Multivariable Hypergeometric Functions

Eric M. Opdam

Abstract. The goal of this lecture is to present an overview of the modern developments around the theme of multivariable hypergeometric functions. The classical Gauss hypergeometric function shows up in the context of differential geometry, algebraic geometry, representation theory and mathematical physics. In all cases it is clear that the restriction to the one variable case is unnatural. Thus from each of these contexts it is desirable to generalize the classical Gauss function to a class of multivariable hypergeometric functions. The theories that have emerged in the past decades are based on such considerations.

1. The Classical Gauss Hypergeometric Function

The various interpretations of Gauss' hypergeometric function have challenged mathematicians to generalize this function. Multivariable versions of this function have been proposed already in the 19th century by Appell, Lauricella, and Horn. Reflecting developments in geometry, representation theory and mathematical physics, a renewed interest in multivariable hypergeometric functions took place from the 1980's. Such generalizations have been initiated by Aomoto [1], Gelfand and Gelfand [14], and Heckman and Opdam [19], and these theories have been further developed by numerous authors in recent years.

The best introduction to this story is a recollection of the role of the Gauss function itself. So let us start by reviewing some of the basic properties of this classical function. General references for this introductory section are [24, 12], and [39].

The *Gauss hypergeometric series* with parameters $a, b, c \in \mathbf{C}$ and $c \notin \mathbf{Z}_{\leq 0}$ is the following power series in z:

$$F(a, b, c; z) := \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n n!} z^n \,. \tag{1}$$

The *Pochhammer symbol* $(a)_n$ is defined by $(a)_n = a(a+1)\ldots(a+n-1)$ for $n \geq 1$, and $(a)_0 = 1$. This series is easily seen to be convergent when $|z| < 1$.

Gauss proved a number of remarkable facts about this function. He showed that

Proposition 1.1. *The hypergeometric series $F(a,b,c;z)$ and any two additional hypergeometric series whose 3-tuples of parameters are equal to (a,b,c) modulo $\mathbf{Z}^3$, satisfy a nontrivial linear relation with coefficients in the ring of polynomials in a, b, c, and z.*

A hypergeometric series whose parameters are $(a\pm 1,b,c)$, $(a,b\pm 1,c)$ or $(a,b,c\pm 1)$ is called *contiguous* to $F(a,b,c;z)$. Gauss worked out the basic cases of the relations between $F(a,b,c;z)$ and two of its contiguous functions, known as the contiguity relations of Gauss. Using such relations, he proved the famous "Gauss summation formula":

Lemma 1.2. *When $c \notin \{0,-1,-2,\dots\}$, and $\mathrm{Re}(c-a-b)>0$, then*

$$F(a,b,c;1)=\frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\,.$$

When we differentiate the series (1) we obtain

$$\frac{d}{dz}F(a,b,c;z)=\frac{ab}{c}F(a+1,b+1,c+1;z)\,. \tag{2}$$

As a special case of proposition 1.1 there exists a linear second order differential equation with polynomial coefficients for the series (1). By an easy direct computation one finds:

Proposition 1.3. *The Gauss series $F(a,b,c;z)$ satisfies the equation*

$$z(1-z)f''+(c-(1+a+b)z)f'-abf=0\,. \tag{3}$$

This equation is of Fuchsian type on the projective line $\mathbf{P}^1(\mathbf{C})$, and it has its *singular* points at $z=0,1$ and ∞. Locally in a neighborhood of any regular point $z_0\in\mathbf{C}\backslash\{0,1\}$ the space of holomorphic solutions to (3) will be two dimensional. This shows that we can continue any locally defined holomorphic solution of (3) holomorphically to any simply connected region in $\mathbf{C}\backslash\{0,1\}$. In particular, the series (1) has such holomorphic continuations. This leads us in a natural way to consider the *monodromy representation* of the Gauss hypergeometric function. Choose a regular base point z_0, and consider the associated two dimensional complex vector space V_{z_0} of solutions to (3). For each element $\gamma\in\Pi_1(\mathbf{C}\backslash\{0,1\},z_0)$ consider the operator $\mu(\gamma)\in\mathrm{End}(V_{z_0})$ representing the effect in V_{z_0} of analytic continuation of a local solution along a closed loop representing γ. This is easily seen to be a *representation*

$$\Pi_1(\mathbf{C}\backslash\{0,1\},z_0)\to\mathrm{GL}(V_{z_0})\,. \tag{4}$$

This representation is very fundamental to the subject. The monodromy representation has important interpretations in algebraic geometry (Picard-Schwarz map) and representation theory (quantum Schur-Weyl duality), as we will see later.

1.1. Behavior at the singular points and monodromy

We can compute the monodromy representation of the hypergeometric function explicitly. This is based on the summation lemma 1.2. We need to study the behavior of the solutions of (3) near the singular points. By substitution into the hypergeometric equation (3) we find that apart from

$$w_{0,1}(z) = F(a,b,c;z)\,, \tag{5}$$

also the expression

$$w_{0,2}(z) = z^{1-c}F(1-c+b, 1-c+a, 2-c; z) \tag{6}$$

gives us a solution of (3), locally defined in sectors of a punctured disk centered at $z=0$. Treating the other singular points similarly, we obtain

$$\begin{aligned} w_{1,1}(z) &= z^{-a}F(a, a-c+1, a+b-c+1; 1-z^{-1})\,, \\ w_{1,2}(z) &= z^{-b}(1-z^{-1})^{c-a-b}F(c-a, 1-a, c-a-b+1; 1-z^{-1}) \end{aligned} \tag{7}$$

at $z=1$, and

$$\begin{aligned} w_{\infty,1}(z) &= z^{-a}(1-z^{-1})^{-a}F(a, c-b, a-b+1; (1-z)^{-1})\,, \\ w_{\infty,2}(z) &= z^{-b}(1-z^{-1})^{-b}F(b, c-a, b-a+1; (1-z)^{-1}) \end{aligned} \tag{8}$$

at $z=\infty$. This gives us a basis of local solutions in the vicinity of each of the singular points, at least when we assume that the parameters a, b and c do not differ by integers. Each of these 6 solutions can be expressed in 4 ways in terms of hypergeometric series (1), and together these constitute Kummer's 24 solutions of the hypergeometric differential equation. When the numbers a, b and c have integer differences, logarithmic terms are usually necessary to describe the local solutions at some of the singular points. This is an important phenomenon called *resonance*. We shall ignore this phenomenon for sake of simplicity.

