Workshop On
PROBABILITY THEORY and its APPLICATIONS

Peter Eichelsbacher, Matthias Löwe

Contents

Preface

This preprint is a collection of abstracts of talks held on the workshop on " PROBABILITY THEORY and its APPLICATIONS" held on 11th of October 1996 in Bielefeld. This workshop was supported by the University of Bielefeld and the SFB 343 "Diskrete Strukturen in der Mathematik".

We thank all partcipants of the workshop for their coming.

Glauber spin-glass dynamics

M. Grunwald*

1 Introduction, Definition of the model

Spin glass models (i.e. a type of disordered lattice models) where introduced in [AnEd75] in the theoretical physicists literature to explain the behaviour of certin diluted magnetic alloys. A from a physical point of view well understood mean-field caricature of this Edwardson Anderson spin glass is the Sherrington-Kirkpatrick (SK) model. This model is "solvable" with different – mostly non mathematically rigorous (but neverseless highly sophisticated) – techniques:

- Most prominent with the mysterious replica-trick,
- by means off the "cavity-method" or
- by using a *dynamical approach* ([SoZi83],[Som87]).

Also some progress was made in the mathematical understanding of the SK-model [BoPi97], the model is still fare away from beeing rigorously solved.

We want to explain mathematically the very first steps off the dynamical approach. Using Large Deviations techniques we prove convergence results for the Glauber spin glass dynamics in the thermodynamic limit, i.e. in the limit "number of variables ($\sim$ spins)" $\to \infty$. We need a lot of notations to describe the LD results (see [Gru96] for the LDP's and M. Grunwald in [BoPi97] for the charactarisation of the limiting dynamic in the general case). We therefore only state the physical interesting convergence results in this proceedings.

With the same physical motivation in mind A. Guionnet and G. Ben Arous studied a Langevin dynamic for (a soft spin version of the) SK-model [BeGu95]. For an overview of this results (with a good understanding of the limiting dynamics) have a look in B. Ben Arous and A. Guionnet in [BoPi97].

Also the dynamical approach was invented at the beginning as a way to get results for the statical behaviour of the SK-model the dynamics has an interessed of its own (see e.g. M. Mezard in [BoPi97]).

Let's be more explicit. The SK-model is defined by a random Hamiltonian

$$\mathrm{U}_N^J(\sigma) := -\frac{1}{2\sqrt{N}} \sum_{i \neq j=1}^{N} J^{i,j} \cdot \sigma_i \sigma_j, \quad \sigma \in \mathcal{E}^N$$

on the state space $\mathcal{E}^N := \{-1,1\}^N$, where for $i > j$ the $J^{i,j}$ are chosen independently according to a centered Gaussian distribution with variance one, and $J^{j,i} := J^{i,j}$ – the coupling-matrix

*Fachbereich Mathematik, TU-Berlin, D-10623 Berlin, e-mail: grunwald@math.tu-berlin.de

J is supposed to be symmetric. Another way to describe the coupling-matrix is to say that the pairs $(J^{i,j}, J^{j,i})$ are i.i.d. centered Gaussian for $i > j$ with covariance

$$\mathbf{E}_J(J^{i,j})^2 = \mathbf{E}_J(J^{j,i})^2 = 1, \quad \mathbf{E}_J(J^{i,j} \cdot J^{j,i}) = \alpha \tag{1}$$

with $\alpha := 1$. For fixed coupling J the Gibbs measures at temperature β are defined by

$$\mu^J_{\beta,N}(\{\sigma\}) := \frac{1}{Z^J_{\beta,N}} \exp[-\beta \cdot \mathrm{U}^J_N(\sigma)], \quad \sigma \in \mathcal{E}^N$$

with the partition-function $Z^J_{\beta,N} := \sum_{\sigma\in\mathcal{E}^N} \exp[-\beta \cdot \mathrm{U}^J_N(\sigma)]$ as the normalization. In the dynamical approach a Glauber dynamics for these measures is introduced, i.e. a Markov process with state space $\mathcal{E}^N$ for which the $\mu^J_{\beta,N}$ are reversible invariant measures. Define the field at site i to be

$$\mathrm{h}^J_i(\cdot) : \mathcal{E}^N \to \mathbb{R}, \quad \sigma \mapsto \mathrm{h}^J_i(\sigma) := \frac{1}{\sqrt{N}} \sum_{j=1, j\neq i}^{N} J^{i,j} \cdot \sigma_j.$$

Then the simplest possible choice for the process is a single spin-flip jump process with state space $\mathcal{E}^N$, where the i-th spin changes its state $\sigma_i \to -\sigma_i$ with the transition rate

$$c_i(\sigma) := \frac{1}{1 + \exp[2\beta\sigma_i \mathrm{h}^J_i(\sigma)]}.$$

We denote by $\sigma^i := (\sigma_1, \cdots, \sigma_{i-1}, -\sigma_i, \sigma_{i+1}, \cdots)$ the configuration with the i-th coordinate being flipped. Then the transition semi-group $\mathrm{P}_{\sigma',\sigma}(t)$ belonging to the process defined by the rates c_i satisfies the forward equation

$$\frac{d}{dt}\mathrm{P}_{\sigma',\sigma}(t) = \sum_{i=1}^{N} \Big(c_i(\sigma^i)\mathrm{P}_{\sigma',\sigma^i}(t) - c_i(\sigma)\mathrm{P}_{\sigma',\sigma}(t) \Big)$$

with the initial condition $\mathrm{P}_{\sigma',\sigma}(0) = \delta_{\sigma',\sigma}$. This process can be constructed as a measure on $(\mathcal{D}_{\mathcal{E}}[0,T])^N$, where $\mathcal{D}_{\mathcal{E}}[0,T] =: X$ is the Skorohod space of càdlàg-functions $u : [0,T] \to \mathcal{E}$. For each $x \in \mathcal{D}_{\mathcal{E}}[0,T]^N$

$$t \mapsto \mathrm{h}^J_i(x(t)), \quad x(t) = (x_1(t), \cdots, x_N(t)) \in \mathcal{E}^N,$$

defines a function in the Skorohod space $\mathcal{D}_{\mathbb{R}}[0,T] =: Y$ of real-valued functions on $[0,T]$. We are considering the law of the combined Markov process

$$((\mathrm{h}_1(x), x_1), \ldots, (\mathrm{h}_N(x), x_N)) \in (Y \times X)^N =: Z^N$$

as a measure on the product of the space of paths with the space of the fields, i.e. as a measure on $(Y \times X)^N$. Let $q \in \mathcal{M}(\mathcal{D}_{\mathcal{E}}[0,T])$ be the Markov process starting with some distribution on $\mathcal{E}$ and jumping with rate one between both states – q will play a role like the Wiener measure for diffusions, i.e. q will be used to formulate the process of interacting spins as absolutely continuous transformation of a "more-independent" measure. We can write the interacting process on the time interval $[0,T]$ via Girsanov formula [Bré81, Theorem VI.3] as

$$p^{N,\mathrm{G},J}(B) := \int_{X^N} \mathbb{1}_B(\mathrm{h}^J(x), x) \cdot \exp\left[\sum_{i=1}^{N} \mathrm{G}(\mathrm{h}^J_i(x), x_i)\right] q^{\otimes N}(dx) \tag{2}$$

for some measurable set $B \subset (Y \times X)^N =: Z^N$, where the Girsanov exponent has the simple form

$$\begin{aligned} \mathrm{G}(y,x) \quad := \quad & \sum_{s:x(s)\neq x(s-)} \ln \frac{1}{1+\exp[2\beta y(s-)x(s-)]} \\ & + \int_0^T \frac{\exp[2\beta y(s)x(s)]}{1+\exp[2\beta y(s)x(s)]}\, ds \end{aligned} \tag{3}$$

for $(y,x) \in Y \times X$. The coupling-matrix J appears in the Eq. (2) only through the field h^J. The idea of Sompolinsky and Zippelius is to average these measures with respect to the distribution of the J's. The physical reasoning behind this is that the "individual dynamics" still evolves with fixed coupling and therefore some information of the distribution of the quenched system can be gained in the limit as $T \to \infty$.

At this point different dynamical models can be defined by choosing different distributions for the coupling-matrix J. The physical SK-model corresponds to the symmetric choice $\alpha = 1$ for the couplings J as in Eq. (1). But every choice $\alpha \in [-1,1]$ is allowed which means that the coupling-matrix J is becoming less and less symmetric up to a complete anti-symmetric matrix for $\alpha = -1$. The most important other choice is the asymmetric case $\alpha = 0$, what implies that the Gaussian variables $J^{i,j}$ and $J^{j,i}$ are independent. The influence of the coupling has been studied in [RSZ89]. In the asymmetric model the proofs are much simpler. It is the dependence in the coupling-matrix J for $\alpha \neq 0$ which makes life harder in the symmetric case.

Let's now continue to define the annealed spin-glass dynamics. Following the ideas of [SoZi83] we average the measures $p^{N,\mathrm{G},J}$ with respect to the distribution of the J defined in Eq. (1) for some fixed $\alpha \in [-1,1]$. Interchanging the order of integration leads to

$$p_\alpha^{N,\mathrm{G}}(B) := \int_{X^N} \int_{Y^N} \mathbf{I}_B(y;x) \cdot \exp\left[\sum_{i=1}^N \mathrm{G}(y_i,x_i)\right] \nu_{\alpha,x}^N(dy)\, q^{\otimes N}(dx), \tag{4}$$

where $\nu_{\alpha,x}^N$ stands for the conditional distribution of the fields $\mathrm{h}^J(x) \in Y^N$ under J for given $x \in X^N$. Since the $J^{i,j}$ are centered Gaussian, $\nu_{\alpha,x}^N$ is a centered Gaussian measure on Y^N with covariance

$$\int_{Y^N} y_i(s)y_i(t)\, \nu_{\alpha,x}^N(dy) = \frac{1}{N} \sum_{j=1, j\neq i}^N x_j(s)x_j(t) \quad \text{for} \quad i \in \{1,\cdots,N\}, \tag{5}$$

for the diagonal elements, and

$$\int_{Y^N} y_i(s)y_j(t)\, \nu_{\alpha,x}^N(dy) = \frac{\alpha}{N} x_j(s)x_i(t) \quad \text{for} \quad i,j \in \{1,\cdots,N\}, i\neq j$$

for the off-diagonal elements. Recall that $Z = Y \times X$ and denote by

$$\Theta_N^Z : Z^N \longrightarrow \mathcal{M}_1(Z), \quad z \longmapsto \frac{1}{N}\sum_{i=1}^N \delta_{z_i}$$

the empirical measure. The measures we study are the distributions of the empirical measures under $p_\alpha^{N,\mathrm{G}}$, i.e. the measures

$$P_\alpha^{N,\mathrm{G}} := \Theta_N^Z(p_\alpha^{N,\mathrm{G}}), \tag{6}$$

on $\mathcal{M}_1(Z)$, where $\mathrm{f}(\nu) = \nu \circ \mathrm{f}^{-1}$ denotes the image of a measure ν under a map f. In the thermodynamic limit a full large deviations principle **(LDP)** holds in the asymmetric case $\alpha = 0$ and the upper large deviations bound holds in the symmetric model and every model $\alpha \neq 0$ (that is enough to proof convergence results).

