) has at least one solution, see Proposition VIII.10. We can then apply the Lagrange multiplier rule (see for example the book of Zeidler [14]) which tells us that there exists a \( \lambda \in \mathcal{B}(\mathfrak{h}) \) such that
\[ \forall (\delta f, \delta \alpha, \delta \gamma), \quad D\mathcal{E}(f, \alpha, \gamma)(\delta f, \delta \alpha, \delta \gamma) + \text{Tr}[D\mathcal{C}(f, \alpha, \gamma)(\delta f, \delta \alpha, \delta \gamma) \lambda] = 0, \]
that is to say
\[ 2\text{Re}(\delta f^* \partial_f \mathcal{E}(f, \gamma, \alpha) + \delta \alpha^* \partial_{\alpha^*} \mathcal{E}(f, \gamma, \alpha)) + \text{Tr}[\partial_{\gamma} \mathcal{E}(f, \gamma, \alpha) \delta \gamma] \n+ \text{Tr}[(\delta \gamma + \delta \gamma \gamma + \delta \gamma - (\delta \alpha^* \otimes 1_\mathfrak{h})(1_\mathfrak{h} \otimes \alpha) - (\alpha^* \otimes 1_\mathfrak{h})(1_\mathfrak{h} \otimes \delta \alpha)) \lambda] = 0. \]
This is equivalent to Equations (VIII.68), (VIII.69) and
\[ \lambda\left(\frac{1}{2} + \gamma\right) + \left(\frac{1}{2} + \gamma\right)\lambda = M(\gamma, \vec{u}) + (\vec{G} + \vec{k}f) \cdot (\vec{G} + \vec{k}f)^* \quad (VIII.74) \]
Using again Proposition VIII.10 we get that Equation (VIII.74) is equivalent to Equation (VIII.71).

For the invertibility of \( A(\lambda) \) note that
\[ A(\lambda) = \frac{1}{4}(\vec{k} \otimes 1 + 1 \otimes \vec{k})^2 + (|\vec{k}| - \vec{k} \cdot \vec{p} + \lambda - \frac{1}{2}|\vec{k}|^2 - |\vec{k}| + \vec{k} \cdot \vec{p}) \vee 1 \]
\[ \geq \left(\frac{\sigma}{2} - \lambda - (|\vec{k}|^2/2 + |\vec{k}| - \vec{p} \cdot \vec{k})\|B\|_1 \right) 1 \vee 1 . \]
For \( M(\gamma, \vec{u}) \), \( M(\gamma, \vec{u}) = \frac{1}{2}|\vec{k}|^2 + |\vec{k}| - \vec{k} \cdot \vec{u} + \vec{k} \cdot \vec{u} \geq \sigma/2 \) if \( \gamma \geq 0 \) and \( |\vec{u}| < 1/2. \)

Let us recall a well known expression for the solution of the Sylvester or Lyapunov equation.

**Proposition VIII.10.** Let \( A \) and \( B \) be bounded self-adjoint operators on a Hilbert space. Suppose \( A \geq a 1 \) with \( a > 0. \) Then the equation
\[ AX +XA = B \]
for \( X \) a bounded operator has a unique solution \( \chi_A(B) = \int_0^\infty e^{-tA}Be^{-tA}dt. \)
If \( B \) a trace class operator then the solution \( X \) is also trace class.

**Proof.** Indeed, \( \chi_A(B) \) is a solution because
\[ A\chi_A(B) + \chi_A(B)A = \int_0^\infty e^{-tA}(AB + BA)e^{-tA}dt \]
\[ = -\int_0^\infty \frac{d}{dt}(e^{-tA}Be^{-tA})dt = B . \]
Conversely, suppose that \( AX +XA = B, \) then
\[ \chi_A(B) = \int_0^\infty e^{-tA}(AX +XA)e^{-tA}dt \]
\[ = -\int_0^\infty \frac{d}{dt}(e^{-tA}Xe^{-tA})dt = X , \]
and thus any solution \( X \) is equal to \( \chi_A(B) \). Hence the solution is unique. \( \square \)

## A Equivalent Characterizations of Centered Quasifree Density Matrices

In this appendix we give various equivalent characterizations of quasifree states. In particular we remark that (ii) in Lemma A.1 below corresponds to the definition of quasifree states in terms of Wick’s Theorem.

**Documenta Mathematica 18 (2013) 1481–1519**
Lemma A.1. Let $\rho \in c\mathcal{Q}\mathcal{M}$ be a centered density matrix and denote $\langle A \rangle_\rho := \text{Tr}_\mathcal{F}[\rho A]$. Then (i) $\iff$ (ii) $\iff$ (iii), where

(i) $\rho \in c\mathcal{Q}\mathcal{F}$ is centered and quasifree;

(ii) All odd correlation functions and all even truncated correlation functions of $\rho$ vanish, i.e., for all $N \in \mathbb{N}$ and $\varphi_1, \ldots, \varphi_{2N} \in \mathfrak{h}$, let either $b_n := a^*(\varphi_n)$ or $b_n := a(\varphi_n)$, for all $1 \leq n \leq 2N$. Then $\langle b_1 \cdots b_{2N-1} \rangle_\rho = 0$ and

$$\langle b_1 b_2 \cdots b_{2N} \rangle_\rho = \sum_{\pi \in \mathfrak{P}_{2N}} \langle b_{\pi(1)} b_{\pi(2)} \rangle_\rho \cdots \langle b_{\pi(2N-1)} b_{\pi(2N)} \rangle_\rho,$$

(A.75)

where $\mathfrak{P}_{2N}$ denotes the set of pairings, i.e., the set of all permutations $\pi \in \mathfrak{S}_{2N}$ of $2N$ elements such that $\pi(2n-1) < \pi(2n + 1)$ and $\pi(2n-1) < \pi(2n)$, for all $1 \leq n \leq N - 1$ and $1 \leq n \leq N$, respectively.

(iii) There exist two commuting quadratic, semibounded Hamiltonians

$$H = \sum_{i,j} \left\{ B_{i,j} a^*(\psi_i) a(\psi_j) + C_{i,j} a^*(\psi_i) a^*(\psi_j) + \overline{C}_{i,j} a(\psi_i) a(\psi_j) \right\},$$

(A.76)

$$H' = \sum_{i,j} \left\{ B'_{i,j} a^*(\psi_i) a(\psi_j) + C'_{i,j} a^*(\psi_i) a^*(\psi_j) + \overline{C'_{i,j}} a(\psi_i) a(\psi_j) \right\},$$

(A.77)

with $B = B^* \geq 0$, $C = C^T \in L^2(\mathfrak{h})$, where $\{\psi_i\}_{i \in \mathbb{N}} \subseteq \mathfrak{h}$ is an orthonormal basis, such that $\exp(-H - \beta H')$ is trace class, for all $\beta < \infty$, and

$$\langle A \rangle_\rho = \lim_{\beta \to \infty} \left\{ \frac{\text{Tr}_\mathcal{F}[A \exp(-H - \beta H')]}{\text{Tr}_\mathcal{F}[\exp(-H - \beta H')]} \right\},$$

(A.78)

for all $A \in \mathcal{B}(\mathfrak{H})$.

Eq. (II.36) and the vanishing (ii) of the truncated correlation functions of a centered quasifree state imply that any quasifree state $\rho \in \mathcal{Q}\mathcal{F}$ is completely determined by its one-point function $\langle a(\varphi) \rangle_\rho$ and its two-point function (one-particle reduced density matrix).

Acknowledgements

The third author was supported for part of this work by DFG Grant No. Ba-1477/5-1 and also by the European Research Council under the European Community’s Seventh Framework Program (FP7/2007-2013)/ERC grant agreement 20285.
REFERENCES


Volker Bach  
Technische Universität Braunschweig  
Institut für Analysis und Algebra  
Rebenring 31, A 14  
38106 Braunschweig  
Germany  
v.bach@tu-bs.de

Sébastien Breteaux  
Technische Universität Braunschweig  
Institut für Analysis und Algebra  
Rebenring 31, A 14  
38106 Braunschweig  
Germany  
sebastien.breteaux@ens-cachan.org

Tim Tzaneteas  
Technische Universität Braunschweig  
Institut für Analysis und Algebra  
Rebenring 31, A 14  
38106 Braunschweig  
Germany  
timmy.tzaneteas@uni-tuebingen.de
Abstract. Let \( f : X \to B \) be a projective surjective morphism between quasi-projective varieties. The goal of this paper is the study of the Chow groups of \( X \) in terms of the Chow groups of \( B \) and of the fibres of \( f \). One of the applications concerns quadric bundles. When \( X \) and \( B \) are smooth projective and when \( f \) is a flat quadric fibration, we show that the Chow motive of \( X \) is “built” from the motives of varieties of dimension less than the dimension of \( B \).

2010 Mathematics Subject Classification: 14C15, 14C25, 14C05, 14D99
Keywords and Phrases: Algebraic cycles, Chow groups, quadric bundles, motives, Chow–Künneth decomposition.

Contents

1. Surjective morphisms and motives 1526
2. On the Chow groups of the fibres 1530
3. A generalisation of the projective bundle formula 1532
4. On the motive of quadric bundles 1534
5. On the motive of a smooth blow-up 1536
6. Chow groups of varieties fibred by varieties with small Chow groups 1540
7. Applications 1547
References 1551
For a scheme $X$ over a field $k$, $\text{CH}_i(X)$ denotes the rational Chow group of $i$-dimensional cycles on $X$ modulo rational equivalence. Throughout, $f : X \to B$ will be a projective surjective morphism defined over $k$ from a quasi-projective variety $X$ of dimension $d_X$ to an irreducible quasi-projective variety $B$ of dimension $d_B$, with various extra assumptions which will be explicitly stated. Let $h$ be the class of a hyperplane section in the Picard group of $X$. Intersecting with $h$ induces an action $\text{CH}_i(X) \to \text{CH}_{i-1}(X)$ still denoted $h$. Our first observation is Proposition 1.6: when $B$ is smooth, the map

$$
\bigoplus_{i=0}^{d_X-d_B} h^{d_X-d_B-i} \circ f^* : \bigoplus_{i=0}^{d_X-d_B} \text{CH}_{i-1}(B) \to \text{CH}_i(X)
$$

is injective for all $l$ and a left-inverse can be expressed as a combination of the proper pushforward $f_*$, the refined pullback $f^*$ and intersection with $h$. It is then not too surprising that, when both $B$ and $X$ are smooth projective, the morphism of Chow motives

$$
\bigoplus_{i=0}^{d_X-d_B} h^{d_X-d_B-i} \circ f^* : \bigoplus_{i=0}^{d_X-d_B} \text{CH}_{i-1}(B) \to \text{CH}_i(X)
$$

is split injective; see Theorem 1.4. By taking a cohomological realisation, for instance by taking Betti cohomology if $k \subseteq \mathbb{C}$, we thus obtain that the map

$$
\bigoplus_{i=0}^{d_X-d_B} h^{d_X-d_B-i} \circ f^* : \bigoplus_{i=0}^{d_X-d_B} H^{n-2i}(B, \mathbb{Q}) \to H^n(X, \mathbb{Q})
$$

is split injective for all $n$ and thus realises the left-hand side group as a sub-Hodge structure of the right-hand side group. This observation can be considered as a natural generalisation of the elementary fact that a smooth projective variety $X$ of Picard rank 1 does not admit a non-constant dominant map to a smooth projective variety of smaller dimension.

Let now $\Omega$ be a universal domain containing $k$, that is an algebraically closed field containing $k$ which has infinite transcendence degree over its prime subfield. Let us assume that there is an integer $n$ such that the fibres $X_b$ of $f$ over $\Omega$-points $b$ of $B$ satisfy $\text{CH}_i(X_b) = \mathbb{Q}$ for all $l < n$. If $f$ is flat, then Theorem 3.2 shows that (1) is surjective for all $l < n$. When $X$ and $B$ are both smooth projective, we deduce in Theorem 4.2 a direct sum decomposition of the Chow motive of $X$ as

$$
\text{h}(X) \cong \bigoplus_{i=0}^{d_X-d_B} \text{h}(B)(i) \oplus M(n),
$$

where $M$ is isomorphic to a direct summand of the motive of some smooth projective variety $Z$ of dimension $d_X - 2n$. This notably applies when $X$ is a projective bundle over a smooth projective variety $B$ to give the well-known isomorphism $\bigoplus_{i=0}^{d_X-d_B} \text{h}(B)(i) \cong \text{h}(X)$. Such a morphism is usually shown to be an isomorphism by an existence principle, namely Manin’s identity principle. Here, we actually exhibit an explicit inverse to that isomorphism. The
same arguments are used in Theorem 5.3 to provide an explicit inverse to the smooth blow-up formula for Chow groups. More interesting is the case when a smooth projective variety \( X \) is fibred by complete intersections of low degree. For instance, the decomposition (2) makes it possible in Corollary 4.4 to construct a Murre decomposition (see Definition 4.3) for smooth projective varieties fibred by quadrics over a surface, thereby generalising a result of del Angel and Müller–Stach [5] where a Murre decomposition was constructed for 3-folds fibred by conics over a surface, and also generalising a previous result [27] where, in particular, a Murre decomposition was constructed for 4-folds fibred by quadric surfaces over a surface. Another consequence of the decomposition (2) is that rational and numerical equivalence agree on smooth projective varieties \( X \) fibred by quadrics over a curve or a surface defined over a finite field; see Corollary 4.8. It should be mentioned that our approach bypasses the technique of Gordon–Hanamura–Murre [8], where Chow–Künneth decompositions are constructed from relative Chow–Künneth decompositions. In our case, we do not require the existence of a relative Chow–Künneth decomposition, nor do we require \( f \) to be smooth away from finitely many points as is the case in [8]. Finally, it should be noted, for instance if \( X \) is complex smooth projective fibred by quadrics, that (2) actually computes some of the Hodge numbers of \( X \) without going through a detailed analysis of the Leray–Serre spectral sequence.

More generally, we are interested in computing, in some sense, the Chow groups of \( X \) in terms of the Chow groups of \( B \) and of the fibres of \( f \). Let us first clarify what is meant by “fibres”. We observed in [27, Theorem 1.3] that if \( B \) is smooth and if a general fibre of \( f \) has trivial Chow group of zero-cycles (i.e. if it is spanned by the class of a point), then \( f_* : CH_0(X) \rightarrow CH_0(B) \) is an isomorphism with inverse a rational multiple of \( h^{d_X-d_B} o f^* \). We thus see that, as far as zero-cycles on the fibres are concerned, it is enough to consider only the general fibre. For that matter, we show in Proposition 2.4 that, provided the ground field is a universal domain, it is actually enough that a very general fibre have trivial Chow group of zero-cycles. However, if one is willing to deal with positive-dimensional cycles, it is no longer possible to ignore the Chow groups of some of the fibres. For instance, if \( \tilde{X}_Y \rightarrow X \) is a smooth blow-up along a smooth center \( Y \subseteq X \), then \( CH_1(\tilde{X}_Y) \) is isomorphic to \( CH_1(X) \oplus CH_0(Y) \), although a general fibre of \( \tilde{X}_Y \rightarrow X \) is reduced to a point and hence has trivial \( CH_1 \). One may argue that a smooth blow-up is not flat. Let us however consider as in [1] a complex flat conic fibration \( f : X \rightarrow \mathbb{P}^2 \), where \( X \) is smooth projective. All fibres \( F \) of \( f \) satisfy \( CH_0(F) = \mathbb{Q} \). A smooth fibre \( F \) of \( f \) is isomorphic to \( \mathbb{P}^1 \) and hence satisfies \( CH_1(F) = \mathbb{Q} \). A singular fibre \( F \) of \( f \) is either a double line, or the union of two lines meeting at a point. In the latter case \( CH_1(F) = \mathbb{Q} \oplus \mathbb{Q} \). This reflects in \( CH_1(X) \) and, as shown in [1], \( CH_1(X)_{\text{hom}} \) is isomorphic to the Prym variety attached to the discriminant curve of \( f \). This suggests that a careful analysis of the degenerations of \( f \) is required in order to derive some precise information on the Chow groups of \( X \). On another perspective, the
following examples show what kind of limitation is to be expected when dealing with fibres with non-trivial Chow groups. For instance, consider a complex Enriques surface \( S \) and let \( T \to S \) be the 2-covering by a \( K3 \)-surface \( T \). We know that \( \text{CH}_0(S) = \mathbb{Q} \), and the fibres \( T_s \) are disjoint union of two points and hence satisfy \( \text{CH}_0(T_s) = \mathbb{Q} \oplus \mathbb{Q} \). However, we cannot expect \( \text{CH}_0(T) \) to be correlated in some way to the Chow groups of \( S \) and of the fibres, because a theorem of Mumford \cite{17} says that \( \text{CH}_0(T) \) is infinite-dimensional in a precise sense. Another example is given by taking a pencil of high-degree hypersurfaces in \( \mathbb{P}^n \). Assume that the base locus \( Z \) is smooth. By blowing up \( Z \), we get a morphism \( \mathbb{P}_2 \to \mathbb{P}^1 \). In that case, \( \text{CH}_0(\mathbb{P}^1) = \mathbb{Q} \) and the \( \text{CH}_0 \) of the fibres is infinite-dimensional, but \( \text{CH}_0(\mathbb{P}_2) = \mathbb{Q} \).

Going back to the case where the general fibre of \( f : X \to B \) has trivial Chow group of zero-cycles, we see that \( \text{CH}_0(X) \) is supported on a linear section of dimension \( d_B \). We say that \( \text{CH}_0(X) \) has niveau \( d_B \). More generally, Laterveer \cite{15} defines a notion of niveau on Chow groups as follows. For a variety \( X \), the group \( \text{CH}_i(X) \) is said to have niveau \( \leq n \) if there exists a closed subscheme \( Z \) of \( X \) of dimension \( \leq i + n \) such that \( \text{CH}_i(Z) \to \text{CH}_i(X) \) is surjective, in other words if the \( i \)-cycles on \( X \) are supported in dimension \( i + n \). It can be proved \cite{27} Theorem 1.7 that if a general fibre \( F \) of \( f : X \to B \) is such that \( \text{CH}_0(F) \) has niveau \( \leq 1 \), then \( \text{CH}_0(X) \) has niveau \( \leq d_B + 1 \). In that context, a somewhat more precise question is: what can be said about the niveau of the Chow groups of \( X \) in terms of the niveau of the Chow groups of the fibres of \( f : X \to S \)? A statement one would hope for is the following: if the fibres \( X_b \) of \( f : X \to B \) are such that \( \text{CH}_i(X_b) \) has niveau \( \leq n \) for all \( \Omega \)-points \( b \in B \), then \( \text{CH}_i(X) \) has niveau \( \leq n + d_B \). We cannot prove such a general statement but we prove it when some of the Chow groups of the fibres of \( f \) are either spanned by linear sections or have niveau 0, i.e. when they are finite-dimensional \( \mathbb{Q} \)-vector spaces. Precisely, if \( f : X \to B \) is a complex projective surjective morphism onto a smooth quasi-projective variety \( B \), we show that \( \text{CH}_i(X) \) has niveau \( \leq d_B \) in the following cases:

- \( \text{CH}_i(X_b) = \mathbb{Q} \) for all \( i \leq l \) and all \( b \in B(\mathbb{C}) \) (Theorem 6.10);
- \( d_B = 1 \) and \( \text{CH}_i(X_b) \) is finitely generated for all \( i \leq l \) and all \( b \in B(\mathbb{C}) \) (Theorem 6.12);
- \( f \) is smooth away from finitely many points, \( \text{CH}_i(X_b) = \mathbb{Q} \) for all \( i < l \) and \( \text{CH}_i(X_b) \) is finitely generated, for all \( b \in B(\mathbb{C}) \) (Theorem 6.13).

These results, which are presented in Section 6, complement the generalisation of the projective bundle formula of Theorem 6.2 by dropping the flatness condition on \( f \) and by requiring in some cases that the Chow groups of the fibres be finitely generated instead of one-dimensional. Their proofs use standard techniques such as localisation for Chow groups (for that matter, information on the Chow groups of the fibres of \( f \) is extracted from information on the Chow groups of the closed fibres of \( f \) in Section 2), relative Hilbert schemes and a Baire category argument. Let us mention that the assumption of Theorem 6.10 on the singular locus of \( f \) being finite is also required in \cite{8} where...
the construction of relative Chow–K"unneth decompositions is considered. Finally, Theorem 7.1 gathers known results about smooth projective varieties whose Chow groups have small niveau. Together with the results above, in Section 7 we prove some conjectures on algebraic cycles (such as Kimura’s finite-dimensionality conjecture [12], Murre’s conjectures [18], Grothendieck’s standard conjectures [13], the Hodge conjecture) for some smooth projective varieties fibred by very low degree complete intersection, or by cellular varieties over surfaces. For instance, we show the existence of a Murre decomposition for smooth projective varieties fibred by cellular varieties over a curve (Proposition 7.7) and for 6-folds fibred by cubics over a curve (Proposition 7.4), and the standard conjectures for varieties fibred by smooth cellular varieties of dimension \( \leq 4 \) (Proposition 7.3) or by quadrics (Proposition 7.2) over a surface.

**Notations.** We work over a field \( k \) and \( \Omega \) denotes a universal domain that contains \( k \). A variety over \( k \) is a reduced scheme of finite type over \( k \). Throughout, \( f : X \to B \) denotes a projective surjective morphism defined over \( k \) from a quasi-projective variety \( X \) of dimension \( d_X \) to an irreducible quasi-projective variety \( B \) of dimension \( d_B \). Given a scheme \( X \) over \( k \), the group \( \text{CH}_i(X) \) is the \( \mathbb{Q} \)-vector space with basis the \( i \)-dimensional irreducible reduced subschemes of \( X \) modulo rational equivalence. By definition, we set \( \text{CH}_j(X) = 0 \) for \( j < 0 \) and we say that \( \text{CH}_i(X) \) is finitely generated if it is finitely generated as a \( \mathbb{Q} \)-vector space, i.e. if it is a finite-dimensional \( \mathbb{Q} \)-vector space. If \( Z \) is an irreducible closed subscheme of \( X \), we write \([Z]\) for the class of \( Z \) in \( \text{CH}_i(X) \). If \( \alpha \) is the class of a cycle in \( \text{CH}_i(X) \), we write \([\alpha]\) for the support in \( X \) of a cycle representing \( \alpha \). If \( Y \) is a scheme over \( k \) and if \( \beta \) is a cycle in \( \text{CH}_i(X \times Y) \), we define its transpose \( ^t\beta \in \text{CH}_i(Y \times X) \) to be the proper pushforward of \( \beta \) under the obvious map \( \tau : X \times Y \to Y \times X \). If \( X \) and \( Y \) are smooth projective, a cycle \( \gamma \in \text{CH}_i(X \times Y) \) is called a correspondence. The correspondence \( \gamma \) acts both on \( \text{CH}_i(X) \) and \( \text{CH}_i(Y) \) in the following way. Let \( p_X : X \times Y \to X \) and \( p_Y : X \times Y \to Y \) be the first and second projections, respectively. These are proper and flat and we may define, for \( \alpha \in \text{CH}_i(X), \gamma \cdot \alpha := (p_Y)_*(\gamma \cdot p_X^*\alpha) \). Here "\( \cdot \)" is the intersection product on non-singular varieties as defined in [24 §8]. We then define, for \( \beta \in \text{CH}_j(Y) \), \( \gamma \cdot \beta := (\gamma \cdot \beta) \). Given another smooth projective variety \( Z \) and a correspondence \( \gamma' \in \text{CH}_i(Y \times Z) \), the composite \( \gamma' \circ \gamma \in \text{CH}_i(X \times Z) \) is defined to be \((p_{XZ})_* (p_{XY}^* \gamma' \cdot p_Y^* \gamma \circ \gamma') \), where \( p_{XY} : X \times Y \times Z \to X \times Y \) is the projection and likewise for \( p_{XZ} \) and \( p_{YZ} \). The composition of correspondences is compatible with the action of correspondences on Chow groups [24 §16].

Motives are defined in a covariant setting and the notations are those of [27]. Briefly, a Chow motive (or motive, for short) \( M \) is a triple \((X, \|X, n)\) where \( X \) is a variety of pure dimension \( d_X \), \( p \in \text{CH}_d(X \times X) \) is an idempotent \((p \circ p = p)\) and \( n \) is an integer. The motive of \( X \) is denoted \( \mathfrak{h}(X) \) and, by definition, is the motive \((X, \Delta_X, 0)\) where \( \Delta_X \) is the class in \( \text{CH}_{d_X}(X \times X) \) of the diagonal in \( X \times X \). We write \( \mathbf{I} \) for the unit motive \((\text{Spec } k, \Delta_{\text{Spec } k}, 0) = \mathfrak{h}(\text{Spec } k) \). With our covariant setting, we have \( \mathfrak{h}(\mathbb{P}^1) = \mathbf{I} \oplus \mathbf{I}(1) \). A morphism between two
motives $(X, p, n)$ and $(Y, q, m)$ is a correspondence in $q \circ \text{CH}_{d_X+n-m}(X \times Y) \circ p$. If $f : X \to Y$ is a morphism, $\Gamma_f$ denotes the graph of $f$ in $X \times Y$. By abuse, we also write $\Gamma_f \in \text{CH}_d(X \times Y)$ for the class of the graph of $f$. It defines a morphism $\Gamma_f : h(X) \to h(Y)$. By definition we have $\text{CH}_i(X, p, n) = p_* \text{CH}_{i-n}(X)$ and $H_i(X, p, n) = p_* H_{i-2n}(X)$, where we write $H_i(X) := H^{2d-i}(X(C), \mathbb{Q})$ for singular homology when $k \subseteq C$, or $H_i(X) := H^{2d-i}(\mathbb{Q}_X)$ for $t$-adic homology ($\neq \text{char } k$) otherwise.

Given an irreducible scheme $Y$ over $k$, $\eta_Y$ denotes the generic point of $Y$. If $f : X \to B$ and if $Y$ is a closed irreducible subscheme of $B$, $X_{\eta_Y}$ denotes the fibre of $f$ over the generic point of $Y$ and $X_{\eta_Y}$ denotes the fibre of $f$ over a geometric generic point of $Y$.

1. Surjective morphisms and motives

Let us start by recalling a few facts about intersection theory. Let $f : X \to Y$ be a morphism of schemes defined over $k$ and let $l$ be an integer. If $f$ is proper, then there is a well-defined proper pushforward map $f_* : \text{CH}_i(X) \to \text{CH}_i(Y)$; see [7] §1.4. If $f$ is flat, then there is a well-defined flat pullback map $f^* : \text{CH}_i(Y) \to \text{CH}_i(X)$; see [7] §1.7. Pullbacks can also be defined in the following two situations. On the one hand, if $D$ is a Cartier divisor with support $\iota : [D] \hookrightarrow X$, there is a well-defined Gysin map $\iota^* : \text{CH}_1(X) \to \text{CH}_1(\{D\})$ and the composite $\iota_* \circ \iota^* : \text{CH}_1(X) \to \text{CH}_1(\{D\})$ does not depend on the linear equivalence class of $D$, that is, there is a well-defined action of the Picard group $\text{Pic}(X)$ on $\text{CH}_1(X)$; see [7] §2. For instance, if $X$ is a quasi-projective variety given with a fixed embedding $X \hookrightarrow \mathbb{P}^N$, then there is a well-defined map $h : \text{CH}_r(X) \to \text{CH}_{r-1}(X)$ given by intersecting with a hyperplane section of $X$. More generally, if $\tau : Y \hookrightarrow X$ is a locally complete intersection of codimension $r$, then there is a well-defined Gysin map $\tau^* : \text{CH}_1(X) \to \text{CH}_1(Y)$; see [7] §6. For $n > 0$, we write $h^0 = \text{id} : \text{CH}_1(X) \to \text{CH}_1(X)$ and $h^n$ for the $n$-fold composite $h \circ \ldots \circ h : \text{CH}_1(X) \to \text{CH}_{1-n}(X)$. By functoriality of Gysin maps [7] §6.5, if $\iota^* : H^n \hookrightarrow X$ denotes a linear section of codimension $n$, then the composite map $h \circ \ldots \circ h$ coincides with $\iota^n \circ (\iota^*)^*$. When $X$ is smooth projective, we write $\Delta_{H^n}$ for the diagonal inside $H^n \times H^n$, and the correspondence $\Gamma_{\iota^*} \circ \Gamma_{\iota^*} = (\iota^n \times \iota^n)_* [\Delta_{H^n}] \in \text{CH}_{d_X-n}(X \times X)$ induces a map $\text{CH}_1(X) \to \text{CH}_{1-n}(X)$ that coincides with the map $h^n$; see [7] §16. By abuse, we also write $h^n = \Gamma_{\iota^*} \circ \Gamma_{\iota^*}$ for $n > 0$ and $h^0 := [\Delta_X]$. On the other hand, if $f : X \to Y$ is a morphism to a non-singular variety $Y$ and if $x \in \text{CH}_r(X)$ and $y \in \text{CH}_s(Y)$, then there is a well-defined refined intersection product $x \cdot f \cdot y \in \text{CH}_r(|x| \cap f^{-1}(|y|))$, where “$\cap$” denotes the scheme-theoretic intersection; see [7] §8. The pullback $f^* y$ is then defined to be the proper pushforward of $[X] \cdot f \cdot y$ in $\text{CH}_s(X)$. Let us denote $\gamma_f : X \to X \times Y$ the morphism $x \mapsto (x, f(x))$. Because $Y$ is non-singular, this morphism is a locally complete intersection morphism and the pullback $f^*$ is by definition $\gamma_f^* \circ p_Y^*$, where $p_Y : X \times Y \to Y$ is the projection and $\gamma_f^*$ is the Gysin map; see [7] §8.
Finally, if \( f \) is flat, then this pullback map coincides with flat pullback [7, Prop. 8.1.2].

We have the following basic lemma.

**Lemma 1.1.** Let \( f : X \to B \) be a projective surjective morphism between two quasi-projective varieties. Let \( X' \hookrightarrow X \) be a linear section of \( X \) of dimension \( \geq d_B \). Then \( f|_{X'} : X' \to B \) is surjective.

**Proof.** Let \( X \hookrightarrow \mathbb{P}^N \) be an embedding of \( X \) in projective space and let \( H \hookrightarrow \mathbb{P}^N \) be a linear subspace such that \( X' \) is obtained as the pullback of \( X \) along \( H \hookrightarrow \mathbb{P}^N \). The linear subvariety \( H \) has codimension at most \( d_X - d_B \) in \( \mathbb{P}^N \) while a geometric fibre of \( f \) has dimension at least \( d_X - d_B \). Thus every geometric fibre of \( f \) meets \( H \) and hence \( X' \). It follows that \( f|_{X'} \) is surjective. \( \square \)

**Lemma 1.2.** Let \( f : X \to B \) be a projective surjective morphism to a smooth quasi-projective variety \( B \). Then there exists a positive integer \( n \) such that, for all \( i, f_* h^{dx-d_B} \circ f^* : \text{CH}_i(B) \to \text{CH}_i(B) \) is multiplication by \( n \). If moreover \( X \) is smooth and \( B \) is projective, then

\[
\Gamma_f \circ h^{dx-d_B} \circ \Gamma_f = n \cdot \Delta_B \in \text{CH}_{dn}(B \times B).
\]

**Proof.** Let \( \iota' := \iota^{dx-d_B} : H' \hookrightarrow X \) be a linear section of \( X \) of dimension \( d_B \).

We first check, for lack of reference, that \( (f \circ \iota')^* = (\iota')^* \circ f^* \) on Chow groups. Here, \( f \circ \iota' \) and \( f \) are morphisms to the non-singular variety \( B \) and as such the pullbacks \( (f \circ \iota')^* \) and \( (\iota')^* \) are the ones of [7] §8, while \( \iota' \) is the inclusion of a locally closed intersection and as such the pullback \( (\iota')^* \) is the Gysin pullback of [7] §6.5. Let \( \sigma \in \text{CH}^*(B) \), then \( (\iota')^* f^* \sigma = (\iota')^* \gamma_f^*(\sigma) = (\gamma_f \circ \iota')^*(X \times \sigma) \). By the second equality follows from the functoriality of Gysin maps [7] §6.5. Since \( \gamma_f \circ \iota' = (\iota' \times \text{id}_B) \circ \gamma_{\iota(B)} \), we get by using functoriality of Gysin maps once more that \( (\iota')^* f^* \sigma = \gamma_{\iota(B)}^*(\iota' \times \text{id}_B)^*(\sigma) \). Now, we have \( (\iota' \times \text{id}_B)^*((X \times \sigma) = (\iota')^*(X) \times \sigma = [H'] \times \sigma \); see [7] Example 6.5.2. We therefore obtain that \( (\iota')^* f^* \sigma = \gamma_{\iota(B)}^*((H') \times \sigma) : (f \circ \iota')^* \sigma \), as claimed.

Thus, since in addition both \( f \) and \( \iota' \) are proper, we have by functoriality of proper pushforward \( (f \circ \iota')_* (f \circ \iota')^* = f_* \iota'_*(\iota')^* f^* = f_* h^{dx-d_B} \circ f^* \). By Lemma 1.1 the composite morphism \( g := f \circ \iota' \) is generically finite, of degree \( n \) say. It follows from the projection formula [7] Prop. 8.1.1(c)] and from the definition of proper pushforward that, for all \( \gamma \in \text{CH}_i(B),

\[
f_* h^{dx-d_B} \circ f^* \gamma = g_* ([H'] \cdot g^* \gamma) = g_* ([H'] \cdot \gamma) = n[B] \cdot \gamma = n \gamma.
\]

Assume now that \( X \) and \( B \) are smooth projective. In that case, we have \( \Gamma_f \circ h^{dx-d_B} \circ \Gamma_f = \Gamma_g \circ \iota^\Gamma_g := (p_{1,3})_*(p_{1,2}^! \Gamma_g \cdot p_{2,3}^! \Gamma_g) \), where \( p_{i,j} \) denotes projection from \( B \times H' \times B \) to the \((i,j)\)-th factor. By refined intersection, we see that \( \Gamma_g \circ \iota^\Gamma_g \) is supported on \( (p_{1,3})((\Gamma_g \times B) \cap [B \times B]) \), which itself is supported on the diagonal of \( B \times B \). Thus \( \Gamma_f \circ h^{dx-d_B} \circ \Gamma_f \) is a multiple of \( \Delta_B \). We have already showed that \( f_* h^{dx-d_B} \circ f^* = (\Gamma_f \circ h^{dx-d_B} \circ \Gamma_f)_* \) acts by multiplication by \( n \) on \( \text{CH}_i(B) \). Therefore, \( \Gamma_f \circ h^{dx-d_B} \circ \Gamma_f = n \cdot \Delta_B. \) \( \square \)
The following lemma is reminiscent of \cite[Prop. 3.1.(a)]{1628-CharlesVial}.

**Lemma 1.3.** Let $f : X \to B$ be a projective morphism to a smooth quasi-projective variety $B$. Then, for all $i$, $f_* \circ h^i \circ f^* : \operatorname{CH}(B) \to \operatorname{CH}_{i+d_X-d_B -1}(B)$ is the zero map for all $l < d_X - d_B$. If moreover $X$ is smooth and $B$ is projective, then

$$
\Gamma_f \circ h^i \circ t \Gamma_f = 0 \in \operatorname{CH}_{d_X-l}(B \times B) \text{ for all } l < d_X - d_B.
$$

**Proof.** By refined intersection \cite[\S 8]{1628-CharlesVial}, the pullback $f^* \alpha$ is represented by a well-defined class in $\operatorname{CH}_{i+d_X-d_B - l}(f^{-1}([\alpha]))$ for any cycle $\alpha \in \operatorname{CH}(B)$. It follows that $h^i \circ f^* \alpha$ is represented by a well-defined class in $\operatorname{CH}_{i+d_X-d_B - l}(f^{-1}([\alpha]))$. Since $f|_{f^{-1}([\alpha])} : f^{-1}([\alpha]) \to [\alpha]$ is proper, we see by proper pushforward that $f_* \circ h^i \circ f^* \alpha$ is represented by a well-defined cycle $\beta \in \operatorname{CH}_{i+d_X-d_B - l}(\{[\alpha]\})$. But then, $\dim \{[\alpha]\} = i$ so that if $l < d_X - d_B$, then $\operatorname{CH}_{i+d_X-d_B - l}(\{[\alpha]\}) = 0$.

Let us now assume that $X$ and $B$ are smooth projective. Let $\iota : H^l \to X$ be a linear section of $X$ of codimension $l$, and let $h^l$ be the class of $(\iota \times \iota)(\Delta_{H^l})$ in $\operatorname{CH}_{d_X-l}(X \times X)$. By definition we have $\Gamma_f \circ h^l \circ t \Gamma_f = (p_{1,4})_*(p_{1,2}^!(\Gamma_f) \cdot p_{2,3}^! h^l \cdot p_{3,4}^! \Gamma_f)$, where $p_{i,j}$ denotes projection from $B \times X \times X$ to the $(i,j)$-th factor. These projections are flat morphisms, therefore by flat pullback we have $p_{1,2}^!(\Gamma_f) = [\Gamma_f \times X \times B], p_{2,3}^! h^l = [B \times \Delta_{H^l} \times B]$ and $p_{3,4}^! \Gamma_f = [B \times X \times \Gamma_f]$. By refined intersection, the intersection of the closed subschemes $\iota \Gamma_f \times X \times B, B \times \Delta_{H^l} \times B$ and $B \times X \times \Gamma_f$ of $B \times X \times X \times B$ defines a $(d_X - l)$-dimensional class supported on their scheme-theoretic intersection $\{(f(h), h, h, f(h)) : h \in H^l\} \subset B \times X \times X \times B$. Since $f$ is projective, this is a closed subset of dimension $d_X - l$. Also its image under the projection $p_{1,4}$ has dimension at most $d_B$, which is strictly less than $d_X - l$ by the assumption made on $l$. The projection $p_{1,4}$ is a proper map and hence, by proper pushforward, we get that $(p_{1,4})_*\{(f(h), h, h, f(h)) \in B \times X \times X \times B : h \in H^l\} = 0$. \hfill $\Box$

**Theorem 1.4.** Let $f : X \to B$ be a surjective morphism of smooth projective varieties over $k$. Consider the following two morphisms of motives

$$
\Phi := \bigoplus_{i=0}^{d_X-d_B} h^{d_X-d_B-i} \circ \Gamma_f : \bigoplus_{i=0}^{d_X-d_B} h(B)(i) \longrightarrow h(X)
$$

and

$$
\Psi := \bigoplus_{i=0}^{d_X-d_B} \Gamma_f \circ h^i : h(X) \longrightarrow \bigoplus_{i=0}^{d_X-d_B} h(B)(i).
$$

Then $\Phi \circ \Phi$ is an automorphism.

**Proof.** The endomorphism $\Psi \circ \Phi : \bigoplus_{i=0}^{d_X-d_B} h(B)(i) \to \bigoplus_{i=0}^{d_X-d_B} h(B)(i)$ can be represented by the $(d_X - d_B + 1) \times (d_X - d_B + 1)$-matrix whose $(i,j)^{th}$-entries are the morphisms

$$
(\Psi \circ \Phi)_{i,j} = \Gamma_f \circ h^{d_X-d_B-(j-i)} \circ \Gamma_f : h(B)(j-1) \to h(B)(i-1).
$$

By Lemma 1.2, there is a non-zero integer $n$ such that the diagonal entries satisfy $(\Psi \circ \Phi)_{i,i} = n \cdot \text{id}_{h(B)(i-1)}$. By Lemma 1.3, $(\Psi \circ \Phi)_{i,j} = 0$ as soon as
There is a nilpotent endomorphism of $\bigoplus_{i=0}^{d_X-d_B} h(B)(i)$ with $N^{d_X-d_B+1} = 0$. Let us define

$$\Xi := (n \cdot \text{id} - n \cdot N)^{-1} = \frac{1}{n} \cdot (\text{id} + N + N^2 + \cdots + N^{d_X-d_B}).$$

It then follows that $\Xi$ is the inverse of $\Psi \circ \Phi$. □

In the situation of Theorem 1.4, the morphism

$$\Theta := \Xi \circ \Psi$$

then defines a left-inverse to $\Phi$ and the endomorphism

$$p := \Phi \circ \Theta = \Phi \circ \Xi \circ \Psi \in \text{End}(h(X))$$

is an idempotent.

**Proposition 1.5.** With the notations above, the idempotent $p \in \text{End}(h(X)) = \text{CH}_{d_X}(X \times X)$ satisfies $p = \iota^* p$. Moreover, the morphism $\Psi \circ p : (X, p) \to \bigoplus_{i=0}^{d_X-d_B} h(B)(i)$ is an isomorphism with inverse $p \circ \Phi \circ \Xi$.