When we want to understand the monodromy in terms of the local basis $w_{0,1}$, $w_{0,2}$, it is sufficient to find the relations with the other local bases (7) and (8) on a common domain. So let us write

$$w_{0,1} = c_1 w_{1,1} + c_2 w_{1,2}\,. \tag{9}$$

Since, when $\mathrm{Re}(c-a-b) > 0$, we have $w_{1,2}(1) = 0$, we obtain from the summation formula (2) that

$$c_1 = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\,. \tag{10}$$

By application of the Kummer transformation rules one can similarly deduce that

$$c_2 = \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}\,. \tag{11}$$

We can similarly deal with the transition to the basis (8).

All this shows us how the Kummer transformations together with the Gauss summation formula make it possible to obtain explicitly the matrices of the monodromy representation. It is a very special feature of the hypergeometric equation.

1.2. The Euler integral

There is another, more geometric way of thinking about the monodromy representation. It is based on the representation of local solutions by means of integrals over *twisted cycles*. The basic form of such a representation is the Euler integral:

Theorem 1.4. *When* $\mathrm{Re}(c) > \mathrm{Re}(a) > 0$*, and* $|z| < 1$*, then*

$$F(a,b,c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 t^{a-1}(1-t)^{c-a-1}(1-tz)^{-b}dt\,.$$

A proof of this theorem can be given by using the binomial expansion of $(1-tz)^{-b}$, and applying the Euler beta-integral formula.

The Euler integral gives rise to a new understanding of what we saw in the previous subsection. Let us first of all remark that we can replace the integration domain $[0,1]$ by any closed cycle C in $\mathbf{C}\backslash\{0,1,\frac{1}{z}\}$, provided that the integrand

$$t^{a-1}(1-t)^{c-a-1}(1-tz)^{-b} \tag{12}$$

is univalued on C. Such a cycle is called a twisted cycle for the coefficient system defined by (12). A famous example of a twisted cycle is the Pochhammer contour (see figure 1) around the points 0 and 1.

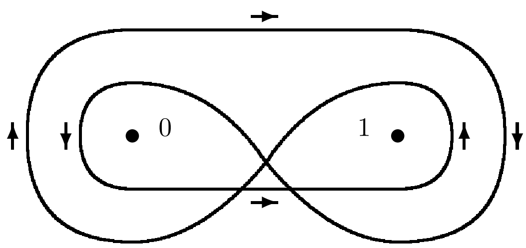

FIGURE 1. The Pochhammer contour

This has the advantage that we can remove the condition $\mathrm{Re}(c) > \mathrm{Re}(a) > 0$. Moreover, we obtain a linear map from the space of homology classes of twisted cycles in $Y_z = \mathbf{C}\backslash\{0,1,\frac{1}{z}\}$, to the space V_z of germs of local solutions at z:

$$\begin{aligned} H_1^{\mathrm{twist}}(Y_z) &\to V_z \\ C &\to \int_C t^{a-1}(1-t)^{c-a-1}(1-tz)^{-b}dt\,. \end{aligned} \tag{13}$$

For generic parameters, $H_1^{\mathrm{twist}}(Y_z)$ is two dimensional and this map is an isomorphism.

Put $X = \mathbf{C}\backslash\{0,1\}$ and $Y = \mathbf{C}^2\backslash\{t=0, t=1, zt=1\}$, and consider the projection $\pi\colon Y \to X$, $\pi(z,t) = z$ on the first coordinate.

$$\begin{array}{ccc} Y & \ni & (z,t) \\ {\scriptstyle\pi}\downarrow & & \downarrow \\ X & \ni & z \end{array} \tag{14}$$

This projection is a fibration with fiber $\pi^{-1}(z) = Y_z$. We define a vector bundle $H_1^{\mathrm{twist}}(Y/X)$ over X whose fiber at z is the twisted homology group $H_1^{\mathrm{twist}}(Y_z)$. An element $C_{z_0} \in H_1^{\mathrm{twist}}(Y_{z_0})$ naturally defines a twisted cycle in every fiber $H_1^{\mathrm{twist}}(Y_z)$ if z is sufficiently close to z_0. Such local sections of $H_1^{\mathrm{twist}}(Y/X)$ are called *flat*, and this natural notion of flat local sections defines an integrable connection on the bundle $H_1^{\mathrm{twist}}(Y/X)$. This is the Gauss Manin connection of the fibration π (with respect to the twisting by the local coefficient system). The "flat continuation" of elements of $H_1^{\mathrm{twist}}(Y_{z_0})$ defines a monodromy representation of $\Pi_1(X, z_0)$ in $\mathrm{GL}(H_1^{\mathrm{twist}}(Y_{z_0}))$.

In short, for generic parameters the isomorphism (13) interprets the monodromy action on the local solution space of the hypergeometric differential equation as the monodromy of the (twisted) Gauss-Manin connection of the fibration (14).

When the parameters a, b and c are rational, we have a projection of the space of 1-cycles of the Riemann surface Z_z of (12) to the space of twisted 1-cycles in the fiber Y_z. Variation of z in the base space X should be thought of as a variation of moduli of the surface Z. The hypergeometric functions are now interpreted as period integrals, considered as functions of the moduli of Z. This point of view gives rise to modular interpretations of X (or certain local compactifications of it) via the *Schwarz map* S. This is the *multivalued* map on X defined by taking the projective ratio

$$S(z) := (\phi_1(z) : \phi_2(z)) \tag{15}$$

of two linearly independent solutions of the hypergeometric differential equation. Its branches are related to each other by the action of the projective monodromy group Γ^+. The S-image of the upper half plane $X^+ \subset X$ is a circular triangle T called the Schwarz triangle. The vertices of T are the S-images of 0, 1, and ∞, and by (5), (7) and (8) its angles are

$$(1-c)\pi, \ (c-a-b)\pi, \ \text{and} \ (a-b)\pi \tag{16}$$

respectively. In order to avoid degeneracies we now assume that (a,b,c) is such that contiguous parameters give equivalent monodromy representations (this is true when $a, b \not\equiv 0, c$ modulo $\mathbf{Z}$). Applying contiguity relations repeatedly we can reduce T so that its angles are nonnegative, and that the sum of two angles is at most π. This ensures that the Schwarz map is a bijection from X^+ to T.