1.1 Statement of the convergence results

In this section we want to state the limiting results which can be derived from the LDP's for the measures $P_\alpha^{N,\mathrm{G}}$.

1.1.1 The $\alpha = 0$ case

In the asymmetric case, we prove that the rate function $\mathrm{S}_0^{\mathrm{G}}$ belonging to the LDP of the measures $P_0^{N,\mathrm{G}}$ has a unique minimum $\nu_0 \in \mathcal{M}_1(Y \times X)$ given as the (unique) solution of the self-consistent equation

$$\nu_0 = \exp[\mathrm{G}]\mathcal{N}_{\pi_X(\nu_0)} \otimes q,$$

where $\mathcal{N}_{\pi_X(\nu_0)}$ is a centered Gaussian measure on $\mathcal{D}_{\mathbb{R}}[0,T]$ with the covariance constructed from the X-marginal $\pi_X(\nu_0)$ of the measure ν_0. For a fixed function $u \in Y$ the absolutely continuous transformation

$$\frac{dq^u}{dq}(x) := \exp[\mathrm{G}(u,x)]$$

of the measure q with the help of the Girsanov exponent G defines a jump process which starts at time $t = 0$ with the same distribution on $\{-1, 1\}$ as q but which has a changed jump-rate. Depending on the "field" $u(t-)$ at (or better a little bit before) time t the rate for a jump $x(t-) \to -x(t)$ under q^u is $\frac{1}{1+\exp[2\beta u(t-)x(t-)]}$. Therefore

$$\nu_0 = \int_Y \int_X \delta_{(y,x)}\, q^y(dx)\, \mathcal{N}_{\pi_X(\nu_0)}(dy)$$

describes a single spin jumping influenced by a (independent) centered Gaussian field y. The covariance of these Gaussian field has to be found sefconsistent.

Via an Borel-Cantelli argument (see [BeGu95], Theorem 2.7) the large deviations result together with the uniqueness of the minimum ν_0 implies the following quenched weak convergence result for the asymmetric spin-glass dynamics.

Corollary 1.1 *Let the $J^{i,j}$ be chosen independently. Then for almost all J*

$$P^{N,\mathrm{G},J} \overset{\mathrm{w}}{\Longrightarrow} \delta_{\nu_0} \text{ in } \mathcal{M}_1(Z),$$

where δ_{ν_0} is the Dirac-measure at ν_0.

Also propagation of chaos results (as in [BeGu95]) can be derived.

1.1.2 The general case ($\alpha \in [-1, 1]$)

The interesting feature in the symmetric model (or in any model with $\alpha \neq 0$) is the appearance of a new order parameter, which is introduced to handle the (weak) dependence in the Gaussian fields with covariance Eq. (5) resulting from the dependence in the coupling matrix J. This c order parameter has a close connection to the physical response function of the dynamics.

As in the asymmetric case the rate function $\mathrm{S}_\alpha^{\mathrm{G}}$ for the LDP of the $P_\alpha^{N,\mathrm{G}}$, $\alpha \in [-1,1]$, has a unique minimum ν_α. In these general case the description of the minima of the rate $\mathrm{S}_\alpha^{\mathrm{G}}$ is given in terms of a self-consistent equation for the covariance of the X-marginal of the measure and – in addition to the parameters needed in the asymmetric model $\alpha = 0$ – the *response function.* These self-consistent equation is

$$\nu_\alpha = \int_Z \delta_{(y+\alpha \mathrm{R}_{\nu_\alpha}\cdot x,x)} \exp[\mathrm{G}(y+\alpha \mathrm{R}_{\nu_\alpha}\cdot x, x)]\, \mathcal{N}_{\pi_X(\nu_\alpha)}(dy) q(dx),$$

where $\mathcal{N}_{\pi_X(\nu_\alpha)}$ is defined as in the asymmetric case and the response function R_{ν_α} applied to some function $u \in Y$ at time t fulfills

$$(\mathrm{R}_{\nu_\alpha}\cdot u)_t = \frac{d}{d\lambda}\int_Z x(t) \exp[\mathrm{G}(y+\alpha \mathrm{R}_{\nu_\alpha}\cdot x + \lambda u, x)]\, \mathcal{N}_{\pi_X(\nu_\alpha)}(dy) q(dx)\Big|_{\lambda=0}.$$

This is not enough to definition R_{ν_α} but shows the physical meaning of R_{ν_α}; $(\mathrm{R}_{\nu_\alpha}\cdot u)_t$ describes how the expectation $\mathbf{E}_{\nu_\alpha} x(t)$ "responds" to a change in an external field u. The process described by ν_α is no longer a particle moving in an independent Gaussian field – the distribution of the field is depending on the path x through the shift given by $\mathrm{R}_{\nu_\alpha}\cdot x$. An additinal property is that R_{ν_α} is *causal*, i.e.

$$u(s) = 0 \text{ for } s \le t \Longrightarrow (\mathrm{R}_{\nu_\alpha}\cdot u)_t = 0 \text{ for } s \le t.$$

The shift in the field is only depending on the past of the $x \in X$. We generalize Corollary 1.1:

Corollary 1.2 *Let the $(J^{i,j}, J^{j,i})$, $i > j$ be chosen independently as defined in Eq. (1) with some $\alpha \in [-1,1]$. Then for almost all J*

$$P^{N,\mathrm{G},J} \overset{\mathrm{w}}{\Longrightarrow} \delta_{\nu_\alpha} \text{ in } \mathcal{M}_1(Z),$$

where δ_{ν_α} is the Dirac-measure at ν_α.

References

[AnEd75] P. .W. Anderson, S. F. [1975]: *Theory of Spin Glasses,* J. Phys. F5, 965-974.

[BeGu95] G. Ben Arous, A. Guionnet [1995]: *Large deviations for Langevin spin glass dynamics,* Probab. Theory Relat. Fields **102**, 455-509.

[BeGu94a] G. Ben Arous, A. Guionnet [1994]: *Annealed and Quenched Propagation of Chaos for Asymmetric Spin Glass Dynamics,* preprint.

[BeGu94b] G. Ben Arous, A. Guionnet [1994]: *Symmetric Langevin spin glass dynamics,* preprint.

[Bré81] P. Brémaud [1981]: *Point Processes and Queues,* Springer series in statistics.

[BoPi97] A. Bovier, P Picco, eds.: [1997] *Mathematics of spin glasses and neural networks,* "Progress in Probability", Birkhäuser.

[Gru96] M. Grunwald [1996]: *Sanov results for Glauber spin-glass dynamics,* Probab. Theory Relat. Fields **106**.

[RSZ89] H. RIEGER, M. SCHRECKENBERG, J. ZITTARTZ [1989]: *Glauber dynamics of the asymmetric SK-model,* Z. Phys. **B, 74**, 527-538.

[ShKi75] D. SHERRINGTON, S. KIRKPATRICK [1975]: *Solvable model of a Spin Glass,* Phys. Rev. Lett **35**, 1972-1796.

[SoZi83] H. SOMPOLINSKY, A. ZIPPELIUS [1983]: *Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses*, Phys. Rev. **25**, 6860.

[Som87] H. J. SOMMERS [1987]: *Path-integral approach to Ising-spin-glass dynamics,* Phys. Rev. Let. **58**, 1268.

A CENTRAL LIMIT THEOREM FOR A ONE-DIMENSIONAL WEAK POLYMER MEASURE

Wolfgang König
Katholieke Universiteit Nijmegen

1 The Model

Let $(S_i)_{i\in\mathbb{N}_0}$ be simple random walk on $\mathbb{Z}$, starting at the origin. Let P be its law and let E be expectation w.r.t. P. For $n \in \mathbb{N}$, and a parameter $\beta \in (0,\infty)$ define a new path law Q_n^β by setting

$$dQ_n^\beta = \exp\left(-\beta \sum_{x\in\mathbb{Z}} \ell_n(x)^2\right) dP/Z_n^\beta \tag{1.1}$$

with Z_n^β the appropriate normalizing constant and

$$\ell_n(x) := \#\{\, k \le n \,|\, S_k = x \,\} \tag{1.2}$$

the n-step local times. The quantity

$$\sum_{x\in\mathbb{Z}} \ell_n(x)^2 = \sum_{i,j=0}^{n} 1_{\{S_i=S_j\}} \tag{1.3}$$

is called the *self-intersection local time*. The measure Q_n^β suppresses paths having many self-intersections until time n and is called the *n-polymer measure with strength of repellence β*, sometimes also called the *Domb-Joyce model*. The family $(Q_n^\beta)_n$ is not consistent. It tends to spread out the path and is therefore called *self-repellent*.

The main question is: How does S_n typically behave under Q_n^β as $n \to \infty$? Traditionally, one is interested in the boundary case $\beta = \infty$ of self-avoiding paths, but our particular question in this talk will be: How does S_n behave under $Q_n^{\beta_n}$ if $\beta_n \downarrow 0$ and $\beta_n n^{\frac{3}{2}} \to \infty$?

2 Heuristic Calculation

We give non-rigorous scaling arguments which give an impression of the scale on which S_n runs under $Q_n^{\beta_n}$ if $\beta_n \downarrow 0$. First we turn to the boundary case $\beta_n = n^{-\frac{3}{2}}$. Here we have, in the spirit of the usual invariance principle,

$$n^{-\frac{3}{2}} \sum_{x\in\mathbb{Z}} \ell_n(x)^2 = \int_{\mathbb{R}} L_n(1,y)^2\, dy \longrightarrow \int_{\mathbb{R}} L(1,y)^2\, dy \tag{2.1}$$

where $L_n(1,y) = n^{-\frac{1}{2}}\ell_n(\lceil yn^{-\frac{1}{2}}\rceil)$ and $(L(T,y))_{T\ge 0, y\in\mathbb{R}}$ denotes the Brownian local time process characterized by the so-called *occupation times formula*

$$\int_A L(T,y)\, dy = \int_0^T 1_{\{B_s\in A\}}\, ds \qquad (A \subset \mathbb{R} \text{ mb.}). \tag{2.2}$$

It can be deduced from (2.1) that $(n^{-\frac{1}{2}}S_{\lceil tn \rceil})_{t\in[0,1]}$ tends, under $Q_n^{n^{-\frac{3}{2}}}$, towards the Edwards model on $[0,1]$ (see Brydges and Slade (1995)). The *Edwards model* on $[0,T]$ is defined as the transformed Brownian path measure

$$d\widetilde{Q}_T^\beta = \exp\left(-\beta \int_{\mathbb{R}} L(T,y)^2\, dy\right) d\widetilde{P}/\widetilde{Z}_T^\beta, \tag{2.3}$$

where $\widetilde{P}$ is the Wiener measure and $\widetilde{Z}_T^\beta$ is the appropriate normalizing constant.