**Proof.** The second claim consists of the following identities: $\Psi \circ p \circ \Phi \circ \Xi = \Psi \circ \Phi \circ \Xi = \Psi \circ \Phi \circ \Xi = \text{id} \circ \text{id} = \text{id}$ and $p \circ \Phi \circ \Xi \circ \Psi \circ p = p \circ p \circ p = p$. As for the first claim, we have

$$p = \frac{1}{n} \cdot \text{id} \circ \left( 1 + N + \cdots + N^{d_X-d_B} \right) \circ \Psi.$$  

Recall that $N = \text{id} - \frac{1}{n} \cdot \Psi \circ \Phi$, so that it is enough to see that $\iota^* (\Phi \circ \Psi) = \Phi \circ \Psi$. A straightforward computation gives

$$\Phi \circ \Psi = \sum_{i=0}^{d_X-d_B} h^{d_X-d_B-i} \circ \iota_f \circ \Gamma_f \circ h^i.$$  

We may then conclude by noting that the correspondence $h \in \text{CH}_{d_X-1}(X \times X)$ satisfies $h = \iota^* h$. □

Finally, let us conclude with the following counterpart of Theorem 1.4 that deals with the Chow groups of quasi-projective varieties.

**Proposition 1.6.** Let $f : X \to B$ be a projective surjective morphism to a smooth quasi-projective variety $B$. Then the map

$$\Phi_* = \bigoplus_{i=0}^{d_X-d_B} h^{d_X-d_B-i} \circ f^* : \bigoplus_{i=0}^{d_X-d_B} \text{CH}_{d_X-i}(B) \to \text{CH}_i(X)$$  

is split injective and its left-inverse is a polynomial function in $f_*, f^*$ and $h$.  

---

Documenta Mathematica 18 (2013) 1521–1553
Proof. Thanks to Lemma 1.2 and to Lemma 1.3 there is a non-zero integer \( n \) such that 

\[
f_* \circ h^i \circ f^* : \text{CH}_i(B) \to \text{CH}_{i+d_X-d_B-1}(B)
\]

is multiplication by \( n \) if \( i = d_X - d_B \) and is zero if \( i < d_X - d_B \). Let us write \( \Psi_* \) for \( \bigoplus_{j=0}^{d_X-d_B} f_* \circ h^j : \text{CH}(X) \to \bigoplus_{j=0}^{d_X-d_B} \text{CH}_{i-j}(B) \). In order to prove the injectivity of \( \Phi_* \), it suffices to show that the composite

\[
\Psi_* \circ \Phi_* : \bigoplus_{i=0}^{d_X-d_B} \text{CH}_{i-i}(B) \to \text{CH}(X) \to \bigoplus_{j=0}^{d_X-d_B} \text{CH}_{i-j}(B)
\]

is an isomorphism. But then, as in the proof of Theorem 1.4, we see that \( \Psi_* \circ \Phi_* \) can be represented by a lower triangular matrix whose diagonal entries’ action on \( \text{CH}_{i-i}(B) \) is given by multiplication by \( n \). \( \square \)

Remark 1.7. Note that the conclusion of Proposition 1.6 also holds for a flat and projective surjective morphism \( f : X \to B \) of quasi-projective varieties.

2. On the Chow groups of the fibres

In this section, we fix a universal domain \( \Omega \). The following statement was communicated to me by Burt Totaro.

Lemma 2.1. Let \( f : X \to B \) be a morphism of varieties over \( \Omega \) and let \( F \) be a geometric generic fibre of \( f \). Then there is a subset \( U \subseteq B(\Omega) \) which is a countable intersection of nonempty Zariski open subsets such that for each point \( b \in U \), there is an isomorphism from the field \( \Omega \) to the field \( \Omega(B) \) such that this isomorphism turns the scheme \( X_b \) over \( \Omega \) into the scheme \( F \) over \( \Omega(B) \). In other words, a very general fibre of \( f \) is isomorphic to \( F \) as an abstract scheme. Consequently, for each point \( p \in U \), \( \text{CH}_i(X_b) \) is isomorphic to \( \text{CH}_i(F) \) for all integers \( i \).

Proof. There exist a countable subfield \( K \subset \Omega \) and varieties \( X_0 \) and \( B_0 \) defined over \( K \) together with a \( K \)-morphism \( f_0 : X_0 \to B_0 \) such that \( f = f_0 \times_{\text{Spec}K} \text{Spec} \Omega \). Let us define \( U \subseteq B(\Omega) \) to be \( \bigcap_{\eta \in B_0} \big( (B_0)_{\eta_B_0} \big)_{\Omega(\Omega)} \), where the intersection runs through all proper \( K \)-subschemas \( Z_0 \) of \( B_0 \). Note that there are only countably many such subschemes of \( B_0 \) and that \( U \) is the set of \( \Omega \)-points of \( B = B_0 \times_{\text{Spec}K} \text{Spec} \Omega \) that do not lie above a proper Zariski-closed subset of \( B_0 \). Let now \( b : \text{Spec} \Omega \to B \) be a \( \Omega \)-point of \( B \) that lies in \( U \), i.e. a point \( b \) such that the composite map \( \beta : \text{Spec} \Omega \to B \) is dominant, or equivalently such that the composite map \( \beta \) factors as \( \eta_{B_0} \circ \alpha \) for some morphism \( \alpha : \text{Spec} \Omega \to \text{Spec} K(B_0) \), where \( \eta_{B_0} : \text{Spec} K(B_0) \to B_0 \) is the generic point of \( B_0 \). Since \( X \) is pulled back from \( X_0 \) along \( B \to B_0 \), we see that \( X_b \) the fibre of \( f \) at \( b \) is the pull back of the generic fibre \( (X_0)_{\eta_{B_0}} \) along \( \alpha \). Consider then \( \eta_{\Omega(B)} : \text{Spec} \Omega(B) \to B \) a geometric generic point of \( B \) such that \( X_{\Omega(B)} = F \). Since the composite map \( \text{Spec} \Omega(B) \to B \) factors through \( \eta_{B_0} : \text{Spec} K(B_0) \to B_0 \), we see as before that \( F \) is the pull-back of the generic fibre \( (X_0)_{\eta_{B_0}} \) along some morphism.
$\alpha' : \text{Spec} \Omega(B) \to \text{Spec} K(B_0)$. The fields $\Omega(B)$ and $\Omega$ are algebraically closed fields of infinite transcendence degree over $K(B_0)$ and there thus exists an isomorphism $\Omega(B) \cong \Omega$ fixing $K(B_0)$. Hence, the fibre $X_b$ identifies with $F$ after pullback by the isomorphism $\text{Spec} \Omega \cong \text{Spec} \Omega(B)$ over $\text{Spec} K(B_0)$.

The last statement follows from the fact that the Chow groups of a variety $X$ over a field only depend on $X$ as a scheme. Precisely, if one denotes $\psi_b : X_b \to F$ an isomorphism of schemes, then the proper pushforward map $(\psi_b)_* : \text{CH}_i(X_b) \to \text{CH}_i(F)$ is an isomorphism with inverse $(\psi_b^{-1})_* : \text{CH}_i(F) \to \text{CH}_i(X_b)$. \hfill $\Box$

The following lemma will be useful to refer to.

**Lemma 2.2.** Let $f : X \to B$ be a projective surjective morphism defined over $\Omega$ onto a quasi-projective variety $B$. Assume that $\text{CH}_i(X_b) = \mathbf{Q}$ (resp. $\text{CH}_i(X_b)$ is finitely generated) for all $b \in B(\Omega)$. Then $\text{CH}_i(X_{\eta_0}) = \mathbf{Q}$ (resp. $\text{CH}_i(X_{\eta_0})$ is finitely generated) for all irreducible subvarieties $D$ of $X$.

**Proof.** Let $D$ be an irreducible subvariety of $B$ and let $\eta_D \to D$ be a geometric generic point of $D$. By Lemma 2.1 applied to $X_D := X \times_B D$, there is a closed point $d \in D$ such that $\text{CH}_i(X_{\eta_d})$ is isomorphic to $\text{CH}_i(X_d)$. By assumption $\text{CH}_i(X_d) = \mathbf{Q}$ (resp. $\text{CH}_i(X_d)$ is finitely generated). Therefore $\text{CH}_i(X_{\eta_d}) = \mathbf{Q}$ (resp. $\text{CH}_i(X_{\eta_d})$ is finitely generated), too. By a norm argument for Chow groups, the pullback map $\text{CH}_i(X_{\eta_0}) \to \text{CH}_i(X_{\eta_d})$ is injective. Hence $\text{CH}_i(X_{\eta_0}) = \mathbf{Q}$ (resp. $\text{CH}_i(X_{\eta_0})$ is finitely generated). \hfill $\Box$

The following definition is taken from Laterveer [15].

**Definition 2.3.** Let $X$ be a variety over $k$. The Chow group $\text{CH}_i(X)$ is said to have niveau $\leq r$ if there exists a closed subscheme $Y \subset X$ of dimension $i+r$ such that the proper pushforward map $\text{CH}_i(Y_\Omega) \to \text{CH}_i(X_\Omega)$ is surjective.

**Proposition 2.4.** Let $f : X \to B$ be a generically smooth, projective and dominant morphism onto a smooth quasi-projective variety $B$ defined over $\Omega$. Let $n$ be a non-negative integer. The following statements are equivalent.

1. If $F$ is a general fibre, then $\text{CH}_0(F)$ has niveau $\leq n$;
2. If $F$ is a very general fibre, then $\text{CH}_0(F)$ has niveau $\leq n$;
3. If $F$ is a geometric generic fibre, then $\text{CH}_0(F)$ has niveau $\leq n$.

**Proof.** The implication $(1) \Rightarrow (2)$ is obvious. Let us prove $(2) \Rightarrow (3)$. Let $X_b$ be a very general fibre and $F$ a geometric generic fibre of $f$, and, by Lemma 2.1 let $\psi_b : X_b \to F$ be an isomorphism of schemes. Assume that there is a closed subscheme $Z$ of dimension $\leq r$ in $X_b$, for some integer $r$, such that the proper pushforward $\text{CH}_0(Z) \to \text{CH}_0(X_b)$ is surjective. Then, denoting $Z'$ the image of $Z$ in $F$ under $\psi_b$, functoriality of proper pushforwards implies that $\text{CH}_0(Z') \to \text{CH}_0(F)$ is surjective. We may then conclude by noting that the subscheme $Z'$ has dimension $\leq r$ in $F$.

As for $(3) \Rightarrow (1)$, let $Y$ be a subvariety of $F$ defined over $\Omega_B$ such that $\text{CH}_0(Y) \to \text{CH}_0(F)$ is surjective. The technique of decomposition of the diagonal of Bloch–Srinivas [2] gives $\Delta_F = \Gamma_1 + \Gamma_2 \in \text{CH}_{\text{dim} F}(F \times F)$, where $\Gamma_1$ is
supported on $F \times Y$ and $\Gamma_2$ is supported on $D \times F$ for some divisor $D$ in $F$. Consider a Galois extension $K/\Omega(B)$ over which the above decomposition and the morphism $Y \to F$ are defined, and consider an étale morphism $U \to B$ with $\Omega(U) = K$ such that $f$ restricted to $U$ is smooth. Let $u$ be a $\Omega$-point of $U$ and let $X_u$ be the fibre of $f$ over $u$. Then the decomposition $\Delta_F = \Gamma_1 + \Gamma_2$ specialised \cite{4} [20.3] on $X_u \times X_u$ to a similar decomposition, where $\Gamma_1|_{X_u \times X_u}$ is supported on $X_u \times Y_u$ and $\Gamma_2|_{X_u \times X_u}$ is supported on $D_u \times X_u$. Letting it act on zero-cycles, we see that $\text{CH}_0(X_u)$ is supported on $Y_u$. \hfill \Box

**Remark 2.5.** When $n = 0$ or $n = 1$, the statements of Proposition 2.4 are further equivalent to $\text{CH}_0(F)$ having niveau $\leq n$ for $F$ the generic fibre of $f$. Indeed, if $X$ is a smooth projective variety such that $\text{CH}_0(X)$ has niveau $\leq 1$, then $\text{CH}_0(X)$ is supported on a one-dimensional linear section \cite{3} Proposition 1.6]. In particular, $\text{CH}_0(X)$ is supported on a one-dimensional subvariety of $X$ which is defined over a field of definition of $X$. Note that, for general $n$, it is a consequence of the Lefschetz hyperplane theorem and of the Bloch–Beilinson conjectures that if $\text{CH}_0(X)$ has niveau $\leq n$, then $\text{CH}_0(X)$ is supported on an $n$-dimensional linear section of $X$.

3. A GENERALISATION OF THE PROJECTIVE BUNDLE FORMULA

We establish a formula that is analogous to the projective bundle formula for Chow groups. Our formula holds for flat morphisms, rather than Zariski locally trivial morphisms as is the case for the projective bundle formula. However, since a flat morphism does not have any local sections in general, it only holds with rational coefficients.

**Proposition 3.1.** Let $f : X \to B$ be a flat projective surjective morphism of quasi-projective varieties. Let $l \geq 0$ be an integer. Assume that

$$\text{CH}_{l-i}(X_{\eta_{B_i}}) = \mathbb{Q}$$

for all $0 \leq i \leq \min(l, d_B)$ and for all closed irreducible subschemes $B_i$ of $B$ of dimension $i$, where $\eta_{B_i}$ is the generic point of $B_i$. Then the map

$$\Phi_* : \bigoplus_{i=0}^{d_X - d_B} h^{d_X - d_B - i} \circ f^* : \bigoplus_{i=0}^{d_X - d_B} \text{CH}_{l-i}(B) \to \text{CH}_l(X)$$

is surjective.

**Proof.** The case when $d_B = 0$ is obvious. Let us proceed by induction on $d_B$. We have the localisation exact sequence

$$\bigoplus_{D \in B^1} \text{CH}_l(X_D) \to \text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B}) \to 0,$$

where the direct sum is taken over all irreducible divisors of $B$. If $l \geq d_B$, let $Y$ be a linear section of $X$ of dimension $l$. By Lemma \cite{1} f|_Y : Y \to B is surjective. The restriction map $\text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B})$ is the direct limit of the flat pullback maps $\text{CH}_l(X) \to \text{CH}_l(X_U)$ taken over all open subsets $U$ of
B; see [3, Lemma 1A.1]. Therefore $\text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B})$ sends the class of $Y$ to the class of $Y_{\eta_B}$ inside $\text{CH}_{l-d_B}(X_{\eta_B})$. But then this class is non-zero because the restriction to $\eta_B$ of a linear section of $Y$ of dimension $d_B$ has positive degree. Furthermore, if $[B]$ denotes the class of $B$ in $\text{CH}_{d_B}(B)$, then the class of $Y$ is equal to $h^{d_X-i} \circ f^*[B]$ in $\text{CH}_l(X)$. Thus, since by assumption $\text{CH}_{l-d_B}(X_{\eta_B}) = \mathbb{Q}$, the composite map

$$\text{CH}_{d_B}(B) \xrightarrow{h^{d_X-i} \circ f^*} \text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B})$$

is surjective.

Consider now the fibre square

$$\begin{array}{ccc}
X_D & \xrightarrow{j_D} & X \\
\downarrow f_D & & \downarrow f \\
D & \xrightarrow{j_D} & B
\end{array}$$

Then $f_D : X_D \to D$ is flat and its fibres above points of $D$ satisfy the assumptions of the theorem. Therefore, by the inductive assumption, we have a surjective map

$$\bigoplus_{i=0}^{d_X-d_B} h^{d_X-d_B-i} \circ f_D^* : \bigoplus_{i=0}^{d_X-d_B} \text{CH}_{l-i}(D) \twoheadrightarrow \text{CH}_l(X_D).$$

Furthermore, since $f$ is flat and $j_D$ is proper, we have the formula [7, Prop. 1.7 & Th. 6.2]

$$j_D^* \circ h^{d_X-d_B-i} \circ f_D^* = h^{d_X-d_B-i} \circ f^* \circ j_D^* : \text{CH}_{l-i}(D) \to \text{CH}_l(X).$$

Hence, the image of $\Phi_*$ contains the image of

$$\bigoplus_{D \in B^1} \bigoplus_{i=0}^{d_X-d_B} j_D^* \circ h^{d_X-d_B-i} \circ f_D^* : \bigoplus_{D \in B^1} \bigoplus_{i=0}^{d_X-d_B} \text{CH}_{l-i}(D) \twoheadrightarrow \text{CH}_l(X).$$

Altogether, this implies that the map $\Phi_*$ is surjective. \hfill \Box

We can now gather the statements and proofs of Propositions 1.6 and 3.1 into the following.

**Theorem 3.2.** Let $f : X \to B$ be a flat and projective surjective morphism onto a quasi-projective variety $B$ of dimension $d_B$. Let $l \geq 0$ be an integer. Assume that

$$\text{CH}_{l-i}(X_b) = \mathbb{Q} \text{ for all } 0 \leq i \leq \min(l, d_B) \text{ and for all points } b \text{ in } B(\Omega).$$

Then the map

$$\Phi_* = \bigoplus_{i=0}^{d_X-d_B} h^{d_X-d_B-i} \circ f^* : \bigoplus_{i=0}^{d_X-d_B} \text{CH}_{l-i}(B) \to \text{CH}_l(X)$$

is surjective.
is an isomorphism. Moreover the map
\[ \Psi_* = \bigoplus_{i=0}^{d_X - d_B} f_* \circ h^i : \text{CH}_l(X) \to \bigoplus_{i=0}^{d_X - d_B} \text{CH}_{l-1}(B) \]
is also an isomorphism.

\[ \square \]

Proof. Let \( B_i \) be an irreducible closed subscheme of \( B \) of dimension \( i \) with \( 0 \leq i \leq \min(l, d_B) \). Since \( \text{CH}_{l-i}(X_b) = \mathbb{Q} \) for all points \( b \in B(\Omega) \), Lemma 4.2 gives \( \text{CH}_{l-i}(X_{\eta_B}) = \mathbb{Q} \). Thus the theorem follows from a combination of Proposition 1.6 (and Remark 1.7) and Proposition 3.1.

4. On the motive of quadric bundles

Let us first recall the following result.

Proposition 4.1 (Corollary 2.2 in [27]). Let \( m \) and \( n \) be positive integers. Let \( Y, q \) be a motive over \( k \) such that \( \text{CH}_i(Y, q^i) = 0 \) for all \( i < n \) and \( \text{CH}_j(Y, q^j) = 0 \) for all \( j < m \). Then there exist a smooth projective variety \( Z \) over \( k \) of dimension \( d_X - m - n \) and an idempotent \( r \in \text{End}(\mathcal{h}(Z)) \) such that \( (Y, q) \) is isomorphic to \( (Z, r, n) \).

The main result of this section is the following theorem.

Theorem 4.2. Let \( f : X \to B \) be a flat morphism of smooth projective varieties over \( k \). Assume that there exists a positive integer \( n \) such that \( \text{CH}_i(X_b) = \mathbb{Q} \) for all \( 0 \leq i \leq n \) and for all points \( b \in B(\Omega) \). Then there exists a smooth projective variety \( Z \) of dimension \( d_X - 2n \) and an idempotent \( r \in \text{End}(\mathcal{h}(Z)) \) such that the motive of \( X \) admits a direct sum decomposition

\[ \mathcal{h}(X) \cong \bigoplus_{i=0}^{d_X - d_B} \mathcal{h}(B)(i) \oplus (Z, r, n). \]

Proof. With the notations of Theorem 1.4 and its proof the endomorphism \( \Psi \circ \Phi \in \text{End} \left( \bigoplus_{i=0}^{d_X - d_B} \mathcal{h}(B)(i) \right) \) admits an inverse denoted \( \Xi \). Proposition 1.5 then states that \( p := \Phi \circ \Xi \circ \Psi \in \text{End}(\mathcal{h}(X)) \) is a self-dual idempotent such that \( (X, p) \cong \bigoplus_{i=0}^{d_X - d_B} \mathcal{h}(B)(i) \). By Theorem 3.2 \( (p_\Omega)_*: \text{CH}_l(X_\Omega) \to \text{CH}_l(\mathcal{X}_\Omega) \) is an isomorphism for all \( l < n \). It follows that \( \text{CH}_l(X_\Omega, p_\Omega) = \text{CH}_l(\mathcal{X}_\Omega) \) for all \( l < n \) and thus that \( \text{CH}_l(X_\Omega, \mathcal{id}_\Omega - p_\Omega) = 0 \) for all \( l < n \). Because \( p = 1_p \), we also have \( \text{CH}_l(X_\Omega, \mathcal{id}_\Omega - 1_p) = 0 \) for all \( l < n \). Proposition 1.1 then yields the existence of a smooth projective variety \( Z \) of dimension \( d_X - 2n \) such that \( (X, \mathcal{id} - p) \) is isomorphic to a direct summand of \( \mathcal{h}(Z)(n) \).

Our original motivation was to establish Murre’s conjectures [18] for smooth projective varieties fibred by quadrics over a surface. The importance of Murre’s conjectures was demonstrated by Jannsen who proved [11] that these hold true for all smooth projective varieties if and only if Bloch and Beilinson’s conjecture holds true. In our covariant setting, Murre’s conjectures can be stated as follows.
(A) There exist mutually orthogonal idempotents $\pi_0, \ldots, \pi_d \in \text{CH}^d(X \times X)$ adding to the identity such that $(\pi_i)_* H_i(X) = H_i(X)$ for all $i$. We say that $X$ has a Chow–Künneth decomposition.

(B) $\pi_0, \ldots, \pi_{2l-1}, \pi_{2l+1}, \ldots, \pi_{2d}$ act trivially on $\text{CH}^l(X)$ for all $l$. 

(C) $F^i \text{CH}^l(X) := \text{Ker} (\pi_{2l}) \cap \ldots \cap \text{Ker} (\pi_{2l+i-1})$ doesn’t depend on the choice of the $\pi_j$’s. Here the $\pi_j$’s are acting on $\text{CH}^l(X)$.

(D) $F^1 \text{CH}^l(X) = \text{CH}^l(X)_{\text{hom}} := \text{Ker} (\text{CH}^l(X) \to H_{2l}(X))$.

**Definition 4.3.** A variety $X$ that satisfies conjectures (A), (B) and (D) is said to have a Murre decomposition.

In the particular case when $f$ is a flat morphism whose geometric fibres are quadrics, Theorem 4.2 implies the following corollary. We write $\lfloor a \rfloor$ for the greatest integer which is smaller than or equal to the rational number $a$.

**Corollary 4.4.** Let $f : X \to B$ be a flat morphism of smooth projective varieties over $k$. Assume that $\text{CH}^l(X_b) = \mathbb{Q}$ for all $0 \leq l < \frac{d_X - d_B}{2}$ and for all points $b \in B(\Omega)$. For instance, the geometric fibres of $f$ could either be quadrics or complete intersection of dimension 4 and bidegree $(2, 2)$. Then

- If $d_B = 1$, then $X$ is Kimura finite-dimensional [12].
- If $d_B \leq 2$, then $X$ has a Murre decomposition.
- If $d_B = 3$, $d_X - d_B$ is odd and $B$ has a Murre decomposition, then $X$ has a Murre decomposition.

**Proof.** By Theorem [12] there is a variety $Z$ and an idempotent $r \in \text{End}(h(Z))$ such that the motive of $X$ admits a direct sum decomposition

$$h(X) \cong \bigoplus_{i=0}^{d_X - d_B} h(B)(i) \oplus (Z, r, \lfloor \frac{d_X - d_B + 1}{2} \rfloor),$$

where

$$d_Z = \begin{cases} d_B - 1 & \text{if } d_X - d_B \text{ is odd;} \\ d_B & \text{if } d_X - d_B \text{ is even.} \end{cases}$$

Thus, we only need to note that any direct summand of the motive of a curve is finite-dimensional [12] and that any direct summand of the motive of a surface has a Murre decomposition [27, Theorem 3.5]. Finally, let us mention that, when $d_B = 1$, it is not necessary to assume $f$ to be flat to conclude that $X$ is Kimura finite-dimensional; see Propositions [7.3] and [7.5] below.

**Remark 4.5.** Examples of 3-folds having a Murre decomposition include products of a curve with a surface [19], 3-folds rationally dominated by a product of curves [28] and uniruled 3-folds [5].

---

1 Actually if $f$ is flat and if its closed geometric fibres are quadrics, then all of its geometric fibres are quadrics. Conversely if the geometric fibres of $f$ are quadrics of dimension $d_X - d_B$, then $f$ is flat.
Remark 4.6 (The case of smooth families). Suppose \( f : X \to B \) is a smooth morphism between smooth projective varieties with geometric fibres being quadric hypersurfaces. Iyer [10] showed that \( f \) is étale locally trivial and deduced that \( f \) has a relative Chow–Künneth decomposition. By using the technique of Gordon–Hanamura–Murre [8], it is then possible to prove that

\[
h(X) \cong \bigoplus_{i=0}^{d_X-2} h(B)(i) \oplus h(B)(d_X-d_B) \quad \text{if \( d_X - d_B \) is odd;}
\]
\[
h(X) \cong \bigoplus_{i=0}^{d_X-2} h(B)(i) \oplus h(B)(\frac{d_X-d_B}{2}) \quad \text{if \( d_X - d_B \) is even.}
\]

Remark 4.7. Suppose \( f : X \to S \) is a complex morphism from a smooth projective 3-fold \( X \) to a smooth projective surface \( S \) whose fibres are conics. In that case, Nagel and Saito [20] identify (up to direct summands isomorphic to \( \mathbb{I} \) or \( \mathbb{I}(1) \)) the motive \((Z,r)\) in the proof of Corollary 4.3 with the \( h_1 \) of the Prym variety \( P \) attached to a double-covering of the discriminant curve \( C \) of \( f \).

If now \( f : X \to S \) is a flat complex morphism from a smooth projective variety \( X \) to a smooth projective surface \( S \) whose fibres are odd-dimensional quadrics, then, because the motive of a curve is Kimura finite-dimensional and by the Lefschetz (1,1)-theorem, one would deduce an identification of the \( h_1 \) of \((Z,r)\) with \( h_1(P) \) from an isomorphism of Hodge structures \( H^1(Z,r) \cong H^1(P) \). Here, \( P \) again is the Prym variety attached to a double-covering of the discriminant curve \( C \) of \( f \). Such an identification is currently being investigated by J. Bouali [4] by generalising the methods of [20].

**Corollary 4.8.** Let \( f : X \to B \) be a flat dominant morphism between smooth projective varieties defined over a finite field \( F \) whose geometric fibres are quadrics. If \( d_B \leq 2 \), then numerical and rational equivalence agree on \( X \).

**Proof.** As in the proof of Corollary 4.3 there is a direct sum decomposition

\[
h(X) \cong \bigoplus_{i=0}^{d_X-2} h(B)(i) \oplus (Z,r, \left\lfloor \frac{d_X-d_B+1}{2} \right\rfloor)
\]

for some smooth projective variety \( Z \), which is a curve if \( d_X - d_B \) is odd and a surface if \( d_X - d_B \) is even. Now the action of correspondences preserves numerical equivalence so that if \( \alpha \) denotes the isomorphism from \( h(X) \) to the right-hand side of (3) and if \( \beta \) denotes its inverse, then we have \( CH_l(X)_{\text{num}} = \beta_\ast \alpha_{\ast} CH_l(X)_{\text{num}} \) for all \( l \). In particular, \( CH_l(X)_{\text{num}} = \beta_\ast \left( \bigoplus_{i=0}^{d_X-2} CH_{l-i}(B)_{\text{num}} \oplus r_\ast CH_{l-m}(Z)_{\text{num}} \right) \), where \( m = \left\lfloor \frac{d_X-d_B+1}{2} \right\rfloor \). The corollary then follows from the fact that for any smooth projective variety \( Y \) defined over a finite field the groups \( CH_0(Y)_{\text{num}}, CH^1(Y)_{\text{num}} \) and \( CH^0(Y)_{\text{num}} \) are zero. \( \square \)

5. On the motive of a smooth blow-up

Let \( X \) be a smooth projective variety over a field \( k \) and let \( j : Y \hookrightarrow X \) be a smooth closed subvariety of codimension \( r \). We write \( \tau : \tilde{X}_Y \to X \) for the blow-up of \( X \) along \( Y \). Manin [10] showed by an existence principle that the natural map, which is denoted \( \Phi \) below, \( h(X) \oplus \bigoplus_{i=1}^{r-1} h(Y)(i) \to h(\tilde{X}_Y) \) is
an isomorphism of Chow motives. Here, we make explicit the inverse to this isomorphism. An application is given by Proposition 5.4. The results of this section will not be used in the rest of the paper.

We have the following fibre square

\[
\begin{array}{ccc}
D & \xrightarrow{\tilde{j}} & \tilde{X}_Y \\
\tau_D \downarrow & & \downarrow \tau \\
Y & \xrightarrow{j} & X
\end{array}
\]

where \( \tilde{j} : D \to \tilde{X}_Y \) is the exceptional divisor and where \( \tau_D : D \to Y \) is a \( \mathbb{P}^{r-1} \)-bundle over \( Y \). Precisely \( D = \mathbb{P}(N_{Y/X}) \) is the projective bundle over \( Y \) associated to the normal bundle \( N_{Y/X} \) of \( Y \) inside \( X \). The tautological line bundle on \( D = \mathbb{P}(N_{Y/X}) \) is \( \mathcal{O}_{\mathbb{P}(N_{Y/X})}(-1) = \mathcal{O}_{\tilde{X}_Y}(D)|_D \). Let \( H_{i-1} \subset \ldots \subset H_i \subset \ldots \subset H \subset D \) be linear sections of \( D \) corresponding to the relatively ample line bundle \( \mathcal{O}_{\mathbb{P}(N_{Y/X})}(1) \), where \( H^i \) has codimension \( i \). Thus, if \( D \hookrightarrow \mathbb{P}^M \times Y \) is an embedding over \( Y \) corresponding to \( \mathcal{O}_{\mathbb{P}(N_{Y/X})}(1) \), then \( H^i \) denotes the smooth intersection of \( D \) with \( L^i \times Y \) for some linear subspace \( L^i \) of codimension \( i \) inside \( \mathbb{P}^M \). Let us write \( \iota^i : H^i \hookrightarrow D \) for the inclusion maps.

Let us define the morphism of motives

\[
\Phi := \tau^* \oplus \bigoplus_{i=1}^{r-1} \iota^i \circ h^{r-1-i} \circ \iota^i \tau_D^* : \text{CH}_l(X) \oplus \bigoplus_{i=1}^{r-1} \text{CH}_{l-i}(Y) \to \text{CH}_l(\tilde{X}_Y).
\]

Here, \( h^l \) is the correspondence \( \Gamma_{l,i} \circ \iota^i \Gamma_{l,i} \); it coincides with the \( l \)-fold composite of \( h := \Gamma_{l,i} \circ \iota^i \Gamma_{l,i} \) with itself.

On the one hand, we have the following blow-up formula for Chow groups; see [10].

**Proposition 5.1.** The induced map

\[
\Phi_* = \tau^* \oplus \bigoplus_{i=1}^{r-1} j_* h^{r_1-1-i} \tau_D^* : \text{CH}_l(X) \oplus \bigoplus_{i=1}^{r-1} \text{CH}_{l-i}(Y) \to \text{CH}_l(\tilde{X}_Y)
\]

is an isomorphism. \( \square \)

On the other hand, we define

\[
\Psi := \Gamma_{l,r} \oplus \bigoplus_{i=1}^{r-1} (-1) \cdot \iota^i \tau_D \circ h^{i-1} \circ \iota^i j^* : \mathfrak{h}(\tilde{X}_Y) \to \mathfrak{h}(X) \oplus \bigoplus_{i=1}^{r-1} \mathfrak{h}(Y)(i).
\]
Let \((\Psi \circ \Phi)_{i,j}\) be the \((i,j)\)th component of \(\Psi \circ \Phi\), where \(h(X)\) is by definition the \(0\)th coordinate of \(h(X) \oplus \bigoplus_{i=1}^{r-1} h(Y)(i)\). Thus, if \(i, j \neq 0\), then \((\Psi \circ \Phi)_{i,j}\) is a morphism \(h(Y)(j) \to h(Y)(i)\); if \(i \neq 0\), then \((\Psi \circ \Phi)_{i,0}\) is a morphism \(h(Y)(j) \to h(X)\); and \((\Psi \circ \Phi)_{0,j}\) is a morphism \(h(X) \to h(X)\).

The following lemma shows that \(\Psi \circ \Phi\) is a lower triangular matrix with invertible diagonal elements.

**Lemma 5.2.** We have

\[
(\Psi \circ \Phi)_{i,j} = \begin{cases} 
0 & \text{if } i < j \\
\Delta_X & \text{if } i = j = 0 \\
\Delta_Y & \text{if } i = j > 0
\end{cases}
\]

**Proof.** The proposition consists of the following relations:

1. \(\Gamma_r \circ \Gamma_r = \Delta_X\).
2. \(\Gamma_{r_1} \circ h^{i-1} \circ \Gamma_j \circ \Gamma_{r_2} \circ h^{r-1} \circ \Gamma_{r_2} = 0\) for all \(1 \leq i < j \leq r - 1\).
3. \(\Gamma_{r_1} \circ h^{i-1} \circ \Gamma_j \circ \Gamma_{r_2} \circ h^{r-1} \circ \Gamma_{r_2} = -\Delta_Y\) for all \(1 \leq i \leq r - 1\).
4. \(\Gamma \circ \Gamma_{r_1} \circ h^{r-1} \circ \Gamma_{r_2} = 0\) for all \(1 \leq i \leq r - 1\).

Let us establish them. The morphism \(\tau\) is a birational morphism so that the identity (1) follows from the projection formula as in the proof of Lemma 1.2. The proof of (4) is a combination of the fact that \(\circ j = j \circ \tau_D\) and Lemma 1.3. As for (2) and (3), we claim that

\[
\Gamma_j \circ \Gamma_j = -\Gamma_i \circ \Gamma_i = -h \in \text{CH}_{\Delta_1}(D \times D).
\]

Indeed, the action of \(h\) on \(\text{CH}_r(D)\) is given by intersecting with the class of \(H\). Also, by \cite{7} Prop. 2.6], the map \(j^* \gamma^*_a : \text{CH}_r(D) \to \text{CH}_{r-1}(D)\) is given by intersecting with the class of \(D_{ij}\) which is precisely \(-h\). The same arguments for the smooth blow-up of \(X \times Z\) along \(Y \times Z\), together with Manin’s identity principle, yield the claim.

In view of the above claim, (2) follows from Lemma 1.3 and (3) follows from Lemma 1.2.

Thus the endomorphism

\[
N := \begin{pmatrix}
\Delta_X & 0 & 0 & \cdots & 0 \\
0 & \Delta_Y & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \Delta_Y
\end{pmatrix} - \Psi \circ \Phi
\]

is a nilpotent endomorphism of \(h(X) \oplus \bigoplus_{i=1}^{r-1} h(Y)(i)\) of index \(\leq r - 1\), i.e. \(N^{r-1} = 0\). The morphism

\[
\Theta := (\text{id} + N + \ldots + N^{r-2}) \circ \Psi
\]

thus gives a left-inverse to \(\Phi\).
The main result of this section is then the following theorem.

**Theorem 5.3.** The morphism $\Theta$ is the inverse of $\Phi$.

**Proof.** Let $p := \Phi \circ \Theta \in \text{End}(\mathfrak{h}(\tilde{X}_Y))$. Because $\Theta$ is a left-inverse to $\Phi$, we see that $p$ is an idempotent. The motive of $\tilde{X}_Y$ thus splits as

$$\mathfrak{h}(\tilde{X}_Y) = (\tilde{X}_Y, p) \oplus (\tilde{X}_Y, \text{id} - p).$$

As a consequence of Proposition [5.1] and Lemma [5.2] we obtain that

$$\text{CH}_*(\tilde{X}_Y, p) = \text{CH}_*(\tilde{X}_Y).$$

Actually, since $(\tilde{X} \times Z)_{Y \times Z}$ canonically identifies with $\tilde{X}_Y \times Z$, we get, thanks to [7] Prop. 16.1.1, that $\Phi \times \Delta_Z$ induces an isomorphism of Chow groups as in Proposition [5.1] and that $\Theta \times \Delta_Z$ is a left-inverse to $\Phi \times \Delta_Z$. Thus, we also have that

$$\text{CH}_*(\tilde{X}_Y \times Z, p \times \Delta_Z) = \text{CH}_*(\tilde{X}_Y \times Z).$$

Therefore

$$\text{CH}_*(\tilde{X}_Y \times Z, \Delta_{\tilde{X}_Y} \times \Delta_Z - p \times \Delta_Z) = 0.$$

By Manin’s identity principle it follows that

$$p = \Delta_{\tilde{X}_Y}.$$ 

In other words, $\Theta$ is not only a left-inverse to $\Phi$, it is also the inverse of $\Phi$. □

Let us now use Theorem [5.3] to study the birational invariance of some groups of algebraic cycles attached to smooth projective varieties. For a smooth projective variety $X$ over $k$, we write $\text{Griff}(X)$ for its Griffiths group $\text{CH}_1(X)_{\text{hom}}/\text{CH}_1(X)_{\text{alg}}$. We also write, when $k \subseteq \mathbb{C}$,

$$T^i(X) := \text{Ker}(A.J^i : \text{CH}^i(X)_{\text{hom}} \to J^i(X))$$

for the kernel of Griffiths’ Abel–Jacobi map to the intermediate Jacobian $J^i(X)$ which is a quotient of $H^{2i-1}(X, \mathbb{C})$.

If $\pi : \tilde{X} \to X$ is a birational map, the projection formula implies that $\Gamma_\pi \circ \Gamma^i_\pi = \Delta_X$; see Lemma [12]. Thus $\pi_* \pi^*$ acts as the identity on $\text{CH}_1(X, \text{Griff}(\tilde{X}), \text{Griff}^2(\tilde{X}), T^2(\tilde{X}), \text{CH}^1(\tilde{X})_{\text{hom}}$ and $\text{CH}^0(\tilde{X})$.

**Proposition 5.4.** Let $\pi : \tilde{X} \to X$ be a birational map between smooth projective varieties. Then $\pi_* \pi^*$ acts as the identity on $\text{CH}_0(\tilde{X}, \text{Griff}(\tilde{X}), \text{Griff}^2(\tilde{X}), T^2(\tilde{X}), \text{CH}^1(\tilde{X})_{\text{hom and \text{CH}^0(\tilde{X})}}$.

**Proof.** By resolution of singularities, there are morphisms $f : Y \to \tilde{X}$ and $g : Y \to X$, which are composite of smooth blow-ups, such that $g = \pi \circ f$. The groups considered in the proposition behave functorially with respect to the action of correspondences. Therefore it is enough to prove the proposition when $\pi$ is a smooth blow-up $\tilde{X}_Y \to X$ as above. First note that $\Psi_i := \mathfrak{h}(\tilde{X}_Y) \to \mathfrak{h}(Y)(i)$ acts as zero on the groups considered in the proposition when $1 \leq i \leq r - 1$, for dimension reasons. Having in mind that the first column and the
first row of \( N \) are zero, and expanding \( \Phi \circ \Theta \), we see that \( \Phi \circ \Theta \) acts like \( \pi^* \pi_* \) on the groups of the proposition. By Theorem 5.3, the correspondence \( \Phi \circ \Theta \) acts as the identity on \( \text{CH}_l(\tilde{X}) \). Thus \( \pi^* \pi_* \) acts as the identity on the groups of the proposition.

\( \square \)

**Remark 5.5.** As a consequence of Theorem 5.3, we obtain an explicit Chow–Künneth decomposition for a smooth blow-up \( \tilde{X}_Y \) in terms of Chow–Künneth decompositions of \( X \) and \( Y \). Precisely, assume that \( X \) and \( Y \) are endowed with Chow–Künneth decompositions \( \{ \pi^i_X, 0 \leq i \leq 2 \dim X \} \) and \( \{ \pi^i_Y, 0 \leq i \leq 2 \dim Y \} \), respectively. Let us define

\[
\pi^i_{\tilde{X}_Y} := \Phi \circ (\pi^i_X \oplus \bigoplus_{j=1}^{r-1} \pi^{i-2j}_Y) \circ \Theta \in \text{End}(\mathfrak{h}(\tilde{X}_Y)).
\]

Then \( \{ \pi^i_{\tilde{X}_Y}, 0 \leq i \leq 2 \dim \tilde{X}_Y \} \) is a Chow–Künneth decomposition of \( \tilde{X}_Y \).