By Schwarz' reflection principle, Γ^+ is realized explicitly as the normal subgroup of index two of holomorphic maps in the group Γ generated by the inversions in the edges of the Schwarz triangle T. By proper choice of the basis ϕ_1, ϕ_2 in (15),

T can be realized as a geodesic triangle in one of the three standard geometries. If the angle sum σ of T exceeds π, we can realize T as a geodesic triangle in $\mathbf{D}^+ := \mathbf{P}^1(\mathbf{C})$ (spherical case). When $\sigma < \pi$, we can realize T as a geodesic triangle in the upper half plane $\mathbf{D}^- := \mathbf{H}$ (hyperbolic case). Finally, when $\sigma = \pi$, we can realize T as a Euclidean triangle in $\mathbf{D}^0 := \mathbf{C}$.

We call T elementary when its angles are of the form $\frac{\pi}{n}$ with $n \in \{2, 3, \ldots\}$. By elementary geometry in the natural geometric domain $\mathbf{D}^\epsilon$ ($\epsilon = \pm, 0$) of T, the group Γ^+ is a *discrete* subgroup of the group of isometries $\mathrm{Aut}(\mathbf{D}^\epsilon)$ if and only if T is finitely tesselated by copies of an elementary Schwarz triangle. When T is elementary, then its closure in its geometric domain $\mathbf{D}^\epsilon$ will be a fundamental domain for the action of Γ on $\mathbf{D}^\epsilon$.

When T is elementary, we can therefore find a holomorphic inverse J of S that extends to $\mathbf{D}^\epsilon$ by adding the points of finite branching order. The map J is automorphic for Γ^+ and realizes an isomorphism

$$J \colon \Gamma^+ \backslash \mathbf{D}^\epsilon \xrightarrow{\sim} \tilde{X} \tag{17}$$

where $\tilde{X}$ is obtained from X by adding the points corresponding to the points of finite branching order.

In the simplest case we consider $a = b = \frac{1}{2}$ and $c = 1$. All angles of T are 0 now. In this case the Euler integral solutions of the hypergeometric differential equation are in fact the classical *elliptic integrals*. We find that Z_z is a double cover of $\mathbf{P}^1(\mathbf{C})$ branched in 0, 1, ∞ and $\frac{1}{z}$, and this is an elliptic curve (with marked point of order two). The projective monodromy group is

$$\Gamma(2) = \left\{ g \in \mathrm{PSL}(2, \mathbf{Z}) \,\middle|\, g \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ modulo } 2 \right\}. \tag{18}$$

In this case, the inverse J is the *lambda invariant* that maps the quotient $\Gamma(2)\backslash\mathbf{H}^+$ isomorphically to $X = \mathbf{C}\backslash\{0, 1\}$. It is an isomorphism between two natural models of the moduli space of elliptic curves with marked points of order two.

We should think of the base space $X = \mathbf{C}\backslash\{0, 1\}$ as the space of configurations of four points in $\mathbf{P}^1(\mathbf{C})$, i.e. the space of positions of four distinct points on the projective line, up to the simultaneous action of projective transformations on these points. In this way it is natural to generalize the above interpretation of the Euler integral to more general configuration spaces of geometric objects. This is a fruitful point of view for generalizing hypergeometric functions, which has led to applications in algebraic geometry.

Configuration spaces are also natural to consider in mathematical physics, of course, and the hypergeometric functions in mathematical physics arise in this way also as functions of configurations. I shall return to these matters later.

One variable generalizations like the functions ${}_pF_q$ and q-deformations like the basic hypergeometric series are certainly important, but we shall restrict ourselves to discussing multivariable generalizations in this overview.

2. Generalizations of Euler's Hypergeometric Integral

The first generalization of hypergeometric functions that comes to mind when we consider the Euler integral is the Lauricella F_D function. Let $X(n)$ denote the space of $n \geq 4$ distinct, marked points $x_1, \ldots, x_n$ in $\mathbf{P}^1(\mathbf{C})$, modulo the action of $\mathrm{PGL}(2, \mathbf{C})$. The space $X(4)$ is nothing but the base space $X = \mathbf{C}\backslash\{0, 1\}$ we considered in the previous section, since we can send the first three points to 0, 1 and ∞ by a uniquely determined fractional linear map, leaving the fourth point as a free variable in X.

Let μ be an n tuple of complex numbers with $\sum \mu_i = 2$. Let w_μ denote the multivalued $(1, 0)$-form

$$w_\mu = \prod_{i=1}^{n} (t - x_i)^{-\mu_i} dt \tag{19}$$

on $Y_x := \mathbf{P}^1(\mathbf{C}) - \{x_1, \ldots, x_n\}$. For any twisted cycle C in Y_x with respect to this form, we define the the following hypergeometric integral

$$I_C(x) := \int_C w_\mu \,. \tag{20}$$

These integrals are solutions of the Lauricella hypergeometric equations of "type D" when we fix $x_{n-2} = 0$, $x_{n-1} = 1$, and $x_n = \infty$ and think of the $I_C(x)$ as functions of the remaining $n-3$ variables. It is known that this is an $n-2$ dimensional space $V(\mu)$ of multivalued functions on $X(n)$. Choose a base point $b \in X(n)$. The map $C \to I_C(x)$ defines an isomorphism $H_1^{\mathrm{twist}}(Y_b)$ and the space $V(\mu)(U)$ where $x \in U$ and U is a suitable neighborhood of b.

The analog of the Schwarz map in this context was studied by Picard ($n = 5$) [35], Terada [37], Deligne and Mostow [8, 9] and others. It was shown that when $\mu_i \in (0, 1)$, the space of twisted cycles $H_1^{\mathrm{twist}}(Y_x)$ carries an hermitian intersection form M of signature $(1, n-3)$, invariant for monodromy. Hence, for a suitable choice of basis C_i of $H_1^{\mathrm{twist}}(Y_b)$, the image of the Picard-Schwarz map

$$PS\colon X(n) \;\to\; \mathbf{P}^{n-3}(\mathbf{C}) \tag{21}$$

$$x \;\to\; (I_{C_1}(x) : \cdots : I_{C_{n-2}}(x)) \tag{22}$$

is inside the set $B = \{z = (z_1 : \cdots : z_{n-2}) | M(z, z) > 0\}$. The space B is isomorphic to the unit ball in $\mathbf{C}^{n-3}$.