In the case $\beta_n \downarrow 0$ and $\beta_n n^{\frac{3}{2}} \to \infty$, we make the following ansatz. For $\alpha \in [0,\frac{1}{2}]$, Stirling's formula yields

$$P\left(S_n = \mathcal{O}(n^{1-\alpha})\right) = e^{\mathcal{O}(n^{1-2\alpha})}. \tag{2.4}$$

If $S_n = \mathcal{O}(n^{1-\alpha}) = \mathcal{O}(\#\{0,\dots,S_n\})$, then $\ell_n(x) = \mathcal{O}(n^\alpha)$ since the path puts $n+1$ hits onto $\mathcal{O}(n^{1-\alpha})$ sites. Thus

$$\beta_n \sum_{x\in\mathbb{Z}} \ell_n(x)^2 = \beta_n \mathcal{O}(n^{1-\alpha})\mathcal{O}(n^\alpha)^2 = \beta_n \mathcal{O}(n^{1+\alpha}). \tag{2.5}$$

Now choose α such that probability and interaction are on the same scale, i.e., $n^{-3\alpha} = \mathcal{O}(\beta_n)$, and we get that $S_n = \mathcal{O}(n\beta_n^{\frac{1}{3}})$ under $Q_n^{\beta_n}$. We shall see indeed later that this is true and we shall make the assertion precise.

3 Central limit theorem for fixed β

We give a central limit theorem for S_n under Q_n^β with fixed $\beta > 0$. First we give a heuristic argument for the asymptotic normality.

It turned out (see K. (1993)) that the rate function

$$\Phi_\beta(\theta) = \lim_{n\to\infty} \frac{1}{n} \log Q_n^\beta(S_n \approx \theta n) \qquad (\theta \in (0,1)) \tag{3.1}$$

is real-analytic and possesses a strict maximum in a certain point $\theta^*(\beta) \in (0,1)$. This fact immediately implies the law of large numbers

$$\lim_{n\to\infty} Q_n^\beta\left(\left|\frac{|S_n|}{n} - \theta^*(\beta)\right| \geq \varepsilon\right) = 0 \qquad (\varepsilon > 0). \tag{3.2}$$

(Because of the symmetry, this means that the distribution of $\frac{S_n}{n}$ under Q_n^β tends weakly towards $\frac{1}{2}(\delta_{-\theta^*(\beta)} + \delta_{\theta^*(\beta)})$.)

Also it turned out that Φ_β has a strictly negative curvature in its maximum point, i.e., the number

$$\sigma^*(\beta) := -\frac{1}{\Phi_\beta''(\theta^*(\beta))} \tag{3.3}$$

is positive. The following heuristic calculation using a Taylor expansion for Φ_β around $\theta^*(\beta)$ indicates the asymptotic normality of $\frac{|S_n|-\theta^*(\beta)n}{\sqrt{n}}$ with variance

$\sigma^*(\beta)^2$:

$$Q_n^\beta\left(\frac{|S_n|-\theta^*(\beta)n}{\sqrt{n}}\le C\right)\simeq\int_{-\infty}^C Q_n^\beta\left(\frac{S_n}{n}\approx\theta^*(\beta)+\frac{x}{\sqrt{n}}\right)dx$$

$$\simeq\int_{-\infty}^C e^{n\Phi_\beta\left(\theta^*(\beta)+\frac{x}{\sqrt{n}}\right)}\,dx\simeq\int_{-\infty}^C e^{\frac{x^2}{2}\Phi_\beta''(\theta^*(\beta))}\,dx. \tag{3.4}$$

(Here $\simeq$ denotes equality up to a factor $e^{o(n)}$.) Indeed, we have:

Theorem (K. 1994): *For all $\beta\in(0,\infty)$ and all $C\in\mathbb{R}$,*

$$\lim_{n\to\infty}Q_n^\beta\left(\frac{|S_n|-\theta^*(\beta)n}{\sqrt{n}\sigma^*(\beta)}\le C\right)=\mathcal{N}((-\infty,C]). \tag{3.5}$$

An analytic characterization of $\theta^*(\beta)$ and $\sigma^*(\beta)$ is given as follows. Pick $r\in\mathbb{R}$ and define $\mathbb{N}\times\mathbb{N}$ matrices by

$$\begin{aligned} P(i,j) &= \binom{i+j-2}{i-1}\left(\tfrac{1}{2}\right)^{i+j-1}, \\ A_{r,\beta}(i,j) &= P(i,j)e^{-\beta(i+j-1)^2+r(i+j-1)}. \end{aligned} \tag{3.6}$$

Then P is a symmetric doubly-stochastic matrix; in fact it is the transition matrix of a critical branching process with geometric offspring distribution and one immigrant per time unit. The matrix $A_{r,\beta}$, regarded as an operator on the sequence space $l^2(\mathbb{N})$, is a Hilbert-Schmidt operator and possesses a largest l^2-eigenvalue

$$\lambda_\beta(r)=\sup_{\|\tau\|_{l^2}=1}\langle\tau,A_{r,\beta}\tau\rangle_{l^2}. \tag{3.7}$$

The function $r\mapsto\lambda_\beta(r)$ is real-analytic and strictly increasing. By λ_β^{-1} we denote its inverse function. Then we have

$$r^*(\beta)=\lambda_\beta^{-1}(1),\quad\theta^*(\beta)=(\lambda_\beta^{-1})'(1),\quad\sigma^{*2}(\beta)=(\lambda_\beta^{-1})''(1). \tag{3.8}$$

4 Asymptotics for $\beta\downarrow 0$

Theorem (vd Hofstad/den Hollander 1995): *There are constants a^*, b^*, $c^*\in(0,\infty)$ such that, as $\beta\downarrow 0$,*

$$r^*(\beta)\sim a^*\beta^{\frac{2}{3}},\qquad\theta^*(\beta)\sim b^*\beta^{\frac{1}{3}},\qquad\sigma^*(\beta)\to c^*. \tag{4.1}$$

The constants are characterized in the following way. For $a\in\mathbb{R}$, define a Sturm-Liouville operator $\mathcal{L}_a$ on $(L^2\cap C^2)([0,\infty))$ by putting

$$\mathcal{L}_a x(u)=ux''(u)+x'(u)+(2au-4u^2)x(u). \tag{4.2}$$

The number

$$\rho(a) = \sup_{\|x\|_{L^2}=1} \langle x, \mathcal{L}_a x \rangle_{L^2} \tag{4.3}$$

is the largest L^2-eigenvalue of $\mathcal{L}_a$. Also the map $a \mapsto \rho(a)$ is real-analytic and strictly increasing with inverse function ρ^{-1}. The constants above are characterized as

$$a^* = \rho^{-1}(0), \qquad b^* = (\rho^{-1})'(0), \qquad c^{*2} = (\rho^{-1})''(0). \tag{4.4}$$

The core of the proof is the following scaling limit assertion: For all $a \in \mathbb{R}$,

$$\begin{aligned} \beta^{-\frac{1}{3}}\left(\lambda(a\beta^{\frac{2}{3}},\beta) - 1\right) &= \sup_{\|\tau\|_{l^2}=1} \beta^{-\frac{1}{3}}\left(\langle \tau, A_{a\beta^{\frac{2}{3}},\beta}\tau\rangle_{l^2} - \|\tau\|_{l^2}^2\right) \\ &\xrightarrow{\beta\downarrow 0} \sup_{\|x\|_{L^2}=1} \langle x, \mathcal{L}_a x\rangle_{L^2} = \rho(a). \end{aligned} \tag{4.5}$$

In order to derive (4.5), introduce the scaling $x_\beta(u) = \beta^{-\frac{1}{6}}\tau(\lceil u\beta^{-\frac{1}{3}}\rceil)$ and use the notion of epi-convergence.

The numerical values of the constants in (4.4) are

$$a^* \approx 2.19, \qquad b^* \approx 1.11, \qquad c^* \approx 0.7. \tag{4.6}$$

As a corollary, the law of large numbers for the Domb-Joyce model is continuous at $\beta = 0$ (in the sense that $\lim_{\beta\downarrow 0}\theta^*(\beta) = 0 = \theta^*(0)$), but the central limit theorem is not, in the sense that $\lim_{\beta\downarrow 0}\sigma^*(\beta)^2 = c^{*2} < 1 = \sigma^*(0)$.

5 The weak central limit theorem

Couple the limits $\beta \downarrow 0$ and $n \to \infty$. Work in preparation:

Theorem (vd Hofstad/den Hollander/K. 1996/7): *If $\beta_n \downarrow 0$ and $\beta_n n^{\frac{3}{2}} \to \infty$, then for all $C \in \mathbb{R}$,*

$$\lim_{n\to\infty} Q_n^{\beta_n}\left(\frac{|S_n| - \theta^*(\beta_n)n}{\sqrt{n}\sigma^*(\beta_n)} \le C\right) = \mathcal{N}((-\infty, C]). \tag{5.1}$$

Corollary: *If $p \in (0, \frac{3}{2})$ and $\beta_n = \beta n^{-p}$ then, under $Q_n^{\beta_n}$,*

$$\begin{aligned} |S_n| &\sim \beta^{\frac{1}{3}} b^* n^{1-\frac{p}{3}}, \\ |S_n| - E_n^{\beta_n}(|S_n|) &\sim c^*\sqrt{n}\mathcal{N}, \\ \ell_n(x) &\asymp n^{\frac{p}{3}}, \\ \beta_n \textstyle\sum_x \ell_n(x)^2 &\asymp n^{1-\frac{2p}{3}}. \end{aligned} \tag{5.2}$$

The core of the weak central limit theorem stated above is that, for any $C \in \mathbb{R}$,

$$\lim_{n\to\infty} e^{r^*(\beta_n)n}\beta_n^{-\frac{1}{3}} E\left(e^{-\beta_n \sum_{x\in\mathbb{Z}} \ell_n^2(x)} 1_{\{|S_n|-\theta^*(\beta_n)n \le C\sigma^*(\beta_n)\sqrt{n}\}}\right) = K\mathcal{N}((-\infty, C]) \tag{5.3}$$

for some $K \in (0,\infty)$. For the continuous analogue, the Edwards model (see (2.3)) it was shown in van der Hofstad, den Hollander and K. (1995) that, for any $C \in \mathbb{R}$,

$$\lim_{T\to\infty} e^{a^*\beta^{\frac{2}{3}}T} \widetilde{E}\left(e^{-\beta\int_{\mathbb{R}} L(T,y)^2 dy} 1_{\{|B_T|-b^*\beta^{\frac{1}{3}}T \le Cc^*\sqrt{T}\}}\right) = K'\mathcal{N}((-\infty, C]). \tag{5.4}$$

where $K' \in (0,\infty)$ is structurally *different* from K.

6 One main tool

In order to attack the proof of the weak central limit theorem one is faced with the problem to control the joint distribution of $(\ell_n(x))_{x\in\mathbb{Z}}$ and S_n. A solution is given in terms of a certain Markov representation of the local times. This was developed in one of the two by now famous papers that introduced the Ray-Knight theorems for Brownian local times which is the continuous analogue.