### 6. Chow groups of varieties fibred by varieties with small Chow groups

In this section, we consider a projective surjective morphism \( f : X \to B \) onto a quasi-projective variety \( B \). We are interested in obtaining information on the niveau of the Chow groups of \( X \), under the assumption that the Chow groups of the fibres of \( f \) over generic points of closed subvarieties of \( B \) are either trivial (= \( \mathbb{Q} \)) or finitely generated. Contrary to Section 3, we do not assume that \( f \) is flat. Let \( l \) be a non-negative integer and let \( d_X \) and \( d_B \) be the dimensions of \( X \) and \( B \), respectively. In §6.1, we assume that \( \text{CH}_{l-i}(X_{\eta_D}) = \mathbb{Q} \) for all \( 0 < i \leq d_B \) and all irreducible subvarieties \( D \subset B \) of dimension \( i \), and we deduce in Lemma 6.1 by a localisation sequence argument that \( \text{CH}_l(X) \) is spanned by \( \text{CH}_l(X_b) \) for all closed points \( b \) in \( B \) and by \( \text{CH}_l(H) \), where \( H \hookrightarrow X \) is a linear section of \( X \) of dimension \( d_X + l \). We then move on to study the subspace of \( \text{CH}_l(X) \) spanned by \( \text{CH}_l(X_b) \) for all closed points \( b \) in \( B \) in the case when \( \text{CH}_l(X_b) = \mathbb{Q} \) in §6.2, and in the case when \( \text{CH}_l(X_b) \) finitely generated in §6.3. The results of §§6.1 & 6.2 are combined into Proposition 6.5 while the results of §§6.1 & 6.3 are combined into Propositions 6.8 and 6.9 in §6.4. Finally, in §6.5, we use Lemma 2.1 to give statements that only involve the Chow groups of closed fibres when \( f \) is defined over the complex numbers; see Theorems 6.10, 6.12 and 6.13.

#### 6.1. Some general statements.

**Lemma 6.1.** Let \( f : X \to B \) be a projective surjective morphism onto a quasi-projective variety \( B \) and let \( H \hookrightarrow X \) be a linear section of dimension \( \geq l + d_B \). Assume that \( \text{CH}_{l-i}(X_{\eta_D}) = \mathbb{Q} \) for all \( 0 < i \leq d_B \) and all irreducible subvarieties \( D \subset B \) of dimension \( i \). Then the natural map \( \bigoplus_{b \in B} \text{CH}_l(X_b) \oplus \text{CH}_l(H) \to \text{CH}_l(X) \) is surjective.
Lemma

Here is an improvement of Lemma 6.1:

Let $X \to B$ be a projective surjective morphism onto a quasi-projective variety $B$ and let $H \hookrightarrow X$ be a linear section of dimension $\geq l + d_B$. Assume that:

- $\text{CH}_{l-i}(X_{\eta_B}) = 0$ for all $i$ such that $0 < i < d_B$ and all irreducible subvarieties $D_i \subset B$ of dimension $i$.
- $\text{CH}_{d_B}(X_{\eta_B})$ is finitely generated.

Then there exist finitely many closed subschemes $Z_j$ of $X$ of dimension $l$ such that the natural map $\bigoplus_j \text{CH}_l(Z_j) \oplus \bigoplus_{b \in B} \text{CH}_B(X_b) \oplus \text{CH}_l(H) \to \text{CH}_l(X)$ is surjective.

Proof. We prove the proposition by induction on $d_B$. If $d_B = 0$ then the statement is obvious. Let us thus consider a morphism $f : X \to B$ and a linear section $\iota : H \hookrightarrow X$ as in the statement of the proposition with $d_B > 0$.

By Lemma 6.1, $f$ restricted to $H$ is surjective. We have the localisation exact sequence

$$\bigoplus_{D \in B^1} \text{CH}_l(X_D) \to \text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B}) \to 0.$$ 

Here, $B^1$ denotes the set of codimension-one closed irreducible subschemes of $B$. For any irreducible codimension-one subvariety $D \subset B$, the restriction of $\iota$ to $D \to B$ defines a linear section $\iota_D : H_D \hookrightarrow X_D$ of dimension $\geq l + d_B - 1$ of $X_D$. The restriction of $f : X \to B$ to $D \to B$ defines a surjective morphism $X_D \to D$ which together with the linear section $\iota_D$ satisfies the assumptions of the proposition. Therefore, by the induction hypothesis applied to $X_D \to D$, the map

$$\bigoplus_{D \in B^1} \text{CH}_l(X_D) \to \text{CH}_l(X_D)$$

is surjective. This yields an exact sequence

$$\bigoplus_{b \in B} \text{CH}_l(X_b) \oplus \bigoplus_{D \in B^1} \text{CH}_l(H_D) \to \text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B}) \to 0.$$ 

Since each of the proper inclusion maps $H_D \to X$ factors through $\iota : H \to X$, we see that the map $\bigoplus_{D \in B^1} \text{CH}_l(H_D) \to \text{CH}_l(X)$ factors through $\iota_* : \text{CH}_l(H) \to \text{CH}_l(X)$. In order to conclude, it is enough to prove that the composite map

$$\text{CH}_l(H) \to \text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B})$$

is surjective. If $l < d_B$, then this is obvious. Let us then assume that $l \geq d_B$. Let $Y$ be an irreducible subvariety of $H$ of dimension $l$ such that the composite $Y \hookrightarrow X \to B$ is dominant. Because $\text{CH}_{l-d_B}(X_{\eta_B}) = 0$ it is enough to see that the class of $Y$ in $\text{CH}_l(X)$ maps to a non-zero element in $\text{CH}_{l-d_B}(X_{\eta_B})$. But, as in the proof of Proposition 6.1, $[Y]$ maps to $[Y_{\eta_B}] \neq 0 \in \text{CH}_{l-d_B}(X_{\eta_B})$. □
Lemma 6.2. Varieties fibred by varieties with Chow groups generated by a linear section.

As in the proof of Lemma 6.1 we have the localisation exact sequence

$$\bigoplus_{D \in B^1} \text{CH}_l(X_D) \to \text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B}) \to 0.$$  

Each of the morphisms $X_D \to D$ satisfies the assumptions of Lemma 6.1 and by the same arguments as in the proof of Lemma 6.1 we get that the image of the map $\bigoplus_{b \in B} \text{CH}_l(X_b) \oplus \text{CH}(H) \to \text{CH}_l(X)$ contains the image of the map $\bigoplus_{D \in B^1} \text{CH}_l(X_D) \to \text{CH}_l(X)$. Let now $Z_j$ be finitely many closed subschemes of $X_{\eta_B}$ whose classes $[Z_j] \in \text{CH}_{l-d_B}(X_{\eta_B})$ generate $\text{CH}_{l-d_B}(X_{\eta_B})$. By surjectivity of the map $\text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B})$ there are cycles $\alpha_j \in \text{CH}_l(X)$ that map to $[Z_j]$. If $Z_j$ is the support in $X$ of any representative of $\alpha_j$, we then have a surjective map $\bigoplus_j \text{CH}_l(Z_j) \to \text{CH}_l(X) \to \text{CH}_{l-d_B}(X_{\eta_B})$. It is then clear that the map $\bigoplus_j \text{CH}_l(Z_j) \oplus \bigoplus_{b \in B} \text{CH}_l(X_b) \oplus \text{CH}(H) \to \text{CH}_l(X)$ is surjective. \hfill \qed

6.2. Varieties fibred by varieties with Chow groups generated by a linear section.

**Lemma 6.3.** Let $f : X \to B$ be a projective surjective morphism onto a quasi-projective variety $B$. Assume that $\text{CH}_l(X_b) = \mathbb{Q}$ for all closed points $b \in B$. Then, if $H \hookrightarrow X$ is a linear section of dimension $\geq l + d_B$, we have

$$\text{Im} \left( \bigoplus_{b \in B} \text{CH}_l(X_b) \to \text{CH}_l(X) \right) \subseteq \text{Im} \left( \text{CH}_l(H) \to \text{CH}_l(X) \right).$$

**Proof.** Let $b$ be a closed point of $B$ and fix $H \hookrightarrow X$ a linear section of dimension $\geq l + d_B$. The morphism $f$ restricted to $H$ is surjective; see Lemma 6.1. Let $Z_l$ be an irreducible closed subscheme of $X$ of dimension $l$ which is supported on $X_b$. Since $f|_H : H \to B$ is a dominant projective morphism, its fibre $H_b$ over $b$ is non-empty and has dimension $\geq l$. By assumption $\text{CH}_l(X_b) = \mathbb{Q}$, so that a rational multiple of $[Z_l]$ is rationally equivalent to an irreducible closed subscheme of $H_b$ of dimension $l$. Therefore $[Z_l] \in \text{CH}_l(X_b)$ belongs to the image of the natural map $\text{CH}_l(H_b) \to \text{CH}_l(X_b)$. Thus the image of $\text{CH}_l(X_b) \to \text{CH}_l(X)$ is contained in the image of $\text{CH}_l(H) \to \text{CH}_l(X)$. \hfill \qed

**Remark 6.4.** It is interesting to decide whether or not it is possible to parametrise such $l$-cycles by a variety of dimension $d_B$; see Proposition 6.7.

**Proposition 6.5.** Let $f : X \to B$ be a projective surjective morphism onto a quasi-projective variety $B$. Assume that $\text{CH}_{l-i}(X_{D_i}) = \mathbb{Q}$ for all $0 \leq i \leq d_B$ and all irreducible subvarieties $D_i \subset B$ of dimension $i$. Then, if $H \hookrightarrow X$ is a linear section of dimension $\geq l + d_B$, the pushforward map $\text{CH}_l(H) \to \text{CH}_l(X)$ is surjective. In particular, $\text{CH}_l(X)$ has niveau $\leq d_B$.

**Proof.** This is a combination of Lemma 6.1 and Lemma 6.3. \hfill \qed
6.3. An argument involving relative Hilbert schemes. Let \( f : X \to B \) be a generically smooth, projective morphism defined over the field of complex numbers \( \mathbb{C} \) onto a smooth quasi-projective variety \( B \). Let \( B^\circ \subseteq B \) be the smooth locus of \( f \) and let \( f^\circ : X^\circ \to B^\circ \) be the pullback of \( f : X \to B \) along the open inclusion \( B^\circ \hookrightarrow B \) so that we have a cartesian square

\[
\begin{array}{ccc}
X^\circ & \xrightarrow{f^\circ} & X \\
\downarrow & & \downarrow f \\
B^\circ & \xrightarrow{f} & B
\end{array}
\]

We assume that there is a non-negative integer \( l \) such that for all closed points \( b \in B^\circ(\mathbb{C}) \) the cycle class map \( \text{CH}_l(X_b) \to H_{2l}(X_b) \) is an isomorphism.

Let \( \pi_d : \text{Hilb}^d(X/B) \to B \) be the relative Hilbert scheme whose fibres over the points \( b \) in \( B \) parametrise the closed subschemes of \( X_b \) of dimension \( l \) and degree \( d \), and let \( p_d : \mathcal{C}^d \to \text{Hilb}^d(X/B) \) be the universal family over \( \text{Hilb}^d(X/B) \); see \([14, \text{Theorem 1.4}]\). We have the following commutative diagram, where all the morphisms involved are proper:

\[
\begin{array}{ccc}
\mathcal{C}^d & \xrightarrow{q_d} & X \\
\downarrow p_d & & \downarrow \pi_d \\
\text{Hilb}^d(X/B) & \xrightarrow{\pi_d} & B
\end{array}
\]

We then consider the disjoint unions \( \text{Hilb}^d(X/B) := \bigsqcup_{d \geq 0} \text{Hilb}^d(X/B) \) and \( \text{C}_l := \bigsqcup_{d \geq 0} \mathcal{C}^d \), and denote \( \pi : \text{Hilb}(X/B) \to B \), \( p : \text{C}_l \to \text{Hilb}(X/B) \) and \( q : \text{C}_l \to X \) the corresponding maps.

Let us then denote

\[ \text{Irr}^l(X/B) := \{ \mathcal{H} : \mathcal{H} \text{ is an irreducible component of } \text{Hilb}^d(X/B) \text{ for some } d \} \]

For a subset \( \mathcal{E} \subseteq \text{Irr}^l(X/B) \), we define the following closed subscheme of \( B^\circ \):

\[ Z_\mathcal{E} := B^\circ \cap \bigcap_{\mathcal{H} \in \mathcal{E}} \pi(\mathcal{H}). \]

We say that a finite subset \( \mathcal{E} \) of \( \text{Irr}^l(X/B) \) is spanning at a point \( t \in Z_\mathcal{E}(\mathbb{C}) \) if \( H_{2l}(X_t) \) is spanned by the set \( \{ cl(q_*[p^{-1}(u)]) : u \in \mathcal{H}, \mathcal{H} \in \mathcal{E}, \pi(u) = t \} \).

Note that, given \( \mathcal{H} \in \text{Irr}^l(X/B) \) and \( u, u' \in \mathcal{H} \) such that \( \pi(u) = \pi(u') = t \), \( cl(q_*[p^{-1}(u)]) = cl(q_*[p^{-1}(u')]) \in H_{2l}(X_t) \) if \( u \) and \( u' \) belong to the same connected component in \( \pi^{-1}(t) \).

Claim. Let \( \mathcal{E} \) be a finite subset of \( \text{Irr}^l(X/B) \) that is spanning at a closed point \( t \in Z_\mathcal{E}(\mathbb{C}) \). Then, for all points \( s \in Z_\mathcal{E}(\mathbb{C}) \) belonging to an irreducible component of \( Z_\mathcal{E} \) that contains \( t \), \( H_{2l}(X_s) \) is spanned by the set \( \{ cl(q_*[p^{-1}(v)]) : v \in \mathcal{H}, \mathcal{H} \in \mathcal{E}, \pi(v) = s \} \).
Indeed, consider any finite subset $E$ of $\text{Irr}_l(X/B)$. The local system of $\mathbb{Q}$-vector spaces $R_{2l}(f_0)_*\mathbb{Q}$ on $B^*$ restricts to a local system $(R_{2l}(f_0)_*\mathbb{Q})|_{Z_E}$ on $Z_E$. If $t$ is a complex point on $Z_E$, let $r_t$ be the rank of the subspace of $H_{2l}(X_t)$ spanned by $\{cl(q_*[p^{-1}(u)]) : u \in \mathcal{H}, \mathcal{H} \in E, \pi(u) = t\}$. If we see this latter set as a set of sections at $t$ of the local system $R_{2l}(f_0)_*\mathbb{Q}$, then these sections extend locally to constant sections of the local system $(R_{2l}(f_0)_*\mathbb{Q})|_{Z_E}$ on $Z_E$. This shows that the rank $r_t$ is locally constant. If $E$ is spanning at the point $t \in Z_E(C)$, then $r_t$ is maximal, equal to $\dim_{\mathbb{Q}} H_{2l}(X_t)$. The subset of $Z_E(C)$ consisting of points $s$ in $Z_E(C)$ for which $r_s = \dim_{\mathbb{Q}} H_{2l}(X_s)$ is therefore both open and closed in $Z_E$. It contains then the irreducible components of $Z_E$ that contain $t$.

**Lemma 6.6.** There exists a finite subset $E$ of $\text{Irr}_l(X/B)$ such that $B^o = Z_E$ and such that $E$ is spanning at every point $t \in B^o(C)$.

**Proof.** By working component-wise, we may assume that $B$ is irreducible. By assumption on $f : X \to B$, $H_{2l}(X_t)$ is spanned by algebraic cycles on $X_t$ for all points $t \in B^o(C)$. Thus, for all points $t \in B^o(C)$, there is a finite subset $E_t$ of $\text{Irr}_l(X/B)$ that is spanning at $t$. For each point $t$, choose an irreducible component $Y_{E_t}$ of $Z_{E_t}$ that contains $t$. According to the claim above, $E_t$ is spanning at every point $s \in Y_{E_t}(C)$. Now, we have $B^o(C) = \prod_{t \in B^o(C)} Y_{E_t}(C)$. Since there are only countably many finite subsets of $\text{Irr}_l(X/B)$ and since $Z_E$ has only finitely many irreducible components, we see that the latter union is in fact a countable union. This yields that $B^o = Y_{E_t}$ for some finite subset $E$ of $\text{Irr}_l(X/B)$ that is spanning at every point in $Y_{E_t}(C)$. We then conclude that $B^o = Z_E$ and that $E$ is spanning at every point $t \in Y_{E_t}(C) = B^o(C)$. 

**Proposition 6.7.** Let $f : X \to B$ be a generically smooth and projective morphism defined over $C$ onto a smooth quasi-projective variety $B$. Let $B^o \subseteq B$ be the smooth locus of $f$. Assume that there is an integer $l \leq d_X - d_B$ such that for all closed points $b \in B^o(C)$ the cycle class map $CH_l(X_b) \to H_{2l}(X_b)$ is an isomorphism. Then $\text{Im} \left( \bigoplus_{b \in B^o} CH_l(X_b) \to CH_l(X) \right)$ is supported on a closed subvariety of $X$ of dimension $d_B + l$.

If moreover $X$ is smooth, then there exist a smooth quasi-projective variety $\tilde{B}$ of dimension $d_B$ and a correspondence $\Gamma \in CH_{d_B+1}(\tilde{B} \times X)$ such that $\Gamma_* : CH_0(\tilde{B}) \to CH_l(X)$ is well-defined and

$$\text{Im} \left( \bigoplus_{b \in B^o} CH_l(X_b) \to CH_l(X) \right) \subseteq \text{Im} \left( \Gamma_* : CH_0(\tilde{B}) \to CH_l(X) \right).$$

**Proof.** By Lemma 6.6, there exists a finite set $E$ of irreducible components of $\text{Hilb}_l(X/B)$ such that $B^o = Z_E$ and such that for all points $t \in B^o(C)$ the set $\{cl(q_*[p^{-1}(u)]) : u \in \mathcal{H}, \mathcal{H} \in E, \pi(u) = t\}$ spans $H_{2l}(X_t)$. Denote $\mathcal{H}_t$ the irreducible components of $\text{Hilb}_l(X/B)$ that belong to $E$ and let $\tilde{H}_i \to H_i$ be resolutions thereof. For all $i$, pick a smooth linear section $\tilde{B}_i \to \tilde{H}_i$ of dimension $d_B$. Lemma 6.6 shows that $r_i : \tilde{B}_i \to \tilde{H}_i \to H_i \to B$ is surjective and a refinement of its proof shows that, for all points $b \in B(C)$, $r_i^{-1}(b)$ contains a point in every connected component of $\tilde{H}_{i,b}$. Consider then $p_i : (\mathcal{C}_i)|_{\tilde{B}_i} \to \tilde{B}_i$.
the pullback of the universal family \( p : C \to \text{Hilb}_l(X/B) \) along \( \tilde{B}_i \hookrightarrow \tilde{H}_i \to H_i \hookrightarrow \text{Hilb}_l(X/B) \). For each \( i \), we have the following picture

\[
\begin{array}{ccc}
(C_l)|_{\tilde{B}_i} & \xrightarrow{q_i} & X \\
p_i & \downarrow & \\
\tilde{B}_i & & \\
\end{array}
\]

and we have

\[
\text{Im} \left( \bigoplus_{b \in B^p} \text{CH}_l(X_b) \to \text{CH}_l(X) \right) \subseteq \sum_i \text{Im} \left( (q_i)_* : \text{CH}_l((C_l)|_{\tilde{B}_i}) \to \text{CH}_l(X) \right)
\]

so that the group on the left-hand side is supported on the union of the scheme-theoretic images of the morphisms \( q_i \).

If \( X \) is smooth, we define \( \Gamma_i \in \text{CH}_{d_{l+1}}(\tilde{B}_i \times X) \) to be the class of the image of \((C_l)|_{\tilde{B}_i}\) inside \( \tilde{B}_i \times X \). Because \( q : C^d_l \to X \) is proper for all \( d \geq 0 \), \( \Gamma_i \) has a representative which is proper over \( X \). It is therefore possible [7, Remark 16.1] to define maps \((\Gamma_i)_* : \text{CH}_0(\tilde{B}_i) \to \text{CH}_l(X)\) for all \( i \). In fact, we have \((\Gamma_i)_* = (q_i)_* p_i^*\). Finally, we define \( \tilde{B} \) to be the disjoint union of the \( \tilde{B}_i \)’s and \( \Gamma \in \text{CH}_{d_{l+1}}(\tilde{B} \times X) \) to be the class of the disjoint union of the correspondences \( \Gamma_i \).

\[
\square
\]

### 6.4. Complex varieties fibred by varieties with small Chow groups.

From now on, the base field \( k \) is assumed to be the field of complex numbers \( \mathbb{C} \).

**Proposition 6.8.** Let \( f : X \to C \) be a generically smooth, projective morphism defined over \( C \) to a smooth curve. Assume that

- \( \text{CH}_l(X_c) \) is finitely generated for all closed points \( c \in C \),
- \( \text{CH}_l(X_c) \to \text{H}_2(X_c) \) is an isomorphism for a general closed point \( c \in C \),
- \( \text{CH}_{l-1}(X_\eta) \) is finitely generated, where \( \eta \) is the generic point of \( C \).

Then \( \text{CH}_l(X) \) has niveau \( \leq 1 \).

**Proof.** We have the localisation exact sequence

\[
\bigoplus_{c \in C} \text{CH}_l(X_c) \longrightarrow \text{CH}_l(X) \longrightarrow \text{CH}_{l-1}(X_\eta) \longrightarrow 0.
\]

Let \( Z_1, \ldots, Z_n \) be irreducible closed subschemes of \( X_\eta \) of dimension \( l - 1 \) that span \( \text{CH}_{l-1}(X_\eta) \) and let \( Z_1, \ldots, Z_n \) be closed subschemes of \( X \) of dimension \( l \) that restrict to \( Z_1, \ldots, Z_n \) in \( X_\eta \). Then by flat pullback the class of \( Z_j \) in \( \text{CH}_l(X) \) maps to the class of \( Z_j \) in \( \text{CH}_{l-1}(X_\eta) \) so that the composite map \( \bigoplus_{j=1}^n \text{CH}_l(Z_j) \to \text{CH}_l(X) \to \text{CH}_{l-1}(X_\eta) \) is surjective.

Let \( U \subseteq C \) be a Zariski-open subset of \( C \) such that for all closed points \( c \in U \) the cycle class map \( \text{CH}_l(X_c) \to \text{H}_2(X_c) \) is an isomorphism. Up to shrinking \( U \), we may assume that \( f|_U : X|_U \to U \) is smooth. We may then apply
Proposition 6.7 to get a closed subscheme \( \iota : D \hookrightarrow X \) of dimension \( l + 1 \) such that \( \iota_* \CH_l(D) \supseteq \Im \left( \bigoplus_{c \in U} \CH_l(X_c) \to \CH_l(X) \right) \).

As such, we have a surjective map

\[
\bigoplus_{j=1}^n \CH_l(Z_j) \oplus \bigoplus_{c \in C \setminus U} \CH_l(X_c) \oplus \CH_l(D) \to \CH_l(X)
\]

and it is straightforward to conclude.

The next proposition is a generalisation of Proposition 6.8 to the case when the base variety \( B \) has dimension greater than 1.

**Proposition 6.9.** Let \( f : X \to B \) be a generically smooth, projective morphism defined over \( C \) to a smooth quasi-projective variety \( B \). Assume that the singular locus of \( f \) in \( B \) is finite and let \( U \) be the maximal Zariski-open subset of \( B \) over which \( f \) is smooth. Assume also that

- \( \CH_l(X_b) \) is finitely generated for all closed points \( b \in B \),
- \( \CH_l(X_b) \to H_{2l}(X_b) \) is an isomorphism for all closed points \( b \in U \),
- \( \CH_{l-i}(X_{\eta^{b_i}}) = \mathbb{Q} \) for all \( i \) such that \( 0 < i < d_B \) and all irreducible subvarieties \( D_i \subset B \) of dimension \( i \).
- \( \CH_{l-d_B}(X_{\eta^{b_i}}) \) is finitely generated.

Then \( \CH_l(X) \) has niveau \( \leq d_B \).

**Proof.** Let \( H \hookrightarrow X \) be a linear section of dimension \( \geq l + d_B \). The restriction of \( f \) to \( H \) is surjective. Thanks to Lemma 6.2 there are finitely many closed subschemes \( Z_j \) of \( X \) of dimension \( l \) such that the natural map \( \bigoplus_j \CH_l(Z_j) \oplus \bigoplus_{b \in B} \CH_l(X_b) \oplus \CH_l(H) \to \CH_l(X) \) is surjective. By Proposition 6.7 there exists a closed subscheme \( \iota : \tilde{B} \hookrightarrow X \) of dimension \( d_B + l \) such that the image of the map \( \bigoplus_{b \in U} \CH_l(X_b) \to \CH_l(X) \) is contained in the image of the map \( \iota_* : \CH_l(\tilde{B}) \to \CH_l(X) \). Therefore the map

\[
\bigoplus_j \CH_l(Z_j) \oplus \CH_l(\tilde{B}) \oplus \bigoplus_{b \in B \setminus U} \CH_l(X_b) \oplus \CH_l(H) \to \CH_l(X)
\]

is surjective. It is then straightforward to conclude.

### 6.5. The Main Results

The field of complex numbers is a universal domain and in view of Section 2 we restate some of the results above in a more comprehensive way.

First, we deduce from Proposition 6.5 the following.

**Theorem 6.10.** Let \( f : X \to B \) be a complex projective surjective morphism onto a quasi-projective variety \( B \). Assume that \( \CH_i(X_b) = \mathbb{Q} \) for all \( i \leq l \) and all closed point \( b \in B \). Then \( \CH_l(X) \) has niveau \( \leq d_B \) for all \( i \leq l \).

**Proof.** By Lemma 2.2 \( \CH_i(X_{\eta^{b_i}}) = \mathbb{Q} \) for all \( i \leq l \) and all irreducible subvarieties \( D \) of \( X \). Proposition 6.5 implies that \( \CH_l(X) \) has niveau \( \leq d_B \).
7.1. Varieties with small Chow groups. In this section, we review the known results about varieties with Chow groups having small niveau; see Definition 2.3. Varieties are defined over an algebraically closed field \(k\) of characteristic zero and \(\Omega\) denotes a universal domain over \(k\). In that case, Grothendieck’s
standard conjectures for a smooth projective variety $X$ over $k$ reduce to the Lefschetz standard conjecture for $X$; see Kleiman [13].

**Theorem 7.1.** Let $X$ be a smooth projective variety of dimension $d$. Assume that the Chow groups $\text{CH}_*(X_{\Omega}), \ldots, \text{CH}_i(X_{\Omega})$ have niveau $\leq n$.

- If $n = 3$ and $l = \left\lceil \frac{d-2}{2} \right\rceil$, then $X$ satisfies the Hodge conjecture.
- If $n = 2$ and $l = \left\lceil \frac{d-2}{2} \right\rceil$, then $X$ satisfies the Lefschetz standard conjecture.
- If $n = 1$ and $l = \left\lceil \frac{d-2}{2} \right\rceil$, then $X$ has a Murre decomposition.
- If $n = 1$ and $l = \left\lfloor \frac{d-2}{2} \right\rfloor$, then $X$ is Kimura finite-dimensional.

**Proof.** The fourth item is proved in [26]. It is also proved there that if $X$ is as in the third item, then $X$ has a Chow–Künneth decomposition. That such a decomposition satisfies Murre’s conjectures (B), (C) and (D) is proved in [25, §4.4.2]. The first item is proved in [15]. We couldn’t find a reference for the proof of the second item so we include a proof here.

Since it is enough to prove the conclusion of the theorem for $X_{\Omega}$, we may assume that $X$ is defined over $\Omega$. Laterveer used the assumptions on the niveau of the Chow groups to show [15, 1.7] that the diagonal $\Delta_X$ admits a decomposition as follows: there exist closed and reduced subschemes $V_j, W^j \subset X$ with $\dim V_j \leq j + 2$ and $\dim W^j \leq n - j$, there exist correspondences $\Gamma_j \in \text{CH}_n(X \times X)$ for $0 \leq j \leq \left\lfloor \frac{d}{2} \right\rfloor$ and $\Gamma' \in \text{CH}_n(X \times X)$ such that each $\Gamma_j$ is in the image of the pushforward map $\text{CH}_n(V_j \times W^j)$, $\Gamma'$ is in the image of the pushforward map $\text{CH}_n(X \times W^j)$, and

$$\Delta_X = \Gamma_0 + \ldots + \Gamma_{\left\lfloor \frac{d}{2} \right\rfloor} + \Gamma'.$$

Given $j$ such that $0 \leq j \leq \left\lceil \frac{d+2}{2} \right\rceil$, let $\tilde{V}_j$ and $\tilde{W}^j$ denote desingularisations of $V_j$ and $W^j$ respectively. The action of $\Gamma_j$ on $H^k(X)$ then factors through $H^k(\tilde{V}_j)$ and through $H_{2n-k}(W^j)$. On the one hand, we have $H_{2n-k}(W^j) = H^{k-2j}(W^j)$ and hence if $k \leq 2j + 1$ then the action of $\Gamma_j$ on $H^k(X)$ factors through the $H^k$ or the $H^0$ of a smooth projective variety. Since the Lefschetz standard conjecture is true in degrees $\leq 1$, it follows that the action of $\Gamma_j$ on $H^k(X)$ factors through the $H_0$ or the $H_1$ of a smooth projective variety. On the other hand, we have $H^k(\tilde{V}_j) = H_{4+2j-k}(\tilde{V}_j)$ and hence if $k \geq 2j + 2$ then $\Gamma_j$ factors through the $H_0$, the $H_1$ or the $H_2$ of a smooth projective variety. Concerning the action of $\Gamma'$ on $H^k(X)$, it factors through $H_{2n-k}(W^{\left\lfloor \frac{d-2}{2} \right\rfloor})$ which vanishes for dimension reasons if $k < n$ when $n$ is odd and if $k < n - 1$ when $n$ is even. When $n$ is even and $k = n - 1$, the action of $\Gamma'$ on $H^k(X)$ factors through the $H_1$ of a curve. Indeed this follows from a combination of the fact that it factors through $H_{n+1}(W^{\left\lfloor \frac{d+1}{2} \right\rfloor}) = H^1(W^{\left\lfloor \frac{d+1}{2} \right\rfloor})$ and of the validity of the Lefschetz standard conjecture in degree 1.

By the Lefschetz hyperplane theorem, we get that for $k < n$ the cohomology groups $H^k(X)$ are generated algebraically (that is through the action of correspondences) by the $H_0$ of points, the $H_1$ of curves and the $H_2$ of surfaces. We may then conclude with [25, Proposition 3.19].
7.2. Varieties fibred by low-degree complete intersections. As explained by Esnault–Levine–Viehweg in the introduction of [6], it is expected from general conjectures on algebraic cycles, that if \( Y \subset \mathbb{P}^n_k \) is a complete intersection of multidegree \( d_1 \geq \ldots \geq d_r \geq 2 \), then \( \text{CH}_l(Y) = \mathbb{Q} \) for all \( l < \lfloor \frac{n - \sum d_i}{2} \rfloor \), see also Paranjape [22] and Schoen [23]. If there is no proof of the above for the moment, the following theorem however was proved.

**Theorem 7.2** (Esnault–Levine–Viehweg [6]). Let \( Y \subset \mathbb{P}^n_k \) be a complete intersection of multidegree \( d_1 \geq \ldots \geq d_r \geq 2 \).

- If either \( d_1 \geq 3 \) or \( r \geq l + 1 \), assume that \( \sum_{i=1}^r \left( \frac{l + d_i}{l + 1} \right) \leq n \).
- If \( d_1 = \ldots = d_r = 2 \) and \( r \leq l \), assume that \( \sum_{i=1}^r \left( \frac{l + d_i}{l + 1} \right) = r(l + 2) \leq n - l + r - 1 \).

Then \( \text{CH}_l(Y) = \mathbb{Q} \) for all \( 0 \leq l' \leq l \). \( \square \)

Let us consider \( f : X \to B \) a dominant morphism between smooth projective complex varieties whose closed fibres are complete intersections. Theorem 6.10, together with Theorem 7.2, shows that the niveau of the first Chow groups of \( X \) have niveau \( \leq \dim B \). When \( X \) is fibred by very low-degree complete intersections, we can thus expect \( X \) to satisfy the assumptions of Theorem 7.1.

In the remainder of this paragraph, we inspect various such cases.

#### 7.2.1. Varieties fibred by quadric hypersurfaces

Let \( Q \subset \mathbb{P}^n \) be a quadric hypersurface. Then \( \text{CH}_l(Q) = \mathbb{Q} \) for all \( l < \frac{\dim Q}{2} \).

**Proposition 7.3.** Let \( f : X \to B \) be a dominant morphism between smooth projective complex varieties whose closed fibres are quadric hypersurfaces.

- If \( \dim B \leq 1 \), then \( X \) is Kimura finite-dimensional and satisfies Murre’s conjectures.
- If \( \dim B \leq 2 \), then \( X \) satisfies Grothendieck’s standard conjectures.
- If \( \dim B \leq 3 \), then \( X \) satisfies the Hodge conjecture.

**Proof.** The fibres of \( f \) have dimension \( \geq \dim X - \dim B \), so that \( X \) satisfies the assumptions of Theorem 6.10 with \( l = \lfloor \frac{\dim X - \dim B - 1}{2} \rfloor \). Thus the Chow groups \( \text{CH}_0(X), \text{CH}_1(X), \ldots, \text{CH}_\lfloor \frac{\dim X - \dim B - 1}{2} \rfloor(X) \) have niveau \( \leq \dim B \). We can therefore conclude by Theorem 7.1. \( \square \)

#### 7.2.2. Varieties fibred by cubic hypersurfaces

Let \( X \subset \mathbb{P}^n \) be a cubic hypersurface. Then

- \( \text{CH}_0(X) = \mathbb{Q} \) for \( \dim X \geq 2 \).
- \( \text{CH}_1(X) = \mathbb{Q} \) for \( \dim X \geq 5 \).
- \( \text{CH}_2(X) = \mathbb{Q} \) for \( \dim X \geq 8 \).

Note that Theorem 7.2 only gives \( \text{CH}_2(X) = \mathbb{Q} \) for \( \dim X \geq 9 \). The bound on the dimension of \( X \) was improved to \( \dim X = 8 \) by Otwinowska [21].

**Proposition 7.4.** Let \( f : X \to B \) be a dominant morphism between smooth projective complex varieties whose closed fibres are cubic hypersurfaces.
If \( \dim X = 6 \) and \( \dim B = 1 \), then \( X \) satisfies Grothendieck’s standard conjectures and has a Murre decomposition.

If \( \dim X = 7 \) and \( \dim B \leq 2 \), then \( X \) satisfies the Hodge conjecture.

If \( \dim X = 9 \) and \( \dim B \leq 1 \), then \( X \) satisfies the Hodge conjecture.

Proof. We use Theorem 6.10 as in the proof of Proposition 7.3. In the first case, we get that \( \text{CH}_0(X) \) and \( \text{CH}_1(X) \) have niveau \( \leq 1 \). In the second case we get that \( \text{CH}_0(X) \) and \( \text{CH}_1(X) \) have niveau \( \leq 2 \) and in the third case we get that \( \text{CH}_0(X) \), \( \text{CH}_1(X) \) and \( \text{CH}_2(X) \) have niveau \( \leq 1 \). We can then conclude in all three cases by Theorem 7.1.

7.2.3. Varieties fibred by complete intersections of bidegree \((2,2)\). Let \( X \subset \mathbb{P}^n \) be the complete intersection of two quadrics. By Theorem 7.2, \( \text{CH}_0(X) = \mathbb{Q} \); and if \( \dim X \geq 4 \), then \( \text{CH}_1(X) = \mathbb{Q} \).

Proposition 7.5. Let \( f : X \to B \) be a dominant morphism between smooth projective complex varieties whose closed fibres are complete intersections of bidegree \((2,2)\).

- If \( \dim B \leq 1 \) and \( \dim X \leq 5 \), then \( X \) is Kimura finite-dimensional.
- If \( \dim B \leq 1 \) and \( \dim X \leq 6 \), then \( X \) satisfies Murre’s conjectures.
- If \( \dim B \leq 2 \) and \( \dim X \leq 6 \), then \( X \) satisfies Grothendieck’s standard conjectures.
- If \( \dim B \leq 3 \) and \( \dim X \leq 7 \), then \( X \) satisfies the Hodge conjecture.

Proof. The variety \( X \) satisfies the assumptions of Theorem 6.10 with \( l = 1 \) for \( \dim X - \dim B \geq 4 \) and with \( l = 0 \) in any case. Thus the Chow group \( \text{CH}_0(X) \) has niveau \( \leq d_B \) and \( \text{CH}_1(X) \) has niveau \( \leq d_B \) for \( \dim X - \dim B \geq 4 \). We can therefore conclude by Theorem 7.1.

7.2.4. Varieties fibred by complete intersections of bidegree \((2,3)\). Let \( X \subset \mathbb{P}^n \) be the complete intersection of a quadric and of a cubic. If \( \dim X \geq 6 \), then Hirschowitz and Iyer [9] showed \( \text{CH}_l(X) = \mathbb{Q} \) for \( l \leq 1 \). (The result of Esnault–Levine–Viehweg only says that \( \text{CH}_l(X) = \mathbb{Q} \) for \( l \leq 1 \) when \( \dim X \geq 7 \).)

Proposition 7.6. Let \( f : X \to C \) be a dominant morphism from a smooth projective complex variety \( X \) to a smooth projective complex curve \( C \) whose closed fibres are complete intersections of bidegree \((2,3)\) of dimension 6. Then \( X \) satisfies the Hodge conjecture.

Proof. By Theorem 6.10 we see that the Chow groups \( \text{CH}_0(X) \) and \( \text{CH}_1(X) \) have niveau \( \leq 1 \). We can thus conclude by Theorem 7.1.

7.3. Varieties fibred by cellular varieties. Let \( f : X \to B \) be a complex dominant morphism from a smooth projective variety \( X \) to a smooth projective variety \( B \) whose closed fibres are cellular varieties (not necessarily smooth). In other words, \( X \) is a smooth projective complex variety fibred by cellular varieties over \( B \). For example, if \( \Sigma \subset B \) is the singular locus of \( f \), then \( X \) could be such that \( X|_{B \setminus \Sigma} \) is a rational homogeneous bundle over \( B \setminus \Sigma \) (e.g.
a Grassmann bundle) and the closed fibres of \( f \) over \( \Sigma \) (the degenerate fibres) are toric. That kind of situation is reminiscent of the setting of \([8]\).

**Proposition 7.7.** Let \( f : X \to B \) be a dominant morphism between smooth projective complex varieties whose closed fibres are cellular varieties.

- Assume \( B \) is a curve, then \( X \) is Kimura finite-dimensional and \( X \) satisfies Murre’s conjectures.
- Assume \( \dim B \leq 2 \) and \( \dim X \leq 6 \). If \( f \) is connected and smooth away from finitely many points in \( B \), then \( X \) satisfies Grothendieck’s standard conjectures.
- Assume \( \dim B \leq 3 \) and \( \dim X \leq 7 \). If \( f \) is connected and smooth away from finitely many points in \( B \), then \( X \) satisfies the Hodge conjecture.

**Proof.** The Chow groups of cellular varieties are finitely generated. The first statement thus follows from Theorems 6.12 and 7.1. Let us now focus on the cases when \( \dim B \) is either 2 or 3. It is a consequence of Mumford’s theorem \([17]\) that a connected smooth projective complex variety with finitely generated Chow group of zero-cycles actually has Chow group of zero-cycles generated by a point. Thus the second and third statements follow from Theorem 6.13 with \( l = 1 \), and from Theorem 7.1. \( \square \)

**Acknowledgements**

Thanks to Reza Akhtar and Roy Joshua for asking a question related to the Chow–Künneth decomposition of a smooth blow-up, to Sergey Gorchinskiy for suggesting parts of Proposition 2.4, to Mingmin Shen for a discussion related to the proof of Lemma 6.6, and to Burt Totaro for suggesting the statement of Lemma 2.1. Finally, many thanks to the referee for helpful comments and suggestions.

This work was supported by EPSRC Early Career Fellowship number EP/K005545/1.