The main theorem of [8] asserts that the projective monodromy group $\Gamma(\mu) \subset PU(1, n-3)$ is discrete if there exist $m_{i,j} \in \mathbf{N} \cup \infty$ such that $1 - \mu_i - \mu_j = \frac{1}{m_{i,j}}$ or $\frac{2}{m_{i,j}}$ when $\mu_i = \mu_j$. Moreover, the image of PS is dense in B, and PS^{-1} extends holomorphically to B and gives an isomorphism

$$PS^{-1}\colon \Gamma(\mu)\backslash B \to \Sigma\backslash\tilde{X}(n) \tag{23}$$

where Σ is the group of permutations of points x_i with equal weights μ_i, and $\tilde{X}(n)$ is some quasi projective local compactification of $X(n)$.

This is a delightful generalization of the theory of the Schwarz map. At the same time it is clear that it is not the end of the story! Other generalizations of the hypergeometric function can be obtained easily by considering hypergeometric integrals associated with configuration spaces of hyperplanes in $\mathbf{P}^n(\mathbf{C})$. For example, Yoshida obtained the modular interpretation of the configuration space $X(3,6)$ of 6 lines in $\mathbf{P}^2(\mathbf{C})$ in his book [39]. Other work in this direction was done by Couwenberg [7], working with the root system type hypergeometric functions that will be discussed in section 3. Many open problems remain in this direction.

2.1. The Gelfand-Kapranov-Zelevinskii-hypergeometric function

This hypergeometric function (sometimes called A-hypergeometric function) was introduced in [15]. It is in fact a very general class of hypergeometric functions that resembles the case of Lauricella functions. The classical generalizations of the Gauss hypergeometric function like ${}_pF_q$, the Lauricella type functions, and Horn's hypergeometric functions all occur as special cases of the GKZ-systems.

The GKZ-hypergeometric functions are defined by means of a deceptively simple system of differential equations. Let $A \subset \mathbf{Z}^n$ be a finite generating subset of $\mathbf{Z}^n$. Assume that A lies inside a rational hyperplane. In other words, there exists a linear function $h\colon \mathbf{Z}^n \to \mathbf{Z}$ such that $h(A) = 1$. Let $L \subset \mathbf{Z}^A$ denote the lattice of relations in A, thus

$$L := \{(a_\omega) \in \mathbf{Z}^A | \sum_{\omega\in A} a_\omega \omega = 0\}\,. \tag{24}$$

For $a \in L$, define a constant coefficient partial differential operator $\Box_a$ on $\mathbf{C}^A$ by

$$\Box_a := \prod_{a_\omega > 0} \left(\frac{\partial}{\partial x_\omega}\right)^{a_\omega} - \prod_{a_\omega < 0} \left(\frac{\partial}{\partial x_\omega}\right)^{a_\omega}\,. \tag{25}$$

Note that $\Box_a$ is homogeneous, since $\sum a_\omega = 0$ for every $a \in L$ (apply h to the relation defined by a).

Also define, for every $i = 1, \ldots, n$,

$$Z_i = \sum_{\omega\in A} \omega_i x_\omega \left(\frac{\partial}{\partial x_\omega}\right)\,. \tag{26}$$

When $(\gamma_1, \ldots, \gamma_n) \in \mathbf{C}^n$ is given, we define the following system of differential equations for functions on $\mathbf{C}^A$:

Definition 2.1. (GKZ-system of equations)

$$(1)\ \Box_a f = 0\ \forall a \in L, \quad (2)\ Z_i f = \gamma_i f\ \forall i = 1, \ldots, n\,. \tag{27}$$

It is known that the system is *holonomic*, i.e. the system has finite dimensional local solutions spaces. It was shown in [15] that the dimension of the local solution space at a regular point is at least equal to the volume of the convex hull of A inside the rational hyperplane containing A, with equality in the nonresonant case. However, an exact general formula for the dimension doesn't seem to be known [17]. The monodromy representation is also not known in general. Gelfand, Kapranov and Zelevinskii [16, 17] have shown that in the non-resonant case the solutions of the system can be represented by generalized Euler integrals. The GKZ-hypergeometric function contains the hypergeometric functions on Grassmannians which were defined previously by Gelfand and Gelfand [14]. Also the hypergeometric integrals studied by Aomoto [1] are of GKZ-type. It was shown by Batyrev [2] that the period integrals of Calabi-Yau hypersurfaces satisfy a system of GKZ-hypergeomtric equations.

3. Analogs of Spherical Functions on Symmetric Spaces

In this section we shall discuss a different kind of multivariable hypergeometric function, the hypergeometric function associated to root systems. This theory is based on other aspects of the Gauss hypergeometric function, namely its role in the representation theory of groups like $\mathrm{SL}(2, \mathbf{R})$.

It is well known that Bessel functions of the half integer order $n/2$ show up as the radial eigenfunctions of the Laplace operator Δ of the Euclidean space $\mathbf{R}^n$. This has a generalization to hypergeometric functions, and this provides the basis of a theory of multivariable hypergeometric functions that is natural in relation to representation theory of reductive algebraic groups. The hypergeometric functions of this kind are called "hypergeometric functions associated to root systems" [19, 18, 20, 33], and are closely related to what is called "Macdonald-Cherednik theory" nowadays.

Let us review the basic construction of these functions. A *Riemannian symmetric space* X is a Riemannian manifold such that at every point p of X, the local geodesic inversion $i_p\colon \exp(tv) \to \exp(-tv)$ extends to a global isometry of X. With this assumption it follows simply that X is complete, and that the group G of isometries of X acts transitively on X. We choose a base point x_0 in X, and denote by K the stabilizer group of x_0 in G. The Lie group G acts transitively on X, and K is a compact subgroup of G which is pointwise fixed for the involution $g \to i_{x_0} g i_{x_0}$ of G.

The Euclidean spaces $\mathbf{R}^n$ are the simplest examples of such spaces. These are examples of *flat* symmetric spaces, by which we mean that the sectional curvature of these spaces is 0. Any simply connected Riemannian symmetric space is a product of factors with constant sectional curvature.