Define the successive times of steps $d \to d+1$ by $T_{0,d} = 0$ and

$$T_{k,d} = \inf\{\, n > T_{k-1,d} \,|\, S_{n-1} = d, S_n = d+1 \,\} \tag{6.1}$$

and the number of steps $x \to x+1$ until time $t \in \mathbb{N}$ by

$$m(t,x) = \#\{\, n \in \{1,\dots,t\} \,|\, S_{n-1} = x, S_n = x+1 \,\}. \tag{6.2}$$

Then the local times, observed at the random time $T_{k,d}$, are described by

$$\ell_{T_{k,d}}(x) = \begin{cases} m(T_{k,d},x) + m(T_{k,d},x-1) - 1 & \text{if } x \in \{1,\dots,d\} \\ m(T_{k,d},x) + m(T_{k,d},x-1) & \text{otherwise.} \end{cases} \tag{6.3}$$

The distribution of $(m(T_{k,d},x)_{x\in\mathbb{Z}}$ is described by

Lemma (Knight 1963) *Fix $k, d \in \mathbb{N}$.*

1. $(m(T_{k,d}, d-x))_{x\in\{0,\dots,d\}}$ *is a homogeneous Markov chain on $\mathbb{N}$ with kernel* P defined in (3.6).

2. $(m(T_{k,d}, d+x))_{x\in\mathbb{N}_0}$ *is a homogeneous Markov chain on $\mathbb{N}_0$ with kernel*

$$P^\star(i,j) = 1_{i\ge 1}\, P(i,j+1) + 1_{i=0}1_{j=0}.$$

3. *The two chains are independent.*

Using a certain summing technique, it is possible to describe the local times $(\ell_n(x))_{x\in\mathbb{Z}}$ at times n with $0 \le S_{n-1} < S_n$, and an analogous description exists for times n with $0 \le S_n < S_{n-1}$.

References

[1] D.C. Brydges and G. Slade, The diffusive phase of a model of self-interacting walks. Prob. Theory Relat. Fields 103, 285-315, 1995.

[2] R. van der Hofstad and F. den Hollander, Scaling for a random polymer, Commun. Math. Phys. 169, 397-440, 1995.

[3] R. van der Hofstad, F. den Hollander and W. König, Central limit theorem for the Edwards model. Preprint 1995. To appear in Ann. Probab.

[4] R. van der Hofstad, F. den Hollander and W. König, Central limit theorem for a weakly interacting random polymer, *in preparation*.

[5] F.B. Knight, Random walks and a sojourn density process of Brownian motion, Transactions of the AMS 109, 56-86, 1963.

[6] W. König, The drift of a one-dimensional self-repellent random walk with bounded increments, Probab. Theory and Relat. Fields 96 (1993), 521–543.

[7] W. König, A central limit theorem for a one-dimensional polymer measure. Preprint 1994. To appear in Ann. Probab.

AMS 1991 subject classifications. 60F05, 60J15, 60J55.

Keywords and phrases. Domb-Joyce model, Knight's theorem for local times of simple random walk, time change and scaling, spectral analysis, central limit theorem.

Address: Katholieke Universiteit Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands.

E-mail address: koenig@sci.kun.nl

ON ESTIMATES OF THE SPECTRAL GAP OF A DISCRETE-TIME MARKOV CHAIN

CHRISTIAN MEISE

ABSTRACT. Well-known inequalities for the spectral gap of a Markov chain, like Poincaré's and Cheeger's Inequality, do not perform well if the transition graph of the Markov chain is strongly connected. By using the discrete Itô formula new bounds on the the spectral gap of a discrete-time Markov chain are derived. The lower bounds are given in terms of mean hitting times to subsets of the state space.

1. INTRODUCTION AND RESULTS

Let X be a finite set. Any transition kernel Q on X determines a linear operator on the set of complex-valued functions defined on X by the definition

$$[Qf](x) = \sum_{y \in X} f(y)Q(x,y), \ x \in X. \tag{1.1}$$

Moreover we have $\|Qf\| \leq \|f\|$, where $\|f\| = \sup\{\, |f(x)| \mid x \in X \,\}$ denotes the supremum norm.

If Q is time-reversible with respect to a probability measure μ the operator $[Q\cdot]$ defined in (1.1) is self-adjoint, so that all the eigenvalues of Q are real. Diaconis and Stroock give spectral estimates for the case of a time-reversible Markov chain [3]. Refer to [4] for the case of non-reversible Markov chains.

Throughout this article we assume that Q is irreducible. Then there exists a uniquely determined stationary measure μ that charges every point.

Instead of Q one may consider the discrete generator of the Markov chain $L := I - Q$. Then there is a one-to-one correspondence between the eigenvalues of L and Q. By the spectral gap of the Markov chain Q we denote the quantity

$$\operatorname{gap}(Q) = \inf{}^{+} \{\, |\lambda| \,:\, \lambda \text{ eigenvalue of } L \,\}.$$

In the time-reversible case λ is the smallest positive eigenvalue of L.

Denote by $T_y = \inf\{\, n \geq 0 \,|\, Y_n = y \,\}$ the first hitting time to the point y. Then we have

Proposition 1.2. *(Hitting Bound) For any eigenvalue $\beta \neq 1$ of Q we have*

$$|1 - \beta| \geq \frac{1}{\sup_{x,y \in X} E_x(T_y)}. \tag{1.3}$$

The above result states that in the case of little connection times the euclidean distance from 1 to the eigenvalue β is large.

We have to pay a slightly worse constant in order to state the following bound which suggests to consider mean hitting times from a point to at least one half of all the points.

Proposition 1.4. *Let Q be an irreducible transition kernel on the finite set X. Then for any eigenvalue $\beta \neq 1$ we have*

$$|1-\beta| \geq \frac{1}{2 \sup \left\{ E_x(T_A) \ : \ |A| = \left\lceil \frac{|X|}{2} \right\rceil , \ x \in X \right\}}. \tag{1.5}$$

In the time-reversible case the factor 2 in the denominator in (1.5) vanishes.

Proposition 1.6. *Let Q be irreducible and time-reversible.*
Then for any eigenvalue $\beta < 1$ we have

$$1-\beta \geq \frac{1}{\sup \left\{ E_x(T_A) \ : \ |A| = \left\lceil \frac{|X|}{2} \right\rceil , \ x \in X \right\}}.$$

In the following we want to state the coupling bound which can also be found in [6].

Let $\Omega_1 = \Omega_2 = X^{\mathbf{N}}$. The coordinate process of the coupled chain is given by $Y_n(\omega) = (Y_n^1(\omega), Y_n^2(\omega)) = (\omega_n^1, \omega_n^2),\ \omega \in \Omega_1 \times \Omega_2$. Let T be the hitting time to the diagonal in $X \times X$, i.e.

$$T = \inf\{ \, n \geq 0 \ | \ Y_n^1 = Y_n^2 \, \}.$$

For the eigenvector f to the eigenvalue $\beta \neq 1$ of the transition kernel Q we define $g : X \times X \to \mathbb{R}$ by $g(y_1, y_2) := f(y_1) - f(y_2),\ y_1, y_2 \in X$.

Set

$$\widetilde{Z}_n = g(Y_n) + (1-\beta) \sum_{k=0}^{n-1} g(Y_k) \tag{1.7}$$

and introduce the following space of martingale measures

$$\widetilde{\mathcal{P}} = \widetilde{\mathcal{P}}(s_0) := \left\{ \, \widetilde{P} \ | \ \widetilde{Z}_0 = g(s_0) \, \widetilde{P} - a.s., \ (\widetilde{Z}_n)_n \text{ is a martingale under } \widetilde{P} \, \right\}.$$

Note that any coupling kernel $\widetilde{Q}$ of Q induces a measure $\widetilde{P}$ such that $(\widetilde{Z}_n)_n$ is a $\widetilde{P}$-martingale. By using the martingale property of $(\widetilde{Z}_n)_n$ for measures $\widetilde{P} \in \widetilde{\mathcal{P}}$, we can show

Theorem 1. *(Coupling Bound) Let Q be time-reversible. Then for any eigenvalue $\beta \neq 1$ we have*

$$1-\beta \geq \inf_{\bar{x} \in X \times X} \frac{1}{\inf_{\widetilde{P} \in \widetilde{\mathcal{P}}(\bar{x})} E_{\widetilde{P}}(T)}. \tag{1.8}$$

In [1] (compare also [2]) Chen and Wang give a lower bound for the spectral gap of a reversible continuous-time Markov process $(X_t)_t$. More precisely they show that under certain assumptions $gap(\Omega) \geq (\max_{x_1 \neq x_2} E_{x_1,x_2}(T))^{-1}$. Here Ω is the operator on the set of measurable bounded functions given by $\Omega f(x) = \int_S (f(y) - f(x)) q(x, dy)$ and q is the Q-matrix of $(X_t)_t$. Therefore we can say that theorem 1 is the discrete-time version of the result from Chen and Wang.

2. Proofs

2.1. **Proof of the Hitting Bound.** For any eigenfunction f to the eigenvalue β the sequence $(Z_n)_n$ given by

$$Z_n = f(Y_n) + (1-\beta)\sum_{k=0}^{n-1} f(Y_k) \tag{2.1}$$

is a martingale with respect to the measure that is induced by any initial distribution and the transition kernel Q. By using the martingale property (2.1) of $(Z_n)_n$ the following Lemma is derived.

Lemma 2. *Let f be an eigenfunction and suppose that $E_x(T_y) < \infty$. Under these assumptions we have the integral representation*

$$f(x) = f(y) + (1-\beta)E_x\left(\sum_{k=0}^{T_y-1} f(Y_k)\right).$$

Proof: By the martingale property of $(Z_n)_n$ under P_x we have

$$f(x) \quad = \quad E_x(f(Y_{\min(T_y,n)})) + (1-\beta)E_x\left(\sum_{k=0}^{\min(T_y,n)-1} f(Y_k)\right).$$

Let n tend to infinity and apply Lebesgue's convergence theorem twice. This yields the assertion. □

The next Lemma deals with a property concerning f.