**References**


Charles Vial  
DPMMS  
University of Cambridge  
Wilberforce Road  
Cambridge  
CB3 0WB  
UK  
c.vial@dpmms.cam.ac.uk


**Invariants of Upper Motives**

**Olivier Haution**

Received: June 4, 2013
Revised: September 20, 2013
Communicated by Alexander Merkurjev

**Abstract.** Let $H$ be a homology theory for algebraic varieties over a field $k$. To a complete $k$-variety $X$, one naturally attaches an ideal $H_X(k)$ of the coefficient ring $H(k)$. We show that, when $X$ is regular, this ideal depends only on the upper Chow motive of $X$. This generalizes the classical results asserting that this ideal is a birational invariant of smooth varieties for particular choices of $H$, such as the Chow group. When $H$ is the Grothendieck group of coherent sheaves, we obtain a lower bound on the canonical dimension of varieties. When $H$ is the algebraic cobordism, we give a new proof of a theorem of Levine and Morel. Finally we discuss some splitting properties of geometrically unirational field extensions of small transcendence degree.

2010 Mathematics Subject Classification: 14C25
Keywords and Phrases: upper motives, canonical dimension, Grothendieck group, algebraic cobordism

1. **Introduction**

The canonical dimension of a smooth complete algebraic variety measures to which extent it can be rationally compressed. In order to compute it, one usually studies the $p$-local version of this notion, called canonical $p$-dimension ($p$ is a prime number). In this paper, we consider the relation of $p$-equivalence between complete varieties, constructed so that $p$-equivalent varieties have the same canonical $p$-dimension. This essentially corresponds to the relation of having the same upper motive with $F_p$-coefficients (see Remark 2.2). For example two complete varieties $X$ and $Y$ are $p$-equivalent, for any $p$, as soon as there are rational maps $X \dashrightarrow Y$ and $Y \dashrightarrow X$. In order to obtain restrictions on the possible values of the canonical $p$-dimension of a variety, one is naturally led to study invariants of $p$-equivalence. We give a systematic way to produce such invariants (and in particular, birational invariants), starting from a homology theory. We provide examples related to $K$-theory and cycle...
modules. We then describe the relation between two such invariants of a complete variety $X$: its index $n_X$, and the integer $d_X$ defined as the g.c.d. of the Euler characteristics of the coherent sheaves of $O_X$-modules. The latter invariant contains both arithmetic and geometric informations; this can be used to give bounds on the possible values of the index $n_X$ (an arithmetic invariant) in terms of the geometry of $X$. For instance, a smooth, complete, geometrically rational (or merely geometrically rationally connected, when $k$ has characteristic zero) variety of dimension $< p - 1$ always has a closed point of degree prime to $p$. Some consequences of this statement are given in Corollary 7.9.

An immediate consequence of the Hirzebruch-Riemann-Roch theorem is that, for a smooth projective variety $X$ and a prime number $p$, we have

$$\dim X \geq (p - 1)(v_p(n_X) - v_p(\chi(X, O_X)))$$

(we denote by $v_p(m)$ the $p$-adic valuation of the integer $m$). When the characteristic of the base field is different from $p$, a result of Zainoulline states that $X$ is incompressible in case of equality (this follows by taking $p$-adic valuations in [Zai10, Corollary (A)]). Here we improve this result, and obtain in Proposition 5.4 the following bound on the canonical $p$-dimension of a regular complete variety $X$ over a field characteristic not $p$ (and also a weaker statement in characteristic $p$)

$$\text{cdim}_p X \geq (p - 1)(v_p(n_X) - v_p(d_X)).$$

We always obtain some bound on the canonical $p$-dimension, even when it is not equal to the dimension. By contrast, the result of [Zai10], or more generally any approach based on the degree formula (see [Mer03]), does not directly say anything about the canonical dimension of varieties which are not $p$-incompressible. One can sometimes circumvent this problem by exhibiting a smooth complete variety $Y$ of dimension $\text{cdim}_p(X)$ which is $p$-equivalent to $X$, and use the degree formula to prove that $Y$ is $p$-incompressible (see [Mer03, §7.3] for the case of quadrics). But this requires to explicitly produce such a variety $Y$, and to find an appropriate characteristic number for $Y$.

Let us mention that the bound above tends to be sharp only when $\text{cdim}_p(X)$ is not too large, compared with $p$ (see Example 6.7).

Another aspect of the technique presented here concerns its application to the algebraic cobordism $\Omega$ of Levine and Morel. Their construction uses two distinct results known in characteristic zero:

(a) the resolution of singularities [Hir64], and
(b) the weak factorisation theorem [AKMW02, W1003].

Some care is taken in the book [LM07] to keep track of which result uses merely (a) or the combination of (a) and (b) most deeper results actually use both. In [Hau12b], we proved that the existence of a weak form of Steenrod operations (which may be considered as a consequence of the existence of algebraic cobordism) only uses (a) Moreover, we proved that it suffices, for the purpose of the construction of these operations, to have a $p$-local version of resolution of...
singularities, which has been recently obtained in characteristic different from \( p \) by Gabber. By contrast, it is not clear what would be a \( p \)-local version of the weak factorisation theorem. In the present paper, we extend the list of results using only \([a]\) by proving the theorem below, whose original proof in \([LM07, \text{Theorem 4.4.17}]\) was based on \([b]\) (and \([a]\)). Our approach moreover allows us to give in Proposition 4.8 a \( p \)-local version of this statement, in terms of \( p \)-equivalence.

**Theorem.** Assume that the base field \( k \) admits resolution of singularities. The ideal \( M(X) = \Omega(k) \) generated by classes of smooth projective varieties \( Y \), of dimension \(< \dim X \) and admitting a morphism \( Y \to X \), is a birational invariant of a smooth projective variety \( X \).

The structure of the paper is as follows. In Section 2 we define the notion of \( p \)-equivalence, and describe its relation with canonical \( p \)-dimension. In Section 3 we introduce a (non-exhaustive) set of conditions on a pair of functors which are expected to be satisfied by a pair homology/cohomology of algebraic varieties. We verify these conditions for \( K \)-theory, cycles modules, and algebraic cobordism. In Section 4 we explain how to construct an invariant of \( p \)-equivalence classes starting from any pair satisfying the conditions of Section 3. We discuss each of the three situations mentioned above. In Section 5 we provide a relation between \( n_X \) and \( d_X \), which are invariants produced by the method of Section 3. This yields the lower bound for canonical dimension. In Section 6 we provide examples, and compute the lower bound in some specific situations. In Section 7 we adopt the point of view of function fields, and consider splitting properties of geometrically unirational field extensions.

2. \( p \)-equivalence

We denote by \( k \) a fixed base field. A variety will be an integral, separated, finite type scheme over \( \text{Spec} k \). The function field of a variety \( X \) will be denoted by \( k(X) \). When \( K \) is a field containing \( k \), we also denote by \( K \) its spectrum, and for a variety \( X \) we write \( X_K \) for \( X \times_k K \). The letter \( p \) will always denote a prime number.

A prime correspondence, or simply a correspondence, \( Y \rightsquigarrow X \) is a diagram of varieties and proper morphisms \( Y \leftarrow Z \to X \), where the map \( Y \to Z \) is generically finite. The degree of this map is the multiplicity of the correspondence. Following \([KM13, \text{§3}]\), we say that two varieties \( X \) and \( Y \) are \( p \)-equivalent if there are correspondences \( Y \rightsquigarrow X \) and \( X \rightsquigarrow Y \) of multiplicities prime to \( p \). It is equivalent to require that each of the schemes \( X_{k(Y)} \) and \( Y_{k(X)} \) have a closed point of degree prime to \( p \).

Note that a rational map \( Y \dashrightarrow X \) between complete varieties gives rise to a correspondence \( Y \rightsquigarrow X \) of multiplicity 1 (by taking for \( Z \) the closure in \( X \times_k Y \) of the graph of the rational map). Thus two complete varieties \( X \) and \( Y \) are \( p \)-equivalent, for all \( p \), as soon as there are rational maps \( Y \dashrightarrow X \) and \( X \dashrightarrow Y \). Therefore we can use \( p \)-equivalence to find birational invariants (as
We will in general restrict our attention to complete regular varieties. In this case the relation of $p$-equivalence becomes transitive (this can be proved using [KM06, Lemma 3.2]). A consequence of a theorem of Gabber [LO, X, Theorem 2.1] is that, when the characteristic of $k$ is not $p$, any complete variety is $p$-equivalent to a projective regular variety (of the same dimension).

Let $X$ be a complete regular variety. Its canonical $p$-dimension $\text{cdim}_p(X)$ can be defined as the least dimension of a closed subvariety $Z \subset X$ admitting a correspondence $X \rightarrowtail Z$ of multiplicity prime to $p$ [KM06, Corollary 4.12]. It is proven in [KM13, Lemma 3.6] that smooth $p$-equivalent varieties have the same canonical $p$-dimension. A slight modification of the arguments used there, together with Gabber’s theorem, yields the following statement.

**Proposition 2.1.** Assume that the characteristic of $k$ is different from $p$. Let $X$ be a complete regular variety. Then $\text{cdim}_p(X)$ is the least dimension of a complete regular variety $p$-equivalent to $X$.

**Remark 2.2 (Upper motives).** Let $X$ be a complete smooth variety. A summand of the Chow motive of $X$ with $\mathbb{F}_p$-coefficients is called upper if it is defined by a projector of multiplicity 1 $\in \mathbb{F}_p$ [Kar13, Definition 2.10]. If two smooth complete varieties have a common upper summand with $\mathbb{F}_p$-coefficients, then they are $p$-equivalent. The converse is true if the varieties are geometrically split and satisfy Rost nilpotence [Kar13, Corollary 2.15].

### 3. Ring theories and modules

We denote by $\text{Ab}$ the category of abelian groups. Let $\mathcal{V}$ be a full subcategory of the category of varieties and proper morphisms, and $\mathcal{R} \text{eg}$ a full subcategory of $\mathcal{V}$, which unless otherwise specified (in §3.2) will consist of the regular varieties in $\mathcal{V}$. Let $R$ and $H$ be two (covariant) functors $\mathcal{V} \rightarrow \text{Ab}$. For a morphism $f$ of $\mathcal{V}$, we denote by $f_\ast$ either of the two corresponding morphisms in $\text{Ab}$. We consider the following conditions on $R$ and $H$.

**Conditions 3.1.** Let $X \in \mathcal{V}$ and $S \in \mathcal{R} \text{eg}$. Let $f: X \rightarrow S$ be a proper, dominant, generically finite morphism of degree $d$.

1. **(R1)** The group $R(S)$ has a structure of an associative ring with unit $1_S$.
2. **(R2)** It is possible to find an element $u \in R(X)$ such that the element $f_\ast(u - d \cdot 1_S)$ is nilpotent in the ring $R(S)$.
3. **(H1)** There is a morphism of abelian groups
   \[ R(X) \otimes H(S) \rightarrow H(X) \ ; \ x \otimes s \mapsto x \cdot f_\ast s. \]
4. **(H2)** The group $H(S)$ has a structure of a left $R(S)$-module such that
   \[ f_\ast(x \cdot f_\ast s) = (f_\ast x) \cdot s. \]
We now describe some classical examples of such functors $R$ and $H$. In all cases, $R$ and $H$ will correspond to cohomology theories, and satisfy additional properties which are logically irrelevant here. We tried to provide minimal conditions making the proof of Lemma 4.1 below work.

3.1. Quillen $K$-theory. (See [Qui73, §7]) Let $\mathcal{V}$ be the category of varieties and proper morphisms, and let $X \in \mathcal{V}$. We let $R(X)$ be the Grothendieck group $K_0'(X)$ of the category of coherent sheaves of $\mathcal{O}_X$-modules, and $H(X)$ be the $m$-th $K$-group $K_m(X)$ of this category.

We denote by $K_m(X)$ the $m$-th $K$-group of the category of locally free coherent sheaves on $X$. The tensor product induces a morphism

\[(1) \quad K_m(X) \otimes K'_0(X) \to K'_m(X) \ ; \ x \otimes y \mapsto x \cap y.\]

When $S$ is regular, the map $- \cap [O_S]$ induces an isomorphism

\[(2) \quad \varphi_S : K_m(S) \to K'_m(S).\]

With $m = 0$, the combination of (2) and (1) gives (R1) (here $1_S = [O_S]$).

When $f : X \to S$ is a morphism, we have a morphism $f^* : K_m(S) \to K_m(X)$, and we can define

\[K_0'(X) \otimes K'_m(S) \to K'_m(X) \ ; \ x \otimes s \mapsto x \cdot f^* s = f^* \circ \varphi^{-1}_S(s) \cap x,\]

proving (H1). Then (H2) follows from the projection formula.

We prove (R2) for $u = [O_X]$. The element $x = f_\ast u - d \cdot [O_S]$ belongs to the kernel of the restriction to the generic point morphism $K'_0(S) \to K'_0(k(S))$ (see e.g. [Ham13, Lemma 2.4]). This amounts to saying that its unique antecedent $y \in K_0(S)$ under (2) (with $m = 0$) has rank zero. Thus for any $n$, its $n$-th power $y^n$ belongs to the $n$-th term of the gamma filtration. The image by (2) (with $m = 0$) of this term is contained in the $n$-th term of the topological filtration [SGA6, Exposé X, Corollaire 1.3.3]. The latter vanishes when $n > \dim S$, hence so does $x^n$.

3.2. Chow groups and cycle modules. Let us sketch how Chow groups and cycle modules can be made to fit into this framework, although the situation is somewhat degenerate. Let $\mathcal{V}$ be the category of varieties and proper morphisms, and $R$ be the Chow group $\text{CH}$. The property (R2) is satisfied with $u = [X]$, since $f_\ast [X] = d \cdot [S]$.

Let $M$ be a cycle module [Ros96, Definition 2.1]. We let $H(-) = A_\ast(-; M)$ be the Chow group with coefficients in $M$ [Ros96, p.356].

When $M$ is Quillen $K$-theory, the conditions (R1), (H1), and (H2) are verified in [Gil81, §8] using Bloch’s formula.

Taking for $\text{Reg}$ the subcategory of smooth varieties, (R1) is classical. When $X \to S$ is a morphism, with $S$ a smooth variety, the pairing (H1)

\[\text{CH}(X) \otimes A_\ast(S; M) \to A_\ast(X; M)\]

is defined by sending $x \otimes s$ to $g^*(x \times_k s)$. Here we use the cross product of [Ros96, §14], and $g^*$ is the pull-back along the regular closed embedding.
The map $g^\ast$ is easily seen to be $\Omega(S)$-linear using the projection formula. This gives an action on the colimit

\[ \Omega(X) \otimes \Omega(S) \rightarrow \Omega(X), \]

proving $\text{(H2)}$. Then property $\text{(H2)}$ follows formally from the projection formula in the smooth case.

For any $T \in \mathcal{V}$, we consider the subgroup $\Omega(T)^{(n)}$ of $\Omega(T)$ generated by the images of $g_\ast$, where $g$ runs over the projective morphisms $W \rightarrow T$ whose image has codimension $\geq n$ in $T$. When $T = S$ is smooth, one checks, using reduction to the diagonal and the moving lemma [LM07, Proposition 3.3.1], that the subgroups $\{\Omega(S)^{(n)}, n \geq 0\}$ define a ring filtration on $\Omega(S)$. Since $\Omega(S)^{(\dim X + 1)} = 0$, any element of $\Omega(S)^{(1)}$ is nilpotent. Let $u$ be the class in $\Omega(X)$ of any resolution of singularities of $X$. Since $f$ separable, the element $f_\ast u - d \cdot 1_S$ vanishes when restricted to some non-empty open subvariety of $S$ by [LM07, Lemma 4.4.5]. By the localisation sequence [LM07, Theorem 3.2.7], this means that this element belongs to $\Omega(S)^{(1)}$, proving $\text{(R2)}$.

Let us mention that the pair $(R, H) = (\Omega(-), \Omega(-)^{(m)})$ satisfies Conditions 3.1 with $f$ separable in $\text{(R2)}$ (where $\Omega(T)^{(m)} = \Omega(T)^{(\dim T - m)}$). Indeed the main point is to see that the map $\Omega(X) \otimes \Omega^{(n)}(S) \rightarrow \Omega^{(m)}(X)$.

This can be seen using the fact that pull-backs along morphisms of smooth varieties respect the filtration by codimension of supports, a consequence of
the moving lemma mentioned above. When \( n = 1 \), this can also be proved directly, using the fact that \( f \) is dominant and the localisation sequence.

4. The subgroup \( H_X(k) \)

In this section, \( (R, H) \) will be a pair of functors \( \mathcal{V} \to \text{Ab} \) satisfying Conditions 3.1. We assume that the base \( k \) belongs to \( \mathcal{V} \), and therefore to \( \text{Reg} \). If \( X \in \mathcal{V} \) is complete, its structural morphism \( x: X \to k \) is then in \( \mathcal{V} \). We consider the subgroup

\[
H_X(k) = \text{im} \left( x_*: H(X) \to H(k) \right) \subset H(k).
\]

In all examples considered in this paper, \( H_X(k) \) is actually an \( R(k) \)-submodule of \( H(k) \).

**Lemma 4.1.** Let \( S \in \text{Reg} \) and \( X \in \mathcal{V} \). Let \( f: X \to S \) be a proper, dominant, generically finite morphism of degree \( d \). Then the map

\[
f_*: H(X) \otimes \mathbb{Z}[1/d] \to H(S) \otimes \mathbb{Z}[1/d]
\]

is surjective.

**Proof.** Using \([R2]\), choose \( u \in R(X) \) such that \( f_*u - d \cdot 1_S \) is nilpotent. The element \( f_*u \) is then invertible in the ring \( R(S) \otimes \mathbb{Z}[1/d] \). The lemma follows, since we have by \([H2]\), for any \( s \in H(S) \),

\[
f_* \left( u \cdot f_* \left( (f_*u)^{-1} \cdot s \right) \right) = s. \]

Proof. We denote by \( \mathbb{Z}_{(p)} \) the subgroup of \( \mathbb{Q} \) consisting of those fractions whose denominator is prime to \( p \).

**Proposition 4.2.** Let \( X, Y \) be complete varieties, with \( X \in \mathcal{V} \) and \( Y \in \text{Reg} \). Let \( Y \to X \) be a correspondence of multiplicity prime to \( p \) (resp. let \( Y \to X \) be a rational map). Then, as subgroups of \( H(k) \otimes \mathbb{Z}_{(p)} \) (resp. \( H(k) \)),

\[
H_Y(k) \otimes \mathbb{Z}_{(p)} \subset H_X(k) \otimes \mathbb{Z}_{(p)} \quad \text{(resp. } H_Y(k) \subset H_X(k) \text{)}.
\]

**Proof.** Let \( Y \to Z \to X \) be a diagram giving the correspondence (resp. the correspondence of multiplicity 1 associated with the rational map). By \( [\text{Lemma 4.1}] \) we have

\[
H_Y(k) \otimes \mathbb{Z}_{(p)} \subset H_Z(k) \otimes \mathbb{Z}_{(p)} \quad \text{(resp. } H_Y(k) \subset H_Z(k) \text{)}.
\]

Since there is a proper morphism \( Z \to X \), we have

\[
H_Z(k) \subset H_X(k). \quad \square
\]

**Corollary 4.3.** Assume that two complete varieties \( Y, X \in \text{Reg} \) are \( p \)-equivalent. Then, as subgroups of \( H(k) \otimes \mathbb{Z}_{(p)} \), we have

\[
H_Y(k) \otimes \mathbb{Z}_{(p)} = H_X(k) \otimes \mathbb{Z}_{(p)}.
\]

**Corollary 4.4.** Let \( Y, X \in \text{Reg} \) be two complete varieties. Assume that there are rational maps \( Y \to X \) and \( X \to Y \). Then \( H_Y(k) = H_X(k) \) as subgroups of \( H(k) \).

We now come back to the examples of theories given in \( \text{Section 3} \).
4.1. CHOW GROUPS. We have $\text{CH}(k) = \mathbb{Z}$, and for a complete variety $X$,

$$\text{CH}_X(k) = n_X \mathbb{Z},$$

where $n_X$ is the index of $X$, defined as the g.c.d. of the degrees of closed points of $X$. We denote the $p$-adic valuation of $n_X$ by

$$n_p(X) = v_p(n_X).$$

We will need the following slightly more precise version of Proposition 4.2 when $R = H = \text{CH}$.

**Proposition 4.5.** Let $X$ and $Y$ be complete varieties, with $Y$ regular. Let $Y \hookrightarrow X$ be a correspondence of multiplicity $m$. Then

$$n_X \mid m \cdot n_Y.$$

**Proof.** Let $Y \xleftarrow{f} Z \xrightarrow{g} X$ be a diagram giving the correspondence. Let $y \in \text{CH}_0(Y)$ be such that $\text{deg}(y) = n_Y$. Using 3.2, we have

$$n_X \mid \text{deg} \circ g_*([Z] \cdot f y) = \text{deg} \circ f_*([Z] \cdot f y) = \text{deg}(f_*([Z]) \cdot y) = \text{deg}(m \cdot y) = m \cdot n_Y. \quad \square$$

4.2. CYCLE MODULES. Let $M$ be a cycle module, $H(-) = A_*(-; M)$, $R = \text{CH}$ (see §3.2). We have $A_*(k; M) = M(k)$. For a complete variety $X$, we claim that the subgroup $H_X(k)$ of $M(k)$ is the image of the morphism

$$(4) \quad \bigoplus_{L/k \in \mathcal{F}_X} M(L) \rightarrow M(k),$$

where $\mathcal{F}_X$ is the class of finite field extensions $L/k$ such that $X(L) \neq \emptyset$. Indeed the image of $A_0(X; M) \rightarrow A_0(k; M)$ vanishes when $n > 0$. The group $A_0(X; M)$ is by definition a quotient of the direct sum of the groups $M(k(x))$, over all closed points $x$ of $X$. Moreover, for each such $x$, the extension $k(x)/k$ belongs to $\mathcal{F}_X$, and the map $M(k(x)) \rightarrow A_0(X; M) \rightarrow A_0(k; M) = M(k)$ is the transfer for the finite field extension $k(x)/k$. Conversely, if $L/k \in \mathcal{F}_X$, then there is a closed point $x$ of $X$ such that $k(x)/k$ is a subextension of $L/k$. Thus $M(L) \rightarrow M(k)$ factors through $M(k(x)) \rightarrow M(k)$. This proves the claim, and additionally shows that $\mathcal{F}_X$ may be replaced in (4) by the set of residue fields at closed points of $X$.

**Corollary 4.3** asserts that, when $X$ is smooth (or merely regular when $M$ is Quillen $K$-theory), the subgroup $H_X(k) \otimes \mathbb{Z}_p$ of $M(k) \otimes \mathbb{Z}_p$ only depends on the $p$-equivalence class of $X$.

**Remark 4.6.** The group $A_0(X; M)$ itself is known to be a birational invariant of a complete smooth variety $X$ (see [Ros96 Corollary 12.10] and [KM13 Appendix RC]).
4.3. Grothendieck group. We have $K'_0(k) = \mathbb{Z}$, and when $X$ is a complete variety,

$$(K'_0)_X(k) = d_X \mathbb{Z},$$

for a uniquely determined positive integer $d_X$. The integer $d_X$ is the g.c.d. of the integers

$$\chi(X, \mathcal{G}) = \sum_i (-1)^i \dim_k H^i(X, \mathcal{G}),$$

where $\mathcal{G}$ runs over the coherent sheaves of $O_X$-modules. We denote the $p$-adic valuation of $d_X$ by

$$d_p(X) = v_p(d_X).$$

Let us record for later reference a consequence of Proposition 4.2.

**Proposition 4.7.** Let $X$ and $Y$ be complete varieties, with $Y$ regular. Let $Y \to X$ be a correspondence of multiplicity prime to $p$. Then

$$d_p(X) \leq d_p(Y).$$

4.4. Algebraic cobordism. We say that two varieties $X$ and $Y$ are separably $p$-equivalent if there are diagrams of projective morphisms $Y \leftarrow Z \rightarrow X$ and $X \leftarrow Z' \rightarrow Y$ such that $f$ and $g$ are both separable and generically finite of degrees prime to $p$. For a projective variety $X$, and an integer $n \geq 0$, we write $\Omega_X(k)^{(n)}$ for the subgroup of $\Omega(k)$ generated by classes of smooth projective varieties $Y$ admitting a morphism $Y \to X$ with image of codimension $\geq n$. In particular $\Omega_X(k)^{(0)} = \Omega_X(k)$.

**Proposition 4.8.** Assume that $k$ admits resolution of singularities. Let $X$ and $Y$ be smooth projective varieties of the same dimension, which are separably $p$-equivalent. Then, as subgroups of $\Omega(k) \otimes \mathbb{Z}(p)$, for $n \geq 0$,

$$\Omega_X(k)^{(n)} \otimes \mathbb{Z}(p) = \Omega_Y(k)^{(n)} \otimes \mathbb{Z}(p).$$

**Proof.** Let $d$ be the common dimension. The proposition follows by applying Corollary 4.3 (more precisely the analog statement for separably $p$-equivalent varieties) with $R = \Omega$ and $H = \Omega(-)^{(d-n)}$ (defined at the end of §3.3).

For a projective variety $X$, let $M(X) \subset \Omega(k)$ be the ideal generated by the classes of smooth projective varieties $Y$ of dimension $< \dim X$ admitting a morphism $Y \to X$. We have $M(X) \subset \Omega_X(k)^{(1)}$; when $k$ admits weak factorisation, this inclusion is an equality [LM07, Theorem 4.4.16].

**Proposition 4.9.** Assume that $k$ admits resolution of singularities. Let $X$ and $Y$ be two smooth projective varieties of the same dimension. If there are rational maps $X \dashrightarrow Y$ and $Y \dashrightarrow X$, then $\Omega_X(k)^{(n)} = \Omega_Y(k)^{(n)}$ as subgroups of $\Omega(k)$, for any $n$. Moreover, $M(X) = M(Y)$ as ideals of $\Omega(k)$.

**Proof.** The first statement follows by taking $R = \Omega$ and $H = \Omega(-)^{(\dim X-n)}$ in Corollary 4.4 (which only requires the validity of Conditions 3.1 for $f$ birational, and in particular separable).
Let $V$ be a smooth projective variety of dimension $d$. The subgroup $\Omega_{<d}(k) \subset \Omega(k)$ generated by the classes of smooth projective varieties of dimension $< d$ is a direct summand of $\Omega(k)$ (indeed $\Omega(k)$ is the quotient of the free group generated by the classes of smooth projective varieties, by relations respecting the dimensional grading). Let $\pi_{<d}: \Omega(k) \to \Omega_{<d}(k)$ be the projection. The subgroup $\pi_{<d}\Omega_V(k(1))$ is generated by the elements $\pi_{<d}[U]$, where $U$ runs over the smooth projective varieties admitting a non-dominant morphism to $V$. The element $\pi_{<d}[U]$ is equal to $[U]$ if $\dim U < d$, and vanishes otherwise. So $\pi_{<d}\Omega_V(k(1))$ is the subgroup of $\Omega(k)$ generated by the classes of smooth projective varieties of dimension $< d$ admitting a morphism to $V$ (necessarily non-dominant), while $M(V)$ is the ideal generated by the same elements. Thus $M(V)$ is the ideal of $\Omega(k)$ generated by $\pi_{<d}\Omega_V(k(1))$, and the second statement follows from the first, with $n = 1$.

The corollary below was proved in [LM07, Theorem 4.4.17, 1.] under the additional assumption that $k$ admits weak factorisation.

**Corollary 4.10.** Assume that $k$ admits resolution of singularities. The ideal $M(X)$ of $\Omega(k)$ is a birational invariant of a smooth projective variety $X$.

## 5. Relations between $n_X$ and $d_X$

We have the obvious relation $d_X \mid n_X$. When $\dim X = 0$, we have $n_X = d_X$. A consequence of the next theorem is that $d_p(X) = n_p(X)$ when $\dim X < p - 1$.

**Theorem 5.1.** Let $X$ be a complete variety. Assume that one of the following conditions holds.

(i) The characteristic of $k$ is not $p$.
(ii) We have $\dim X < p(p - 1)$.
(iii) The variety $X$ is regular and (quasi-)projective over $k$.

Then

$$n_p(X) \leq d_p(X) + \left\lfloor \frac{\dim X}{p - 1} \right\rfloor.$$  

**Proof.** For (i) and (ii) we may assume that $X$ is projective over $k$ by Chow’s lemma. Indeed if $X' \to X$ is an envelope [Ful98, Definition 18.3], then it follows from [Ful98, Lemma 18.3] that $n_X = n_{X'}$ and $d_X = d_{X'}$. The group $K_p^0(X)$ is generated by the classes $[O_Z]$, where $Z$ runs over the closed subvarieties of $X$. Therefore $d_X$ is the g.c.d. of the integers $\chi(Z, O_Z)$, for $Z$ as above. In particular, we can find a closed subvariety $Z$ of $X$ such that

$$d_p(X) = v_p(\chi(Z, O_Z)).$$

We now claim that, under the assumption (ii) or (iii) we have,

$$n_p(Z) \leq v_p(\chi(Z, O_Z)) + \left\lfloor \frac{\dim Z}{p - 1} \right\rfloor.$$  

This will conclude the proof, since $n_p(X) \leq n_p(Z)$, and $\dim Z \leq \dim X$. 

Documenta Mathematica 18 (2013) 1555–1572
If we assume \(\text{(iii)}\) then \(\dim Z < p(p-1)\), and \(\text{(ii)}\) follows from [Hau12a Proposition 9.1]. But the argument of loc. cit. can also be used in the situation \(\text{(i)}\).

Namely, we let \(\tau_a : K_0^c(-) \to \text{CH}_a(-) \otimes \mathbb{Q}\) be the map of [Ful98 Theorem 18.3] (it is the homological Chern character, denoted \(\text{ch}_a\) in [Hau12a]). Let \(z : Z \to k\) be the structural morphism of \(Z\). Then by [Ful98 Theorem 18.3 (1)]

\[
\chi(Z, \mathcal{O}_Z) = \tau_0 \circ z_*[\mathcal{O}_Z] = z_* \circ \tau_0[\mathcal{O}_Z].
\]

If \(\text{(i)}\) holds, then [Hau12a Theorem 4.2] says that the element \(p^{\dim Z/(p-1)} \cdot \tau_0[\mathcal{O}_Z] \in \text{CH}_0(Z) \otimes \mathbb{Q}\) belongs to the image of \(\text{CH}_0(Z) \otimes \mathbb{Z}_p\), hence can written as \(b \otimes \lambda^{-1}\), with \(\lambda\) an integer prime to \(p\), and \(b \in \text{CH}_0(Z)\). Thus:

\[
\left[\frac{\dim Z}{p-1}\right] + v_p(z_* \circ \tau_0[\mathcal{O}_Z]) = v_p(z_*b) \geq \nu_p(Z).
\]

Using \(\text{(i)}\), this gives \(\text{(ii)}\).

Now we assume \(\text{(iii)}\). We identify the groups \(K_0^c(X)\) and \(K_0(X)\) using \(\text{(2)}\), and take \(a \in K_0(X)\). Let \(x : X \to k\) be the structural morphism of \(X\). Since \(x\) is quasi-projective and \(X\) regular, the morphism \(x\) is a local complete intersection (i.e. factors as a regular closed embedding followed by a smooth morphism); let \(T_x \in K_0(X)\) be its virtual tangent bundle. We apply the Grothendieck-Riemann-Roch theorem [Ful98 Theorem 18.2], and get in \(Q = K'_0(k) \otimes Q = \text{CH}(k) \otimes Q\) the equalities

\[
x_*a = \text{ch} \circ x_*a = x_* \circ \text{Td}(T_x) \circ \text{ch} a.
\]

Here \(\text{ch}\) is the Chern character, with components \(\text{ch}^n : K_0(X) \to \text{CH}^n(X) \otimes \mathbb{Q}\), and \(\text{Td} = \sum_n \text{Td}^n\) is the Todd class. By [Hau12a Lemma 6.3], the morphism

\[
p^{\dim X/(p-1)} \cdot \text{Td}^n(T_x) : \text{CH}^*(X) \otimes Q \to \text{CH}^{*+n}(X) \otimes Q
\]

sends the image of \(\text{CH}^*(X) \otimes \mathbb{Z}_p\) to the image \(\text{CH}^{*+n}(X) \otimes \mathbb{Z}_p\). The degree zero component of \(p^{\dim X/(p-1)} \cdot \text{Td}(T_x) \circ \text{ch} a\) is

\[
\sum_{n=0}^{\dim X} (p^{\dim X-n)/(p-1)} \cdot \text{Td}^{\dim X-n}(T_x) \circ (p^{\dim X/(p-1)} \cdot \text{ch}^n) \in \text{CH}_0(X) \otimes Q.
\]

By [Lemma 5.2] and the remark above, this element belongs to the image of \(\text{CH}_0(X) \otimes \mathbb{Z}_p \to \text{CH}_0(X) \otimes Q\). Using \(\text{(2)}\), we obtain,

\[
\left[\frac{\dim X}{p-1}\right] + v_p(x_*a) \geq \nu_p(X).
\]

The statement follows, since we can choose \(a\) such that \(v_p(x_*a) = \nu_p(X)\). \hfill \Box

**Lemma 5.2.** Let \(X\) be a variety, and \(\text{ch} : K_0(X) \to \text{CH}(X) \otimes \mathbb{Q}\) the Chern character. Then for all integers \(n\) and elements \(a \in K_0(X)\), we have

\[
p^{\dim X/(p-1)} \cdot \text{ch}^n a \in \text{im} \left(\text{CH}^n(X) \otimes \mathbb{Z}_p \to \text{CH}^n(X) \otimes \mathbb{Q}\right).
\]
Proof. This follows from the splitting principle. In more details, we proceed exactly as in [Hau12a, Lemma 6.3], using that ch factors through the operational Chow ring tensored with $\mathbb{Q}$, and that for a line bundle $L$,

$$\text{ch}(L) = \sum_{n \geq 0} c_1(L)^n n!.$$ 

□

Remark 5.3. From the proof, we see that the statement of Lemma 5.2 may be improved: the exponent $[n/(p - 1)]$ can be replaced by $[(n - 1)/p - 1]$.

Proposition 5.4. Let $X$ be a complete regular variety. Then

$$\text{cdim}_p(X) \geq \begin{cases} (p - 1) \cdot (n_p(X) - d_p(X)) & \text{if } k \text{ has characteristic } \neq p, \\ \begin{cases} (p - 1) \cdot \min(p, n_p(X) - d_p(X)) & \text{if } k \text{ has characteristic } = p. \end{cases} \end{cases}$$

Proof. Let $Z \subset X$ be a closed subvariety admitting a correspondence $X \to Z$ of multiplicity prime to $p$, and such that $\dim Z = \text{cdim}_p(X)$. By Proposition 4.7 we have $d_p(Z) \leq d_p(X)$. Since there is a morphism $Z \to X$, we have $n_p(X) \leq n_p(Z)$. We conclude by applying Theorem 5.1 (i), (ii) to the complete variety $Z$. □

6. Examples

In view of Section 5, it may seem desirable to find conditions on a complete variety $X$ that give upper bounds for $d_p(X)$.

Proposition 6.1. Let $X$ be a complete smooth variety. Assume that there is a field extension $l/k$, and a complete smooth $l$-variety $Y$ such that $X \times_k Y$ is a rational $l$-variety. Then $d_X = 1$.

Proof. It will be sufficient to prove that $\chi(X, O_X) = \pm 1$. While doing so, we may extend scalars, and thus assume that $k = l$. The variety $Z = X \times_k Y$ is then rational. Since the coherent cohomology groups of the structure sheaf are birational invariants of a complete smooth variety [CR11, Theorem 3.2.8], so is its Euler characteristic. The structure sheaf of the projective space has Euler characteristic equal to 1, hence $\chi(Z, O_Z) = 1$. Since $\chi(Z, O_Z) = \chi(X, O_X) \cdot \chi(Y, O_Y)$ by [Ful98, Example 15.2.12], we are done. □

Corollary 6.2. Let $X$ be a complete, smooth, geometrically rational variety. Then $d_X = 1$.

Example 6.3 (Projective homogeneous varieties). Let $X$ be a complete, smooth, geometrically connected variety, which is homogeneous under a semi-simple linear algebraic group. Then $X$ is geometrically rational. Thus by Corollary 6.2 we have $d_X = 1$.

Proposition 6.4. Assume that $k$ has characteristic zero. Let $X$ be a complete, smooth, geometrically rationally connected variety. Then $d_X = 1$.

Proof. We proceed as in the proof of Proposition 6.1 and assume that $X$ is rationally connected. Then the groups $H^i(X, O_X)$ vanish for $i > 0$ by [Deb01, Corollary 4.18, a)], and therefore $\chi(X, O_X) = 1$. □
Remark 6.5 (Decomposition of the diagonal). The statement of Proposition 6.4 is more generally true (still in characteristic zero, for $X$ smooth complete) under the assumption that the diagonal decomposes, i.e. that there is a zero-cycle $z$ on $X$ whose degree $N$ is not zero, and a non-empty open subvariety $U$ of $X$, such that the cycles $[U] \times_k z$ and $N \cdot [\Gamma_U]$ are rationally equivalent on $U \times_k X$, where $\Gamma_U \subset U \times_k X$ is the graph of $U \to X$. This is the case when $\text{CH}_0(X_\Omega) = \mathbb{Z}$, where $\Omega$ is an algebraic closure of $k(X)$ (see e.g. the introduction of [Esn03]).

Example 6.6 (Complete intersections). Let $H$ be a complete intersection of hypersurfaces in $\mathbb{P}^n$ of degrees $\delta_1, \cdots, \delta_m$, with $\delta_1 + \cdots + \delta_m \leq n$. When $H$ is smooth, it is Fano, hence geometrically rationally chain connected. If, in addition, $k$ has characteristic zero, the variety $H$ is geometrically rationally connected, and Proposition 6.4 shows that $\chi(H, \mathcal{O}_H) = 1$. Alternatively, a direct computation shows that this is true in general (in any characteristic, for possibly singular $H$). This can be used to produce other sufficient conditions on a complete variety $X$ for the equality $d_X = 1$, namely:

— $X$ becomes isomorphic to such an $H$ after extension of the base field,

—or $X$ is smooth and becomes birational to such a smooth $H$ after extension of the base field.

Example 6.7 (Hypersurfaces). Let $H$ be a regular hypersurface of degree $p$ and dimension $\geq p - 1$. By Example 6.6 we have $d_p(H) = 0$. Assume that $H$ has no closed point of degree prime to $p$. Then by Proposition 6.4 we have $\text{cdim}_p(H) \geq p - 1$.

In case $\dim H = p - 1$, this bound is optimal, and $H$ is $p$-incompressible (see also [Mer03, § 7.3] and [Zai10, Example 6.4]). A more general statement was proved in [Hau12a, Proposition 10.1].

When $p = 2$, we have $\text{cdim}_2(H) \geq 1$. This bound is sharp when $\dim H = 2$, as can be seen by taking for $H$ an anisotropic smooth projective Pfister quadric surface. In general this is far from being sharp: in characteristic not two, it is known that $\text{cdim}_2(H) = 2^n - 1$, where $n$ is such that $2^n - 1 \leq \dim H < 2^{n+1} - 1$ (see [Mer03, § 7]).

Example 6.8 (Separation). Let $X$ and $Y$ be complete varieties, with $X$ regular. Assume that the degree of any closed point of $Y$ is divisible by $p$, that $d_p(X) = 0$ (see in particular Example 6.6 and Proposition 6.4), and that $\dim Y < p - 1 \leq \dim X$.

Then the degree of any closed point of $Y_{k(X)}$ is divisible by $p$. (Indeed, assuming the contrary, there would be a correspondence $X \sim Y$ of degree prime to $p$. Then $d_p(Y) = 0$ by Proposition 4.7, hence $n_p(Y) = 0$ by Theorem 5.1.)

Proposition 6.9. Let $f : Y \dashrightarrow X$ be a rational map between complete regular varieties. Let $F$ be the generic fiber of $f$, considered as a $k(X)$-variety. Then

(i) If $d_p(F) = 0$ then $d_p(X) = d_p(Y)$.
(ii) We have $n_p(X) \leq n_p(Y) \leq n_p(X) + n_p(F)$.
Proof. Let us prove (i). The map $f$ induces a correspondence $Y \dashv Z \overset{h}{\to} X$ with $g$ birational. By Lemma 4.1 we have $d_Y = d_Z$.