Let us assume from now on that X has sectional curvature -1. The local structure of X can be described by introducing "polar coordinates". Let A be a maximal flat totally geodesic submanifold through x_0. Then $A \simeq \mathbf{R}^r$ for some positive integer r. This dimension r is called the *rank* of X. It turns out that the group

$$W := \frac{\{k \in K \mid k \text{ stabilizes } A\}}{\{k \in K \mid k \text{ fixes } A \text{ pointwise}\}} \tag{28}$$

is a finite crystallographic reflection group acting on A. All K orbits intersect A, and this intersection is an orbit of W. The local structure of X is determined completely by the behavior of the function $\delta(x) := \mathrm{Vol}(Kx)$ on A. This function takes the form (see [21])

$$\delta(x) = \prod_{\alpha \in R} |\sinh \alpha(x)|^{\frac{1}{2}m_\alpha} \tag{29}$$

for a certain finite set of linear functions R on A, and certain non-negative integer labels m_α for the elements of R. Clearly δ has to be W invariant, implying that set R is W-stable and the labels are W-invariant. In fact, the orthogonal reflections in the hyperplanes $H_\alpha := \{x \mid \alpha(x) = 0\}$ ($\alpha \in R$) generate the reflection group W. Hence the local structure of X is determined by R and the labels m_α. We call R the *root system* of X, and the m_α are called the root multiplicities.

The analogs of the Laplace operator on $\mathbf{R}^n$ are the G invariant differential operators on X. The algebra of such operators is denoted by $\mathbf{D}(X)$. In the case $\mathbf{R}^n$ this is the operator algebra generated by the Laplace operator Δ. Although $\mathbf{D}(X)$ is in general no longer generated by a single operator, its structure is amazingly simple (cf. [21]):

Theorem 3.1. $\mathbf{D}(X)$ *is a polynomial algebra of rank* r *over* $\mathbf{C}$.

The analogs on X of the Bessel function of half integer order are the *elementary spherical functions* on X. An elementary spherical function ϕ on X (with origin x_0) is an eigenfunction of the algebra $\mathbf{D}(X)$ which is moreover K-invariant. In other words, there is an algebra homomorphism $\lambda\colon \mathbf{D}(X) \to \mathbf{C}$ such that

$$\Delta\phi = \lambda(\Delta)\phi \ \forall \Delta \in \mathbf{D}(X)\,. \tag{30}$$

Such a function depends on the "radial" variables $x \in A \simeq \mathbf{R}^n$ only. As in the case of spherical waves on $\mathbf{R}^n$, we derive the differential equations for $\phi(x)$ by separation of the radial and rotational variables. This reduces the equations (30) to a system of W-invariant equations on $A \simeq \mathbf{R}^n$. The simplest equation of this type is the second order equation that is derived from the Laplace-Beltrami operator $\Delta_{\mathrm{LB}} \in \mathbf{D}(X)$ of X. Its radial part $L = L(R, m_\alpha)$ has the following form on A:

$$L(R, m_\alpha)\phi = \Delta_A \phi + \sum_{\alpha \in R} m_\alpha \frac{\cosh(\alpha)}{\sinh(\alpha)} \alpha(\nabla_A \phi)\,, \tag{31}$$

where Δ_A is the Laplace operator of the Euclidean space A, and $\nabla_A \phi$ denotes the gradient vector of ϕ in A. When the rank r of X equals 1, the eigenfunction equations (30) reduce to the eigenfunction equation for L. In this case, the root system R has the form $R = \{\pm\beta, \pm 2\beta\}$. When we use $z = -\sinh^2(\alpha(x))$ as a new coordinate, we obtain the hypergeometric differential equation (3) whose parameters a, b and c can be expressed in terms of m_β, $m_{2\beta}$, and the eigenvalue λ. In order to obtain the full three parameter family of hypergeometric functions we have to abandon the rank one symmetric spaces altogether, and allow arbitrary complex values for the labels m_β and $m_{2\beta}$. The situation is similar to the case spherical waves on $\mathbf{R}^n$: only Bessel functions of half integer order allow such a geometric interpretation.

How can we imitate this step in higher rank situations? The operator L is a certain W-invariant deformation of Δ_A, which has the remarkable property that it defines a *completely integrable system.* That is to say, the algebra of W-invariant differential operators commuting with L contains a polynomial algebra of rank r (the radial parts of the operators $\Delta \in \mathbf{D}(X)$).

The crucial step towards the theory of hypergeometric function for root systems is the insight that this property of complete integrability is not lost when we choose arbitrary complex coefficients m_α in (31) instead of the positive integer labels dictated by the local structure of X (see [20, 34] and the references therein):

Theorem 3.2. *In the algebra of linear partial differential operators with polynomial coefficients in the unknowns m_α, the commutant algebra of the operator L given by* (31) *is isomorphic to a polynomial algebra $\mathbf{D}(R, m_\alpha)$ of rank r.*

This theorem is not merely an interpolation from the classical cases bases of the theory of Riemannian symmetric spaces X. In general, for a given root system R, "nature" has only given us finitely many symmetric spaces X with a root system of type R. Theorem 3.2 is therefore rather surprising, and points at something new. We will explore this in the next subsection.

Anyway, it is now clear how we should define the hypergeometric functions associated to a root system:

Theorem 3.3. *Let $m = (m_\alpha)$ be a set of complex root labels. Given a character λ of the algebra $\mathbf{D}(R, m_\alpha)$, the system of hypergeometric differential equations is defined on the complexification $A_{\mathbf{C}}$ by*

$$\Delta\phi = \lambda(\Delta)\phi \ \forall \Delta \in \mathbf{D}(R, m_\alpha)\,. \tag{32}$$

The system is invariant for W and for the lattice of translations T on which all the roots take values in $2\pi i$.

The system is regular at the regular points of the action of the affine reflection group $W \ltimes T$, and locally its solution space has dimension $|W|$ at regular points. There is a unique holomorphic solution $F_R(\lambda, m_\alpha; x)$ defined in a neighborhood of the origin in $A_{\mathbf{C}}$, normalized in the origin by $F_R(\lambda, m_\alpha; 0) = 1$. It is called *the hypergeometric function associated with R.*

This function has very elegant properties. Its monodromy representation has been determined, at least for generic parameters. The fundamental group is of the regular orbit space of $W \ltimes T$ acting on $A_{\mathbf{C}}$ is the affine braid group associated with W. The monodromy representations factors through an affine Hecke algebra quotient of the group algebra of the braid group [19].

Considering its origin, it is not surprising that the root system hypergeometric function can be used as the kernel of a deformation of the Fourier transform, generalizing the harmonic analysis of zonal spherical functions [33, 5]. This harmonic analysis contains a lot of combinatorial information about root systems, already in the polynomial case (corresponding to zonal polynomials on the compact form of X). We return to this issue in the next subsection. The spectral analysis using these functions is also related to the dynamics of the integrable models of "Calogero-Moser" type in mathematical physics.