Lemma 3. *Suppose that the stationary measure μ of Q charges every point, i.e. $\mu(y) > 0$ for all $y \in X$. Then for any $x \in X$ there exists $y \in X$ such that $|f(x) - f(y)| > |f(x)|$.*

Proof: First note that $E_\mu(f) = 0$. Let $x \in X$ such that $|f(x)| > 0$. Instead of using the $\mathbb{R}$-basis $1, i$ in the vector space of complex numbers we take $f(x)$ and v, where $v \neq 0$ is a complex number orthogonal to $f(x)$. For any $y \in X$ there are (uniquely determined real) coefficients $\alpha_{y,1}, \alpha_{y,2}$ such that $f(y) = \alpha_{y,1}f(x) + \alpha_{y,2}v$. ¿From $E_\mu(f) = 0$ we conclude

$$\begin{aligned} 0 &= \sum_{y\in X} f(y)\mu(y) \\ &= \sum_{y\neq x}(\alpha_{y,1}\mu(y)f(x) + \alpha_{y,2}\mu(y)v) + f(x)\mu(x) \end{aligned}$$

By uniqueness of the coefficients

$$\mu(x) + \sum_{y\neq x} \alpha_{y,1}\mu(y) = 0.$$

Consequently there exists $y \in X$ such that $\alpha_{y,1} < 0$. This yields the assertion because

$$\begin{aligned} |f(x) - f(y)| &= |f(x) - \alpha_{y,1}f(x) - \alpha_{y,2}v| \\ &\geq |1-\alpha_{y,1}||f(x)| > |f(x)|, \end{aligned}$$

where we have used the orthogonality of $f(x)$ and v in the last but one inequality. $\square$

Now we are able to prove Proposition 1.2

Observe first that for finite X the mean hitting time $E_x(T_y)$ is finite. More precisely for $N := (Id - Q(< y, y >))^{-1}$ where $Q(< y, y >)$ is the submatrix obtained by removing the y-th row and column. Then, cm. [5]

$$\sup_{x \in X} E_x(T_y) = \|N(1, \dots, 1)^t\|_\infty .$$

Without loss of generality we may assume that f satisfies

$$\sup\{\, |f(z)| \,:\, z \in X \,\} = 1.$$

Choose $x \in X$ such that $|f(x)| \geq 1$. By Lemma 3 there exists $y \in X$ satisfying $|f(x) - f(y)| > |f(x)|$.

The integral representation (2) yields

$$f(x) = f(y) + (1-\beta) E_x \left(\sum_{k=0}^{T_y - 1} f(Y_k) \right).$$

Since $|f(Y_k)| \leq 1$, we conclude

$$\begin{aligned} |1-\beta| &\geq \frac{|f(x) - f(y)|}{E_x(T_y)} \geq \frac{1}{E_x(T_y)} \\ &\geq \frac{1}{\sup_{x,y \in X} E_x(T_y)}. \qquad \square \end{aligned}$$

2.2. **Proof of Proposition 1.4.** At first we state a preparatory Lemma.

Lemma 4. *Suppose that f is an eigenfunction to the eigenvalue $\beta \neq 1$ of an irreducible transition kernel Q on the finite state space X. Then there exists $x \in X$ such that at least one half of the points do not lie in the circle with radius $\frac{1}{2}\|f\|_\infty$ centered at $f(x)$.*

Proof: Without loss in generality let $\|f\|_\infty = 1$. Let $x \in X$ such that $|f(x)| = 1$. If there are at least $\left\lceil \frac{|X|}{2} \right\rceil$ points outside the circle of radius $\frac{1}{2}$ centered at $f(x)$ nothing has to be shown. Otherwise there are $\left\lceil \frac{|X|}{2} \right\rceil$ many points lying in the circle centered at $f(x)$. We apply Lemma 3 and obtain another point $y \in X$ such that $|f(y) - f(x)| > |f(x)| = 1$. Thus the circles centered at $f(x)$ and $f(y)$ of radius $\frac{1}{2}$ are disjoint so that y can be chosen to be the point of interest. $\square$

Next let f be an eigenvector to β. Then consider the (complex) martingale

$$\widetilde{Z}_n = f(Y_n) + (1-\beta) \sum_{k=0}^{n-1} f(Y_k).$$

Assume that $\|f\|_\infty = 1$ and choose $x \in X$ according to Lemma 4. Let

$$T = \inf \left\{ n \geq 0 \,:\, |f(Y_n) - f(x)| \geq \frac{1}{2} \right\}.$$

Applying the optional stopping theorem yields

$$f(x) - E_x(f(Y_{T\wedge n})) = (1-\beta)E_x(\sum_{m=0}^{T\wedge n-1} f(Y_m)).$$

Use Lebesgue's convergence theorem for the conclusion

$$f(x) - E_x(f(Y_T)) = (1-\beta)E_x(\sum_{m=0}^{T-1} f(Y_m)).$$

Next take absolute values on both sides

$$\begin{aligned} |f(x) - E_x(f(Y_T))| &= |1-\beta||E_x(\sum_{m=0}^{T-1} f(Y_m))| \\ &\le |1-\beta|E_x(T) \end{aligned}$$

By definition of T for any $\omega \in \{T < \infty\}$ we have $|f(x) - f(Y_T(\omega))| \geq \frac{1}{2}$ which yields the result. □

Since the arguments used in the proof for the time-reversible Hitting Bound is very similar to the previous proof for the non-reversible Hitting Bound. Therefore it is omitted. The proof for the coupling bound can be found in [6].

References

1. M.F. Chen, F.Y. Wang, *Applications of coupling method to the first eigenvalue on manifold*, Science in China, Technical Report, No. 215 (1994).
2. M.F. Chen, *Optimal Markovian Couplings and Applications*, Acta Math. Sinica (NS) **10** (1994) (Nr. 3).
3. P. Diaconis, D. Stroock, *Geometric Bounds For Eigenvalues of Markov Chains*, The Annals of Applied Probability, No. 1, 36-61 (1991).
4. J.A. Fill, *Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process*, The Annals of Applied Probability, No. 1, 62-87 (1991).
5. J.G. Kemeny, J.L. Snell, *Finite Markov chains*, D. Van Nostrand Company, Inc. (1960).
6. C. Meise, *Lower bounds for the spectral gap of Markov chains*, Preprint 96-019, SFB 343, Universität Bielefeld (1996).
7. E. Seneta, *Non-negative matrices and Markov chains*, Springer-Verlag, New York (1981).

(Christian Meise) Fakultät für Mathematik, Universität Bielefeld, Postfach 10 01 31, D-33501 Bielefeld, Germany

E-mail address, Christian Meise: meise@mathematik.uni-bielefeld.de

UNIQUENESS OF BROWNIAN MOTION ON NESTED FRACTALS

VOLKER METZ (UNIVERSITÄT BIELEFELD)

The desire to study fractals arose in physics in the early 1980s. Fractals are believed to provide good models for disordered media, although no natural medium has a fractal structure on the sub-molecular scale. One way to study fractals is to consider their "Laplacians" or in probabilistic terms their "Brownian motions" whose generators are the "Laplacians".

To simplify the notation we will restrict ourselves to an example, the Vicsek snowflake, a nested fractal. Roughly speaking a nested fractal is a finitely ramified fractal with many symmetries. A fractal is called finitely ramified, if it becomes disconnected when finitely many appropriate points are removed. The Vicsek set can be constructed by five similitudes $\psi_1, \ldots, \psi_5 : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\psi_1(x) := \frac{x}{3} + \frac{2}{3}(1,1), \quad \psi_2(x) := \frac{x}{3} + \frac{2}{3}(0,1), \quad \psi_3(x) := \frac{x}{3},$$
$$\psi_4(x) := \frac{x}{3} + \frac{2}{3}(1,0), \quad \psi_5(x) := \frac{x}{3} + \frac{1}{3}(1,1).$$

They contract the unit square $[0,1]^2$ by $1/3$ and arrange five such small copies in a chessboard pattern inside $[0,1]^2$. Now, the transformation is repeated on the remaining squares infinitely often. More precisely, define $\Psi(M) := \bigcup_{i=1}^5 \psi_i(M)$ for $M \subset \mathbb{R}^2$, then the fractal is given by

$$X := \bigcap_{n\in\mathbb{N}} \Psi^n([0,1]^2).$$

In our example the copies $\psi_1(X), \ldots, \psi_5(X)$ overlap in 4 points, the corners of $\psi_5(X)$. When one of these points is removed the fractal becomes disconnected.

Our fractal has Hausdorff dimension $\ln 5/\ln 3$, so its two dimensional Lebesgue measure is 0 which means that its "holes" (cut into the chessboard) are dense in $[0,1]^2$. In this sense it is very inhomogeneous (no manifold), thus there is no direct way to construct a "Brownian motion" on it. Here a "Brownian motion" is meant to be a diffusion with strong Markov property that is invariant under the symmetries of X and thereby uniquely determined up to a constant time scaling. Instead, we will define a sequence of more and more refined "grids" that approximate X and consider a scaled sequence of random walks on them to define a limiting process. The main problem is to find a scaling constant which guarantees a nontrivial limit. That is, the limit has to be irreducible and the particle has to travel across X in finite, positive time.

Let $x_1, \ldots, x_4$ denote the corners of $[0,1]^2$. They are the unique fixed points of $\psi_1, \ldots, \psi_4$. We define the vertices of the n–generation "grid" by

$$\begin{aligned} V_0 &:= \{x_1, \ldots, x_4\}, \\ V_n &:= \Psi^n(V_0) \qquad\qquad (n \in \mathbb{N}) \end{aligned}$$

because they contain all (eventually disconnecting) points and the set of all vertices, $V := \bigcup_n V_n$, is dense in X with respect to the Euclidean topology of $\mathbb{R}^2$. Let $\mathcal{G}$ denote the symmetry group of X. In our case it is generated by certain reflections. For $x, y \in V_0$ with $x \neq y$ define $L_{x,y}$ to be the set of points equidistant from x and y. Then $\mathcal{G}$ is generated by the reflections in $L_{x,y}$ for all $x, y \in V_0$, $x \neq y$. A symmetric, $\mathcal{G}$–invariant random walk on V_0 can be defined via its generator, or alternatively via its bilinear form, here called a Dirichlet form $\mathcal{E}_0$. Choose transition weights $c : V_0^2 \to \mathbb{R}_+$, symmetric, 0 on the diagonal, and $\mathcal{G}$–invariant, that is, $c(g(x), g(y)) = c(x, y)$ for all $x, y \in V_0$ and $g \in \mathcal{G}$. Define

$$\text{(D)} \qquad \mathcal{E}_0(f, g) := \frac{1}{2} \sum_{x,y \in V_0} (f(x) - f(y))(g(x) - g(y))c(x, y)$$

for $f, g : V_0 \to \mathbb{R}$. This Dirichlet form defines a symmetric Markov chain on V_0 with one step transition probability $p(x, y) = c(x, y)/(\sum_{y \in V_0} c(x, y))$. The self–similarity of X is used to define a random walk on V_1 by five coupled copies of the one on V_0. It is determined by the Dirichlet form $\mathcal{E}_1$ which is the image of $\mathcal{E}_0$ under the coupling map Ψ.

$$\text{(C)} \qquad \mathcal{E}_1(f, g) := \Psi(\mathcal{E}_0)(f, g) := \sum_{i=1}^{5} \mathcal{E}_0(f \circ \psi_i, g \circ \psi_i)$$

for $f, g : V_1 \to \mathbb{R}$. The copies $\psi_i(V_0)$, $1 \leq i \leq 5$, overlap in $\psi_5(V_0) \subset V_1$ where the walker decides independently with equal probability to choose one of the neighbouring copies to continue its journey, because every copy of $\mathcal{E}_0$ in the above sum has the same weight. The idea now is to find a scaling constant $\gamma > 0$ such that $(\gamma^{-n}\Psi^n(\mathcal{E}_0))_{n \in \mathbb{N}}$ defines the desired limit. Once γ and a suitable $\mathcal{E}_0$ have been found the construction of the limit is well known (see the definition (L) below).