We now prove the equality $d_p(X) = d_p(Z)$. We have a cartesian square

$$
\begin{array}{ccc}
F & \xrightarrow{\rho} & Z \\
\downarrow{l} & & \downarrow{h} \\
k(X) & \xrightarrow{n} & X
\end{array}
$$

Since $d_p(F) = 0$, there is $v \in K'_0(F) \otimes \mathbb{Z}(p)$ such that

$$
l_*v = 1 \in \mathbb{Z}(p) = K'_0(k(X)) \otimes \mathbb{Z}(p).
$$

Since $\rho^*: K'_0(Z) \to K'_0(F)$ is surjective (by the localisation sequence), we can find $u \in K'_0(Z) \otimes \mathbb{Z}(p)$ such that $\rho^*u = v$. Then we have

$$
\eta^* \circ h_*u = l_* \circ \rho^*u = l_*v = 1 = \eta^*[\mathcal{O}_X].
$$

It follows that the element $h_*u - [\mathcal{O}_X]$ is in the kernel of $\eta^*$, hence is nilpotent; therefore $h_*u$ is invertible in $K'_0(X) \otimes \mathbb{Z}(p)$. We can now conclude that $d_p(X) = d_p(Z)$, as in Lemma 4.1.

The first inequality in (ii) follows from Proposition 4.5. To prove the second, note that $F \subset Y_{k(X)}$ and use Lemma 7.5 below. \hfill \Box

Example 6.10. Let $p$ be an odd prime. Let $H$ be a regular hypersurface of degree $p$ and dimension $p - 1$, with $n_p(H) = 1$. Let $X$ be a regular complete variety admitting a rational map $X \dashv H$, whose generic fiber $F$ is also a hypersurface of degree $p$ and dimension $p - 1$.

We have $d_p(F) = d_p(H) = 0$ by Example 6.6 and therefore $d_p(X) = 0$ by Proposition 6.9 (i). Moreover $n_p(F) \leq 1$ by Theorem 5.1, hence by Proposition 6.9 (ii) we have $1 \leq n_p(X) \leq 2$.

We use Proposition 5.4. If $n_p(X) = 2$, then $\text{cdim}_p(X) = \text{dim} X = 2(p - 1)$. If $n_p(X) = 1$, then $\text{cdim}_p(X) \geq p - 1$. This bound is sharp: if $n_p(F) = 0$, then $X$ and $H$ are $p$-equivalent, hence have the same canonical $p$-dimension; but $\text{cdim}_p(H) = \text{dim} H = p - 1$.

7. Invariants of function fields

Definition 7.1. Let $K/k$ be a finitely generated field extension. Let $M$ be a regular projective variety such that $k(M)$ contains $K$ as a $k$-subalgebra with $[k(M):k]$ finite and prime to $p$. We define $d_p(K/k) := d_p(M)$ and $n_p(K/k) := n_p(M)$. By Lemma 7.4 below, and Corollary 4.3, these integers do not depend on the choice of $M$. If there is no such $M$, we set $d_p(K/k) = n_p(K/k) = \infty$ (by Gabber’s theorem, this may only happen in characteristic $p$).

An alternative definition of $n_p(K/k)$ can be found in [Mer03, Remark 7.7]. Note that, when $X$ is a complete variety, we have (by Proposition 4.7 and Proposition 4.5)

$$
d_p(X) \leq d_p(k(X)/k) \quad \text{and} \quad n_p(X) \leq n_p(k(X)/k),
$$

\text{Documenta Mathematica 18 (2013) 1555–1572}
with equalities when \( X \) is regular (by Remark 7.2).

**Remark 7.2.** This process can be used more generally to define invariants \( H_{K/k}(k) \otimes \mathbb{Z}(p) \), when the pair \((R, H)\) satisfies Conditions 3.1.

**Remark 7.3 (Geometrically unirational field extensions).** When \( k \) has characteristic zero, it is possible to have some control on \( d_p(K/k) \) without explicitly introducing a variety \( M \) as in Definition 7.1. Namely assume that there is a field extension \( l/k \) such that the ring \( K \otimes_k l \) is contained in a purely transcendental field extension of \( l \). Then \( d_p(K/k) = 0 \) for any \( p \). (Indeed by Hironaka’s resolution of singularities [Hir64], there is a smooth projective variety \( M \) with function field \( K \). Then \( M \) is geometrically unirational, hence geometrically rationally connected, and \( d_M = 1 \) by Proposition 6.4.)

**Lemma 7.4.** Let \( X_1 \to X \leftarrow X_2 \) be a diagram of varieties and proper morphisms, generically finite of degree prime to \( p \). Then \( X_1 \) is \( p \)-equivalent to \( X_2 \).

**Proof.** Consider the cartesian square

\[
\begin{array}{ccc}
X' & \to & X_2 \\
\downarrow^{g_1} & & \downarrow^{f_2} \\
X_1 & \to & X
\end{array}
\]

We claim that \( X' \) has an irreducible component which is generically finite of degree prime to \( p \) over \( X_1 \); this component then gives a correspondence \( X_1 \leadsto X_2 \) of multiplicity prime to \( p \), and the lemma follows by symmetry.

Since \( f_i \) is generically finite and dominant, and \( X_i \) is integral, the generic fiber of \( f_i \) is the spectrum of \( K_i := k(X_i) \) (for \( i = 1, 2 \)). The ring \( B = K_1 \otimes_k K_2 \) is artinian, and its spectrum is the generic fiber of \( g_2 \). Write \( B = B_1 \times \cdots \times B_n \), with \( B_j \) an artinian local ring with residue field \( L_j \) (for \( j = 1, \cdots, n \)). We have

\[
[K_2 : k(X)] = \dim K_1, B = \sum_{j=1}^n \dim K_1, B_j = \sum_{j=1}^n [L_j : K_1] \cdot \text{length } B_j
\]

(for the last equality, use e.g. [Ful98] Lemma A.1.3], replacing \( A, B, M \) with \( K_1, B_j, B_j \)). But the leftmost integer is prime to \( p \) by hypothesis. It follows that for some \( j \) the integer \([L_j : K_1]\) is prime to \( p \). The closure of \( \text{Spec } L_j \) in \( X' \) then gives the required irreducible component. \( \square \)

**Lemma 7.5.** Let \( X \) and \( M \) be complete varieties, with \( M \) regular. Then

\[
n_p(X, M) \leq n_p(X_{k(M)}) + n_p(M).
\]

**Proof.** Write \( E = k(M) \). Let \( L/E \) be a finite field extension such that \( X_E(L) \neq \emptyset \) and \( v_p[L : E] = n_p(X_E) \). Then the closure of the image of the map induced by an \( L \)-point of \( X_E \)

\[
\text{Spec } L \hookrightarrow X_E = (\text{Spec } E) \times_k X \to M \times_k X
\]
Olivier Haution

gives a correspondence $M \rightarrow X$, whose multiplicity is $[L : E]$. By Proposition 4.5 we have

$$n_p(X) \leq v_p[L : E] + n_p(M) = n_p(X_E) + n_p(M).$$

**Proposition 7.6.** Let $K/k$ be a finitely generated field extension. Let $X$ be a complete variety. Then

$$n_p(X) \leq n_p(X_K) + n_p(K/k).$$

**Proof.** We may assume that there is a projective regular variety $M$ with $[k(M) : K]$ finite and prime to $p$ (otherwise the statement is empty). Since $n_p(X_{k(M)}) \leq n_p(X_K)$, the statement follows from Lemma 7.5.

**Proposition 7.7.** Let $K/k$ be a finitely generated field extension. Then

$$n_p(K/k) \leq d_p(K/k) + \left\lceil \frac{\text{tr.deg}(K/k)}{p-1} \right\rceil.$$

**Proof.** As above, we may assume that there is a smooth projective variety $M$ such that $[k(M) : K]$ is prime to $p$, and we apply Theorem 5.1 (iii) with $X = M$.

**Corollary 7.8.** Let $K/k$ be a field extension of transcendence degree $< p - 1$, with $d_p(K/k) = 0$. Then $n_p(K/k) = 0$.

**Corollary 7.9.** Let $K/k$ be a field extension of transcendence degree $< p - 1$. Assume that $d_p(K/k) = 0$ (see in particular Remark 7.3). Then

(i) The relative Brauer group $\ker(\text{Br}(k) \rightarrow \text{Br}(K))$ has no $p$-primary torsion.

(ii) Let $\alpha$ be a pure symbol in $K_2^M(k)/p$. Assume that $k$ has characteristic zero. If $\alpha_K = 0$ then $\alpha = 0$.

**Proof.** We have $n_p(K/k) = 0$ by Corollary 7.8. To prove (i) let $A$ be central simple $k$-algebra of $p$-primary exponent. Take for $X$ the Severi-Brauer variety of $A$. Then for any field extension $l/k$, the class of $A \otimes_k l$ vanishes in the Brauer group of $l$ if and only if $n_p(X_l) = 0$. By Proposition 7.6 we have $n_p(X_K) = n_p(X)$, and (ii) follows.

To prove (ii) one can take for $X$ a complete generic $p$-splitting variety [SJ06] for $\alpha$, and argue as above.

**Example 7.10.** Let $D$ be a non-trivial $p$-primary central division $k$-algebra, and $K/k$ a splitting field extension for $D$. Assume that $K$ is the function field of a complete, smooth, geometrically rational variety (or more generally that $d_p(K/k) = 0$). Corollary 7.9 (i) says that $\text{tr.deg}(K/k) \geq p - 1$. One may ask whether we always have

$$\text{tr.deg}(K/k) \geq \text{ind}(D) - 1.$$ 

In other words, has the Severi-Brauer variety of $D$ the smallest possible dimension among the complete, smooth, geometrically rational varieties whose function field splits $D$?
Acknowledgements. I would like to thank Nikita Karpenko for his comments on this paper, and the referee for his/her careful reading of the paper and constructive remarks. The first version of this paper has been written while I was employed by the University of Nottingham.

References


Olivier Haution
Mathematisches Institut
Ludwig-Maximilians-Universität München
Theresienstr. 39
D-80333 München
Germany
olivier.haution at gmail.com


DENSITIES OF THE RANEY DISTRIBUTIONS

WOJIECH MŁOTKOWSKI¹, KAROL A. PENSON², KAROL ŻYCZKOWSKI³

Received: October 7, 2012
Revised: November 2, 2013

Communicated by Friedrich Götze

ABSTRACT. We prove that if $p \geq 1$ and $0 < r \leq p$ then the sequence $(m^{p+r})_m (\binom{mp+r}{m}^{\frac{r}{mp+r}})$ is positive definite. More precisely, it is the moment sequence of a probability measure $\mu(p, r)$ with compact support contained in $[0, +\infty)$. This family of measures encompasses the multiplicative free powers of the Marchenko-Pastur distribution as well as the Wigner’s semicircle distribution centered at $x = 2$. We show that if $p > 1$ is a rational number and $0 < r \leq p$ then $\mu(p, r)$ is absolutely continuous and its density $W_{p,r}(x)$ can be expressed in terms of the generalized hypergeometric functions. In some cases, including the multiplicative free square and the multiplicative free square root of the Marchenko-Pastur measure, $W_{p,r}(x)$ turns out to be an elementary function.

2010 Mathematics Subject Classification: Primary 44A60; Secondary 33C20

Keywords and Phrases: Mellin convolution, free convolution, Meijer $G$-function, generalized hypergeometric function.

¹W. M. is supported by the Polish National Science Center grant No. 2012/05/B/ST1/00626.
²K. A. P. acknowledges support from PAN/CNRS under Project PICS No. 4339 and from Agence Nationale de la Recherche (Paris, France) under Program PHYSCOMB No. ANR-08-BLAN-0243-2.
³K. Ż. is supported by the Grant DEC-2011/02/A/ST1/00119 of Polish National Centre of Science.
For \( p, r \in \mathbb{R} \) we define the Raney numbers (or two-parameter Fuss-Catalan numbers) by

\[
A_m(p, r) := \frac{r}{m!} \prod_{i=1}^{m-1} (mp + r - i),
\]

(1)

\( A_0(p, r) := 1 \). We can also write

\[
A_m(p, r) = \binom{mp + r}{m} \frac{r}{mp + r},
\]

(2)

(unless \( mp + r = 0 \)), where the generalized binomial is defined by

\[
\binom{a}{m} := \frac{a(a-1)\ldots(a-m+1)}{m!}.
\]

Let \( B_p(z) \) denote the generating function of the sequence \( \{A_m(p, 1)\}_{m=0}^{\infty} \), the Fuss numbers of order \( p \):

\[
B_p(z) := \sum_{m=0}^{\infty} A_m(p, 1)z^m,
\]

(3)

convergent in some neighborhood of 0. For example

\[
B_2(z) = \frac{2}{1 + \sqrt{1 - 4z}}
\]

(4)

Lambert showed that

\[
B_p(z)^r = \sum_{m=0}^{\infty} A_m(p, r)z^m,
\]

(5)

see [9]. These generating functions also satisfy

\[
B_p(z) = 1 + zB_p(z)^p,
\]

(6)

which reflects the identity \( A_m(p, p) = A_{m+1}(p, 1) \), and

\[
B_p(z) = B_{p-r}(zB_p(z)^r).
\]

(7)

Using the free probability theory (see [28, 18, 6]) it was shown in [16] that if \( p \geq 1 \) and \( 0 \leq r \leq p \) then the sequence \( \{A_m(p, r)\}_{m=0}^{\infty} \) is positive definite, i.e. is the moment sequence of a probability measure \( \mu(p, r) \) on \( \mathbb{R} \). Moreover, \( \mu(p, r) \) has compact support (and therefore is unique) contained in the positive half-line \([0, \infty)\) (for example \( \mu(p, 0) = \delta_0 \)). The measures \( \mu(p, r) \) satisfy some interesting relations, for example

\[
\mu(p_1, r) \boxtimes \mu(1 + p_2, 1) = \mu(p_1 + rp_2, r)
\]

(8)
and
\[ \mu(p, r) \triangleright \mu(p + s, s) = \mu(p + s, r + s), \] (9)
see [16], where “\( \triangleright \)” and “\( \triangleright \)” denotes the multiplicative free and the monotonic convolution (see [17]). A relation analogous to (9) is also satisfied by the three-parameter family of distributions studied by Arizmendi and Hasebe [4].

Among the measures \( \mu(p, r) \) perhaps the most important is the Marchenko-Pastur (called also the free Poisson) distribution
\[ \mu(2, 1) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}} dx \quad \text{on } [0,4], \quad (10) \]
which plays an important role in the theory of random matrices, see [29, 10, 11, 2, 1, 5]. It was proved in [1] that the multiplicative free power \( \mu(2, 1)^{\otimes n} = \mu(n+1, 1) \) is the limit of the distribution of squared singular values of the power \( G^n \) of a random matrix \( G \), when the size of the matrix \( G \) goes to infinity. The moments of \( \mu(2, 1) \), \( A_m(2, 1) = \binom{2m+1}{m+1} / (2m + 1) \), are called Catalan numbers and play an important role in combinatorics, see A000108 in OEIS [24].

In this paper we are going to prove positive definiteness of \( \{A_m(p, r)\}_{m=0}^\infty \) using more classical methods. Namely, we show that if \( p > 1, 0 < r \leq p \) and if \( p \) is a rational number then \( \mu(p, r) \) is absolutely continuous and can be represented as Mellin convolution of modified beta measures. Next we provide a formula for the density \( W_{p,r}(x) \) of \( \mu(p, r) \) in terms of the Meijer \( G \)-function and of the generalized hypergeometric functions (cf. [30, 21], where \( p \) was assumed to be an integer). This allows us to draw graphs of these densities and, in some particular cases, to express \( W_{p,r}(x) \) as an elementary function.

Let us mention that the measures \( \mu(2, 1)^{\otimes p} = \mu(1 + p, 1) \) were also studied by Banica, Belinschi, Capitaine and Collins [5] as a special case of the free Bessel laws. They showed in particular that for \( p > 0 \) this measure is absolutely continuous and its support is \( [0, (p + 1)^{p+1}p^{-p}] \). Liu, Song and Wang [14] found a formula expressing the density of \( \mu(2, 1)^{\otimes n} \), \( n \) natural, as integral of a certain kernel over \( [0,1]^n \). Recently Haagerup and Möller [12] studied a two-parameter family \( \mu_{\alpha,\beta}, \alpha, \beta > 0 \), of probability measures. The measures \( \mu_{\alpha,0} \) coincide with our \( \mu(1 + \alpha, 1) \), but if \( \beta > 0 \) then \( \mu_{\alpha,\beta} \) has noncompact support, so it does not coincide with any of \( \mu(p, r) \). The authors found a formula for the density function of \( \mu_{\alpha,\beta} \), which in the case of \( W_{1+p,1} \) reads as follows:
\[ W_{1+p,1} \left( \frac{\sin^{p+1}((p+1)t)}{\sin t \sin^p(pt)} \right) = \frac{\sin^2 t \sin^{p-1}(pt)}{\pi \sin^2((p+1)t)}, \quad (11) \]
for \( 0 < t < \pi/(p+1) \). It can be used for drawing the graph of \( W_{1+p,1}(x) \) by computer.
1 Preliminaries

For probability measures $\mu_1, \mu_2$ on the positive half-line $[0, \infty)$ the Mellin convolution is defined by

$$(\mu_1 \circ \mu_2)(A) := \int_0^\infty \int_0^\infty 1_A(xy) d\mu_1(x) d\mu_2(y)$$

(12)

for every Borel set $A \subseteq [0, \infty)$. This is the distribution of product $X_1 \cdot X_2$ of two independent nonnegative random variables with $X_i \sim \mu_i$. In particular, $\mu \circ \delta_c (c > 0)$ is the dilation of $\mu$:

$$(\mu \circ \delta_c)(A) = D_c \mu(A) := \mu\left(\frac{1}{c} A\right).$$

If $\mu$ has density $f(x)$ then $D_c \mu$ has density $f(x/c)/c$.

If both the measures $\mu_1, \mu_2$ have all moments $s_m(\mu_i) := \int_0^\infty x^m d\mu_i(x)$ finite then so has $\mu_1 \circ \mu_2$ and

$$s_m(\mu_1 \circ \mu_2) = s_m(\mu_1) \cdot s_m(\mu_2)$$

for all $m$.

If $\mu_1, \mu_2$ are absolutely continuous, with densities $f_1, f_2$ respectively, then so is $\mu_1 \circ \mu_2$ and its density is given by the Mellin convolution:

$$(f_1 \circ f_2)(x) := \int_0^\infty f_1(x/y) f_2(y) \frac{dy}{y}.$$

We will need the following modified beta measures:

**Lemma 1.1.** Let $u, v, l > 0$. Then

$$\frac{\Gamma(u + n/l) \Gamma(u + v)}{\Gamma(u + v + n/l) \Gamma(u)} \infty_{n=0}$$

is the moment sequence of the probability measure

$$b(u + v, u, l) := \frac{l}{B(u, v)} x^{lu-1} (1 - x^l)^{v-1} dx$$

(13)

on $[0, 1]$, where $B$ is the Euler beta function.

**Proof.** Using the substitution $t = x^l$ we obtain:

$$\frac{\Gamma(u + n/l) \Gamma(u + v)}{\Gamma(u + v + n/l) \Gamma(u)} = \frac{B(u + n/l, v)}{B(u, v)} = \frac{1}{B(u, v)} \int_0^1 t^{u+n/l-1} (1 - t)^{v-1} dt$$

$$= \frac{l}{B(u, v)} \int_0^1 x^{lu+n-1} (1 - x^l)^{v-1} dx.$$
Densities of the Raney Distributions

Note that if \( X \) is a positive random variable whose distribution has density \( f(x) \) and if \( l > 0 \) then the distribution of \( X^{1/l} \) has density \( l x^{l-1} f(x') \). In particular, if the distribution of a random variable \( X \) is \( b(u + v, u, 1) \) then the distribution of \( X^{1/l} \) is \( b(u + v, u, l) \). For \( u, l > 0 \) we also define
\[
b(u, u, l) := \delta_1.
\]

2 Applying Mellin convolution

From now on we assume that \( p > 1 \) is a rational number, say \( p = k/l \), with \( 1 \leq l < k \), and that \( 0 < r \leq p \). We will show that then \( A_m(p, r) \) is the moment sequence of a probability measure \( \mu(p, r) \), which can be represented as Mellin convolution of modified beta measures. In particular, \( \mu(p, r) \) is absolutely continuous and we will denote its density by \( W_{p,r} \). The case when \( p \) is an integer was studied in [21, 30].

First we need to express the numbers \( A_m(p, r) \) in a special form.

**Lemma 2.1.** If \( p = k/l \), where \( k, l \) are integers, \( 1 \leq l < k \), and \( 0 < r \leq p \) then
\[
A_m(p, r) = \frac{r}{l \sqrt{2\pi k (p - 1)}} \left( \frac{p}{p - 1} \right)^r \prod_{j=1}^k \Gamma(\beta_j + m/l) \prod_{j=1}^k \Gamma(\alpha_j + m/l) c(p)^m,
\]
where \( c(p) = p^p (p - 1)^{1-p} \),
\[
\alpha_j = \begin{cases} 
\frac{j}{l} & \text{if } 1 \leq j \leq l, \\
\frac{r + j - l}{k - l} & \text{if } l + 1 \leq j \leq k,
\end{cases}
\]
\[
\beta_j = \frac{r + j - 1}{k}, \quad 1 \leq j \leq k.
\]

**Proof.** First we write:
\[
\binom{mp + r}{m} \frac{r}{mp + r} = \frac{r \Gamma(mp + r)}{\Gamma(m + 1) \Gamma(mp - m + r + 1)}.
\]
Now we apply the Gauss’s multiplication formula:
\[
\Gamma(nz) = (2\pi)^{(1-n)/2} n^{nz - 1/2} \Gamma(z) \Gamma \left( z + \frac{1}{n} \right) \Gamma \left( z + \frac{2}{n} \right) \ldots \Gamma \left( z + \frac{n-1}{n} \right)
\]
to get:
\[
\Gamma(mp + r) = \Gamma \left( k \left( \frac{m}{l} + \frac{r}{k} \right) \right) = (2\pi)^{(1-k)/2} k^{mk/l + r - 1/2} \prod_{j=1}^k \Gamma \left( \frac{m}{l} + \frac{r + j - 1}{k} \right),
\]
\[
\Gamma(m + 1) = \Gamma \left( \frac{m + 1}{l} \right) = (2\pi)^{(1-l)/2} m^{m+1/2} \prod_{j=1}^l \Gamma \left( \frac{m + j}{l} \right)
\]
and

$$
\Gamma(mp - m + r + 1) = \Gamma \left( (k - l) \left( \frac{m}{l} + \frac{r + 1}{k - l} \right) \right) = (2\pi)^{(1-k+l)/2} (k - l)^{m(k-l)/l+r+1/2} \prod_{j=l+1}^{k} \Gamma \left( \frac{m}{l} + \frac{r + j - l}{k - l} \right).
$$

It remains to use them in (18).

In order to apply Lemma 1.1 we need to modify enumeration of $\alpha$’s.

**Lemma 2.2.** For $1 \leq i \leq l + 1$ denote

$$
j_i := \left\lfloor \frac{(i-1)k}{l} \right\rfloor + 1,
$$

where $\lfloor \cdot \rfloor$ is the floor function, so that

$$
1 = j_1 < j_2 < \ldots < j_l < k < k + 1 = j_{l+1}.
$$

For $1 \leq j \leq k$ define

$$
\tilde{\alpha}_j = \begin{cases} 
i & \text{if } j = j_i, 1 \leq i \leq l, \\
\frac{r + j - i}{k - l} & \text{if } j_i < j < j_{i+1}. 
\end{cases}
$$

Then the sequence $\{\tilde{\alpha}_j\}_{j=1}^{k}$ is a rearrangement of $\{\alpha_j\}_{j=1}^{k}$. Moreover, if $0 < r \leq p = k/l$ then we have $\beta_j \leq \tilde{\alpha}_j$ for all $j \leq k$.

**Proof.** It is easy to verify the first statement. Assume that $j = j_i$ for some $i \leq l$. We have to show that

$$
\frac{r + j_i - 1}{k} \leq \frac{i}{l},
$$

which is equivalent to

$$
lr + l \left[ \frac{k(i-1)}{l} \right] \leq ki.
$$

The latter is a consequence of the fact that $\left\lfloor x \right\rfloor \leq x$ and the assumption that $r \leq p = k/l$.

Now assume that $j_i < j < j_{i+1}$. We ought to show that

$$
\frac{r + j - 1}{k} \leq \frac{r + j - i}{k - l},
$$

which is equivalent to

$$
lr + lj + k - l - ki \geq 0.
$$
Using the inequality $\lfloor x \rfloor + 1 > x$ we obtain

$$lj + k - l - ki \geq l(j_i + 1) + k - l - ki$$

which completes the proof, as $r > 0$.

Now we are ready to prove the main theorem of this section.

**Theorem 2.3.** Suppose that $p = k/l$, where $k,l$ are integers, $1 \leq l < k$, and that $r$ is a real number such that $0 < r \leq p$. Then there exists a unique probability measure $\mu(p,r)$ such that (1) is its moment sequence. Moreover $\mu(p,r)$ can be represented as the following Mellin convolution:

$$\mu(p,r) = b(\tilde{\alpha}_1, \beta_1, l) \circ \ldots \circ b(\tilde{\alpha}_k, \beta_k, l) \circ \delta_{c(p)},$$

where

$$c(p) := \frac{p^p}{(p - 1)^{p-r}}.$$

Consequently, $\mu(p,r)$ is absolutely continuous and its support is $[0, c(p)]$.

It is easy to see that the density function is positive on $(0, c(p))$. The representation of densities in the form of Mellin convolution of modified beta measures was used in different context in [8], see its Appendix A.

**Example.** For the Marchenko-Pastur measure we get the following decomposition:

$$\mu(2,1) = b(1, 1/2, 1) \circ b(2, 1, 1) \circ \delta_1,$$

where $b(1, 1/2, 1)$ has density $1/(\pi \sqrt{x - x^2})$ on $[0, 1]$, the arcsine distribution with the moment sequence $(\frac{j^m}{m!})$ for $m = 1$, and $b(2, 1, 1)$ is the Lebesgue measure on $[0, 1]$ with the moment sequence $1/(m + 1)$.

**Proof.** In view of Lemma 2.1 and Lemma 2.2 we can write

$$A_m(p,r) = D \prod_{j=1}^k \frac{\Gamma(\beta_j + m/l) \Gamma(\tilde{\alpha}_j)}{\Gamma(\alpha_j + m/l) \Gamma(\beta_j)} \cdot c(p)^m$$

for some constant $D$. Taking $m = 0$ we see that $D = 1$.

Note that a part of the theorem illustrates a result of Kargin [13], who proved that if $\mu$ is a compactly supported probability measure on $[0, \infty)$, with expectation 1 and variance $V$, and if $L_n$ denotes the supremum of the support of the multiplicative free convolution power $\mu^\otimes n$, then

$$\lim_{n \to \infty} \frac{L_n}{n} = eV,$$

where $e = 2.71 \ldots$ is the Euler’s number. The Marchenko-Pastur measure $\mu(2,1)$ has expectation and variance equal to 1 and $\mu(2,1)^\otimes n = \mu(n + 1, 1)$, so
in this case $L_n = (n+1)^{n+1}/n^n$ (this was also proved in [29] and [11]) and (21) holds.

The density function for $\mu(p, r)$ will be denoted by $W_{p,r}(x)$. Since $A_m(p, p) = A_{m+1}(p, 1)$, we have

$$W_{p,p}(x) = x \cdot W_{p,1}(x),$$

for example

$$W_{2,2}(x) = \frac{1}{2\pi} \sqrt{x(4-x)} \quad \text{on } [0,4],$$

which is the semicircle Wigner distribution with radius 2, centered at $x = 2$.

Now we can reprove the main result of [16].

**Theorem 2.4.** Suppose that $p, r$ are real numbers satisfying $p \geq 1$, $0 \leq r \leq p$. Then there exists a unique probability measure $\mu(p, r)$, with compact support contained in $[0, c(p)]$, such that $\{A_m(p, r)\}_{m=0}^{\infty}$ is its moment sequence.

**Proof.** It follows from the fact that the class of positive definite sequence is closed under pointwise limits.

**Remark.** In view of Theorem 2.1 in [5], for every $p > 1$ the measure $\mu(p, 1)$ is absolutely continuous and its support is equal $[0, c(p)]$, see also [14, 12].

### 3 Applying Meijer $G$-function

The aim of this section is to describe the density function $W_{p,r}(x)$ of $\mu(p, r)$ in terms of the Meijer $G$-function (see [19] for example) and consequently, as a linear combination of generalized hypergeometric functions. We will see that in some particular cases $W_{p,r}$ can be represented as an elementary function.

For $p > 1$, $r > 0$ define an analytic function

$$\phi_{p,r}(\sigma) = \frac{r \Gamma((\sigma-1)p + r)}{\Gamma(\sigma)\Gamma((\sigma-1)(p-1) + r + 1)},$$

which is well defined whenever $(\sigma-1)p + r$ is not a nonpositive integer. Note that $\phi_{p,1}(\sigma + 1) = \phi_{p,p}(\sigma)$ and if $m$ is a natural number then

$$\phi_{p,r}(m+1) = \binom{mp + r}{m} \frac{r}{mp + r}.$$

Then we define $W_{p,r}$ as the inverse Mellin transform:

$$W_{p,r}(x) = \frac{1}{2\pi i} \int_{d-i\infty}^{d+i\infty} x^{-\sigma} \phi_{p,r}(\sigma) d\sigma,$$

$x > 0$, if exists, see [25] for details. It turn out that if $p > 1$ is a rational number then $W_{p,r}$ can be expressed in terms of the Meijer $G$-function and its Mellin transform is $\phi_{p,r}$. For the theory of the Meijer $G$-functions we refer to [15, 23, 19].
Densities of the Raney Distributions

Theorem 3.1. Suppose that $p = k/l$, where $k, l$ are integers, $1 \leq l < k$ and $r > 0$. Then $W_{p, r}(x)$ is well defined and

$$W_{p, r}(x) = \frac{rp^r}{x(p-1)^{r+1/2}2\pi} G_k^{k,0}_{k,0} \left( \frac{x^\sigma}{c(p)^l} \bigg| \alpha_1, \ldots, \alpha_k \right), \quad (24)$$

$x \in (0, c(p))$, where $c(p) = p^p(p-1)^{1-p}$ and the parameters $\alpha_j, \beta_j$ are given by (16) and (17). Moreover, $\phi_{p, r}$ is the Mellin transform of $W_{p, r}$, namely

$$\phi_{p, r}(\sigma) = \int_0^c c(p) x^{\sigma-1} W_{p, r}(x) \, dx, \quad \sigma > 1 - r/p.$$  

for $\Re \sigma > 1 - r/p$.

If $0 < r \leq p$ then $W_{p, r}(x) > 0$ for $0 < x < c(p)$ and therefore $W_{p, r}$ is the density function of the probability distribution $\mu(p, r)$.

Proof. Putting $m = \sigma - 1$ in (15) we get

$$\phi_{p, r}(\sigma) = \frac{r(p-1)^{p-r-3/2}}{lp^{p-r}2\pi} \prod_{j=1}^k \Gamma(\beta_j - 1/l + \sigma/l) \prod_{j=1}^k \Gamma(\alpha_j - 1/l + \sigma/l) c(p)^\sigma. \quad (26)$$

Writing the right hand side as $\Phi(\sigma/l - 1/l)c(p)^\sigma$, using the substitution $\sigma = lu + 1$ and the definition of the Meijer $G$-function (see [19] for example), we obtain

$$W_{p, r}(x) = \frac{1}{2\pi i} \int_{d-i\infty}^{d+i\infty} \Phi(\sigma/l - 1/l)c(p)^\sigma x^{-\sigma} \, d\sigma$$

$$= \frac{lc(p)}{2\pi x_1} \int_{d-i\infty}^{d+i\infty} \Phi(u) \left( x'/c(p)^l \right)^{-u} \, du$$

$$= \frac{rp^r}{x(p-1)^{r+1/2}2\pi} G_k^{k,0}\left( \frac{x^\sigma}{c(p)^l} \bigg| \alpha_1, \ldots, \alpha_k \bigg| \beta_1, \ldots, \beta_k \right),$$

where $z = x'/c(p)^l$. Recall that for the Meijer function of type $G_k^{k,0}$ there is no restriction on the parameters and the integral converges for $0 < x < c(p)$ (see 16.17.1 in [19]).

On the other hand, substituting $x = c(p)t^{1/l}$ we can write

$$\int_0^{c(p)} x^{\sigma-1} W_{p, r}(x) \, dx$$

$$= \frac{rp^r}{(p-1)^{r+1/2}2\pi} \int_0^{c(p)} x^{\sigma-2} G_k^{k,0}\left( \frac{x^\sigma}{c(p)^l} \bigg| \alpha_1, \ldots, \alpha_k \bigg| \beta_1, \ldots, \beta_k \right) \, dx,$$

$$= \frac{rp^r c(p)^{\sigma-1}}{l(p-1)^{r+1/2}2\pi} \int_0^1 t^{(\sigma-1)/l-1} G_k^{k,0}\left( \frac{\alpha_1, \ldots, \alpha_k}{\beta_1, \ldots, \beta_k} \right) \, dt.$$
Since $\sum_{j=1}^{k} (\beta_j - \alpha_j) = -3/2 < 0$, so the assumptions of (2.24.2.1) in [23], the third case, are satisfied and therefore the last integral is convergent provided

$$-\frac{r}{k} = -\min \beta_j < |\sigma - 1|/l,$$

(equivalently: $\Re\sigma > 1 - r/p$) and the whole expression is equal to the right hand side of (26).

For the last statement we note that in view of Theorem 2.3, of the uniqueness part of the Riesz representation theorem for linear functionals on $C[0, c(p)]$ and of the Weierstrass approximation theorem, for $0 < r \leq p$ the density function of $\mu(p, r)$ must coincide with $W_{p,r}$.

Now applying Slater’s formula we can express $W_{p,r}$ as a linear combination of hypergeometric functions.

**Theorem 3.2.** For $p = k/l$, with $1 \leq l < k$, $r > 0$, and $x \in (0, c(p))$ we have

$$W_{p,r}(x) = \frac{\gamma(k,l,r) \sum_{h=1}^{k} c(h,k,l,r) \mathcal{K}_{k-1} \left( a(h,k,l,r) \bigg| z \right) z^{(r+h-1)/k-1/l}}{\sqrt{2k\pi}} G_{k,0} \left( z \bigg| \alpha_1, \ldots, \alpha_k \bigg),$$

where $z = x^l/c(p)^l$,

$$\gamma(k,l,r) = \frac{r(p-1)^{p-r-3/2}}{p^{p-r-2k}} \cdot \left( \prod_{j=1}^{h-1} \Gamma \left( \frac{j-h}{l} \right) \prod_{j=h+1}^{k} \Gamma \left( \frac{j-h}{k} \right) \right) \cdot \left( \prod_{j=1}^{l} \Gamma \left( \frac{j-h}{l} \right) \prod_{j=l+1}^{k} \Gamma \left( \frac{j-h}{k} \right) \right),$$

and the parameter vectors of the hypergeometric functions are

$$a(h,k,l,r) = \left( \left\{ \frac{r + h - 1}{k} - \frac{j-l}{l} \right\}_{j=1}^{l} \right),$$

$$b(h,k,l,r) = \left( \left\{ \frac{k + h - j}{k} \right\}_{j=1}^{h-1}, \left\{ \frac{k + h - j}{k} \right\}_{j=h+1}^{k} \right).$$

**Proof.** Putting $z = x^l/c(p)^l$, and hence $x = c(p)z^{1/l}$, we can rewrite (24) as

$$W_{p,r}(x) = \frac{r(p-1)^{p-r-3/2}}{z^{1/l}p^{p-r-2k}\sqrt{2k\pi}} \mathcal{K}_{k,0} \left( z \bigg| \alpha_1, \ldots, \alpha_k \bigg),$$

$x \in (0, c(p))$. Observe that for $1 \leq i < j \leq k$ the difference $\beta_j - \beta_i = (j-i)/k$ is never an integer. Therefore we can apply formula (8.2.2.3) in [23] (see also (16.17.2) in [19] or formula (7) on page 145 in [15]), so that

$$c(h,k,l,r) = \frac{\prod_{j \neq h} \Gamma(\beta_j - \beta_h)}{\prod_{j=1}^{k} \Gamma(\alpha_j - \beta_h)}.$$
which gives (29). For the parameter vectors we have
\[ a(h, k, l, r) = 1 + \beta_h - \alpha_j \]
and
\[ b(h, k, l, r) = 1 + \beta_h - \beta_j, \quad j \neq h, \]
which leads to (30) and (31). Finally, the summand with index \( h \) is in addition multiplied by \( z^{\beta_h - 1/l} \).

Theorem 3.1 and Theorem 3.2 are sufficient for drawing graphs of the functions \( W_{p, r} \) with help of computer programs. In some cases however it is possible to express \( W_{p, r} \) as an elementary function. The most tractable case is \( p = 2 \). We know already that
\[ W_{2,1}(x) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}}, \quad W_{2,2}(x) = \frac{1}{2\pi} \sqrt{x(4-x)}. \]

Now we can give a simple formula for \( W_{2, r} \).

**Corollary 3.3.** For \( p = 2, \ r > 0 \), the function \( W_{2, r} \) is
\[ W_{2, r}(x) = \frac{\sin \left( r \cdot \arccos \sqrt{\frac{x}{4}} \right)}{\pi x^{1-r/2}}, \quad (33) \]
for \( x \in (0, 4) \). If \( 0 < r \leq 2 \) then \( W_{2, r} \) is the density function of the measure \( \mu(2, r) \). In particular for \( r = 1/2 \) and \( r = 3/2 \) we have
\[ W_{2,1/2}(x) = \frac{\sqrt{2} - \sqrt{x}}{2\pi x^{3/4}}, \quad (34) \]
\[ W_{2,3/2}(x) = \frac{(\sqrt{x} + 1) \sqrt{2 - \sqrt{x}}}{2\pi x^{1/4}}. \quad (35) \]

Note that if \( r > 2 \) then \( W_{2, r}(x) < 0 \) for some values of \( x \in (0, 4) \).

**Proof.** We take \( k = 2, \ l = 1 \) so that \( c(2) = 4, \ z = x/4 \) and \( \gamma(2, 1, r) = r 2^r/(8\sqrt{\pi}) \). Using the Euler’s reflection formula and the identity \( \Gamma(1 + r/2) = \Gamma(r/2)r/2 \) we get
\[ c(1, 2, 1, r) = \frac{\Gamma(1/2)}{\Gamma(1-r/2)\Gamma(1+r/2)} = \frac{2\sin(\pi r/2)}{r\sqrt{\pi}}, \]
\[ c(2, 2, 1, r) = \frac{\Gamma(-1/2)}{\Gamma((1-r)/2)\Gamma((1+r)/2)} = -\frac{2\cos(\pi r/2)}{\sqrt{\pi}}. \]
We also need formulas for two hypergeometric functions, namely
\[
2F_1\left(\frac{r}{2}, -\frac{r}{2}; \frac{1}{2}; z\right) = \cos(r \arcsin \sqrt{z}),
\]
\[
2F_1\left(\frac{1+r}{2}, \frac{1-r}{2}; \frac{3}{2}; z\right) = \frac{\sin(r \arcsin \sqrt{z})}{r \sqrt{z}},
\]
see 15.4.12 and 15.4.16 in [19]. Now we can write
\[
W_{2,r}(x) = \sin\left(\pi r/2\right) \cos(r \arcsin \sqrt{x/4}) - \cos\left(\pi r/2\right) \sin(r \arcsin \sqrt{x/4}) = \frac{\sin\left(\pi r/2\right)}{4\pi x^{1-r/2}} - \frac{\sin\left(r \arcsin \sqrt{x/4}\right)}{\pi x^{1-r/2}}.
\]
For the special cases we use the identity \(\sin\left(\frac{1}{2} \arccos(t)\right) = \sqrt{\frac{1-t}{2}}\), which is valid for \(0 \leq t \leq 1\).

Remark. Note that
\[
\frac{W_{2,1}(\sqrt{x})}{2\sqrt{x}} = \frac{1}{4} W_{2,1/2}(\frac{x}{4}) = \frac{\sqrt{4 - \sqrt{x}}}{4\pi x^{3/4}}.
\] (36)

It means that if \(X, Y\) are random variables such that \(X \sim \mu(2,1)\) and \(Y \sim \mu(2,1/2)\) then \(X^2 \sim 4Y\). This can be also derived from the relation \(A_0(2,1/2)4^m = A_{2m}(2,1) = (3^{2n+1})/(4n + 1)\), A048990 in OEIS [24]. Hence A048990 is the moment sequence of the density function (36), \(x \in (0,16)\).