There is no Euler type integral representation known in general, except for $W = S_n$. This has to be considered as a missing link. It makes it more difficult to give geometric meaning to the monodromy representation.

3.1. The Cherednik-Macdonald theory

Complete integrability of a system of differential operators is a rare and delicate property. The integrability of the Laplace-Beltrami operator L of X as in the previous subsection is indeed very special, as it reflects the geometry of X. The fact that the deformations of L in (30) do not destroy the the integrability is therefore remarkable, and it indicates that there should exist a more fundamental structure than the symmetric space X itself. On the algebraic level this structure is well understood. It is Ivan Cherednik's *double affine Hecke algebra* [4]. It simultaneously captures the so-called *spherical convolution algebras* of the p-adic symmetric spaces $X(\mathbf{Q}_p)$ and the algebras of G invariant differential operators $\mathbf{D}(X)$ (with X a real form of the symmetric space) as in the previous subsection.

Let us briefly look at this interesting object when the rank of X equals 1. We follow the nice presentation from [31]. As before, we put $R = \{\pm\beta, \pm 2\beta\}$. The double affine Hecke algebra $\mathcal{H}$ has generators $T_0, T_1, T_1^\vee$ and $T_0^\vee$ over the field K of rational functions in 5 indeterminates $t_i, t_i^\vee$ $(i = 0, 1)$ and q, with relations:

$$\begin{aligned}(T_i - t_i)(T_i + t_i^{-1}) &= 0\,,\\ (T_i^\vee - t_i^\vee)(T_i^\vee + t_i^{\vee -1}) &= 0\,,\\ T_0 T_1 T_1^\vee T_0^\vee &= q\,.\end{aligned}$$

The subalgebra $K\langle T_0, T_1\rangle$ of $\mathcal{H}$ generated by T_0 and T_1 is an ordinary affine Hecke algebra, which has a one dimensional representation ρ defined by

$$\rho(T_0) = q_0,\ \rho(T_1) = q_1\,. \tag{33}$$

The induced module $\mathrm{Ind}_{K\langle T_0,T_1\rangle}^{\mathcal{H}}(\rho)$ is naturally isomorphic to the Laurent polynomial ring $K[X, X^{-1}]$, where $X = T_1 T_1^\vee$. This defines a faithful representation π of $\mathcal{H}$, as operator algebra on $K[X, X^{-1}]$.

For example, the operator $\pi(T_1)$ is a "Lusztig operator"

$$\pi(T_1) = t_1 s_1 + (a_1 + a_1^\vee X^{-1})\frac{1}{1 - X^{-2}}(1 - s_1) \tag{34}$$

where s_1 is the involutive automorphism of $K[X, X^{-1}]$ defined by $s_1(X^n) = X^{-n}$, $a_i = t_i - t_i^{-1}$ and $a_i^\vee = t_i^\vee - t_i^{\vee -1}$. Similarly we have

$$\pi(T_0) = t_0 s_0 + (a_0 + q^{-1} a_0^\vee X)\frac{1}{1 - q^{-2}X^2}(1 - s_0)\,, \tag{35}$$

where $s_0(X^n) = q^{2n}X^{-n}$. These formulas can be checked by some disciplined direct computations. The analog of the radial part of the Laplace-Beltrami operator $L(R, m_\beta)$ of (31) is now given by the operator

$$\Lambda(R, t_i, t_i^\vee, q) := (\pi(Y) - 1)(1 - \pi(Y^{-1})) \tag{36}$$

on $K[X, X^{-1}]$, where $Y = T_1T_0$. The relation with the previous subsection is achieved by a limiting procedure $q \to 1$, after the specializations $t_0 = t_0^\vee = q^{-m_\beta/2}$, $t_1 = q^{-m_{2\beta}}$, and finally $t_1^\vee = 1$. When we formally write $X = \mathrm{e}^\beta$, we find by direct computation that the relation with the operator (31) is given by

$$L(R, m_\alpha) = \lim_{q\to 1} \frac{\Lambda(R, t_i, t_i^\vee, q)}{q - q^{-1}} \,. \tag{37}$$

Let us return to the general rank case now. Intelligible sources for this material are [29] and [23]. The double affine Hecke algebra $\mathcal{H}$ can be defined without too much difficulty. It consists of two dual affine Hecke algebras, whose finite dimensional Hecke-subalgebras are identified. As in the rank 1 case, this algebra has a faithful representation in a Laurent polynomial algebra $K[X_1^{\pm 1} \ldots, X_r^{\pm r}]$, and it contains a rank r polynomial subalgebra

$$\mathcal{L} = K[\Lambda_1(R, t_i, t_i^\vee, q), \ldots, \Lambda_r(R, t_i, t_i^\vee, q)] \,, \tag{38}$$

the "q-analog" of the algebra $\mathbf{D}(R, m_\alpha)$ of theorem 3.2.

The elementary spherical functions on the compact real form X^{comp} of a Riemannian symmetric space X, are the so-called zonal polynomials. In the rank one case these are the well known Jacobi-polynomials. A far reaching generalization of the zonal spherical polynomial is the Macdonald-Koornwinder polynomial [29, 27]. These are by definition the W-invariant polynomial eigenfunctions of the algebra $\mathcal{L}$, suitably normalized. In the one variable case, these are the polynomials of the q-Askey-Wilson scheme. The original zonal polynomials are obtained by the limit transition as described above. Hall-Littlewood polynomials, in their role of the elementary spherical functions on a p-adic symmetric space $X(\mathbf{Q}_p)$, arise as the limit for $q \to 0$ when we put $t_\alpha = \frac{1}{p}$.

The Macdonald polynomials have many interpretations in algebraic combinatorics, in mathematical physics and in representation theory. The polynomials have been instrumental in the solution of various conjectures on the combinatorial properties of reflection groups [32, 4]. Macdonald's "constant term conjectures" [28] are the most prominent among these. Figure 2 (also due to Macdonald) gives an overview of their representation theoretic significance. In addition, one knows in the case of the reflection group $W = S_n$, the symmetric group on n letters, that the polynomials are also the elementary spherical functions of the compact quantum group $U_q(n)$.

Much of the theory of Macdonald polynomials is related to the spectral analysis of the double affine Hecke algebra that generalizes the spherical harmonic analysis related to the compact real form of the symmetric space X.