To find γ we solve a finite dimensional nonlinear eigenvalue problem using ideas of [4] and a nonlinear version of the Perron–Frobenius theory known as Hilbert's projective metric on cones (see [6]). We define the trace $\Phi(\mathcal{E}_1)$ of $\mathcal{E}_1$ on V_0 by

$$\text{(T)} \qquad \Phi(\mathcal{E}_1)(f, f) = \min\{\mathcal{E}_1(g, g) | g : V_1 \to \mathbb{R}, g|_{V_0} = f\}.$$

By polarization this defines a Dirichlet form which determines a symmetric random walk on V_0. The random process inherits the hitting probabilities of $\mathcal{E}_1$ between points of V_0 but rescaled to add up to 1 and the time changed such that successive jumps occur with time difference one (see [2]). This can be deduced from the form of the minimizing element of the variational problem (T), the Dirichlet problem on $V_1 \setminus V_0$ for $f : V_0 \to \mathbb{R}$. Its solution is $H^{\mathcal{E}_1}_{V_1 \setminus V_0} f$ a function with boundary data f on

V_0 and harmonic with respect to $\mathcal{E}_1$ on $V_1 \setminus V_0$. It is unique, if $\mathcal{E}_1$ is irreducible or equivalently, if the corresponding random walk on V_1 is irreducible. The harmonic kernel $H^{\mathcal{E}_1}_{V_1\setminus V_0}$ can be expressed by the above hitting probabilities.

With $\Lambda := \Phi \circ \Psi$ the eigenvalue problem is

$$\exists \mathcal{F}_0 \, \exists \gamma > 0 : \Lambda(\mathcal{F}_0) = \gamma \mathcal{F}_0, \tag{F}$$

where $\mathcal{F}_0$ should be defined by a strictly positive c to ensure that the scaled limit is irreducible. In this equation only V_0 and V_1 are involved instead of $(V_n)_n$ because of the self–similarily of X and the V_n's. Division by γ transforms equality (F) into a fixed point equation. In [3] it is shown that a solution $\mathcal{F}_0$, γ of (F) defines a nontrivial limit via the Dirichlet forms

$$\begin{aligned} \mathcal{E}_n &:= \Psi^n(\mathcal{E}_0) \text{ for all } n \in \mathbb{N}, \\ \mathcal{D}(\mathcal{E}) &:= \{f \in \mathcal{C}(X,\mathbb{R}) | \sup_{n\in\mathbb{N}} \gamma^{-n} \mathcal{E}_n(f|_{V_n}, f|_{V_n}) < \infty\}, \\ \mathcal{E}(f,f) &:= \lim_{n\to\infty} \gamma^{-n} \mathcal{E}_n(f|_{V_n}, f|_{V_n}). \end{aligned} \tag{L}$$

The limit in (L) actually equals the supremum. The existence of $\mathcal{F}_0$ and γ was proved by Lindstrøm in 1990 and its uniqueness up to a constant multiple in 1995 independently by Sabot and Metz (see [4], [7], [5]).

We interpret V_0^2 as a set of (undirected) edges and c as nonnegative $\mathcal{G}$–invariant weights on these edges. This allows only a two parameter family of weights on the edges. Let $\mathbb{D}$ denote the cone of all corresponding Dirichlet forms on V_0 defined by (D), then $\mathbb{D} \simeq \mathbb{R}^2_+$ the standard nonnegative cone in $\mathbb{R}^2$ and $\mathbb{B} := \mathbb{D} - \mathbb{D} \simeq \mathbb{R}^2$. To define an appropriate version of Hilbert's metric we need

$$\mathbb{P} := \{\mathcal{E} \in \mathbb{D} - \mathbb{D} \mid \mathcal{E}(f,f) \geq 0 \text{ for all } f : V_0 \to \mathbb{R}\},$$

the cone of positive semidefinite forms on V_0. It is "better" than $\mathbb{D}$ because Λ is monotone with respect to the partial ordering induced by $\mathbb{P}$ but it is not monotone with respect to the one induced by $\mathbb{D}$. In our example

$$\mathbb{P} \simeq \{(x,y) \in \mathbb{R}^2 \mid x \geq 0, x + y \geq 0\}.$$

The geometry of $\mathbb{P}$ depends on the size of V_0 and its symmetries $\mathcal{G}$. The usual form norm on $\mathbb{B}$ defines a Banach space. For $\mathcal{A}, \mathcal{B} \in \mathbb{B}$ we define the partial ordering induced by $\mathbb{P}$:

$$\mathcal{A} \leq \mathcal{B} \;:\Leftrightarrow\; B - \mathcal{A} \in \mathbb{P}$$

which is equivalent to $\mathcal{A}(f,f) \leq \mathcal{B}(f,f)$ for all $f : V_0 \to \mathbb{R}$. Let $\mathbb{P}^\circ$ denote the interior of $\mathbb{P}$. It consits of all irreducible Dirichlet forms. We compare $\mathcal{A}, \mathcal{B} \in \mathbb{P}^\circ$ by

$$\begin{aligned} M(\mathcal{A}/\mathcal{B}) &:= \inf\{\alpha > 0 \mid \mathcal{A} \leq \alpha\mathcal{B}\}\ , \\ m(\mathcal{A}/\mathcal{B}) &:= \sup\{\alpha > 0 \mid \alpha\mathcal{B} \leq \mathcal{A}\}\ . \end{aligned}$$

Then Hilbert's distance is

$$h(\mathcal{A},\mathcal{B}) := \ln \frac{M(\mathcal{A}/\mathcal{B})}{m(\mathcal{A}/\mathcal{B})}\ .$$

It is not a metric because $h(\alpha\mathcal{A}, \beta\mathcal{B}) = h(\mathcal{A}, \mathcal{B})$ for all $\alpha, \beta > 0$ and $h(\mathcal{A}, \mathcal{B}) = 0$ if and only if $\mathcal{A}$ and $\mathcal{B}$ are linearly dependent. Hence we have to project everything onto $S_1(0)$, the unit sphere in $\mathbb{B}$. Now, $(S_1(0) \cap \mathbb{P}^\circ, h)$ is a complete metric space by [1]. It measures the distance between irreducible, symmetric Markov chains on V_0. The definition of Λ immediately implies that it is increasing with respect to "$\leq$" and positively homogeneous. Thus it is h–nonexpansive and it remains to be shown that there is some kind of contraction.

Theorem. ([5]): *Let $\mathcal{A} \in \mathbb{P}^\circ$, $\mathcal{F}$ a solution of (F) and $\mathcal{A}$ not a multiple of $\mathcal{F}$. Then there exists an $n \in \mathbb{N}$ such that*

$$h(\mathcal{A}, \mathcal{F}) > h(\Lambda^n(\mathcal{A}), \mathcal{F}).$$

The proof of this theorem is indirect and relies on the action of Λ on $\mathbb{D}$ and $\mathbb{P}$ in the following respects: (i) $\mathbb{D}$ and $\mathbb{P}$ are Λ–invariant, (ii) Λ maps $\partial\mathbb{D} \cap \mathbb{P}^\circ$ into $\mathbb{D}^\circ$ (where $\partial\mathbb{D}$ denotes the boundary of $\mathbb{D}$), and (iii) the Fréchet derivative $D\Lambda(\mathcal{F})$ of Λ in $\mathcal{F}$ maps $\partial\mathbb{P} \cap \mathbb{D}$ into $\mathbb{P}^\circ$. Furthermore, we need (iv) the polyhedral structure of $\mathbb{P}$, that is, $\mathbb{P}$ is the intersection of $\dim \mathbb{B}$ half spaces.

Condition (i) guarantees that our fixed point set up is well defined. It says that coupling and taking traces maps symmetric random walks onto processes of the same kind. Property (ii) implies that there cannot be a solution of (F) in $\mathbb{P}^\circ \setminus \mathbb{D}$. Probabilistically it means that taking traces turns an irreducible random walk on V_1 into one which has strictlt positive hitting probabilities between points of V_0. Statement (iii) causes the contraction of Hilbert's distance between $\mathcal{F}$ and another point of $\mathbb{D}^\circ$ in a neighbourhood of $\partial\mathbb{P} \cap \mathbb{D}$ by Λ. This says that taking traces maps nearly reducible random walks into less reducible ones. The uniqueness of a solution of (F) up to constant multiples is a consequence of these three properties. The final weak contraction result additionally needs (iv). This condition entirely depends on $\mathcal{G}$ and has nothing to do with probability theory. It allows to transform our fixed point set up into $\mathbb{R}^{\dim \mathcal{B}}$ equipped with the uniform norm and a nonexpansive map which shows that Hilbert balls are convex but not necessarily strictly convex.

The convergence property in the above theorem implies stability results for the dynamical system defined by Λ/γ on $\mathbb{D}^\circ$. More precisely, one may choose any $\mathcal{E}_0 \in \mathbb{D} \cap \mathbb{P}^\circ$ and $((\Lambda/\gamma)^n(\mathcal{E}_0))_{n\in\mathbb{N}}$ still ends up on the ray spanned by $\mathcal{F}$. One could even start in $\mathbb{P}^\circ \setminus \mathbb{D}$ but this has no Markov process interpretation any more because of eventually negative transition weights.

References

1. G. Birkhoff, *Uniformly semi-primitive multiplicative processes*, Trans. Amer. Math. Soc. **104** (1962), 35–51.
2. M. Fukushima, Y. Oshima, M. Takeda, *Dirichlet Forms and Symmetric Markov Processes*, de Gruyter Stud. Math., vol. 19, Walter de Gruyter, Berlin–New York, 1994.
3. S. Kusuoka, *Diffusion processes on nested fractals*, Statistical Mechanics and Fractals (R. L. Dobrushin, S. Kusuoka, eds.), Springer Lect. Notes Math., vol. 1567, 1993, pp. 39–98.
4. T. Lindstrøm, *Brownian Motion on Nested Fractals*, Mem. Amer. Math. Soc., vol. 83, no. 420, Amer. Math. Soc., Providence, 1990.
5. V. Metz, *Renormalization contracts on nested fractals*, J. reine angew. Math. **480** (1996), 161–175.

6. R. D. Nussbaum, *Hilbert's Projective Metric and Iterated Nonlinear Maps*, Mem. Amer. Math. Soc., vol. 75, no. 391, Amer. Math. Soc., Providence, 1988.
7. C. Sabot, *Éxistence et unicité de la diffusion sur un fractal*, dissertation, Université Pierre et Marie Curie, Laboratoire de Probabilité, Paris, 1995.

NEURAL NETWORKS WITH TEMPORAL ASSOCIATION

MICHAEL SANDMEIER

Abstract. H. Sompolinsky and I. Kanter proposed a Hopfield-type model storing sequences of patterns. Several non-rigorous results are presented in physical literature.
Though an energy function does not exist, we obtain rigorous results concerning the storage capacity, the dynamics and the thermodynamic properties of the model similar to those for the Standard-Hopfield model.