4 Some particular cases

In this part we will see that for \(k = 3\) some densities still can be represented as elementary functions. We need two families of formulas (cf. 15.4.17 in [19]).

Lemma 4.1. For \(c \neq 0, -1, -2, \ldots\) we have
\[
2F_1\left(\frac{c}{2}, \frac{c-1}{2}; c; z\right) = 2^{c-1}(1 + \sqrt{1-z})^{1-c},
\] (37)
\[
2F_1\left(\frac{c+1}{2}, \frac{c-2}{2}; c; z\right) = \frac{c-1}{c}(1 + \sqrt{1-z})^{1-c}(c-1 + \sqrt{1-z}).
\] (38)

Proof. We know that \(2F_1(a, b; c; z)\) is the unique function \(f\) which is analytic at \(z = 0\), with \(f(0) = 1\), and satisfies the hypergeometric equation:
\[
z(1-z)f''(z) + \left[c - (a + b + 1)z\right]f'(z) - abf(z) = 0
\]
(see [3]). Now one can check that this equation is satisfied by the right hand sides of (37) and (38) for given parameters \(a, b, c\).
Now consider $p = 3/2$.

**Theorem 4.2.** Assume that $p = 3/2$. Then for $r = 1/2, 1, 3/2$ we have

$$W_{3/2,1/2}(x) = \frac{(1 + \sqrt{1 - 4x^2/27})^{2/3}}{2^{5/3} x^{1/3}} = \left(1 - \sqrt{1 - 4x^2/27}\right)^{2/3}$$

(39)

$$W_{3/2,1}(x) = 3^{1/2} \frac{1 + \sqrt{1 - 4x^2/27}}{2^{1/3} x^{1/3}}$$

(40)

$$+ \frac{3^{1/2} x^{1/3}}{2^{1/3} x^{1/3}} \left(1 - \sqrt{1 - 4x^2/27}\right)^{2/3}$$

and, finally, $W_{3/2,3/2}(x) = x \cdot W_{3/2,1}(x)$, with $x \in (0, 3\sqrt{3}/2)$.

**Proof.** For arbitrary $r$ we have

$$W_{3/2,r}(x) = \frac{2^{1-2r/3} \sin \left(\frac{2\pi r}{3}\right)}{3^{1/2-r \pi}} \left(\begin{array}{c} 2 + 2r \\ 6 \\ z \\ 3 \\ \end{array}\right) \left(\begin{array}{c} 5 + 2r \\ 6 \\ z \\ 3 \\ \end{array}\right) \left(\begin{array}{c} 7 + 2r \\ 6 \\ z \\ 3 \\ \end{array}\right)$$

where $z = 4x^2/27$. If $r = 1/2$ or $r = 1$ then one term vanishes and in the two others the hypergeometric functions reduce to $\left(\begin{array}{c} 3 \\ 3 \\ \end{array}\right)$.

For $r = 1/2$ we apply (37) to obtain:

$$W_{3/2,1/2}(x) = \frac{z^{1/3}}{2^{1/3} 3^{1/2} \pi} 2 \left(\begin{array}{c} 1 \\ 6 \\ -1 \\ 3 \\ z \\ \end{array}\right) - \frac{z^{1/3}}{2^{1/3} 3^{1/2} \pi} 2 \left(\begin{array}{c} 5 \\ 6 \\ -2 \\ 3 \\ z \\ \end{array}\right)$$

and this yields (39).

For $r = 1$ we use (38):

$$W_{3/2,1}(x) = \frac{z^{1/6}}{2^{1/3} \pi} 2 \left(\begin{array}{c} 5 \\ 6 \\ -2 \\ 3 \\ z \\ \end{array}\right) + \frac{z^{1/6}}{2^{1/3} \pi} 2 \left(\begin{array}{c} 7 \\ 6 \\ -1 \\ 3 \\ z \\ \end{array}\right)$$

Documenta Mathematica 18 (2013) 1573–1596
\[ W = \frac{z^{-1/6}}{4\pi} (1 + \sqrt{1 - z})^{1/3} (3\sqrt{1 - z} - 1) + \frac{z^{1/6}}{4\pi} (1 + \sqrt{1 - z})^{-1/3} (3\sqrt{1 - z} + 1) \]
\[ = \frac{z^{-1/6}}{4\pi} (1 + \sqrt{1 - z})^{1/3} (3\sqrt{1 - z} - 1) + \frac{z^{-1/6}}{4\pi} (1 - \sqrt{1 - z})^{1/3} (3\sqrt{1 - z} + 1). \]

Now we have
\[ (1 + \sqrt{1 - z})^{1/3} (3\sqrt{1 - z} - 1) = - (1 + \sqrt{1 - z})^{1/3} (3 - 3\sqrt{1 - z} - 2) \]
\[ = -3z^{1/3} (1 - \sqrt{1 - z})^{2/3} + 2 (1 + \sqrt{1 - z})^{1/3} \]
and similarly
\[ (1 - \sqrt{1 - z})^{1/3} (3\sqrt{1 - z} + 1) = 3z^{1/3} (1 + \sqrt{1 - z})^{2/3} - 2 (1 - \sqrt{1 - z})^{1/3}. \]

Therefore
\[ W_{3/2,1}(x) = \frac{z^{-1/6}}{2\pi} \left( (1 + \sqrt{1 - z})^{1/3} - (1 - \sqrt{1 - z})^{1/3} \right) \]
\[ + \frac{3z^{1/6}}{4\pi} \left( (1 + \sqrt{1 - z})^{2/3} - (1 - \sqrt{1 - z})^{2/3} \right), \]
which entails (40). The last statement is a consequence of (22).

The dilation \( D_2 \mu(3/2, 1/2) \), with the density \( W_{3/2,1/2}(x/2)/2 \), is known as the Bures distribution, see (4.4) in [26]. The integer sequence
\[ 4^m A_m(3/2, 1/2) = \binom{3m/2 + 1/2}{n} \frac{4^m}{3m + 1}, \]
moments of the density function \( W_{3/2,1/2}(x/4)/4 \) on the interval \((0, 6\sqrt{3})\), appears as A078531 in [24] and counts the number of symmetric noncrossing connected graphs on \( 2n + 1 \) equidistant nodes on a circle. The axis of symmetry is a diameter of a circle passing through a given node, see [7].

The measure \( \mu(3/2, 1) \) is equal to \( \mu(2, 1)^{G_1/2} \), the multiplicative free square root of the Marchenko-Pastur distribution and the integer sequence
\[ 4^m A_m(3/2, 1) = \binom{3m/2 + 1}{n} \frac{4^m}{3m/2 + 1}, \]
moments of the density function \( W_{3/2,1}(x/4)/4 \) on \((0, 6\sqrt{3})\), appears in [24] as A214377.

For the sake of completeness we also include the densities for the sequences \( A_m(3, 1) \) (A001764 in [24]) and \( A_m(3, 2) \) (A006013), which have already appeared in [20, 21].
Proof. Recall that the measure $\mu$ is equal to $\mu(2, 1)^{\otimes 2}$, the multiplicative free square of the Marchenko-Pastur distribution.

**Theorem 4.3.** Assume that $p = 3$. Then for $r = 1, 2, 3$ we have

$$W_{3,1}(x) = \frac{3 \left(1 + \sqrt{1 - 4x/27} \right)^{2/3} - 2x^{1/3}}{2^{4/3}3^{1/2}\pi x^{2/3} \left(1 + \sqrt{1 - 4x/27} \right)^{1/3}}, \quad (41)$$

$$W_{3,2}(x) = \frac{9 \left(1 + \sqrt{1 - 4x/27} \right)^{4/3} - 24x^{2/3}}{2^{5/3}3^{3/2}\pi x^{1/3} \left(1 + \sqrt{1 - 4x/27} \right)^{2/3}}, \quad (42)$$

and, finally, $W_{3,3}(x) = x \cdot W_{3,1}(x)$, with $x \in (0, 27/4)$.

**Proof.** For arbitrary $r$ we have

$$W_{3,r}(x) = \frac{2^{(6-2r)/3} \sin \left(\pi r/3 \right)}{3^{3-r} \pi} \, {}_3F_2 \left( \begin{array}{c} \frac{r}{3}, \frac{3-r}{6}, \frac{1}{3} \\ \frac{2}{3}, \frac{1}{3} \end{array} \right) \left( z \right)^{(r-3)/3} - \frac{2^{(4-2r)/3} \sin \left(\left(1 + r\right)\pi/3 \right)}{3^{3-r} \pi} \, {}_3F_2 \left( \begin{array}{c} \frac{1+r}{3}, \frac{5-r}{6}, \frac{2-r}{3} \\ \frac{4}{3}, \frac{2}{3} \end{array} \right) \left( z \right)^{(r-2)/3} + \frac{r(r-1) \sin \left(\left(1 - r\right)\pi/3 \right)}{2^{1+2r}3^{3/2} \pi} \, {}_3F_2 \left( \begin{array}{c} \frac{2+r}{3}, \frac{7-r}{6}, \frac{4-r}{3} \\ \frac{5}{3}, \frac{4}{3} \end{array} \right) \left( z \right)^{(r-1)/3},$$

where $z = 4x/27$. For $r = 1$ and $r = 2$ we have similar reduction as in the previous proof. Here we will be using only (37).

Taking $r = 1$ we get

$$W_{3,1}(x) = \frac{2^{1/3} \pi^{-2/3}}{3^{1/2} \pi} \, {}_2F_1 \left( \begin{array}{c} \frac{1}{3}, \frac{1}{6} \\ \frac{2}{3} \end{array} \right) \left( z \right) - \frac{z^{-1/3}}{2^{1/3}3^{1/2} \pi} \, {}_2F_1 \left( \begin{array}{c} \frac{5}{6}, \frac{1}{3} \\ \frac{4}{3} \end{array} \right) \left( z \right)$$

$$= \frac{z^{-2/3}}{3^{3/2} \pi} \left( 1 + \sqrt{1 - z} \right)^{1/3} - \frac{z^{-1/3}}{3^{3/2} \pi} \left( 1 + \sqrt{1 - z} \right)^{-1/3}$$

$$= \frac{(1 + \sqrt{1 - z})^{2/3} - z^{1/3}}{3^{3/2} \pi z^{2/3} \left( 1 + \sqrt{1 - z} \right)^{1/3}},$$

which implies (41).

Now we take $r = 2$:

$$W_{3,2}(x) = \frac{z^{-1/3}}{2^{1/3}3^{1/2} \pi} \, {}_2F_1 \left( \begin{array}{c} \frac{1}{3}, \frac{1}{3} \\ \frac{2}{3} \end{array} \right) \left( z \right) - \frac{z^{1/3}}{2^{5/3}3^{1/2} \pi} \, {}_2F_1 \left( \begin{array}{c} 5/6, 1/3 \\ 5/3, 5/3 \end{array} \right) \left( z \right)$$

$$= \frac{z^{-1/3}}{2 \cdot 3^{1/2} \pi} \left( 1 + \sqrt{1 - z} \right)^{2/3} - \frac{z^{1/3}}{2 \cdot 3^{1/2} \pi} \left( 1 + \sqrt{1 - z} \right)^{-2/3}$$

$$= \frac{(1 + \sqrt{1 - z})^{1/3} - z^{2/3}}{2 \cdot 3^{1/2} \pi z^{1/3} \left( 1 + \sqrt{1 - z} \right)^{2/3}},$$

and this gives us (42). Finally we apply (22). □

Recall that the measure $\mu(3, 1)$ is equal to $\mu(2, 1)^{\otimes 2}$, the multiplicative free square of the Marchenko-Pastur distribution.
Figure 1: Raney distributions $W_{3/2,r}(x)$ with values of the parameter $r$ labeling each curve. For $r > p$ solutions drawn with dashed lines are not positive.

5 Graphical representation of selected cases

The explicit form of $W_{p,r}(x)$ given in Theorem 3.2 permits a graphical visualization for any rational $p > 0$ and arbitrary $r > 0$. We shall represent some selected cases in Figures 1–9. These graphs which are partly negative are drawn as dashed curves. In Fig. 1 the graphs of the functions $W_{3/2,r}(x)$ for values of $r$ ranging from 0.9 to 2.3 are given. For $r \leq 3/2$ these functions are positive, otherwise they develop a negative part. In Fig. 2 we represent $W_{3/2, r}(x)$ for $r$ ranging from 2 to 3.4. In Fig. 3 we display the densities $W_{p,p}(x)$ for $p = 6/5, 5/4, 4/3$ and 3/2. All these densities are unimodal and vanish at the extremities of their supports. They can be therefore considered as generalizations of the Wigner’s semicircle distribution $W_{2,2}(x)$, see equation (23). In Fig. 4 we depict the functions $W_{4/3, r}(x)$, for values $r$ ranging from 0.8 to 2.4. Here for $r \geq 1.4$ negative contributions clearly appear. In Fig. 5 and 6 we present six densities expressible through elementary functions, namely $W_{3/2, r}(x)$ for $r = 1/2, 1.3/2$, see Theorem 4.2 and $W_{3, r}(x)$ for $r = 1, 2, 3$, see Theorem 4.3. In Fig. 7 the set of densities $W_{p,1}(x)$ for five fractional values of $p$ is presented. The appearance of maximum near $x = 1$ corresponds to the fact that $\mu(p,1)$ weakly converges to $\delta_1$ as $p \to 1^+$. In Fig. 8 the fine details of densities $W_{p,1}(x)$ for $p = 5/2, 7/3, 9/4, 11/5$, on a narrower range $2 \leq x \leq 4.5$ are presented. In Fig. 9 we display the densities $W_{p,1}(x)$ for $p = 2, 5/2, 3, 7/2, 4$, near the upper edge of their respective supports, for $3 \leq x \leq 9.5$. 
Figure 2: As in Fig. 1 for Raney distributions $W_{5/2,r}(x)$.

Figure 3: Diagonal Raney distributions $W_{p,p}(x)$ with values of the parameter $p$ labeling each curve.
Figure 4: The functions $W_{4/3,r}(x)$ for $r$ ranging from 0.8 to 2.4.

Figure 5: Raney distributions $W_{3/2,r}(x)$ with values of the parameter $r$ labeling each curve. The case $W_{3/2,1}(x)$ represents $MP^{21/2}$, the multiplicative free square root of the Marchenko-Pastur distribution.
Figure 6: Raney distributions $W_{3,r}(x)$ with values of the parameter $r$ labeling each curve. The case $W_{3,1}(x)$ represents $MP^2$, the multiplicative free square of the Marchenko-Pastur distribution.

Figure 7: Raney distributions $W_{p,1}(x)$ with values of the parameter $p$ labeling each curve. The case $W_{3/2,1}(x)$ represents the multiplicative free square root of the Marchenko–Pastur distribution, $MP^{3/2}$. 
Figure 8: Tails of the Raney distributions $W_{p,1}(x)$ with values of the parameter $p$ labeling each curve.

Figure 9: As in Fig. 8 for larger values of the parameter $p$. 
Figure 10: Parameter plane \((p, r)\) describing the Raney numbers. The shaded set \(\Sigma\) corresponds to nonnegative probability measures \(\mu(p, r)\). The vertical line \(p = 2\) and the stars represent values of parameters for which \(W_{p,r}(x)\) is an elementary function. Here MP denotes the Marchenko–Pastur distribution, \(\text{MP}^{s}\) its \(s\)-th free multiplicative power, B-the Bures distribution while SC denotes the semicircle law. For \(p > 1\) the points \((p, p)\) on the upper edge of \(\Sigma\) represent the generalizations of the Wigner semicircle law, see Fig. 3.

The Fig. 10 summarizes our results in the \(p > 0, r > 0\) quadrant of the \((p, r)\) plane, describing the Raney numbers (c.f. Fig. 5.1 in [16] and Fig. 7 in [21]). The shaded region \(\Sigma\) indicates the probability measures \(\mu(p, r)\) (i.e. where \(W_{p,r}(x)\) is a nonegative function). The vertical line \(p = 2\) and the stars indicate the pairs \((p, r)\) for which \(W_{p,r}(x)\) is an elementary function, see Corollary 3.3, Theorem 4.2 and Theorem 4.3. The points \((3/2, 1)\) and \((3, 1)\) correspond to the multiplicative free powers \(\text{MP}^{1/2}\) and \(\text{MP}^{2}\) of the Marchenko-Pastur distribution MP. Symbol B at \((3/2, 1/2)\) indicates the Bures distribution and SC at \((2, 2)\) denotes the semicircle law centered at \(x = 2\), with radius 2.

It is our pleasure to thank M. Bożejko, Z. Burda, K. Górska, I. Nechita and M. A. Nowak for fruitful interactions.

References


Densities of the Raney Distributions


Abstract. Supramenability of groups is characterised in terms of invariant measures on locally compact spaces. This opens the door to constructing interesting crossed product $C^*$-algebras for non-supramenable groups. In particular, stable Kirchberg algebras in the UCT class are constructed using crossed products for both amenable and non-amenable groups.

2010 Mathematics Subject Classification: 43A07, 46L55, 46L35
Keywords and Phrases: Supramenable groups, actions on locally compact spaces, purely infinite $C^*$-algebras and actions.

1. Introduction
A group $\Gamma$ is called amenable if it carries an invariant finitely additive measure $\mu$ with $0 < \mu(\Gamma) < \infty$. This concept introduced by von Neumann [23] was originally motivated by the Hausdorff–Banach–Tarski paradox [10] and has become central in many aspects of group theory and beyond. However, from the point of view of paradoxical decompositions, it is more natural to consider the following question of von Neumann [23, §4]. Let $\Gamma$ be a group acting on a set $X$. Given a subset $A \subseteq X$, when is there an invariant finitely additive measure $\mu$ on $X$ with $0 < \mu(A) < \infty$?
Following Rosenblatt [18], the group $\Gamma$ is called supramenable if there is such a measure for every non-empty subset; it turns out that it is sufficient to consider the case $X = \Gamma$. Supramenability is much stronger than amenability: whilst it holds for commutative and subexponential groups, it fails already for metabelian groups (Example 2.8) and thus is not preserved under extensions. It passes to subgroups and quotients but is not known to be preserved under direct products.
One of the many equivalent characterizations of amenability is the existence of a non-zero invariant Radon measure on any compact Hausdorff $\Gamma$-space (see e.g. [16, 5.4]). It turns out that there is an analogous characterization of supramenability using locally compact spaces. We shall establish this and a couple more characterizations:

**Theorem 1.1.** The following conditions are equivalent for any group $\Gamma$.

(i) $\Gamma$ is supramenable.

(ii) Any co-compact $\Gamma$-action on a locally compact Hausdorff space admits a non-zero invariant Radon measure.

(iii) The Roe algebra $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$ contains no properly infinite projection.

(iv) There is no injective Lipschitz map from the free group $F_2$ to $\Gamma$.

(This also gives a partial answer to the questions raised after Theorem 6.3.1 in [19]. Regarding condition (iii), we recall that $\Gamma$ is amenable if and only if the Roe algebra is itself not properly infinite [17]. Theorem 1.1 is proved below in Propositions 2.7, 3.4 and 5.3.)

We shall leverage the failure of condition (ii) in order to provide interesting $C^*$-algebras through the corresponding reduced crossed product construction. To this end, we need to establish that the locally compact $\Gamma$-space can be assumed to have several additional properties, as follows. Recall that a Kirchberg algebra is a $C^*$-algebra which is simple, purely infinite, nuclear and separable. Kirchberg algebras in the UCT class are completely classified by their $K$-theory, see [15] and [12].

**Theorem 1.2.** Let $\Gamma$ be a countable group. Then $\Gamma$ admits a free, minimal, purely infinite action on the locally compact non-compact Cantor set $K^*$ if and only if $\Gamma$ is non-supramenable. The crossed product $C^*$-algebra $C_0(K^*) \rtimes_{\text{red}} \Gamma$ associated with any such action will be a stable, simple, purely infinite $C^*$-algebra.

If the non-supramenable group $\Gamma$ is amenable, then the associated crossed product $C^*$-algebra $C_0(K^*) \rtimes_{\text{red}} \Gamma$ will be a stable Kirchberg algebra in the UCT class.

In the above, an action is called purely infinite if every compact-open subset is paradoxical in a sense made precise in Definition 4.4 below. It is a strengthening of the failure of condition (ii) in Theorem 1.1. As for the non-compact Cantor set $K^*$, one can realise it e.g. as the usual (compact) Cantor set $K$ with a point removed, or as $\mathbb{N} \times K$.

In the second statement of Theorem 1.2, the amenability of $\Gamma$ is only used to deduce that the action on $K^*$ is amenable. But of course there are many amenable actions of non-amenable groups; therefore, we strive to produce more examples of $C^*$-algebras as above using crossed products for groups $\Gamma$ that are not necessarily amenable. Here is a summary of some progress in that direction.

**Theorem 1.3.** Let $\Gamma$ be a countable group.

(i) If $\Gamma$ contains an infinite exact subgroup, then $\Gamma$ admits a free minimal amenable action on $K$ or on $K^*$; necessarily on $K^*$ if $\Gamma$ is non-exact.
If $\Gamma$ contains an infinite exact subgroup of infinite index, then $\Gamma$ admits a free minimal amenable action on $K^*$.

(ii) If $\Gamma$ contains an element of infinite order, or if $\Gamma$ contains an infinite amenable subgroup of infinite index, then $\Gamma$ admits a free minimal amenable action on $K^*$ such that $K^*$ admits an invariant non-zero Radon measure. The associated crossed product $C_0(K^*) \rtimes_{\text{red}} \Gamma$ is a stably finite simple separable nuclear $\text{C}^*$-algebra in the UCT class.

(iii) If $\Gamma$ contains an exact non-supramenable subgroup, then $\Gamma$ admits a free minimal amenable purely infinite action on $K^*$. The associated crossed product $C_0(K^*) \rtimes_{\text{red}} \Gamma$ is a stable Kirchberg algebra in the UCT class.

Acknowledgements. Work supported in part by the Swiss National Science Foundation, the European Research Council, Danish National Research Foundation (DNRF) through the Centre for Symmetry and Deformation at University of Copenhagen, and The Danish Council for Independent Research, Natural Sciences.

2. Actions on locally compact spaces

It is well-known that a group is amenable if and only if whenever it acts on a compact Hausdorff space, then the space admits an invariant probability measure. We shall here characterise supramenable groups in a similar way by their actions on locally compact Hausdorff spaces.

Let us first recall the basic notions of comparison of subsets of a group $\Gamma$. If $A, B \subseteq \Gamma$, then write $A \sim_\Gamma B$ if there are finite partitions $\{A_j\}_{j=1}^n$ and $\{B_j\}_{j=1}^n$ of $A$ and $B$, respectively, and elements $t_1, t_2, \ldots, t_n \in \Gamma$ such that $A_j = t_j B_j$. Write $A \lessdot_\Gamma B$ if $A \sim_\Gamma B_0$ for some $B_0 \subseteq B$. A set $A \subseteq \Gamma$ is said to be paradoxical if there are disjoint subsets $A_0, A_1$ of $A$ such that $A \sim_\Gamma A_0 \sim_\Gamma A_1$.

Tarski’s theorem says that $A \subseteq \Gamma$ is non-paradoxical if and only if there is an invariant finitely additive measure $\mu$ defined on the entire power set, $P(\Gamma)$, of $\Gamma$, such that $\mu(A) = 1$. (Notice that the empty set is paradoxical.)

The following proposition is well-known. For the weaker case of functions with “$A$-bounded” support, see for example [8, Chap. 1].

Proposition 2.1. Let $\Gamma$ be a group and let $\mu$ be a finitely additive measure on $\Gamma$. Let $V_\mu$ be the subspace of $\ell^\infty(\Gamma)$ consisting of all $f \in \ell^\infty(\Gamma)$ such that $\mu(\text{supp}(f)) < \infty$. It follows that there is a unique positive linear functional $I_\mu: V_\mu \to \mathbb{C}$ such that $I_\mu(1_E) = \mu(E)$ for all $E \subseteq \Gamma$ with $\mu(E) < \infty$. If $\mu$ is $\Gamma$-invariant, then so is $I_\mu$.

Proof. Let $\mathcal{F}$ be the collection of subsets of $\Gamma$ of finite measure. For each $F \in \mathcal{F}$ let $V_F$ be the subspace of $\ell^\infty(\Gamma)$ consisting of all functions with support in $F$. The family $\{V_F\}$ is upwards directed with union $V_\mu$. Using that the set of bounded functions $F \to \mathbb{C}$ that take finitely many values is uniformly dense in $V_F$ it is easy to see that there is a unique positive (necessarily bounded) linear functional $I^{(F)}_\mu: V_F \to \mathbb{C}$ such that $I^{(F)}_\mu(1_E) = \mu(E)$ for all $E \subseteq F$. It follows
that there exists a unique positive (unbounded) linear functional \( I_\mu : V_\mu \to \mathbb{C} \)
that extends all the \( I^{(F)}_\mu \)'s.
When \( \mu \) is translation invariant, the invariance of \( I_\mu \) follows from uniqueness
of \( I_\mu \).

Let \( \Gamma \) be a group which acts on a locally compact Hausdorff space \( X \). We
say that the action is co-compact if there is a compact subset \( K \) of \( X \) such
that \( \bigcup_{t \in \Gamma} t.K = X \). Any minimal action of any group on any locally compact
Hausdorff space is automatically co-compact.

Observe that if \( K \) is as above, and if \( \lambda \) is a non-zero Radon measure on
\( X \), then \( 0 < \lambda(K) < \infty \).

If \( X \) is locally compact and Hausdorff and if \( \Gamma \) acts co-compactly on
\( X \), then there is a compact subset \( K \subseteq X \) such that
\( \bigcup_{t \in \Gamma} t.K^o = X \), where \( K^o \) denotes
the interior of \( K \). This follows easily from the fact that for each compact set
\( K' \subseteq X \) there is another compact set \( K \subseteq X \) such that
\( K' \subseteq K^o \).

Lemma 2.2. Let \( \Gamma \) be a group acting co-compactly on a locally compact Hausdorff space \( X \). For each compact subset \( K \) of \( X \), such that
\( \bigcup_{t \in \Gamma} t.K^o = X \), and for each \( x_0 \in X \), put
\[
A(K, x_0) = \{ t \in \Gamma \mid t.x_0 \in K \}.
\]
If \( A(K, x_0) \) is non-paradoxical in \( \Gamma \) for some \( K \) and \( x_0 \) as above, then there is
a non-zero \( \Gamma \)-invariant Radon measure \( \lambda \) on \( X \).

Proof. By Tarski’s theorem there exists an invariant finitely additive measure
\( \mu \) on \( \Gamma \) such that \( \mu(A(K, x_0)) = 1 \).
Each compact set \( L \subseteq X \) is contained in \( \bigcup_{s \in F} s.K \) for some finite subset \( F \) of
\( \Gamma \). Hence
\[
\mu \left( \{ t \in \Gamma \mid t.x_0 \in L \} \right) \leq \mu \left( \bigcup_{s \in F} s.A(K, x_0) \right) \leq |F|\mu(A(K, x_0)) < \infty.
\]

For each \( f \in C_c(X) \) let \( \hat{f} \in \ell^\infty(\Gamma) \) be given by \( \hat{f}(t) = f(t.x_0) \), \( t \in \Gamma \). In the notation of Proposition \( 2.1 \) it follows from \( 2.1 \) that \( \hat{f} \in V_\mu \). Let \( I_\mu : V_\mu \to \mathbb{C} \)
be the \( \Gamma \)-invariant functional associated with \( \mu \) constructed in Proposition \( 2.1 \).
Define \( \Lambda : C_c(X) \to \mathbb{C} \) by \( \Lambda(f) = I_\mu(\hat{f}) \). As \( I_\mu \) is \( \Gamma \)-invariant, so is \( \Lambda \). By Riesz’ representation theorem there is a Radon measure \( \lambda \) on \( X \) such that
\( \Lambda(f) = \int_X f d\lambda \). As \( \Lambda \) is \( \Gamma \)-invariant, so is \( \lambda \).

If \( f \in C_c(X) \) is such that \( f \geq 1_K \), then \( \hat{f} \geq 1_{A(K, x_0)} \), so \( \Lambda(f) = I_\mu(\hat{f}) \geq \mu(A(K, x_0)) = 1 \). This shows that \( \lambda(K) \geq 1 \), so \( \lambda \) is non-zero.

The action of a group \( \Gamma \) on itself given by left-multiplication extends to an
action of \( \Gamma \) on its beta-compactification, \( \beta \Gamma \). We shall refer to this action as
the canonical action of \( \Gamma \) on \( \beta \Gamma \).

Definition 2.3. Fix a subset \( A \) of \( \Gamma \). Denote by \( K_A \) its closure in \( \beta \Gamma \), and put
\[
X_A = \bigcup_{t \in \Gamma} t.K_A.
\]
As βΓ is Stonian (a.k.a. extremely disconnected), it follows that $K_A$ is compact-open in βΓ. Moreover, $X_A$ is an open Γ-invariant subset of βΓ. In particular, $X_A$ is locally compact and Hausdorff, and Γ acts co-compactly on $X_A$. Note also that $Γ \subseteq X_A$ (if $A \neq ∅$), and that $Γ = X_A$ if and only if $A$ is finite and non-empty.

We state below two easy lemmas that shall be used frequently in what follows.

**Lemma 2.4.** Let Γ be a group.

(i) Let A and B be subsets of Γ. If $A \cap B = ∅$, then $K_A \cap K_B = ∅$.

(ii) For any subset A of Γ we have $A = K_A \cap Γ$.

(iii) If A and B are subsets of Γ, then $A \subseteq B$ if and only if $K_A \subseteq K_B$.

(iv) If K is a compact-open subset of βΓ and if $A = K \cap Γ$, then $K = K_A$.

(v) If $A$ is a subset of Γ and $t \in Γ$, then $t.K_A = K_{tA}$.

(vi) $βΓ$ and ∅ are the only Γ-invariant compact-open subset of βΓ.

**Proof.** (i). As A is an open subset of βΓ we have $K_B \subseteq βΓ \setminus A$. Hence $A \subseteq βΓ \setminus K_B$, which entails that $K_A \subseteq βΓ \setminus K_B$, because $K_B$ is open.

(ii). It is clear that $A \subseteq K_A \cap Γ$. Conversely, if $t \in K_A \cap Γ$, then $\{t\}$ is an open subset of βΓ that intersects $K_A$, so it also intersects A, i.e., $t \in A$.

(iii). The non-trivial implication follows from part (ii).

(iv). It is clear that $K_A \subseteq K$. If $x \in K$ and if $V$ is an open neighbourhood of $x$, then $K \cap V$ is a non-empty open subset of βΓ. As Γ is an open and dense subset of βΓ we deduce that $A \cap V = Γ \cap K \cap V$ is non-empty. As $V$ was an arbitrary open neighbourhood of $x$, we conclude that $x \in K_A$.

(v). As $A \subseteq K_A$ we have $tA \subseteq t.K_A$, so $K_{tA} \subseteq t.K_A$. Applying this inclusion with $t^{-1}$ in the place of t we get

$$t.K_A = t.K_{t^{-1}A} \subseteq t.(t^{-1}.K_{tA}) = K_{tA}.$$  

(vi). This follows from (iv) and (v) and the fact that Γ and ∅ are the only Γ-invariant subsets of Γ.

The lemma above says that there is a one-to-one correspondence between subsets of Γ and compact-open subsets of βΓ. This is not surprising. Indeed the function algebras $ℓ^∞(Γ)$ and $C(βΓ)$ are canonically isomorphic (by the definition of the beta-compactification!). The canonical isomorphism carries the projection $1_A \in ℓ^∞(Γ)$ onto $1_{K_A} \in C(βΓ)$. We shall work with both pictures in this paper.

If $A$ and $B$ are subsets of a group Γ, then write $A \propto Γ B$ if $A$ is $B$-bounded i.e., if $A \subseteq \bigcup_{t \in F} tB$ for some finite subset $F$ of Γ.

**Lemma 2.5.** Let A and B be subsets of a group Γ.

(i) $K_B \subseteq X_A$ if and only if $B \propto Γ A$.

(ii) $X_A = X_B$ if and only if $A \propto Γ B \propto Γ A$.

(iii) If $A \propto Γ B \propto Γ A$, then A is paradoxical if and only if B is paradoxical.

**Proof.** (i). If $B \subseteq \bigcup_{t \in F} tA$ for some finite subset $F$ of Γ, then

$$K_B \subseteq K_{\bigcup_{t \in F} tA} = \bigcup_{t \in F} t.K_A \subseteq X_A.$$
by Lemma 2.4(iii) and (v). Conversely, if $K_B$ is a subset of $X_A = \bigcup_{t \in \Gamma} t.K_A$, then $K_B \subseteq \bigcup_{t \in F} t.K_A$ for some finite subset $F$ of $\Gamma$ by compactness of $K_B$. Hence $B$ is contained in $\bigcup_{t \in F} t.A$ by Lemma 2.4(iii).

(ii). Note that $K_B \subseteq X_A$ if and only if $X_B \subseteq X_A$ because $X_A$ is $\Gamma$-invariant. Thus (ii) follows from (i).

(iii). Suppose that $B \subseteq \bigcup_{t \in F} t.A$ and $A \subseteq \bigcup_{t \in F'} t.B$ for some finite subsets $F$ and $F'$ of $\Gamma$. Suppose that $A$ is non-paradoxical. Then, by Tarski’s theorem, we can find a finitely additive invariant measure $\lambda$ on $\Gamma$ such that $\lambda(A) = 1$. Since $\lambda(B) \leq |F| \lambda(A)$ and $\lambda(A) \leq |F'| \lambda(B)$, we conclude that $0 < \lambda(B) < \infty$, which in turns implies that $B$ is non-paradoxical. \hfill \Box

It follows from Lemma 2.3 vii) and from Lemma 2.5 that if $A$ is a non-empty subset of $\Gamma$, then $X_A$ is compact if and only if $X_A = \beta \Gamma$ if and only if $\Gamma \propto A$.

**Proposition 2.6.** Let $A$ be a non-empty subset of $\Gamma$. There is a non-zero $\Gamma$-invariant Radon measure on $X_A$ if and only if $A$ is non-paradoxical in $\Gamma$.

**Proof.** In the notation of Lemma 2.2 it follows from Lemma 2.4(ii) that $A = A(K_A, e)$, and hence that $X_A$ admits a non-zero $\Gamma$-invariant Radon measure if $A$ is non-paradoxical.

Suppose next that there is a non-zero $\Gamma$-invariant Radon measure $\lambda$ on $X_A$. Then, necessarily, $0 < \lambda(K_A) < \infty$. Let $\Omega_A$ be the set of all $A$-bounded subsets of $\Gamma$, i.e., $E \in \Omega_A$ if and only if $E \propto A$ if and only if $K_E \subseteq X_A$, cf. Lemma 2.5(i). Let $\mu$ be the invariant finitely additive measure on $\Gamma$ defined by

$$
\mu(E) = \begin{cases} 
\lambda(K_E), & E \in \Omega_A, \\
\infty, & E \notin \Omega_A.
\end{cases}
$$

Use Lemma 2.4(i) to see that $\mu$, indeed, is finitely additive. As $\mu(A) = \lambda(K_A)$ by definition we conclude that $A$ is non-paradoxical. \hfill \Box

**Proposition 2.7.** A group $\Gamma$ is supramenable if and only if whenever it acts co-compactly on a locally compact Hausdorff space $X$, then $X$ admits a non-zero $\Gamma$-invariant Radon measure.

**Proof.** If $\Gamma$ is not supramenable and if $A$ is a non-empty paradoxical subset of $\Gamma$, then $\Gamma \acts X_A$ is an example of a co-compact action on a locally compact Hausdorff space that admits no non-zero invariant Radon measure, cf. Proposition 2.6.

The reverse implication follows from Lemma 2.2 (and the remark above that lemma). \hfill \Box

Proposition 2.7 says that every non-supramenable group admits a co-compact action on a locally compact Hausdorff space for which there is no non-zero invariant Radon measures. Towards proving Theorem 1.2 we shall later improve this result and show that one always can choose this space to be the locally compact non-compact Cantor set, and that one further can choose the action to be free, minimal and purely infinite (see Definition 4.4).
Example 2.8. Let $\Gamma$ denote the $ax + b$ group with $a \in \mathbb{Q}^+$ and $b \in \mathbb{Q}$. It is well-known that $\Gamma$ is non-supramenable. In fact, $\Gamma$ contains a free semigroup of two generators. One can take the two generators to be $2x$ and $2x + 1$.

It also follows from Proposition 2.7 that $\Gamma$ is non-supramenable, as the canonical action of $\Gamma$ on $\mathbb{R}$ admits no non-zero invariant Radon measure. We can use Lemma 2.2 to construct paradoxical subsets of $\Gamma$ other than the ones coming from free sub-semigroups. Indeed, applying Lemma 2.2 to the compact set $K = [\alpha, \beta] \subseteq \mathbb{R}$ and to $x_0 = \gamma \in \mathbb{R}$ we find that

$$\{ax + b \mid \alpha \leq a\gamma + b \leq \beta\} \subseteq \Gamma$$

is paradoxical whenever $\alpha, \beta, \gamma \in \mathbb{R}$ and $\alpha < \beta$.

3. A geometric description of supramenable groups

We shall here describe supramenability in terms of the existence of injective quasi-isometric embeddings of free groups.

Definition 3.1. Let $\Gamma$ and $\Lambda$ be groups. A map $f : \Gamma \rightarrow \Lambda$ is said to be Lipschitz if for every finite set $S \subseteq \Gamma$ there is a finite set $T \subseteq \Lambda$ such that

$$(3.1) \quad \forall x, y \in \Gamma : xy^{-1} \in S \implies f(x)f(y)^{-1} \in T.$$  

A map $f : \Gamma \rightarrow \Lambda$ is said to be a quasi-isometric embedding if it is Lipschitz and satisfies that for every finite set $T \subseteq \Lambda$ there is a finite set $S \subseteq \Gamma$ such that

$$(3.2) \quad \forall x, y \in \Gamma : f(x)f(y)^{-1} \in T \implies xy^{-1} \in S.$$  

This definition corresponds to the usual terminology if $\Gamma$ and $\Lambda$ are considered as metric spaces with a right-invariant word metric. If a map $f$ has either of the two properties, the map $\hat{f}$ defined by $\hat{f}(x) = f(x^ {-1})^{-1}$ satisfies its left-invariant analogue and vice versa.

If $\Gamma$ and $\Lambda$ are groups and let $f : \Gamma \rightarrow \Lambda$ be an injective Lipschitz map. Let $A$ be a subset of $\Gamma$. A map $\sigma : A \rightarrow \Gamma$ is said to be a piecewise translation if it is injective and if there is a finite set $S \subseteq \Gamma$ such that $\sigma(x)x^{-1} \in S$ for all $x \in A$. The composition of piecewise translations (when defined) is again a piecewise translation.

If $\sigma : A \rightarrow \Gamma$ is a piecewise translation, then $\sigma(A_0) \sim_\Gamma A_0$ for all $A_0 \subseteq A$. A non-empty subset $A \subseteq \Gamma$ is paradoxical if and only if there are piecewise translations $\sigma^\pm : A \rightarrow A$ with disjoint images. In this case, for each natural number $n$, there are piecewise translations $\sigma_j : A \rightarrow A$, $j = 1, 2, \ldots, n$, with disjoint images.

Lemma 3.2. Let $\Gamma$ and $\Lambda$ be groups and let $f : \Gamma \rightarrow \Lambda$ be an injective Lipschitz map. Let $A$ be a subset of $\Gamma$ and let $\sigma : A \rightarrow \Gamma$ be a piecewise translation.

Documenta Mathematica 18 (2013) 1597–1626
It follows that the map $\tau: f(A) \to \Lambda$, given by $\tau \circ f = f \circ \sigma$, is a piecewise translation.

Proof. Injectivity (and well-definedness) of $\tau$ follows from injectivity of $f$ and of $\sigma$.

There is a finite set $S \subseteq \Gamma$ such that $\sigma(x)x^{-1} \in S$ for all $x \in A$. As $f$ is Lipschitz there is a finite set $T \subseteq \Lambda$ such that $f(x)f(y)^{-1} \in T$ whenever $xy^{-1} \in S$. Hence

$$
\tau(f(x))f(x)^{-1} = f(\sigma(x))f(x)^{-1} \in T
$$

for all $x \in A$, which proves that $\tau$ is a piecewise translation. \hfill \Box

Lemma 3.3. Let $\Gamma$ and $\Lambda$ be groups and let $f: \Gamma \to \Lambda$ be an injective Lipschitz map. Then $f(A)$ is a paradoxical subset of $\Lambda$ whenever $A$ is a paradoxical subset of $\Gamma$.