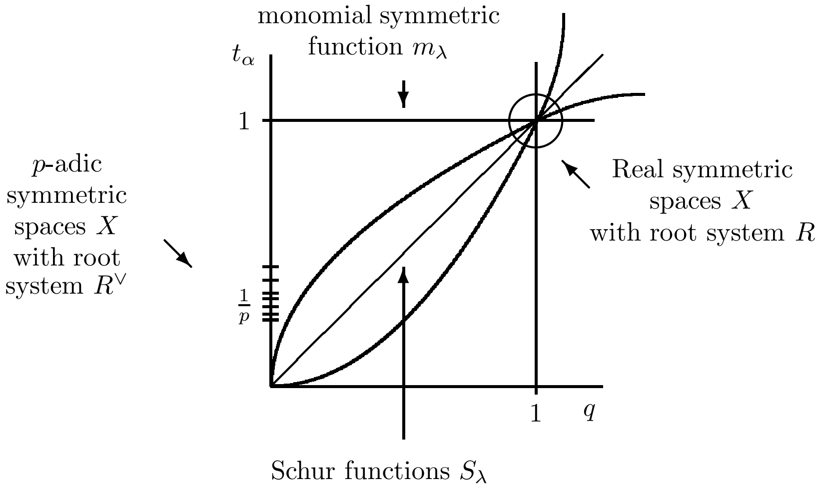

FIGURE 2. Macdonald polynomials and symmetric spaces

It is a major open problem to find the spectral theory of the operator algebra $\mathcal{L}$ introduced in (38) that generalizes the spherical Harish-Chandra transform on real non-compact symmetric spaces and Macdonald's p-adic spherical transform. It has been achieved in the differential limit for q tends to 1 [33, 5], and for the rank 1 double affine Hecke algebra in [25]. In general one does not know how to construct the non-polynomial eigenfunctions at present.

4. Integrable Models and Hypergeometric Functions

We saw that the differential equations for the hypergeometric function associated to a root system is a completely integrable system. This system is also known in mathematical physics as the trigonometric *Calogero-Moser* system. When $W = S_n$, the symmetric group on n letters, this system describes the dynamics of a quantum mechanical system of n particles moving on the real line under the influence of a pair potential that is proportional to the inverse square of the hyperbolic sine of the distance of the particles. The generalization to the algebra of difference equations $\mathcal{L}$ of (38) has the interpretation of making the quantum system relativistic. For $W = S_n$ such a relativistic model was found explicitly by Ruijsenaars [36], and this was extended to general classical root systems by Van Diejen [10]. The spectral problem for Cherednik's operator realization of affine Hecke algebras, discussed in the previous subsection, is of course very important for the dynamics of these related integrable models from mathematical physics.

4.1. The Knizhnik-Zamolodchikov equations

There is another interpretation of hypergeometric functions, in mathematical physics, via the so called Knizhnik-Zamolodchikov equations of conformal field theory [38, 13]. These equations are the differential equations for the n-point correlation functions $\psi(z_1, \ldots, z_n)$ of conformal field theory for a Kac-Moody algebra $\hat{\mathfrak{g}}$. The points $z_1, \ldots, z_n$ are distinct points on the complex line $\mathbf{C}$, and the correlation function takes values in an n-fold tensor product $V_1 \otimes \cdots \otimes V_n$ of representations of the finite dimensional simple Lie algebra $\mathfrak{g}$. The equations have the form

$$(k + h^\vee)\frac{d\psi}{dz_i} = \left(\sum_{\substack{j=1\\ j\neq i}}^{n} \frac{\Omega_{i,j}}{z_i - z_j}\right)\psi \; (\forall i = 1, \ldots, n)\,, \tag{39}$$

where Ω is the symmetric "Casimir tensor" $\Omega = \sum x_i \otimes x^i \in \mathfrak{g} \otimes \mathfrak{g}$ corresponding to the invariant scalar product on $\mathfrak{g}$, and $\Omega_{i,j}$ denotes the action of Ω on the i-th and j-th slot of the n-fold tensor ψ. The number $h^\vee$ is the dual Coxeter number of $\hat{\mathfrak{g}}$, and the complex number k is called the central charge.

These KZ-equations have been studied intensively by mathematicians because of their interesting relation with quantum groups. The system of differential equations is integrable and is invariant for the $\mathfrak{g}$ action on ψ. Hence it defines a monodromy representation of the fundamental group B_n of the regular orbit space of S_n acting on $\mathbf{C}^n$. This monodromy representation has values in the space of $\mathfrak{g}$-intertwiners between tensor products of $\mathfrak{g}$-modules. This defines the structure of a braided tensor category on the tensor category of $\mathfrak{g}$-modules. For generic values of k, this is the representation category of the quantum universal enveloping algebra $U_q(\mathfrak{g})$, where $q = \exp(\frac{\pi i}{k+h^\vee})$ (Drinfeld-Kohno theorem) [11, 26, 22]. When $\mathfrak{g} = \mathfrak{gl}_n$ this is the quantum group version of the classical Schur-Weyl duality.

It is fully justified to view the KZ-equations as generalized hypergeometric equations. In simple examples the equations reduce to (3) and the solutions of (39) allow integral representations that are generalizations of Euler's integral representation 1.4. This gives a beautiful geometric interpretation of the monodromy representation of the KZ-equation, analogous to the isomorphism (13). I refer the reader to the books [38] and [13] for a detailed exposition of this point of view.

The Knizhnik-Zamolodchikov equations have a number of variations and generalizations (allowing more complicated coefficient functions and discretization for example). For these complicated matters the reader is referred to [13].

In the case of trigonometric coefficient functions, we have the following relation to the Macdonald-Cherendnik theory. Take $\mathfrak{g} = \mathfrak{gl}_n$ and let ψ take it values in the zero-weight space of $V^{\otimes n}$, where V is the defining representation of $\mathfrak{g}$. The resulting system of equations can be shown to be equivalent to 3.3 when $W = S_n$. Cherednik has introduced the root system analogs of these (very special) Knizhnik-Zamolodchikov equations [3]. He found an explicit system of first order differential equations, and these equations were shown to be equivalent to (32) in [30].

The representation theoretic meaning via the quantum Schur-Weyl duality and the geometric interpretation of the monodromy representation give the KZ-equation a very rich structure in the case of $W = S_n$. Much of this is missing in the theory for general root systems. It is an open problem to find an interpretation of the monodromy representation as period integrals in the general root system case. However, some progress has been made by Couwenberg and Heckman [7] on the study of a Schwarz map for hypergeometric functions for root systems.