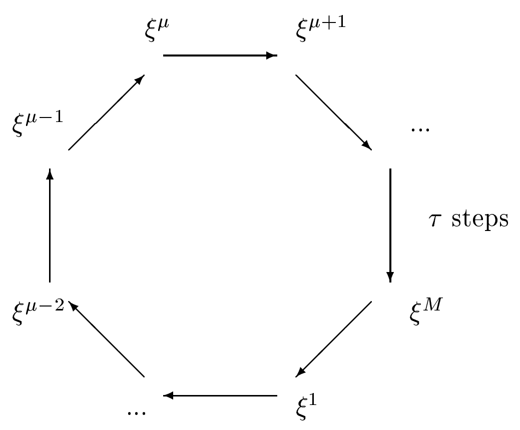
1. The Model

We consider a sequence $(\xi^\nu)_{\nu\in\mathbb{N}}$ of patterns, $\xi^\nu = (\xi_i^\nu)_{i\in\mathbb{N}}$, $\xi_i^\nu \in \{-1,+1\}^{\mathbb{N}}$ (i.i.d.r.v.) with $P(\xi_i^\nu = +1) = P(\xi_i^\nu = -1) = \frac{1}{2}$ for all $\nu,\ i \in \mathbb{N}$.
For a given number N, we define

$$\Xi := \Xi|_N := \{\xi_i^\nu :\ i = 1..N,\ \nu = 1..M(N)\},$$

where N denotes the number of neurons, the complexity of the network, and M:=M(N), depending on N, denotes the number of patterns to be stored. For a fixed $\tau \in \mathbb{N}$, which can be regarded as the time a pattern ξ^μ is presented, we define a periodic sequence (cycle) of patterns through

$$\begin{aligned} \xi : \mathbb{N} &\longrightarrow \{\xi^\nu : \nu = 1..M\} \\ \xi((\mu-1)\cdot\tau + t) &:= \xi^\mu, \quad \textit{for all } t = 1..\tau \textit{ and } \mu = 1..M \\ \xi(t + M\cdot\tau) &:= \xi(t), \quad \textit{for all } t \end{aligned}$$



We study a network with a dynamics given by

$$\sigma_i(t+1) := sgn(\sum_{j\neq i} J_{ij}^{(1)} \cdot \sigma_j(t) + \lambda \cdot J_{ij}^{(2)} \cdot \overline{\sigma_j}(t)) \ \forall i$$

where $J_{ij}^{(1)} := \sum_\nu \xi_i^\nu \xi_j^\nu$, $J_{ij}^{(2)} := \sum_\nu \xi_i^{\nu+1} \xi_j^\nu$, $\overline{\sigma_j}(t) := \sum_{k=1}^{\tau} w(k) \cdot \sigma(t+1-k)$ denotes the selected mean over the past, where w is a distribution on $\{1..\tau\}$ and $\lambda > 1$.
It has been useful to describe the properties of an associative network in terms of an energy function. Such a function does not exist in our model, which will be the main difficulty.

2. The Storage Capacity

The first question is the question of storage capacity. There are several notions of storage that would be reasonable. The choice depends on possible applications.
We will discuss four different notions and calculate the corresponding storage capacity.

2.1. Dynamic Notion of Storage.

Our aim is to retrieve the patterns exactly at each time t.

Definition 1.
We say, a sequence $\{\xi^\nu\}$ is stable, if for all t

$$\sigma_0(t) = \xi(t),$$

where σ_0 denotes the trajectory defined by the dynamics and a given initial state, i.e. assuming exact reconstruction of the first period. We get

Theorem 1.
Under the assumption $M(N) = \frac{N}{\gamma log N}$ the following assertion holds:
If $\gamma > 4 \cdot \frac{1+\lambda^2}{(1-\lambda)^2}$

$$\lim_{N\to\infty} P(\bigcap_t \{\sigma_0(t) = \xi(t)\}) = 1.$$

In the case $w \equiv \delta_\tau$ and $\lambda \to \infty$ we get exactly the same storage capacity as in the corresponding result for the Standard-Hopfield model, $\gamma > 4$. This is reasonable since networks with synapsis $J_{ij}^{(1)}$, resp. $J_{ij}^{(2)}$, have analogue properties.

2.2. Static Notion of Storage.

To get larger capacities we accept a small error between σ_0 and ξ, uniformly for all times t.

Definition 2.
Given a small $\delta > 0$, we say, a sequence $\{\xi^\nu\}$ is stable, if

$$d(\sigma_0(t), \xi(t)) < \delta N$$

for all t, where d denotes the Hamming distance.

We can improve the order of the capacity choosing M proportional to N:

Theorem 2.
There exist $\alpha_c > 0$ *and* $\delta \in (0; \frac{1}{2})$*, such that for all* $\alpha < \alpha_c$ *the following assertion holds:*
If $M = \alpha \cdot N$*, we have*

$$P(\bigcap_t \{d(\xi(t), \sigma_0(t)) < \delta N\})$$

tends to 1 exponentially fast as $N \to \infty$.

Next, we want to generalize these notions using the delay τ to attract the following pattern. Stability is then depending on the states at the end of each period.

2.3. **Dynamic Notion of Storage.** (attracting version)

Definition 3.
We say, a sequence $\{\xi^\nu\}$ *is stable, if*

$$\sigma_0(\mu \cdot \tau) = \xi^\mu$$

for all $\mu = 1..M$.

Nevertheless, we have to bound the error during the whole period since the following dynamics depends on the past. The storage capacity is of the same order as before, but we can improve the numerical values:

Theorem 3.
Assuming $M \leq \frac{N}{\gamma log N}$*, the following assertion holds:*
For $\gamma > 4 \cdot \frac{1+\lambda^2 \cdot \sum_k w^2(k)}{(1+\lambda w(\tau) - \lambda \sum_{k=1}^{\tau-1} w(k))^2}$ *we have*

$$\lim_{N\to\infty} P(\bigcap_\mu \{\sigma_0(\mu \cdot \tau) = \xi^\mu\}) = 1.$$

To see the improvement we look at the case $w \equiv \delta_\tau$: Theorem 3 assumes $\gamma > \frac{1+\lambda^2}{(1+\lambda)^2}$ which is determined by the last step of each period while the choice of γ in Theorem 1 depends on the transition from the actual pattern to the following one.

2.4. **Static Notion of Storage.** (attracting version)

Definition 4.
Given a small $\delta > 0$*, we say, a sequence* $\{\xi^\nu\}$ *is stable, if*

$$d(\sigma_0(\mu \cdot \tau), \xi^\mu) < \delta N$$

for all $\mu = 1..M$.

Theorem 4.
There exist $\alpha_c > 0$ *and* $\delta \in (0; \frac{1}{2})$*, such that for all* $\alpha < \alpha_c$ *the following assertion holds:*
If M=$\alpha \cdot$ *N, we have*

$$P(\bigcap_\mu \{d(\xi^\mu, \sigma_0(\mu \cdot \tau)) < \delta N\})$$

tends to 1 exponentially fast.

Again, the numerical value (of α) is larger since both capacities, of Theorem 2 and Theorem 4, are determined by the transition but the transition in Theorem 4 leads to a larger environment than the corresponding one in Theorem 2.

3. Stochastic Dynamics

Finally, we study the behaviour of the network, substituting the deterministic dynamics by a stochastic one. For simplicity of notation we fix $w \equiv \delta_\tau$. Now, Sompolinsky and Kleinfeld proposed an adiabatically varying energy function of the following form:

Definition 5. *Assuming perfect memory we define a Hamiltonian for the μ-th period:*

$$\begin{aligned} H^\mu &: \{-1,+1\}^N \longrightarrow \mathbb{R} \\ H^\mu(\sigma) &:= -\frac{1}{2N}\sum_i \sum_{j\neq i}(J^{(1)}_{ij}\cdot\sigma_j + \lambda\cdot J^{(2)}_{ij}\cdot\xi^{\mu-1}_j)\cdot\sigma_i. \end{aligned}$$

With the help of this Hamiltonian we define the update by the discrete Glauber dynamics

$$P(\sigma_i \mapsto -\sigma_i) = (1+exp(2\beta h_i\sigma_i))^{-1}$$

or by the Metropolis algorithm

$$P(\sigma\mapsto\sigma') = \begin{cases} 1 & , \quad \Delta H^\mu < 0 \\ e^{-\beta\Delta H^\mu} & , \quad otherwise \end{cases}$$

where $h_i := \sum_{j\neq i}(J^{(1)}_{ij}\cdot\sigma_j + \lambda\cdot J^{(2)}_{ij}\cdot\xi^{\mu-1}_j)$,
$\Delta H^\mu := H^\mu(\sigma) - H^\mu(\sigma')$ and β stands for the inverse temperature. Both rules define a Markov chain on $S_N = \{-1,+1\}^N$ with a stationary distribution given by the finite volume Gibbs measure

$$\mathcal{G}^\mu_{N,\beta}(\sigma) := \frac{1}{Z^\mu_{N,\beta}}e^{-\beta H^\mu(\sigma)}.$$

Define the induced measure

$$\mathcal{D}^\mu_{N,\beta}[\omega] := \mathcal{G}^\mu_{N,\beta}[\omega]\circ\mathcal{M}_N[\omega]^{-1}$$

with

$$\begin{aligned} \mathcal{M}_N : \{-1,+1\}^N &\to \mathbb{R}^M \\ \sigma &\mapsto (m^1_N[\omega](\sigma),..,m^M_N[\omega](\sigma)), \end{aligned}$$

where

$$m^\nu_N[\omega](\sigma) := \frac{1}{N}\sum_{i=1}^N \xi^\nu_i[\omega]\sigma_i$$

(overlap).

Theorem 5. *Let $M = \frac{N}{\gamma logN}$, with $\gamma > \frac{4}{\epsilon^2}$, $0<\epsilon<1$ and $\beta\geq\beta^*$. Then*

$$w-\lim_{N\to\infty}\mathcal{D}^\mu_{N,\beta}[\omega]([a^-(\beta,\epsilon),a^+(\beta,\epsilon)]\cdot e^\mu) = 1$$

for almost all ω, where e^μ denotes the μ-th unit vector of $\mathbb{R}^M$.

Here, $a^-(\beta,\epsilon)$, repsectively $a^+(\beta,\epsilon)$, denotes the largest solution of

$$a = tanh(\beta\cdot(a+\lambda\cdot(1\pm\epsilon)))$$

and $\beta^* := inf\{\beta \ : \ 1 - \frac{\beta}{cosh^2(\beta[a(\beta)+\lambda(1+\epsilon)])} > 0\}$.