Proof. Assume that $A$ is a non-empty paradoxical subset of $\Gamma$. Then there are piecewise translations $\sigma^\pm: A \to A$ with disjoint images. Put $B = f(A)$, and let $\tau^\pm: B \to B$ be given by $\tau^\pm \circ f = f \circ \sigma^\pm$. Then $\tau^\pm$ are piecewise translations (by Lemma 3.2) and $\tau^+(B) \cap \tau^-(B) = \emptyset$ (by injectivity of $f$). This shows that $B$ is paradoxical. \hfill \Box

Proposition 3.4. A group $\Gamma$ is supramenable if and only if there is no injective Lipschitz map $f: \mathbb{F}_2 \to \Gamma$.

Proof. The “only if” part follows from Lemma 3.3 and the fact that $\mathbb{F}_2$ is non-amenable and hence paradoxical.

Suppose that $\Gamma$ is not supramenable and let $A$ be a non-empty paradoxical subset of $\Gamma$. Then we can find four piecewise translations $\sigma^\pm, \tau^\pm: A \to A$ with disjoint images such that there exists $s_0 \in A$ not in the image of any of these four piecewise translations.

Let $a, b$ denote the generators of $\mathbb{F}_2$ and define a map $f: \mathbb{F}_2 \to A$ by induction on the word length $\ell$ as follows. We first take care of length zero by setting $f(e) = s_0$. If $x \in \mathbb{F}_2$ has length $\ell(x) \geq 1$, and if $x = a^{\pm 1}x' \in \mathbb{F}_2$ is a reduced word, then put $f(x) = \sigma^\pm \circ f(x')$. Similarly, if $x = b^{\pm 1}x' \in \mathbb{F}_2$ is a reduced word, then let $f(x) = \tau^\pm \circ f(x')$.

We prove by contradiction that $f$ is injective. Indeed, if not, we can choose two reduced words $x \neq y$ such that $f(x) = f(y)$ and with $\ell(x)$ minimal for these conditions. Because $\sigma^\pm$ and $\tau^\pm$ have disjoint images (not containing $s_0$), there exists $c \in \{a,a^{-1},b,b^{-1}\}$ such that $x = cx'$ and $y = cy'$ are reduced words. But then $f(x') = f(y')$, contradicting the minimality of $\ell(x)$. Therefore $f$ is injective.

Let us finally show that $f$ is Lipschitz. By the remarks below Definition 3.1 we need only show that there is a finite set $T \subseteq \Gamma$ such that (3.1) holds with $S = \{a,a^{-1},b,b^{-1}\}$. This is done by first choosing

$$
T' = \{\sigma^\pm(x)x^{-1} \mid x \in A\} \cup \{\tau^\pm(x)x^{-1} \mid x \in A\}.
$$

Documenta Mathematica 18 (2013) 1597–1626
If $x, y \in F_2$ and $xy^{-1} \in S$, then either $x = cy$ or $y = cx$ for some $c \in S$ and such that $cy$, respectively, $cx$, is reduced. In the former case $f(x) = \sigma^\pm \circ f(y)$ or $f(x) = \tau^\pm \circ f(y)$, hence $f(x)f(y)^{-1} \in T'$. In the latter case $f(y)f(x)^{-1} \in T'$. Finally we can choose $T = T' \cup T'^{-1}$. 

Benjamini and Schramm proved in [5] that there is an injective quasi-isometric embedding from $F_2$ into any non-amenable group $\Gamma$. This gives us the following sequence of inclusions:

\[
\{\text{groups of sub-exponential growth}\} \subseteq \{\text{supramenable groups}\} = \{\text{groups with no injective Lipschitz inclusions of } F_2\} \subseteq \{\text{groups with no quasi-isometric embedding of } F_2\} \subset \{\text{amenable groups}\}.
\]

De Cornulier and Tessera showed in [7, 5.1] that any finitely generated solvable group with exponential growth contains a quasi-isometrically embedded free sub-semigroup on two generators. Solvable groups are amenable and the free group $F_2$ embeds quasi-isometrically in the free semi-group on two generators. Therefore the last inclusion is strict and for solvable groups the first four sets are equal.

We conclude this section by using the theorem of Benjamini and Schramm mentioned above to show that each non-amenable group contains small non-empty paradoxical subsets.

**Definition 3.5.** Let $A$ be a subset of a group $\Gamma$.

(i) Write $A \ll \Gamma$ if $A \preceq \Gamma \setminus B$ for all $A$-bounded subsets $B$ of $\Gamma$.

(ii) Say that $A$ is absorbing if $\bigcap_{t \in S} t A \neq \emptyset$ for all finite sets $S \subseteq \Gamma$.

(iii) Write $A \ll^* \Gamma$ if all $A$-bounded subsets of $\Gamma$ are non-absorbing.

For example, if $A$ is a subgroup of $\Gamma$, then $A \ll \Gamma$ if and only if $|\Gamma : A| = \infty$. A subset of $\Gamma = \mathbb{Z}$ is absorbing if and only if it contains arbitrarily long consecutive sequences (compare (iii) below for the non-commutative case).

We list some elementary properties of the relations defined above.

**Lemma 3.6.** Let $\Gamma$ be a group and let $A, B$ be subsets of $\Gamma$.

(i) If $B \propto A$ and $A \ll \Gamma$, then $B \ll \Gamma$.

(ii) If $A \ll \Gamma$, then $A \ll^* \Gamma$.

(iii) $A$ is absorbing if and only if for every finite subset $F \subseteq \Gamma$ there is $g \in \Gamma$ with $Fg \subseteq A$.

(iv) Suppose $A$ absorbing. Then $\Gamma$ is amenable if and only if there is a left-invariant mean on $\Gamma$ supported on $A$.

**Proof.** (i). As $B \propto A$, i.e., $B$ is $A$-bounded, we have $B \subseteq \bigcup_{s \in S} s A$ for some finite $S \subseteq \Gamma$. Let $n$ be the number of elements in $S$.

Let $C$ be a $B$-bounded subset of $\Gamma$. Then $C$ is also $A$-bounded. By the hypothesis on $A$ there exists a piecewise translation $\sigma_1 : A \rightarrow \Gamma \setminus C$. Since
We can repeat the process and obtain piecewise translations \( \sigma_1, \ldots, \sigma_n : A \rightarrow \Gamma \setminus C \) with disjoint images. These piecewise translations can be assembled to one piecewise translation \( \sigma : \bigcup_{s \in S} sA \rightarrow \Gamma \setminus C \). The restriction of \( \sigma \) to \( B \) will then witness the relation \( B \preceq \Gamma \setminus C \).

(ii). Note first that if \( E \subseteq \Gamma \) is such that \( E \ll \Gamma \), then \( E \) cannot be absorbing. Indeed, there is a piecewise translation \( \sigma : E \rightarrow \Gamma \setminus E \), and \( \sigma(E) \subseteq \bigcup_{s \in S} sE \) for some finite \( S \subseteq \Gamma \). Put \( T = S \cup \{ e \} \). Then \( \bigcap_{t \in T} tE = \emptyset \).

Suppose that \( A \ll \Gamma \) does not hold. Then there exists an absorbing set \( B \subseteq \Gamma \) such that \( B \preceq A \). Hence \( B \preceq \Gamma \) does not hold, which by (i) implies that \( A \ll \Gamma \) does not hold.

(iii). It suffices to observe that for any \( F \subseteq \Gamma \) we have

\[
\{ g \in \Gamma : Fg \subseteq A \} = \bigcap_{t \in F^{-1}} tA.
\]

(iv). This follows from the fact that left \( \text{Følner} \) sets can be arbitrarily translated by right multiplication.

One can show that \( A \ll \Gamma \) does not imply \( A \ll \Gamma \) (there are counterexamples with \( \Gamma = \mathbb{Z} \)). We now turn to the existence of interesting small sets in groups.

**Lemma 3.7.** Let \( \Gamma \) and \( \Lambda \) be groups and let \( f : \Gamma \rightarrow \Lambda \) be an injective quasi-isometric embedding. Then \( f(A) \ll \Lambda \) whenever \( A \ll \Gamma \).

In particular, each non-amenable group \( \Gamma \) contains a non-empty paradoxical subset \( A \) with \( A \ll \Gamma \).

**Proof.** Suppose that \( A \ll \Gamma \). Let \( T \) be a finite subset of \( \Lambda \) and find a finite set \( S \subseteq \Gamma \) such that (3.2) holds. There is a piecewise translation \( \sigma : A \rightarrow \Gamma \setminus \bigcup_{s \in S} sA \). This implies that \( \sigma(x)y^{-1} \notin S \) for all \( x, y \in A \). Let \( \tau : f(A) \rightarrow \Lambda \) be defined by \( \tau \circ f = f \circ \sigma \). Then \( \tau \) is a piecewise translation by Lemma 3.2.

Moreover,

\[
\tau(f(x))f(y)^{-1} = f(\sigma(x))f(y)^{-1} \notin T
\]

for all \( x, y \in A \). Hence \( \tau \) maps \( f(A) \) into \( \Lambda \setminus \bigcup_{t \in T} tf(A) \). This proves that \( f(A) \ll \Lambda \).

If \( \Gamma \) is non-amenable then, as noted above, by the theorem of Benjamini and Schramm in [3] there is an injective quasi-isometric embedding \( f : \mathbb{F}_2 \rightarrow \Gamma \). The free group \( \mathbb{F}_2 \) contains (several) non-amenable subgroups \( H \) of infinite index. For any such subgroup \( H \) we can put \( A = f(H) \). Then \( A \ll \Gamma \) by the first part of the lemma, and \( A \) is paradoxical by Lemma 3.3. \( \square \)

In general, if we are not asking that \( A \) in the lemma above is paradoxical, we have the following:

**Lemma 3.8.** Each infinite group \( \Gamma \) contains an infinite subset \( A \) such that \( A \ll \Gamma \).

**Proof.** Choose a sequence \( x_1, x_2, x_3, \ldots \) of elements in \( \Gamma \) such that

\[
(3.3) \quad x_n \notin \{ x_kx_{\ell}^{-1}x_m : 1 \leq k, \ell, m < n \}
\]
for all \( n \geq 2 \), and put \( A = \{ x_1, x_2, x_3, \ldots \} \). We claim that \(| sA \cap A | \leq 2\) for all \( s \in \Gamma \setminus \{ e \} \). Indeed, if \( sA \cap A \neq \emptyset \) and \( s \neq e \), then \( sx_n = x_m \) for some \( n \neq m \). Suppose that \( k, \ell \in \mathbb{N} \) is another pair such that \( sx_k = x_\ell \). Suppose also that \( k > \ell \). Then \( x_k = x_n s^{-1} x_\ell = x_n x_m^{-1} x_\ell \), whence \( k \leq \max\{ n, m, \ell \} \) by (3.3), i.e. \( k \leq \max\{ n, m \} \). Similarly, if \( \ell > k \), then \( x_\ell = x_m x_n^{-1} x_k \), which entails that \( \ell \leq \max\{ n, m \} \). Now, interchanging the roles of the pairs \( (n, m) \) and \( (k, \ell) \), we conclude that \( \max\{ n, m \} = \max\{ k, \ell \} \) whenever \( sx_n = x_m \) and \( sx_k = x_\ell \). It follows easily from this observation that \(| sA \cap A | \leq 2\).

Let \( F \subset \Gamma \) be finite, choose \( s \in \Gamma \setminus F \), and put \[
H = A \cap \bigcup_{t \in F} s^{-1} t A = \bigcup_{t \in F} (A \cap s^{-1} t A).
\]

Then \( H \) is finite and
\[
(3.4) \quad A \setminus H \sim_{\Gamma} s (A \setminus H) \subseteq \Gamma \setminus \bigcup_{t \in F} t A.
\]

To show that \( A \not\leq_{\Gamma} \Gamma \setminus \bigcup_{t \in F} t A \) it now suffices to show that \[
H \not\leq_{\Gamma} \Gamma \setminus \left( \bigcup_{t \in F} t A \cup s A \right) = \Gamma \setminus \bigcup_{t \in F'} t A,
\]
where \( F' = F \cup \{ s \} \). As \( H \) is finite it suffices to show that \( | \Gamma \setminus \bigcup_{t \in F'} t A | \geq |H| \).

However, one can deduce from (3.3) that \( \Gamma \setminus \bigcup_{t \in F} t A \) is infinite for any finite set \( F' \subset \Gamma \); and this completes the proof.

\[ \square \]

4. Structure of crossed product \( C^* \)-algebras

We review here some, mostly well-known, results about the structure of crossed product \( C^* \)-algebras; and we introduce the notion of purely infinite actions. Recall first that a dynamical system \( \Gamma \curvearrowright X \) is said to be regular if the full and the reduced crossed product \( C^* \)-algebras coincide, i.e., if the natural epimorphism
\[
C_0(X) \rtimes_{\text{full}} \Gamma \twoheadrightarrow C_0(X) \rtimes_{\text{red}} \Gamma
\]
is an isomorphism. Anantharaman-Delaroche proved in [1] that \( \Gamma \curvearrowright X \) is regular if the action is amenable (in the sense of Anantharaman-Delaroche, [1] Definition 2.1]). Matsumura proved later that the converse also holds when \( X \) is compact and the group \( \Gamma \) is exact, [14].

Moreover, the action is amenable if and only if the crossed product \( C^* \)-algebra \( C_0(X) \rtimes_{\text{red}} \Gamma \) is nuclear. Archbold and Spielberg proved in [3] that the full crossed product \( C_0(X) \rtimes_{\text{full}} \Gamma \) is simple if and only if the action \( \Gamma \curvearrowright X \) is minimal, topologically free and regular. Combining this result of Archbold and Spielberg with the results of Anantharaman-Delaroche mentioned above, we get:

**Proposition 4.1 (Anantharaman-Delaroche, Archbold–Spielberg).** Let \( \Gamma \) be a countable group acting on a locally compact Hausdorff space \( X \). Then \( C_0(X) \rtimes_{\text{red}} \Gamma \) is simple and nuclear if and only if \( \Gamma \curvearrowright X \) is minimal, topologically free and amenable.

Documenta Mathematica 18 (2013) 1597–1626
Recall that a **Kirchberg algebra** is a $C^*$-algebra which is simple, purely infinite, nuclear and separable.

The proposition below is essentially contained in [13, Theorem 5 and its proof] by Laca and Spielberg. To deduce simplicity of $C_0(X) \rtimes_{\text{red}} \Gamma$ one can also use Archbold–Spielberg, [3]. Recall that a projection $p$ in a $C^*$-algebra $A$ is said to be **properly infinite** if $p \oplus p \bowtie p$, i.e., $p \oplus p$ is Murray–von Neumann equivalent to a subprojection of $p$.

**Proposition 4.2 (Archbold–Spielberg, Laca–Spielberg).** Let $\Gamma$ be a countable group acting minimally and topologically freely on a metrizable totally disconnected locally compact Hausdorff space $X$. Suppose further that each non-zero projection in $C_0(X)$ is properly infinite in $C_0(X) \rtimes_{\text{red}} \Gamma$. Then $C_0(X) \rtimes_{\text{red}} \Gamma$ is simple and purely infinite.

J.-L. Tu proved in [22] (see also [21]) that the $C^*$-algebra associated with any amenable groupoid belongs to the UCT class. The crossed product $C_0(X) \rtimes_{\text{red}} \Gamma$ can be identified with the $C^*$-algebra of the groupoid $X \rtimes \Gamma$, which is amenable if $\Gamma$ acts amenably on $X$. Hence $C_0(X) \rtimes_{\text{red}} \Gamma$ belongs to the UCT class whenever $\Gamma$ acts amenably on $X$. Combining this deep theorem with Propositions 4.1 and 4.2 we get:

**Corollary 4.3.** Let $X$ be a metrizable totally disconnected locally compact Hausdorff space and let $\Gamma$ be a group acting on $X$. Then $C_0(X) \rtimes_{\text{red}} \Gamma$ is a Kirchberg algebra in the UCT class if and only if the action of $\Gamma$ on $X$ is minimal, topologically free, amenable, and each non-zero projection in $C_0(X)$ is properly infinite in $C_0(X) \rtimes_{\text{red}} \Gamma$.

It would be very desirable to replace the last condition on the projections in $C_0(X)$ with purely dynamical conditions on the $\Gamma$-space $X$. This leads us to the notion of purely infinite actions, which we shall proceed to define.

Let $\Gamma$ be a group acting on a locally compact totally disconnected Hausdorff space $X$. Let $\mathbb{K}(X)$ (or just $\mathbb{K}$ if $X$ is understood) be the algebra of all compact-open subsets of $X$. A set $K \subseteq \mathbb{K}$ is said to be $(X, \Gamma, \mathbb{K})$-paradoxical if there exist pairwise disjoint sets $K_1, K_2, \ldots, K_{n+m} \subseteq \mathbb{K}$ and elements $t_1, t_2, \ldots, t_{n+m} \in \Gamma$ such that $K_j \subseteq K$ for all $j$ and

\[
K = \bigcup_{j=1}^{n} t_j.K_j = \bigcup_{j=n+1}^{n+m} t_j.K_j.
\]

(We can think of this version of paradoxicality as a relative version of the classical notion of paradoxicality from the Hausdorff–Banach–Tarski paradox.)

**Definition 4.4.** An action of a group $\Gamma$ on a totally disconnected Hausdorff space $X$ is said to be **purely infinite** if every compact-open subset of $X$ is $(X, \Gamma, \mathbb{K})$-paradoxical (in the sense defined above).

Let us for a moment turn to the $C^*$-algebra point of view. Let us first fix some (standard) notations for crossed product $C^*$-algebras. Let $\Gamma$ be a group acting on a $C^*$-algebra $A$ (where $A$ for example could be $C_0(X)$ for some
locally compact Hausdorff space $X$). Let the action of $\Gamma$ on $\mathcal{A}$ be denoted by $t \mapsto \alpha_t \in \text{Aut}(\mathcal{A})$, $t \in \Gamma$. If $\mathcal{A} = C_0(X)$ and $\Gamma$ acts on $X$, then the action of $\Gamma$ on $\mathcal{A}$ is given by $\alpha_t(f) = t.f$, where $(t.f)(x) = f(t^{-1}.x)$ for $f \in C_0(X)$ and $x \in X$. Let $u_t, t \in \Gamma$, denote the unitary elements in (the multiplier algebra of) $\mathcal{A} \rtimes_{\text{red}} \Gamma$ that implement the action of $\Gamma$ on $\mathcal{A}$, i.e., $\alpha_t(a) = u_t a u_t^*$ for $a \in \mathcal{A}$ and $t \in \Gamma$. The set of finite sums of the form $\sum_{t \in \Gamma} a_t u_t$, where $a_t \in \mathcal{A}$ and only finitely many $a_t$ are non-zero, forms a uniformly dense $^*$-subalgebra of $\mathcal{A} \rtimes_{\text{red}} \Gamma$.

A projection $p \in \mathcal{A}$ is said to be $($\mathcal{A},$\Gamma)$-

paradoxical if there are pairwise orthogonal subprojections $p_1, p_2, \ldots, p_{n+m}$ of $p$ in $\mathcal{A}$ and elements $t_1, t_2, \ldots, t_{n+m}$ in $\Gamma$ such that

\begin{equation} \label{eq:paradoxical_projection}
    p = \sum_{k=1}^{n} \alpha_{t_k}(p_k) = \sum_{k=n+1}^{n+m} \alpha_{t_k}(p_k).
\end{equation}

We record some facts about paradoxical projections and sets:

(a) If $p$ is a projection in a $C^*$-algebra $\mathcal{A}$ on which a group $\Gamma$ acts, and if $p$ is $(\mathcal{A}, \Gamma)$-paradoxical, then $p$ is properly infinite in $\mathcal{A} \rtimes_{\text{red}} \Gamma$.

(b) If $K$ is a compact-open subset of a locally compact Hausdorff $\Gamma$-space $X$, then $K$ is $(X, \Gamma, K)$-paradoxical if and only if the projection $1_K$ is $(C_0(X), \Gamma)$-paradoxical.

(c) If $A \subseteq \Gamma$, then $A$ is paradoxical if and only if $K_A$ is $(\beta \Gamma, \Gamma, K)\text{-paradoxical}$, which again happens if and only if $1_A$ is a properly infinite projection in $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$ or, equivalently, if and only if $1_{K_A}$ is a properly infinite projection in $C(\beta \Gamma) \rtimes_{\text{red}} \Gamma$.

**Proof.** (a). Let $p_1, p_2, \ldots, p_{n+m} \in \mathcal{A}$ be pairwise orthogonal subprojections of $p$ and $t_1, t_2, \ldots, t_{n+m}$ in $\Gamma$ be such that (4.2) holds. Let $u_t$ be the unitaries in the multiplier algebra of $\mathcal{A}$ that implement the action $\alpha_t, t \in \Gamma$, on $\mathcal{A}$. Define the following two elements in $\mathcal{A} \rtimes_{\text{red}} \Gamma$:

$$
    v = \sum_{j=1}^{n} p_j u_{t_j}, \quad w = \sum_{j=n+1}^{n+m} p_j u_{t_j}.
$$

It then follows from (4.2) that

$$
    v^* v = p = w^* w, \quad v^* w = \perp, \quad v^* w^* \leq p, \quad w w^* \leq p.
$$

This shows that $p$ is a properly infinite projection in $C_0(X) \rtimes_{\text{red}} \Gamma$.

(b). This follows immediately from the fact that there is a bijective correspondence between compact-open subsets of $X$ and projections in $C_0(X)$ (given by $K \leftrightarrow 1_K$).

(c). The first claim, that $A$ is paradoxical if and only if $K_A$ is $(\beta \Gamma, \Gamma, K)$-paradoxical, follows easily from Lemma 2.3. Indeed, $A$ is paradoxical if and only if there exist pairwise disjoint subsets $A_1, A_2, \ldots, A_{n+m} \subseteq A$ and elements
\[ A = \bigcup_{j=1}^{n} t_j A_j = \bigcup_{j=n+1}^{n+m} t_j A_j. \]

Taking closures (relatively to \( \beta \Gamma \)) we obtain that (4.1) holds with \( K = K_A \) and \( K_j = K_{A_j} \). Hence \( K_A \) is \((\beta \Gamma, \Gamma, K)\)-paradoxical. For the reverse implication, if (4.1) holds with \( K = K_A \), then intersect all sets in (4.1) with \( \Gamma \) and use Lemma 2.4 to conclude that (4.3) holds, so \( A \) is paradoxical.

It was shown in [17, Proposition 5.5] that if \( \Gamma \subseteq \Gamma \), then \( 1_A \) is properly infinite in \( C_0(X) \otimes_{\text{red}} \Gamma \) if and only if \( A \) is paradoxical. The natural isomorphism from \( C_0(X) \otimes_{\text{red}} \Gamma \) to \( C(\beta \Gamma) \) maps \( 1_A \) to \( 1_{K_A} \), and this isomorphism preserves the property of being properly infinite in the crossed product by \( \Gamma \).

**Lemma 4.5.** Let \( \Gamma \) be a group acting on a locally compact totally disconnected Hausdorff space \( X \). Consider the following three conditions on a compact-open subset \( K \) of \( X \):

(i) There is no \( \Gamma \)-invariant Radon measure \( \mu \) on \( X \) such that \( \mu(K) > 0 \).

(ii) \( 1_K \) is a properly infinite projection in \( C_0(X) \otimes_{\text{red}} \Gamma \).

(iii) \( K \) is \((X, \Gamma, K)\)-paradoxical.

Then (iii) \( \Rightarrow \) (ii) \( \Rightarrow \) (i).

**Proof.** (ii) \( \Rightarrow \) (i). Assume (i) does not hold. Then there is an invariant Radon measure \( \mu \) on \( X \) such that \( \mu(K) = 1 \). The measure \( \mu \) extends to a densely defined (possibly unbounded) trace \( \tau \) on \( C_0(X) \otimes_{\text{red}} \Gamma \) (cf. [17, Lemma 5.3]) such that \( \tau(1_K) = \mu(K) = 1 \). A properly infinite projection is either zero (or infinite) under any positive trace, so \( 1_K \) cannot be properly infinite in \( C_0(X) \otimes_{\text{red}} \Gamma \).

(iii) \( \Rightarrow \) (ii). This follows from (a) and (b) above.

The proposition below follows immediately from Corollary 4.3 and from “(iii) \( \Rightarrow \) (ii)” of the lemma above.

**Proposition 4.6.** Let \( X \) be a metrizable totally disconnected locally compact Hausdorff space and let \( \Gamma \) be a group which acts on \( X \) in such a way that the action is topologically free, minimal, amenable and purely infinite. Then \( C_0(X) \otimes_{\text{red}} \Gamma \) is a Kirchberg algebra in the UCT class.

**Remark 4.7** (The type semigroup). It is an interesting question if the converse of Proposition 4.6 holds, or to what extend the reverse implications in Lemma 4.5 hold. In other words, if \( \Gamma \) acts freely, minimally and amenable on a metrizable totally disconnected locally compact Hausdorff space \( X \) such that \( C_0(X) \otimes_{\text{red}} \Gamma \) is a Kirchberg algebra, does it then follow that the action is purely infinite? As a Kirchberg algebra has no traces, the space \( X \) cannot have any non-zero \( \Gamma \)-invariant Radon measures if \( C_0(X) \otimes_{\text{red}} \Gamma \) is a Kirchberg algebra.

We can analyze these questions using the relative type semigroup \( S(X, \Gamma, K) \) considered for example in [17, Section 5], see also [24]. This semigroup can be...
defined as being the universal ordered abelian semigroup generated by elements $[K]$, with $K \in \mathbb{K}$, subject to the relations

$$[t.K] = [K], \quad [K \cup K'] = [K] + [K']$$

if $K \cap K' = \emptyset$, \quad $[K] \leq [K']$ if $K \subseteq K'$,

where $K, K' \in \mathbb{K}$ and $t \in \Gamma$. As the set of compact-open sets is closed under forming set differences, we deduce that the ordering on $S(X, \Gamma, \mathbb{K})$ is the algebraic order: If $x, y \in S(X, \Gamma, \mathbb{K})$, then $x \leq y$ if and only if there exists $z \in S(X, \Gamma, \mathbb{K})$ such that $y = x + z$.

In the language of the type semigroup, $K \in \mathbb{K}$ is $(X, \Gamma, \mathbb{K})$-paradoxical if and only if $2[K] \leq [K]$.

Property (i) in Lemma 4.5 holds if and only if there exists a natural number $n$ such that $(n + 1)[K] \leq n[K]$, which in turns implies that $m[K] \leq n[K]$ for all $m \geq 1$. We conclude that the implication "(i) $\Rightarrow$ (iii)" holds for all compact-open subsets $K$ of $X$ if $S(X, \Gamma, \mathbb{E})$ is almost unperforated, that is, for all $x, y \in S(X, \Gamma, \mathbb{K})$ for all integers $n \geq 1$, $(n + 1)x \leq ny$ implies $x \leq y$. In fact, "(i) $\Rightarrow$ (iii)" also holds under the much weaker comparability assumption that there exist integers $n, m \geq 2$ such that for all $x, y \in S(X, \Gamma, \mathbb{K})$, $nx \leq my$ implies $x \leq y$.

It has recently been shown by Ara and Exel in [2] that $S(X, \Gamma, \mathbb{K})$ need not be almost unperforated. They give counterexamples where $\Gamma$ is a free group acting on the (locally compact, non-compact) Cantor set. It is not known if this phenomenon can occur also when the free group acts freely and minimally on the Cantor set. Nonetheless, the example by Ara and Exel indicates that the questions raised in this remark may all have negative answers.

We close this section with two lemmas that shall be used in Section 6. An open subset $E$ of a $\Gamma$-space $X$ is said to be $\Gamma$-full in $X$ if $X = \bigcup_{t \in \Gamma} t.E$.

**Lemma 4.8.** Let $X$ be a totally disconnected locally compact Hausdorff space, let $\Gamma$ be a group acting co-compactly on $X$, and let $Y$ be a closed $\Gamma$-invariant subset of $X$.

(i) If $K'$ is a compact-open $\Gamma$-full subset of $Y$, then there exists a compact-open $\Gamma$-full subset $K$ of $X$ such that $K' = K \cap Y$.

(ii) If $K$ is a $(X, \Gamma, \mathbb{K})$-paradoxical compact-open subset of $X$, then $K \cap Y$ is a $(Y, \Gamma, \mathbb{K})$-paradoxical compact-open subset of $Y$.

**Proof.** (i). There is a compact-open subset $K_0$ of $X$ such that $K' = K_0 \cap Y$. It follows from the assumption that the action of $\Gamma$ on $X$ is co-compact that there exists a compact-open subset $L \subseteq X$ such that $X = \bigcup_{t \in \Gamma} t.L$. Since $K'$ is $\Gamma$-full in $Y$ and $L \cap Y$ is compact there is a finite subset $S \subseteq \Gamma$ such that $L \cap Y \subseteq \bigcup_{t \in S} t.K'$. Put

$$K = K_0 \cup \left(L \setminus \bigcup_{t \in S} t.K'\right).$$

Then $K$ is compact-open, $K \cap Y = K_0 \cap Y = K'$, and $X = \bigcup_{t \in \Gamma} t.K$.

(ii). Follows from the definitions. \qed
Lemma 4.9. Consider the canonical action of $\Gamma$ on $b\Gamma$. Let $A$ and $B$ be subsets of $\Gamma$. Then $K_B$ is a $\Gamma$-full subset of $X_A$ if and only if $A\propto_{\Gamma} B\propto_{\Gamma} A$.

Proof. By definition, $K_B$ is $\Gamma$-full in $X_A$ if and only if $X_B = X_A$. The lemma therefore follows from Lemma 2.5(ii). \hfill \square

5. The Roe algebra

The Roe algebra associated with a group $\Gamma$ is the reduced crossed product $C^*$-algebra $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$. We give below a characterisation of supramenability in terms of the Roe algebra.

Lemma 5.1. Let $p \in \ell^\infty(\Gamma)$ and $q \in \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$ be projections that generate the same closed two-sided ideal in $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$. If $q$ is properly infinite, then so is $p$.

Proof. By the assumption that $p$ and $q$ generate the same closed two-sided ideal in the Roe algebra $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$ it follows that $p \preceq q \otimes 1_n$ and $q \preceq p \otimes 1_m$ relatively to $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$ for some integers $n, m \geq 1$. As $q$ is properly infinite we have $q \otimes 1_k \preceq q$ for all $k \geq 1$. Hence

$$(p \otimes 1_m) \oplus (p \otimes 1_m) \preceq q \otimes 1_{2nm} \preceq q \preceq p \otimes 1_m,$$

so $p \otimes 1_m$ is properly infinite in the Roe algebra. It then follows from [17, Proposition 5.5] that $p$ itself is properly infinite in the Roe algebra. \hfill \square

Remark 5.2 (The Pedersen ideal). Gert K. Pedersen proved that every $C^*$-algebra $A$ contains a smallest dense two-sided ideal, now called the Pedersen ideal in $A$. The Pedersen ideal of $A$ can be obtained as the intersection of all dense two-sided ideals in $A$; and the content of Pedersen’s result is that this ideal is again a dense ideal in $A$.

For every positive element $a$ in $A$ and for every $\varepsilon > 0$ one can consider the “$\varepsilon$-cut-down”, $(a - \varepsilon)_+ \in A$, (which is the positive part of the self-adjoint element $a - \varepsilon \cdot 1_A$ in the unitalization $\tilde{A}$ of $A$). This element $(a - \varepsilon)_+$ belongs to the Pedersen ideal for every positive element $a$ in $A$ and for every $\varepsilon > 0$. If $p \in A$ is a projection and if $0 < \varepsilon < 1$, then $(p - \varepsilon)_+ = (1 - \varepsilon)p$. All projections in $A$ therefore belong to the Pedersen ideal in $A$.

The Pedersen ideal of a unital $C^*$-algebra is the algebra itself, i.e., unital $C^*$-algebras have no proper dense two-sided ideals. The Pedersen ideal of $K(H)$, the compact operators on a Hilbert space $H$, is the algebra of all finite rank operators on $H$. The Pedersen ideal of $C_0(X)$, for some locally compact Hausdorff space $X$, is $C_c(X)$.

An element $a$ in a $C^*$-algebra $A$ is said to be full, if $a$ is not contained in any proper closed two-sided ideal in $A$. If $A$ contains a full projection $p$, then the primitive ideal space of $A$ is (quasi-)compact. This can be rephrased as follows: Whenever $\{I_\alpha\}_{\alpha \in A}$ is an increasing net of closed two-sided ideals in $A$ such that $A = \bigcup_{\alpha \in A} I_\alpha$, then $A = I_\alpha$ for some $\alpha \in A$. Indeed, $\bigcup_{\alpha \in A} I_\alpha$ is a dense two-sided ideal in $A$, which therefore contains the Pedersen ideal, and hence contains all projections in $A$. Thus $p \in I_\alpha$ for some $\alpha$, whence $A = I_\alpha$. 

Documenta Mathematica 18 (2013) 1597–1626
As mentioned in the previous section, it was shown in [17, Proposition 5.5] that if \( A \subseteq \Gamma \), then \( 1_A \) is properly infinite in \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) if and only if \( A \) is paradoxical. We sharpen this result as follows:

**Proposition 5.3.** The following two conditions are equivalent for every group \( \Gamma \):

(i) \( \Gamma \) is supramenable.

(ii) The Roe algebra \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) contains no properly infinite projection.

**Proof.** Suppose that (ii) does not hold. If \( \Gamma \) is non-exact, then \( \Gamma \) is non-amenable and in particular not supramenable, so (i) does not hold. Suppose that \( \Gamma \) is exact and that \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) contains a properly infinite projection \( p \). Let \( \mathcal{I} \) be the closed two-sided ideal in \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) generated by \( p \). Then, by [20, Theorem 1.16], which applies because \( \Gamma \) is assumed to be exact and the action of \( \Gamma \) on \( \ell^\infty(\Gamma) \) is free, \( \mathcal{I} \) is the closed two-sided ideal generated by \( \ell^\infty(\Gamma) \cap \mathcal{I} \). Let \( \mathcal{A} \) be the directed net of all finite subsets of the set of all projections in \( \ell^\infty(\Gamma) \cap \mathcal{I} \); and for each \( \alpha \in \mathcal{A} \) let \( \mathcal{I}_\alpha \) be the closed two-sided ideal in \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) generated by \( \alpha \). Then

\[
\mathcal{I} = \bigcup_{\alpha \in \mathcal{A}} \mathcal{I}_\alpha.
\]

It therefore follows from Remark 5.2 that \( \mathcal{I} = \mathcal{I}_\alpha \) for some \( \alpha \in \mathcal{A} \). Let \( q \in \ell^\infty(\Gamma) \) be the supremum of the projections belonging to \( \alpha \). Then \( \mathcal{I} \) is equal to the closed two-sided ideal generated by \( q \). Hence \( p \) and \( q \) generate the same closed two-sided ideal in the Roe algebra, whence \( q \) is properly infinite by Lemma 5.1. If (ii) holds and if \( A \subseteq \Gamma \), then \( 1_A \) is not a properly infinite projection in \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \), so \( A \) is non-paradoxical, cf. [17, Proposition 5.5] or claim (c) (below Definition 4.4). This shows that \( \Gamma \) is supramenable. \( \square \)

**Remark 5.4.** A projection is said to be infinite if it is Murray–von Neumann equivalent to a proper subprojection of itself. A unital \( C^* \)-algebra is infinite if its unit is an infinite projection, or, equivalently, if it contains a non-unitary isometry. If a unital \( C^* \)-algebra contains an infinite projection, then it is infinite itself. The Roe algebra \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) is therefore infinite whenever \( \Gamma \) is non-supramenable.

The Roe algebra \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) is also infinite whenever \( \Gamma \) contains an element \( t \) of infinite order. Indeed, let \( A = \{ t^n \mid n \geq 0 \} \). Then \( tA \subseteq A \), so \( u_t 1_A u_t^* = 1_A < 1_A \), which shows that \( 1_A \) is an infinite projection in \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \). (As before we let \( (u_t)_{t \in \Gamma} \) denote the unitaries in \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) implementing the action of \( \Gamma \) on \( \ell^\infty(\Gamma) \).)

On the other hand, if \( \Gamma \) is locally finite (an increasing union of finite groups), then \( \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \) is finite. To see this observe first that \( C(X) \rtimes_{\text{red}} \Gamma \) is finite whenever \( \Gamma \) is a finite group acting on a compact Hausdorff space \( X \). Indeed, a unital \( C^* \)-algebra is finite if it admits a separating family of tracial states. A crossed product \( C^* \)-algebra \( C(X) \rtimes_{\text{red}} \Gamma \) is therefore finite if \( X \) admits a separating family of invariant probability measures. (By separating we mean...
that every non-empty open set is non-zero on at least one of the probability measures in the family.) If $\Gamma$ is finite then the family consisting of $\Gamma$-means of every probability measure on $X$ will be such a separating family of invariant probability measures.

Suppose $\Gamma$ is locally finite and write $\Gamma = \bigcup_{n=1}^{\infty} \Gamma_n$, where $\Gamma_1 \subseteq \Gamma_2 \subseteq \Gamma_3 \subseteq \cdots$ is an increasing sequence of finite subgroups of $\Gamma$. Then

$$\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma = \varprojlim \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma_n,$$

since $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma_n$ is isomorphic to $C^*(\ell^\infty(\Gamma) \cup \{u_t \mid t \in \Gamma_n\})$. Any inductive limit of finite $C^*$-algebras is again finite, so $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$ is finite because all $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma_n$ are finite.

It seems plausible that the Roe algebra $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$ is finite if and only if $\Gamma$ is locally finite.

In the proposition below we identify the Roe algebra with $C(\beta \Gamma) \rtimes_{\text{red}} \Gamma$.

If $A \subseteq \Gamma$ and $\{U_i\}_{i \in I}$ is an increasing family of proper open $\Gamma$-invariant subsets of $X_A$, then $\bigcup_{i \in I} U_i$ is also a proper subset of $X_A$. (Otherwise $K_A$ would be contained in one of the $U_i$’s, which again would entail that $U_i = X_A$.) It follows that $X_A$ contains a maximal proper open $\Gamma$-invariant subset, or, equivalently, that $X_A$ contains a minimal (non-empty) closed $\Gamma$-invariant subset. This argument shows moreover that each non-empty closed invariant subset of $X_A$ contains a minimal (non-empty) closed $\Gamma$-invariant subset. Recall from Definition 3.5 the definition of the notion $A \ll^* \Gamma$.

**Proposition 5.5.** Let $A$ be a non-empty subset of $\Gamma$ and let $Z$ be a minimal (non-empty) closed $\Gamma$-invariant subset of $X_A$. Then:

(i) $Z$ is a locally compact totally disconnected Hausdorff space, and the action of $\Gamma$ on $Z$ is free and minimal.

(ii) The action of $\Gamma$ on $Z$ is purely infinite if $A$ is paradoxical.

(iii) The action of $\Gamma$ on $Z$ is amenable if $\Gamma$ is exact.

(iv) If $A \ll^* \Gamma$, then $Z$ is necessarily non-compact; and if $A \ll^* \Gamma$ does not hold, then there exists a minimal (non-empty) closed $\Gamma$-invariant subset of $X_A$ which is compact.

**Proof.** It is clear from its definition that $Z$ is a closed $\Gamma$-invariant subspace of $X_A$, and hence itself a $\Gamma$-space. Being a closed subset of the totally disconnected locally compact Hausdorff space $X_A$, $Z$ is also a totally disconnected locally compact Hausdorff space. By maximality of $U$, the action of $\Gamma$ on $Z$ is minimal. Freeness and amenability of an action of a group on a space pass to any $\Gamma$-invariant subspace. As $\Gamma$ acts freely on $\beta \Gamma$ for all $\Gamma$ we conclude that (i) holds. Moreover, if $\Gamma$ is exact, then the action of $\Gamma$ on $\beta \Gamma$ is amenable (see [6, Theorem 5.1.6]), so (iii) holds.