References

[1] K. Aomoto, *On the structure of integrals of power products of linear functions*, Sci. Papers, Coll. Gen. Education, Univ. Tokyo **27** (1977), 49–61.

[2] V. V. Batyrev, *Variations of the mixed Hodge structure of affine hypersurfaces in algebraic tori*, Duke Math. J. **69(2)** (1993), 349–409.

[3] I. Cherednik, *A unification of Knizhnik-Zamolodchikov equations and Dunkl operators via affine Hecke algebras*, Inv. Math. **106** (1991), 411–432.

[4] ———, *Double affine Hecke algebras and Macdonald conjectures*, Ann. Math. **141** (1995), 191–216.

[5] ———, *Inverse Harish-Chandra transform and difference operators*, Internat. Math. Res. Notices **15** (1997), 733–750.

[6] ———, *Lectures on affine Knizhnik-Zamolodchikov equations, quantum many-body problems, Hecke algebras, and Macdonald theory*, RIMS-1144 (1997).

[7] W. J. Couwenberg, *Complex reflection groups and hypergeometric functions*, Thesis University of Nijmegen (1994).

[8] P. Deligne and G. D. Mostow, *Monodromy of hypergeometric functions and non-lattice integral monodromy*, I.H.E.S. Publ. Math. **63** (1986), 5–89.

[9] ———, *Commensurabilities among lattices in $PU(1, n)$*, Annals of Math studies **132**, Princeton UP (1993).

[10] J. F. van Diejen, *Self-dual Koornwinder-Macdonald polynomials*, Inv. Math. **126** (1996), 319–341.

[11] V. G. Drinfeld, *Quasi-Hopf algebras*, Algebra i Analiz **1(6)** (1989), 114–148, Engl. transl. in Leningrad Math. J. **1** (1990), 1419–1457.

[12] A. Erdéli (eds.), *Higher transcendental functions*, vol. **1**, McGraw-Hill (1953).

[13] P. I. Etingof, I. B. Frenkel and A. A. Kirillov, Jr., *Lectures on Representation Theory and Knizhnik-Zamolodchikov Equations*, Mathematical Surveys and Monographs **58**, A.M.S. (1998).

[14] I. M. Gelfand and S. I. Gelfand, *Generalized hypergeometric functions*, Dokl. Akad. Nauk. SSSR **228 (2)** (1986), 279–283.

[15] I. M. Gelfand, M. M. Kapranov and A. V. Zelevinskii, *Hypergeometric functions and toral manifolds*, Funct. Anal. and its Appl. **23** (1989), 94–106.

[16] ———, *Generalized Euler integrals and A-hypergeometric functions*, Adv. in Math. **84** (1990), 255–271.

[17] ———, *A correction to the paper "Hypergeometric functions and toral manifolds"*, Funct. Anal. and its Appl. **27** (1993), 295–295.

[18] G. J. Heckman, *Dunkl operators*, Séminaire BOURBAKI 49ème année, 1996–97, n° 828 (1997).

[19] G. J. Heckman and E. M. Opdam, *Root systems and hypergeometric functions I*, Comp. Math. **64** (1987), 329–352.

[20] G. J. Heckman, and H. Schlichtkrull, *Harmonic Analysis and Special Functions on Symmetric Spaces*, Academic Press (1994).

[21] S. Helgason, *Groups and Geometric Analysis*, Perspectives in Mathematics **16**, Academic Press, New York (1984).

[22] D. Kazhdan and G. Lusztig, *Tensor structures arising from affine Lie algebras I-IV*, J. A.M.S. **6** (1993), 905–947, 949–1011; **7** (1994), 335–381, 383–453.

[23] A. A. Kirillov, Jr., *Lectures on affine Hecke algebras and Macdonald's conjectures*, Bull. AMS **34**(3) (1997), 251–292.

[24] F. Klein, *Vorlesungen ueber die hypergeometrische Funktion*, Grundlehren der mathematische Wissenschaften, Springer Verlag (1933).

[25] E. Koelink and J. V. Stokman, *The Askey-Wilson transform*, preprint xxx.lanl.gov/ abs/ math.CA/0004053 (2000).

[26] T. Kohno, *Monodromy representations of braid groups and Yang-Baxter equations*, Ann. Inst. Fourier **37(4)** (1987), 139–160.

[27] T. H. Koornwinder, *Askey-Wilson polynomials for root systems of type BC*, Contemp. Math. **138** (1992), 189–204.

[28] I. G. Macdonald, *Some conjectures for root systems*, SIAM J. of Math. An. **13** (1982) 988–1007.

[29] ———, *Affine Hecke algebras and orthogonal polynomials*, Séminaire Bourbaki, Vol. 1994–1995 n° 797.

[30] A. Matsuo, *Integrable connections related to zonal spherical functions*, Invent. Math. **110** (1992), 95–121.

[31] M. Noumi and J. V. Stokman, *Askey-Wilson polynomials: an affine Hecke-algebraic approach*, preprint (2000).

[32] E. M. Opdam, *Some applications of hypergeometric shift operators*, Inv. Math. **98** (1989), 1–18.

[33] ———, *Harmonic analysis for certain representations of graded Hecke algebras*, Acta. Math. **175** (1995), 75–121.

[34] ———, *Lectures on Dunkl operators for real and complex reflection groups*, To appear.

[35] E. Picard, *Sur une extension aux fonctions de deux variables du problème relatif aux fonctions hypergéométriques*, Ann. E.N.S. **10** (1881), 305–321.

[36] S. N. M. Ruijsenaars, *Complete integrability of relativistic Calogero-Moser systems and elliptic function identities*, Comm. Math. Phys. **110** (1987), 191–213.

[37] T. Terada, *Problème de Riemann et fonctions automorphes provenant des fonctions hypergéométriques de plusieurs variables*, J. of Math. Kyoto Univ. **13-3** (1973), 557–578.

[38] A. Varchenko, *Multidimensional hypergeometric functions and representation theory of Lie algebras and quantum groups*, Advanced Series in Mathematical Physics **21**, World Scientific (1995).

[39] M. Yoshida, *Hypergeometric function, my love*, Aspects of Mathematics **E32**, Vieweg, (1997).

KdV Institute for Mathematics,
University of Amsterdam,
Plantage Muidergracht 24,
1018 TV Amsterdam, The Netherlands
E-mail address: `opdam@wins.uva.nl`