Let us prove (ii). Assume that $A$ is paradoxical. Take a non-empty compact-open subset $K$ of $Z$. Then $K$ is $\Gamma$-full in $Z$ by minimality of the action. We
can therefore use Lemma 4.8(i) to find a compact-open subset \( K' \) of \( X_A \) which is \( \Gamma \)-full in \( X_A \) and satisfies \( K = K' \cap Z \). Put \( B = K' \cap \Gamma \). Then \( K_B = K' \) by Lemma 2.4(iv), and \( A \preceq B \preceq A \) by Lemma 4.9. Hence \( B \) is paradoxical by Lemma 4.8(ii). Claim (c) (below Definition 4.4) then states that \( K \) is \( (Z, \Gamma, K) \)-paradoxical.

(iv). Suppose that \( Z \) is compact. As \( Z \subseteq X_A = \bigcup_{t \in \Gamma} t.K_A \), it follows that there exists a finite set \( S \subseteq \Gamma \) such that \( Z \subseteq \bigcup_{s \in S} s.K_A = K_B \), where \( B = \bigcup_{s \in S} sA \).

For each finite set \( T \subseteq \Gamma \) we have
\[
Z = \bigcap_{t \in T} t.Z \subseteq \bigcap_{t \in T} t.K_B = K_{\bigcap_{t \in T} tB}.
\]

This shows that \( \bigcap_{t \in T} tB \) is non-empty, so \( B \) is absorbing. Hence \( A \preceq \Gamma \) does not hold.

Suppose, conversely, that \( A \preceq \Gamma \) does not hold. Find an absorbing subset \( B \) of \( \Gamma \) so that \( B \preceq A \). Then \( K_B \subseteq X_A \) by Lemma 2.4(i). Put \( W = \bigcap_{t \in T} t.K_B \). Then \( W \) is compact and \( \Gamma \)-invariant. Moreover, since
\[
\bigcap_{s \in S} t.K_B = K_{\bigcap_{s \in S} sB} \neq \emptyset
\]
for all finite sets \( S \subseteq \Gamma \), we conclude that \( W \) is non-empty. By the argument above the proposition we can find a minimal (non-empty) closed \( \Gamma \)-invariant subset \( Z \) of \( W \). As \( W \) is compact, so is \( Z \).

It should be remarked that the space \( Z \) from Proposition 5.5 may not be metrizable and that the crossed product \( C_0(Z) \times_{\text{red}} \Gamma \) accordingly may not be separable. It follows from Proposition 5.5 and from Proposition 4.1 that the crossed product \( C_0(Z) \times_{\text{red}} \Gamma \) is simple and nuclear if \( \Gamma \) is exact; and it follows from Proposition 5.5 and Proposition 4.6 that the crossed product is simple, nuclear and purely infinite (but not necessarily a Kirchberg algebra) if \( A \) moreover is paradoxical.

Recall from Lemma 2.3 that \( X_A \) itself is compact if and only if \( X_A = \beta\Gamma \). Clearly, any minimal closed \( \Gamma \)-invariant subset of \( \beta\Gamma \) is compact. However, a minimal closed invariant subset \( Z \) of \( X_A \) can be compact even when \( X_A \) is non-compact. This follows from Proposition 5.5(iv) if we can find an absorbing subset \( A \) of a group \( \Gamma \) such that \( X_A \neq \beta\Gamma \) (or equivalently, such that \( \Gamma \) is not \( A \)-bounded). There are many such examples of \( A \subseteq \Gamma \), e.g. \( \Gamma = \mathbb{Z} \) and \( A = \mathbb{N} \). Perhaps surprisingly it turns out to be a subtle matter to decide when the space \( Z \) is non-discrete. Clearly, if \( A \) is non-empty and finite, then \( Z = X_A = \Gamma \) and we are in the trivial situation of \( \Gamma \) acting on itself. More generally, if \( Z \) has an isolated point, then the \( \Gamma \)-space \( Z \) is conjugate to \( \Gamma \). (Indeed, if \( x_0 \in Z \) is an isolated point, then \( Z = \Gamma.x_0 \) by minimality of the action.)

The example below shows that \( Z \) (in Proposition 5.5) can be discrete (and hence trivial) even when \( A \) is infinite. More precisely, if \( A \subseteq \Gamma \) is such that \( sA \cap A \) is finite for all \( e \neq s \in \Gamma \), then \( Z \) is discrete. On the other hand, if \( A \) is paradoxical, then \( Z \) cannot be discrete by Proposition 5.5(ii).
Example 5.6. Let $\Gamma$ be a countably infinite group and suppose that $A \subseteq \Gamma$ is an infinite set such that $|sA \cap A| < \infty$ for all $s \in \Gamma \setminus \{e\}$. It was shown in Lemma 3.8 (and its proof) that each infinite group $\Gamma$ contains such a subset $A$. Let $1_A \in \ell^\infty(\Gamma) \subseteq \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$ denote indicator function for $A$.

Inside the Roe algebra, $c_0(\Gamma) \rtimes_{\text{red}} \Gamma$ is the smallest non-zero ideal (every other non-zero ideal contains this ideal). If $A \subseteq \Gamma$ is as above, then the following holds in the quotient of the Roe algebra by $c_0(\Gamma) \rtimes_{\text{red}} \Gamma$:

$$\forall x \in \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma : 1_A x 1_A + c_0(\Gamma) \rtimes_{\text{red}} \Gamma \in \ell^\infty(\Gamma) + c_0(\Gamma) \rtimes_{\text{red}} \Gamma.$$  

(5.1) In other words, if $\pi: \ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma \to (\ell^\infty(\Gamma)/c_0(\Gamma)) \rtimes_{\text{red}} \Gamma$ denotes the quotient mapping, then $\pi(1_A)$ is an abelian projection. In particular, the corner algebra $1_A(\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma)1_A$ has a character.

Let us see why (5.1) holds. It suffices to establish (5.1) for $x$ in a dense subset of $\ell^\infty(\Gamma) \rtimes_{\text{red}} \Gamma$, so we may assume that $x = \sum_{t \in F} f_t u_t$, where $F \subseteq \Gamma$ is finite, $f_t \in \ell^\infty(\Gamma)$, and $t \mapsto u_t$ is the canonical unitary representation of the action of $\Gamma$ on $\ell^\infty(\Gamma)$. Now,

$$1_A x 1_A = \sum_{t \in F} 1_A f_t u_t 1_A = \sum_{t \in F} 1_A f_t 1_A u_t = \sum_{t \in F} 1_{A \cap \Gamma} f_t u_t.$$

But $1_{A \cap \Gamma} \in c_0(\Gamma)$ whenever $t \neq e$, so $\pi(1_A x 1_A) = \pi(1_A f_e)$ and $1_A f_e \in \ell^\infty(\Gamma)$. Let $A \subseteq \Gamma$ be as above, consider the open set $X_A$, and let $U$ be a maximal proper open $\Gamma$-invariant subset of $X_A$, cf. Definition 2.3. Since $A$ is infinite and $\Gamma$ is dense in $\beta \Gamma$, we have $\Gamma \subseteq U \subset X_A$. We claim that the minimal $\Gamma$-space $Z = X_A \setminus U$ is, in fact, discrete. More precisely, the compact-open subset $K_A \cap Z$ of $Z$ is a singleton, and hence an isolated point in $Z$. Indeed, the natural epimorphism $C_0(X_A) \rtimes_{\text{red}} \Gamma \to C_0(Z) \rtimes_{\text{red}} \Gamma$ maps $1_{K_A} \rtimes 1_{K_A \cap Z}$ to $C_0(Z) \rtimes_{\text{red}} \Gamma$ onto $1_{K_A \cap Z}$ and its kernel contains $c_0(\Gamma) \rtimes_{\text{red}} \Gamma$. We can therefore deduce from (5.1) that

$$\mathcal{A} := 1_{K_A \cap Z} (C_0(Z) \rtimes_{\text{red}} \Gamma) 1_{K_A \cap Z}$$

is abelian. As $C_0(Z) \rtimes_{\text{red}} \Gamma$ is a simple $C^*$-algebra, so is $\mathcal{A}$, whence $\mathcal{A} \cong \mathbb{C}$. As $C(K_A \cap Z)$ is a sub-$C^*$-algebra of $\mathcal{A}$ this implies that $K_A \cap Z$ is a singleton.

6. Kirchberg algebras arising as crossed products by non-supramenable groups

We show here that every non-supramenable countable group admits a free, minimal, purely infinite\footnote{In the sense of Definition 4.4} action on the locally compact, non-compact Cantor set; and that the action moreover can be chosen to be amenable if the group is exact. The (reduced) crossed product $C^*$-algebra associated with such an action will in the latter case be a stable Kirchberg algebra in the UCT class. Our construction is an adaption of the one from [17, Section 6], where it was shown that every exact non-amenable countable group admits a free, minimal, amenable action on the (compact) Cantor set such that the crossed product $C^*$-algebra is a unital Kirchberg algebra in the UCT class. (One can easily
modify the construction in \[17\] to make the action on the Cantor set purely infinite.)

Fix a countable group $\Gamma$ and a non-empty subset $A$ of $\Gamma$. Let $K_A \subseteq X_A \subseteq \beta \Gamma$ be as in Definition \[23\]. Note that $C_0(X_A)$ is a $\Gamma$-invariant closed ideal in $C(\beta \Gamma)$, and the smallest such which contains the projection $1_{K_A}$. Sometimes we prefer to work with $\ell^\infty(\Gamma)$ rather than $C(\beta \Gamma)$. To avoid confusion we denote by $C_A$ the $\Gamma$-invariant ideal in $\ell^\infty(\Gamma)$ that corresponds to $C_0(X_A)$. Thus $C_A$ is the smallest $\Gamma$-invariant closed ideal in $\ell^\infty(\Gamma)$ which contains $1_A$.

Fix an increasing sequence $\{F_n\}_{n=1}^\infty$ of finite subsets of $\Gamma$ with $\bigcup_{n=1}^\infty F_n = \Gamma$. Put $B_n = \bigcup_{t \in F_n} tA$, and put $p_n = 1_{B_n} \in C_A$. Then $\{p_n\}_{n=1}^\infty$ is an increasing approximate unit for $C_A$ consisting of projections. We shall use and refer to this approximate unit several times in the following.

It is well-known that the canonical action of $\Gamma$ on $\beta \Gamma$ is free, and it was shown by Ozawa (see \[6\] Theorem 5.1.6) that this action is amenable whenever $\Gamma$ is exact. The two lemmas below tell us how these properties can be preserved after passing to a suitable separable sub-$C^*$-algebra of $C_A$. If $A$ is an abelian $C^*$-algebra, then let $\hat{A}$ denote its space of characters, so that $A \cong C_0(\hat{A})$.

It was remarked in \[17\] that if a group $\Gamma$ acts on a compact Hausdorff space $X$, and if there are projections $q_j(t)$ in $C(X)$, for $t \in \Gamma \setminus \{e\}$ and for $j = 1, 2, 3$, such that

$$\alpha_t(q_j(t)) \perp q_j(t), \quad q_1(t) + q_2(t) + q_3(t) = 1,$$

for all $e \neq t \in \Gamma$ and $j = 1, 2, 3$, then $\Gamma$ acts freely on $X$. (As below Definition \[4\] $t \mapsto \alpha_t$ denotes the induced action on $C(X)$ given by $\alpha_t(f)(x) = f(t^{-1}x)$.) Consider now the case where $\Gamma$ acts on a non-compact locally compact Hausdorff space $X$ and where $C_0(X)$ has an increasing approximate unit $\{p_n\}_{n=1}^\infty$ consisting of projections. It is then easy to see that $\Gamma$ acts freely on $X$ if there are projections $q_{j,n}^{(t)}$ in $C_0(X)$, for $t \in \Gamma \setminus \{e\}$, $j = 1, 2, 3$ and $n \in \mathbb{N}$, such that

$$\alpha_t(q_{j,n}^{(t)}) \perp q_{j,n}^{(t)}, \quad q_{1,n}^{(t)} + q_{2,n}^{(t)} + q_{3,n}^{(t)} = p_n,$$

holds for all $e \neq t \in \Gamma$, $j = 1, 2, 3$ and $n \in \mathbb{N}$.

**Lemma 6.1 (cf. \[17\] Lemma 6.3).** There is a countable subset $M'$ of $C_A$ such that if $A$ is any $\Gamma$-invariant sub-$C^*$-algebra of $C_A$ which contains $M'$, then $\{p_n\}_{n=1}^\infty$ (defined above) is an approximate unit for $A$, and $\Gamma$ acts freely on the character space $\hat{A}$ of $A$.

**Proof.** By \[17\] Corollary 6.2 we can find projections $q_j^{(t)}$ in $\ell^\infty(\Gamma)$ for $e \neq t \in \Gamma$ and $j = 1, 2, 3$, such that \[6.1\] holds. The projections

$$q_{j,n}^{(t)} = q_j^{(t)} p_n, \quad e \neq t \in \Gamma, \quad j = 1, 2, 3, \quad n \in \mathbb{N},$$

satisfy \[6.2\]. As $p_n$ belongs to $C_A$ for each $n$, so does each $q_{j,n}^{(t)}$. Let

$$M' = \{q_{j,n}^{(t)} \mid t \in \Gamma, \ n \in \mathbb{N}, \ j = 1, 2, 3\} \subseteq C_A,$$

Documenta Mathematica 18 (2013) 1597–1626.
and let $\mathcal{A}$ be any $\Gamma$-invariant sub-$C^*$-algebra of $\mathcal{C}_A$ that contains $M'$. Then each $p_n$ belongs to $\mathcal{A}$ by (6.2). Moreover, $\{p_n\}_{n=1}^{\infty}$ is an approximate unit for $\mathcal{A}$ because it is an approximate unit for $\mathcal{C}_A$. Finally, $\Gamma$ acts freely on $\hat{A}$ because (6.2) holds.

**Lemma 6.2.** Suppose that $\Gamma$ is an exact countable group. Then there is a countable subset $M''$ of $\mathcal{C}_A$ such that if $\mathcal{A}$ is any $\Gamma$-invariant sub-$C^*$-algebra of $\mathcal{C}_A$ which contains $M''$, then $\Gamma$ acts amenably on $\hat{A}$.

**Proof.** The action of $\Gamma$ on $\beta \Gamma$ is amenable, cf. [3, Theorem 5.1.6]. Let $(m_i)_{i \in I}$ be a net of approximately invariant continuous means $m_i: \beta \Gamma \to \text{Prob}(\Gamma)$ that witnesses the amenability of this action, cf. Definition 2.1 in [1]. Let $\tilde{m}_i$ be the restriction of $m_i$ to $X_A \subseteq \beta \Gamma$. Then $(\tilde{m}_i)_{i \in I}$ is a net of approximate invariant continuous means $\tilde{m}_i: X_A \to \text{Prob}(\Gamma)$ that witnesses the amenability of the action of $\Gamma$ on $X_A$. As $\Gamma$ is countable and $X_A$ is $\sigma$-compact, the approximate mean $(\tilde{m}_i)_{i \in I}$ can be chosen to be countable.

The spectrum $Y := \hat{A}$ of a $\Gamma$-invariant sub-$C^*$-algebra $\mathcal{A}$ of $\mathcal{C}_A$ is a $\Gamma$-invariant quotient of the space $X_A = \hat{C}_A$. We wish to make sure that each of the functions $\tilde{m}_i$ passes to this quotient to yield a continuous function $\tilde{m}_i: Y \to \text{Prob}(\Gamma)$, in which case the net $(\tilde{m}_i)_{i \in I}$ will witness that $\Gamma$ acts amenably on $Y$. This will happen if for all $x, y \in X_A$, for which $\tilde{m}_i^x \neq \tilde{m}_i^y$ for some $i$, there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.

For each $i$ and for each $t \in \Gamma$ let $f_{i,t}: X_A \to \mathbb{R}$ be the continuous function given by $f_{i,t}(x) = \tilde{m}_i^x(\{t\})$, $x \in X_A$. For each $s \in \Gamma$ let $f_{i,t,s}$ be the restriction of $f_{i,t}$ to the compact-open subset $s.K_A$ of $X_A$. The countable set

$$M'' = \{f_{i,t,s} \mid i \in I, s, t \in \Gamma\},$$

will then separate any pair of points $x, y \in X_A$ that are separated by the $\tilde{m}_i$’s.

Recall the definition of paradoxicality of projections from Section 4.

**Lemma 6.3** (cf. [17, Lemma 6.6]). For each projection $p \in \mathcal{C}_A$, which is $(\ell^\infty(\Gamma),\Gamma)$-paradoxical, there is a finite set $M_p \subseteq \mathcal{C}_A$ such that whenever $\mathcal{A}$ is a $\Gamma$-invariant sub-$C^*$-algebra of $\mathcal{C}_A$ which contains $M_p \cup \{p\}$, then $p$ is $(\mathcal{A},\Gamma)$-paradoxical.

**Proof.** Note first that if $p$ in $\mathcal{C}_A$ is $(\ell^\infty(\Gamma),\Gamma)$-paradoxical, then it is also $(\mathcal{C}_A,\Gamma)$-paradoxical. Indeed, the projections $p_j \in \ell^\infty(\Gamma)$ that witness that $p$ is $(\ell^\infty(\Gamma),\Gamma)$-paradoxical, cf. [17,2], will automatically belong to the closed ideal $\mathcal{C}_A$ in $\ell^\infty(\Gamma)$. Accordingly we can let $M_p$ consist of the finitely many projections $p_j$ from [17].

**Lemma 6.4.** Each $\Gamma$-full projection in $\mathcal{C}_A$ is $(\ell^\infty(\Gamma),\Gamma)$-paradoxical if $\mathcal{A} \subseteq \Gamma$ is paradoxical.

**Proof.** Take a $\Gamma$-full projection in $\mathcal{C}_A \subseteq \ell^\infty(\Gamma)$ and write it as $1_B$ where $B \subseteq \Gamma$. Identify $1_B$ with the $\Gamma$-full projection $1_K \in \mathcal{C}_0(X_A)$. Then $K_B$ is $\Gamma$-full in $X_A$, which contains $\Gamma \subseteq \mathbb{N}$.
and so $A \subset \Gamma B \subset \Gamma A$ by Lemma 4.9. Hence $B$ is paradoxical by Lemma 2.5(iii), so $1_B$ is a paradoxical projection, cf. claim (c) below Definition 4.4. □

**Lemma 6.5 (cf. [17, Lemma 6.7]).** Let $T$ be a countable subset of $C_A$. Then there is a countable $\Gamma$-invariant set $Q$ consisting of projections in $C_A$ such that $T \subseteq C^*(Q)$.

**Proof.** The proof is identical with the proof of [17, Lemma 6.7]. Use that $C_A$ is of real rank zero (being a closed two-sided ideal in the real rank zero $C^*$-algebra $\ell_\infty(\Gamma)$).

If the group $\Gamma$ acts on a $C^*$-algebra $A$ and if $p \in A$ is a projection, then we say that $p$ is $\Gamma$-full if $p$ is not contained in any proper closed two-sided $\Gamma$-invariant ideal in $A$.

In the rest of this section we let $t \mapsto \alpha_t$, $t \in \Gamma$, denote the canonical actions of $\Gamma$ on $\ell_\infty(\Gamma)$, respectively, on $C(\beta\Gamma)$, as well as on all their $\Gamma$-invariant sub-$C^*$-algebras.

**Proposition 6.6.** Let $\Gamma$ be a non-supramenable group, and let $A$ be a subset of $\Gamma$. As above, let $C_A$ be the smallest closed $\Gamma$-invariant ideal in $\ell_\infty(\Gamma)$ that contains $1_A$. It follows that there is a separable, $\Gamma$-invariant sub-$C^*$-algebra $A$ of $C_A$ such that

(i) $1_A \in \ell_\infty(\Gamma)$ is a $\Gamma$-full projection in $A$,

(ii) $A$ is generated as a $C^*$-algebra by its projections,

(iii) Every projection in $A$, which is $(\ell_\infty(\Gamma), \Gamma)$-paradoxical, is also $(A, \Gamma)$-paradoxical,

(iv) $\Gamma$ acts freely on the character space $\hat{A}$ of $A$.

Furthermore, if $\Gamma$ is exact, then $A$ can be chosen such that $\Gamma$ acts amenably on $\hat{A}$.

**Proof.** If $\Gamma$ is exact, then let $M = M' \cup M'' \cup \{1_A\}$, and let otherwise $M = M' \cup \{1_A\}$, where $M'$ and $M''$ are as in Lemma 6.1 and Lemma 6.2 respectively. Suppose that $A$ is a $\Gamma$-invariant (separable) sub-$C^*$-algebra of $C_A$ which contains $M$. Then $\Gamma$ acts freely on $\hat{A}$, the approximate unit $\{p_n\}_{n=1}^\infty$ defined above Lemma 6.1 is contained in and is an approximate unit for $A$ by Lemma 6.4 and if $\Gamma$ is exact, then $\Gamma$ acts amenably on $\hat{A}$ by Lemma 6.2.

Let us show that $1_A$ is a $\Gamma$-full projection in any such $C^*$-algebra $A$. Let $I$ be the smallest closed $\Gamma$-invariant ideal in $A$ which contains $1_A$. Then $1_{tA} = \alpha_t(1_A) \in I$ for all $t \in \Gamma$. Since $p_n = 1_{B_n}$ and $B_n = \bigcup_{t \in F_n} tA$ for some finite subset $F_n$ of $\Gamma$, we see that $p_n \leq \sum_{t \in F_n} 1_{tA} \in I$, so $p_n \in I$ for all $n$ as $\{p_n\}_{n=1}^\infty$ is an approximate unit for $A$ we conclude that $I = A$, so $1_A$ is $\Gamma$-full in $A$.

Construct inductively countable $\Gamma$-invariant sets of projections

$$Q_0 \subseteq Q_1 \subseteq Q_2 \subseteq \cdots \subseteq C_A, \quad P_0 \subseteq P_1 \subseteq P_2 \subseteq \cdots \subseteq C_A,$$

Documenta Mathematica 18 (2013) 1597–1626
such that $M \subseteq C^*(Q_0)$, $P_n$ is the set of projections in $C^*(Q_n)$, and such that each projection in $P_n$, which is $(\ell^\infty(\Gamma), \Gamma)$-paradoxical, is $(C^*(Q_{n+1}), \Gamma)$-paradoxical. The set of projections in any separable abelian $C^*$-algebra is countable. Therefore $P_n$ is countable if $Q_n$ is countable.

The existence of $Q_0$ follows from Lemma 6.5. Suppose that $n \geq 0$ and $Q_n, P_n \subseteq C_A$ have been found. The existence of $Q_{n+1}$ then follows from Lemma 6.3 and Lemma 6.6.

Put

$$P_\infty = \bigcup_{n=1}^{\infty} P_n, \quad A = C^*(P_\infty).$$

Then $A$ is a $\Gamma$-invariant separable sub-$C^*$-algebra of $C_A$ which contains $M$. By definition, $A$ is generated by its projections, so (ii) holds.

Observe that each projection in $A$ actually belongs to $P_\infty$. Indeed, if $p$ is a projection in $A$, then, for some $n$, there is a projection $p' \in C^*(P_n)$ such that $\|p - p'\| < 1$. Since $A$ is commutative this entails that $p = p'$. Each projection in $C^*(P_n)$ belongs to $P_{n+1}$, so $p \in P_{n+1}$. Let $p$ be a projection in $A$ which is $(\ell^\infty(\Gamma), \Gamma)$-paradoxical. Then $p \in P_n$ for some $n$. It follows that $p$ is $(C^*(Q_{n+1}), \Gamma)$-paradoxical and hence also $(A, \Gamma)$-paradoxical. Thus (iii) holds.

Recall the definition of the relation $A \preceq \Gamma$ from Definition 6.6.

**Lemma 6.7.** Let $\Gamma$ be a countable group and let $A$ be a non-empty subset of $\Gamma$ such that $A \preceq \Gamma$. Let $A \subseteq C_A$ be as in Proposition 6.6. Then $A$ is non-unital, and so is $A/\mathcal{I}$ for every proper closed $\Gamma$-invariant ideal $\mathcal{I}$ in $A$.

**Proof.** As in the beginning of this section (and as in the proof of Proposition 6.6 above) let $\{p_n\}$ be the approximate unit for $C_A$ and for $A$, given by $p_n = 1_{B_n}$, where $B_n = \bigcup_{t \in F_n} tA$ for some increasing sequence, $\{F_n\}$, of finite subsets of $\Gamma$ whose union is $\Gamma$.

Let $\mathcal{I}$ be a $\Gamma$-invariant closed two-sided ideal in $A$ and suppose that $A/\mathcal{I}$ is unital. Then $p_n + \mathcal{I}$ is a unit for $A/\mathcal{I}$ for some $n$. Now, $A \not\subseteq \Gamma \backslash B_n$ by the assumption that $A \preceq \Gamma$. Hence there is a partition $A_1, \ldots, A_k$ of $A$ and elements $t_1, \ldots, t_k$ in $\Gamma$ such that $\{t_j A_j\}$ are pairwise disjoint subsets of $\Gamma \backslash B_n$.

For each $j$,

$$0 = (1_{t_j A_j} + \mathcal{I})(1_{B_n} + \mathcal{I}) = (1_{t_j A_j} + \mathcal{I})(p_n + \mathcal{I}) = 1_{t_j A_j} + \mathcal{I}.$$}

Hence $1_{t_j A_j}$ belongs to $\mathcal{I}$ for all $j$, so $1_A = \sum_{j=1}^{n} 1_{t_j A_j}$ belongs to $\mathcal{I}$. Finally, because $1_A$ is a $\Gamma$-full projection in $A$, we must have $\mathcal{I} = A$.  

By the *locally compact non-compact Cantor set* $K^*$ we shall mean the unique (up to homeomorphism) locally compact, non-compact, totally disconnected second countable Hausdorff space that has no isolated points. This set arises, for example, from the usual (compact) Cantor set $K$ by removing one point. One can also realise it as the product $\mathbb{N} \times K$, as the local field $Q_p$, etc.
Proof of Theorem 7.2. The “only if” part follows from Proposition 2.7 (and Lemma 4.5). The conclusions about the crossed product $C_0(K^*) \rtimes_{red} \Gamma$ follow from Proposition 4.2 and Lemma 4.5, respectively, from Proposition 4.6.

Let $X_A$ and $C_A$ be as defined in the beginning of this section, and let $A$ be as in Proposition 6.6. The $C^*$-algebra $A$ contains the projection $1_A$, and this projection is furthermore $\Gamma$-full in $A$ by Proposition 6.6. It follows that $A$ contains a maximal proper closed $\Gamma$-invariant ideal $I$ (cf. Remark 5.2). Put $B = A/I$. Let $X = \hat{B}$ and $Y = \hat{A}$.

Then $B \cong C_0(X)$, $A \cong C_0(Y)$, and $X$ is a closed $\Gamma$-invariant subset of $Y$. We know from Proposition 6.6 that $\Gamma$ acts freely on $Y$, and the action is moreover amenable if $\Gamma$ is exact. These properties pass to the subset $X$, so the action of $\Gamma$ on $X$ is free, amenable (if $\Gamma$ is exact), and minimal (by maximality of $I$).

We know from Proposition 6.5 that $A$ is generated by its projections. Hence $B$ is also generated by its projections. Hence $X$ and $Y$ are totally disconnected. As $A$ and $B$ are separable we conclude that $X$ and $Y$ are second countable. We show below that the action of $\Gamma$ on $X$ is purely infinite. This clearly will imply that $X$ has no isolated points (an isolated point is compact-open and of course never paradoxical). If $\Gamma$ is non-amenable, then $A \ll \Gamma$ which by Lemma 6.7 entails that $B$ is non-unital, whence $X$ is non-compact. We arrive at the same conclusion (that $X$ is non-compact) if $\Gamma$ is amenable, because no action of an amenable group on a compact Hausdorff space can be purely infinite. We can therefore conclude that $X$ is homeomorphic to $K^*$.

We finally show that the action of $\Gamma$ on $X$ is purely infinite. Let $K$ be a non-empty compact-open subset of $X$. Then $K$ is $\Gamma$-full in $X$ because $X$ is a minimal $\Gamma$-space. Hence there exists a $\Gamma$-full compact-open subset $K'$ of $Y$ such that $K = K' \cap X$, cf. Lemma 4.8(i). Let $p \in A$ be the projection that corresponds to the projection $1_{K'} \in C_0(Y)$. Then $p$ is $\Gamma$-full in $A$ and therefore also $\Gamma$-full in $C_A$. Hence $p$ is $(\ell^\infty(\Gamma), \Gamma)$-paradoxical because $A$ is paradoxical, cf. Lemma 6.4. By the construction of $A$, cf. Proposition 6.6 it follows that $p$ is $(A, \Gamma)$-paradoxical. Hence $1_{K'}$ is $(C_0(Y), \Gamma)$-paradoxical, so $K'$ is $((Y, \Gamma, \mathbb{K})$-paradoxical, cf. claim (c) (below Definition 4.4). By Lemma 4.8(ii) this implies that $K$ is $(X, \Gamma, \mathbb{K})$-paradoxical. 

7. AMENABLE ACTIONS OF NON-EXACT GROUPS ON LOCALLY COMPACT HAUSSDORFF SPACES

It is well-known that only exact (discrete) group can act amenably on a compact Hausdorff space, cf. [6, Theorem 5.1.7]. We also know that any group (exact or not) admits an amenable action on some locally compact Hausdorff space, for example on the group itself. We shall here address the issue of when a non-exact group admits amenable actions on the locally compact non-compact
Cantor set $K^*$, and whether the second part of Theorem 1.2 can be extended to all countable non-supramenable groups. The goal is to prove Theorem 1.3.

We first recall a notion of induced action. Let $\Gamma_0$ be a subgroup of a countable group $\Gamma$, and suppose that $\Gamma_0$ acts on a locally compact Hausdorff space $X$. Then $\Gamma$ acts by left-multiplication on the first coordinate of $\Gamma \times X$, and $\Gamma_0$ acts on $\Gamma \times X$ by $s(t,x) = (ts^{-1}, s.x)$, for $s \in \Gamma_0$, $(t,x) \in \Gamma \times X$. These two actions commute and thus $\Gamma$ acts on $Y = (\Gamma \times X)/\Gamma_0$. This action is called the induced action. We record some facts about this construction:

**Lemma 7.1.** Let $\Gamma_0 \curvearrowright X$ and $\Gamma \curvearrowright Y$ be as above. Then:

(i) If $X \cong K$, then $Y \cong K$ or $Y \cong K^*$ according to whether $|\Gamma : \Gamma_0|$ is finite or infinite.

(ii) If $X \cong K^*$, then $Y \cong K^*$.

(iii) If $\Gamma_0 \curvearrowright X$ is free, then so is $\Gamma \curvearrowright Y$.

(iv) If $\Gamma_0 \curvearrowright X$ is minimal, then so is $\Gamma \curvearrowright Y$.

(v) If $\Gamma_0 \curvearrowright X$ is amenable, then so is $\Gamma \curvearrowright Y$.

(vi) If $X$ admits a non-zero $\Gamma_0$-invariant Radon measure, then $Y$ admits a non-zero $\Gamma$-invariant Radon measure.

(vii) If $X$ admits a non-zero $\Gamma_0$-invariant Radon measure, then $Y$ admits a non-zero $\Gamma$-invariant Radon measure.

(viii) If $\Gamma_0 \curvearrowright X$ is purely infinite, then so is $\Gamma \curvearrowright Y$.

**Proof.** Let $\pi : \Gamma \times X \to Y$ denote the quotient mapping.

(i) Write $\Gamma = \bigcup_{t \in I} t \Gamma_0$ as a disjoint union of left cosets and put $A = \{t_\alpha \mid \alpha \in I\}$. Then $A \times X$ is a clopen subset of $\Gamma \times X$ which is a transversal for the action of $\Gamma_0$ on $\Gamma \times X$. The restriction of $\pi$ to $A \times X$ therefore defines a homeomorphism $A \times X \to Y$.

(ii) and (iii) follow from (i).

(iv) Suppose that $t.\pi(s,x) = \pi(s,x)$ for some $t \in \Gamma$ and some $(s,x) \in \Gamma \times X$. Then there is $r \in \Gamma_0$ such that $(ts,x) = (sr^{-1}, r.x)$. As $\Gamma_0$ acts freely on $X$ this implies that $r = e$, so $t = e$.

(v) Each $\Gamma \times \Gamma_0$ orbit on $\Gamma \times X$ is dense, if $\Gamma_0$ acts minimally on $X$. Hence each $\Gamma$ orbit on $Y$ is dense.

(vi) Let $m_i : X \to \text{Prob}(\Gamma_0)$, $i \in I$, be approximate invariant continuous means that witness the amenability of the action of $\Gamma_0$ on $X$. This is defined for each $m \in \text{Prob}(\Gamma_0)$ and for each $t \in \Gamma$ define $m_t \in \text{Prob}(\Gamma)$ by $m_t(E) = m(t^{-1}E \cap \Gamma_0)$, for $E \subseteq \Gamma$. Identify $Y$ with $A \times X$ as in (i), and define

$$\tilde{m}_i : A \times X \to \text{Prob}(\Gamma) \quad \text{by} \quad \tilde{m}_i(t_{\alpha} x) = (m_{t_{\alpha}}^x)_{\alpha} \quad x \in X, \alpha \in I.$$ 

Let $s \in \Gamma$ and $(t_{\alpha}, x) \in A \times X$ be given. Then $st_{\alpha} = t_{\beta} r$ for some (unique) $\beta \in I$ and $r \in \Gamma_0$, and $s.(t_{\alpha}, x) = (t_{\beta}, r.x)$. Let $E \subseteq \Gamma$. Then

$$\tilde{m}_i^{(t_{\alpha}, x)}(E) = \tilde{m}_i^{(t_{\beta}, r.x)}(E) = m_{t_{\beta}}^{r.x}(t_{\beta}^{-1}E \cap \Gamma_0) = m_{t_{\beta}}^{r.x}(r_{\alpha}^{-1}s^{-1}E \cap \Gamma_0).$$
The latter expression is close to $m_i^x(t^{-1}s^{-1}E \cap \Gamma_0)$ when $i$ is large by approximate invariance of $(m_i)$, and
\[ s_\ast \tilde{m}_i^{(t,x)}(E) = \tilde{m}_i^{(t_\ast x)}(s^{-1}E) = m_i^x(t^{-1}s^{-1}E \cap \Gamma_0). \]
This shows that $(\tilde{m}_i)_i$ is an approximate invariant continuous mean.

(vii). Identify again $Y$ with $A \times X$ as in (i). Let $\lambda$ be a non-zero $\Gamma_0$-invariant Radon measure on $X$ and let $\mu$ be counting measure on $A$. Then $\mu \otimes \lambda$ is a $\Gamma$-invariant non-zero Radon measure on $A \times X$.

(viii). If $F$ and $F'$ are disjoint compact-open subsets of $Y$ both of which are $(Y,\Gamma)$-paradoxical, then $F \cup F'$ is also $(Y,\Gamma)$-paradoxical. We need therefore only show that $F = \pi(\{t\} \times E)$ is $(Y,\Gamma)$-paradoxical for each $t \in \Gamma$ and for each compact-open subset $E$ of $X$. As $E$ is $(X,\Gamma_0)$-paradoxical there are pairwise disjoint compact-open subsets $E_1, E_2, \ldots, E_{n+m}$ of $X$ and elements $t_1, t_2, \ldots, t_{n+m} \in \Gamma_0$ such that (4.1) holds. Put $F_j = \pi(\{t\} \times E_j)$ and let $s_j \in \Gamma$ be such that $s_j t = t_j$. Then $F_1, F_2, \ldots, F_{n+m}$ are pairwise disjoint compact-open subsets of $F$. As
\[ s_j \pi(t, x) = \pi(s_j t, x) = \pi(t t_j, x) = \pi(t, t_j, x), \]
we see that $s_j F_j = \pi(\{t\} \times t_j, E_j)$. This shows that
\[ F = \bigcup_{j=1}^n s_j F_j = \bigcup_{j=n+1}^{n+m} s_j F_j, \]
and hence that $F$ is $(Y,\Gamma)$-paradoxical. □

Proof of Theorem 1.3. In each case we have a subgroup $\Gamma_0$ of $\Gamma$ which admits a free minimal amenable action on $X$, where $X = K$ or $X = K^\ast$. This action induces a free minimal amenable action of $\Gamma$ on $Y$, where $Y = K$ or $Y = K^\ast$ (according to the conclusions of (ii) and (iii) in Lemma 7.1).

(i). Here $\Gamma_0$ acts on $X = K$ in the prescribed way by [11], or by [17]. The former reference gives a free minimal action on $K$ for every infinite $\Gamma_0$, and this action will be amenable if $\Gamma_0$ is amenable. The latter reference gives a free minimal amenable action on $K$, whenever $\Gamma_0$ is exact and non-amenable. We conclude by inducing the action and recall that a non-exact group cannot act amenable on a compact set. If $|\Gamma : \Gamma_0| = \infty$, then $Y = K_\ast$, see Lemma 7.1.

(ii). It is known that $\Gamma_0 = Z$ has many interesting minimal actions on $X = K_\ast$, see [8]. These actions are necessarily free since every non-trivial subgroup of $Z$ has finite index. Pick such an action; it is of course amenable since $Z$ is amenable. Therefore it induces the desired action of $\Gamma$ on $Y = K_\ast$. If $\Gamma_0$ is infinite and amenable, then $\Gamma_0$ has a free minimal action on $X = K$ by [11], and this action is necessarily amenable. Again, if $|\Gamma : \Gamma_0| = \infty$, then $Y = K_\ast$.

(iii). Here $\Gamma_0$ is exact and non-supramenable. By Theorem 1.2 there is a free minimal amenable purely infinite action of $\Gamma_0$ on $X = K_\ast$. This action induces a free minimal amenable purely infinite action of $\Gamma$ on $Y = K_\ast$. □

Questions 7.2. Let $\Gamma$ be a countably infinite group.
(i) Does $\Gamma$ admit a free minimal amenable action on $K^*$?
In that case, $C_0(K^*) \rtimes_{\text{red}} \Gamma$ is a (non-unital) simple nuclear separable $C^*$-algebra in the UCT class.

(ii) Does $\Gamma$ admit a free minimal amenable action on $K^*$ which leaves invariant a non-zero Radon measure on $K^*$?
In that case, $C_0(K^*) \rtimes_{\text{red}} \Gamma$ is a (non-unital) simple nuclear separable $C^*$-algebra in the UCT class with a densely defined trace (hence it is stably finite).

(iii) Does $\Gamma$ admit a free minimal amenable purely infinite action on $K^*$ if $\Gamma$ is non-supramenable?
In that case, $C_0(K^*) \rtimes_{\text{red}} \Gamma$ is a stable Kirchberg algebra in the UCT class.

It follows from Theorem 1.3 that Question (i) has an affirmative answer if $\Gamma$ contains an element of infinite order, or if $\Gamma$ contains an infinite exact subgroup of infinite index, or if $\Gamma$ contains an infinite exact subgroup and $\Gamma$ itself is non-exact. Question (ii) has an affirmative answer if $\Gamma$ contains an element of infinite order, or if $\Gamma$ contains an infinite amenable subgroup of infinite index. Question (iii) has an affirmative answer if $\Gamma$ contains an exact non-supramenable subgroup.

On the other hand, as for Question (i), we do not even know if every countable amenable group $\Gamma$ admits a free minimal action on $K^*$. If there is an (infinite) set $A \subseteq \Gamma$ such that $A \ll \Gamma$ and such that some minimal $\Gamma$-subspace of $X_A$ is non-discrete, then the construction in Proposition 5.5 and Section 6 will give such an action. Example 5.6 shows that the existence of a non-discrete minimal $\Gamma$-subspace of $X_A$ is not guaranteed, even if $A$ is infinite.

REFERENCES


\footnote{Added in proof: It will be shown in a subsequent paper by the third named author and H. Matui that each infinite countable group $\Gamma$ indeed does contain a subset $A$ with these properties.}


<table>
<thead>
<tr>
<th>Julian Kellerhals</th>
<th>Nicolas Monod</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPFL</td>
<td>EPFL</td>
</tr>
<tr>
<td>SB-EGG</td>
<td>SB-EGG</td>
</tr>
<tr>
<td>Station 8</td>
<td>Station 8</td>
</tr>
<tr>
<td>1015 Lausanne</td>
<td>1015 Lausanne</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Switzerland</td>
</tr>
<tr>
<td><a href="mailto:julian.kellerhals@epfl.ch">julian.kellerhals@epfl.ch</a></td>
<td><a href="mailto:nicolas.monod@epfl.ch">nicolas.monod@epfl.ch</a></td>
</tr>
</tbody>
</table>

Mikael Rørdam  
Department of Mathematics  
University of Copenhagen  
Universitetsparken 5  
DK-2100  
Copenhagen Ø  
Denmark  
rordam@math.ku.dk