References

1. A. Bovier, V. Gayrard; A retrieval phase of the Hopfield model: A rigorous analysis of the overlap distribution; WIAS-preprint No. 161, Berlin (1995)
2. A. Bovier, V. Gayrard, P. Picco; Gibbs states of the Hopfield model in the regime of perfect memory; Probab. Theory Relat. Fields 100, 329-363 (1994)
3. J. L. van Hemmen, R. Kühn; Temporal Association; in Models of neural networks, E. Domany, J. L. van Hemmen, K. Schulten (eds.), Springer, Berlin (1991)
4. J. J. Hopfield; Neural networks and physical systems with emergent collective computational abilities; Proc. nat. Acad. Sci. USA 79, 2554-2558 (1982)
5. C. Newman; Memory capacity in neural networks; Neural Networks 1, 223-238 (1988)
6. H. Sompolinsky, I. Kanter; Temporal association in asymmetric neural networks; Phys. Rev. Lett. 57, 2861-2864 (1986)
7. H. Sompolinsky, D. Kleinfeld; Associative Neural Network Model for the Generation of Temporal Patterns; Biophys. J. 54, 1039-1051 (1988)
8. H. Sompolinsky, D. Kleinfeld; Associative Network Models for Central Pattern Generators; in Methods in Neuronal Modeling, C. Koch, I. Segev, MIT Press (1989)

(Michael Sandmeier) Fakultät für Mathematik, Universität Bielefeld, Postfach 10 01 31, D-33501 Bielefeld, Germany

E-mail address, Michael Sandmeier: `michael@mathematik.uni-bielefeld.de`

LARGE DEVIATIONS FOR HIERACHICALLY INTERACTING MARKOV CHAINS

Boualem Djehiche
and
Alexander Schied

This note announces results that were obtained by Boualem Djehiche (KTH, Stockholm) and the author. The proofs of all results as well as applications to epidemic models can be found in [DS].

1. Model. Let $\mathbb{E}$ denote a Polish space with Borel σ-field $\mathcal{B}_{\mathbb{E}}$. We consider a system of M colonies each consisting of N individuals. Let X_t^{ij} denote the state of the jth individual in the ith colony at time $t \in I := [0, T]$, for some fixed $T > 0$. Assume that, for each (i, j), $(X_t^{ij}, t \in I)$ is an $\mathbb{E}$–valued jump process with jump rate $\gamma^{ij}(\cdot)$ and jump kernel π^{ij}. The jump intensity measure $\gamma^{ij}(\mathbf{x})\, dt$ gives the jump rate of the jth individual in the ith colony when the system is in the state $\mathbf{x} \in \mathbb{E}^{NM}$. The jump size distribution $\pi^{ij}(x, dy)$, $x \in \mathbb{E}$, is such that if a jump of the jth individual in the ith colony at site $x \in \mathbb{E}$ occurs, then it is of the form $x \to dy \subset \mathbb{E}$ with probability $\pi^{ij}(x, dy)$. We are going to assume throughout that $\pi^{ij}(x, dy) = \pi(x, dy)$, for a fixed given probability transition kernel π on $\mathbb{E} \times \mathcal{B}_{\mathbb{E}}$. Consequently, all interaction is guided into the system through the jump rates γ^{ij}, what means that we are dealing with a jump process analogue of a system of interacting diffusions.

For each colony i, we can construct the corresponding flow of empirical distributions

$$\mu_t^i = \frac{1}{N} \sum_{j=1}^{N} \delta_{X_t^{ij}} \qquad (t \in I) \tag{1.1}$$

as stochastic process taking values in the space $\mathcal{M} = \mathcal{M}(\mathbb{E})$ of Borel probability measures on $\mathbb{E}$. The *two-level* empirical process then is given by

$$\xi_t^{M,N} = \frac{1}{M} \sum_{i=1}^{M} \delta_{\mu_t^i} \qquad (t \in I). \tag{1.2}$$

It takes values in the space $\mathcal{M}^{II} = \mathcal{M}(\mathcal{M})$ of probability measures on $\mathcal{M}$. Both $\mathcal{M}$ and $\mathcal{M}^{II}$ will be endowed with the corresponding weak topologies and Borel σ-fields. Of course, as long as the number M of colonies is fixed, ξ_t could be obtained from a standard empirical distribution on the enlarged state space $\{1, \ldots, M\} \times \mathbb{E}$. However, the two-level model allows the number of colonies to grow simultaneously with the total number of particles, which seems to be natural for many models. For this reason, we are interested in an interaction that is guided into the system through the two-level empirical distribution ξ of the state $\mathbf{x} = (x^{ij}) \in \mathbb{E}^{NM}$ of the

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

system, and the empirical distribution μ^i of the corresponding colony. This means that the jump intensities $(\gamma^{ij}(\mathbf{x}))$ have to be of the form

$$\gamma^{ij}(\mathbf{x}) = \gamma(x^{ij}, \mu^i, \xi) \qquad (1 \le i \le M,\ 1 \le j \le N), \tag{1.3}$$

where γ is a fixed function on $\mathbb{E} \times \mathcal{M} \times \mathcal{M}^{II}$. For simplicity, we will make the following assumptions on the jump characteristics π and γ:

Assumption I. *The kernel π possesses the Feller property.*

Assumption II. *γ is a bounded positive and continuous function on $\mathbb{E}\times\mathcal{M}\times\mathcal{M}^{II}$.*

In the sequel we will be interested in the asymptotics of the two-level empirical process $\xi_t = \xi_t^{N,M}$ as $N,\ M \to \infty$. Due to (1.3), $\xi_t^{N,M}$ is a measure-valued Markov process. It has an infinitesimal generator $\mathcal{A}^{N,M}$ acting on bounded measurable functions F on $\mathcal{M}^{II}$ as follows:

$$\begin{aligned} \mathcal{A}^{M,N}F(\xi) = MN \int & \overline{m}(\xi, d\mu, dx, dy) \\ & \Big[F\Big(\xi + \frac{1}{M}\{\delta_{(\mu + \frac{1}{N}(\delta_y - \delta_x))} - \delta_\mu\}\Big) - F(\xi)\Big], \end{aligned} \tag{1.4}$$

where we have used the abbreviations

$$\overline{m}(\xi, d\mu, dx, dy) = \xi(d\mu)\, \mu(dx)\, m(x, \mu, \xi, dy) \tag{1.5}$$

and

$$m(x, \mu, \xi, dy) = \gamma(\xi, \mu, x)\, \pi(x, dy). \tag{1.6}$$

Theorems 4.7.3 and 4.4.2 in Ethier and Kurtz [EK] and Assumption II show that the martingale problem associated with $\mathcal{A}^{N,M}$ is well-posed and that its unique solution is a strong Markov process.

As usual, we will frequently denote the integration of a bounded and measurable function f on $\mathbb{E}$ against a measure $\mu \in \mathcal{M}$ by $\langle \mu, f\rangle$. To distinguish it from an integral of a function F on $\mathcal{M}^{II}$ with respect to some $\xi \in \mathcal{M}^{II}$, we will write $\ll \xi, F \gg$ in the latter case.

Now choose a totally bounded metric d generating the topology of $\mathbb{E}$. See Parthasarathy [Pa], p. 43, for existence. We define a space $\mathcal{F}_0$ of finitely based functions on $\mathcal{M}$ as the set of all functions f that are of the form

$$f(\mu) = \varphi\big(\langle \mu, \psi_1\rangle, \dots, \langle \mu, \psi_k\rangle\big) \qquad (\mu \in \mathcal{M}), \tag{1.7}$$

for some $k \in \mathbb{N}$, $\varphi \in C_c^1\big(\mathbb{R}^k\big)$, the space of continuously differentiable functions having compact support in $\mathbb{R}^k$, and $\psi_1, \dots, \psi_k \in C_u\big(\mathbb{E}\big)$, the space of bounded and uniformly continuous functions on the metric space $(\mathbb{E}, d)$. For the ease of notation we will adopt the convention of rewriting (1.7) by

$$f(\mu) = \varphi\big(\langle \mu, \vec{\psi}\rangle\big) \qquad (\mu \in \mathcal{M}).$$

We will also need a space $\mathcal{F}$ of time dependent test functions on $I \times \mathcal{M}$ being of the form

$$f_t(\mu) = f(t, \mu) = \varphi\big(t, \langle \mu, \vec{\psi}(t) \rangle\big) \qquad (t \in I,\ \mu \in \mathcal{M})$$

with $\varphi \in C_b^1(I \times \mathbb{R}^k)$ and a vector $\vec{\psi} = (\psi_1, \ldots, \psi_k)$ such that each component is a function from I to $C_u(\mathbb{E})$, differentiable with respect to the sup-norm $\|\cdot\|$ on $C_u(\mathbb{E})$. Clearly, functions $f \in \mathcal{F}$ are differentiable with respect to time with derivative

$$\partial_t f(t, \mu) = (\partial_t f_t)(\mu) = (\partial_t \varphi)\big(t, \langle \mu, \vec{\psi}(t) \rangle\big) + (\nabla \varphi)\big(t, \langle \mu, \vec{\psi}(t) \rangle\big) \cdot \langle \mu, \partial_t \vec{\psi}(t) \rangle.$$

Here ∇ denotes the usual Euclidean gradient and "$\cdot$" the Euclidean scalar product. In "space direction" we define difference operators D^N on $\mathcal{F}_0$ by

$$D^N f(\mu, x, y) = N\Big(f\big(\mu + \tfrac{1}{N}(\delta_y - \delta_x)\big) - f(\mu)\Big). \tag{1.8}$$

Then, as $N \to \infty$,

$$D^N f(\mu, x, y) \longrightarrow Df(\mu, x, y) := (\nabla \varphi)\big(\langle \mu, \vec{\psi}(t) \rangle\big) \cdot (\vec{\psi}(y) - \vec{\psi}(x)). \tag{1.9}$$

To any kernel $k(x, \mu, \xi, dy)$ on $\mathbb{E} \times \mathcal{M} \times \mathcal{M}^{II} \times \mathcal{B}_{\mathbb{E}}$ and $\xi \in \mathcal{M}^{II}$ there correspond linear operators $\mathcal{G}_k^N(\xi)$ $(N \in \mathbb{N})$ and $\mathcal{G}_k(\xi)$ on $\mathcal{F}_0$ given by

$$\mathcal{G}_k^N(\xi) f(\mu) = \int \int D^N f(\mu, x, y)\, k(x, \mu, \xi, dy)\, \mu(dx)$$

and

$$\mathcal{G}_k(\xi) f(\mu) = \int \int Df(\mu, x, y)\, k(x, \mu, \xi, dy)\, \mu(dx)$$

With the above notations, for every $f \in \mathcal{F}$,

$$\mathcal{A}^{N,M} \ll \cdot, f \gg\, = \,\ll \cdot, \mathcal{G}_m^N(\cdot) f \gg, \tag{1.10}$$

where m is the kernel defined in (1.6).

2. A large deviation upper bound. Suppose that, for all N and M, we are given some deterministic $\nu^{N,M} \in \mathcal{M}^{II}$ having the form of a two-level empirical distribution as in (1.2). With $\mathcal{P}^{N,M}$ we then denote the law of the solution of the martingale problem associated with the generator $\mathcal{A}^{N,M}$ of (1.4) and starting point $\nu^{N,M}$. Then $\mathcal{P}^{N,M}$ is Borel probability measure on the space $D(I, \mathcal{M}^{II})$ of all càdlàg functions endowed with the Skorohod topology. This topology is independent of an a priori choice of metric on $\mathcal{M}^{II}$ see Jakubowski [Ja].

The law of large numbers of the system as N, $M \to \infty$ has been studied by Dawson [D1]. There the following result has been stated.

Proposition 1.1. *Suppose $\nu^{M,N}$ converges weakly to some $\nu \in \mathcal{M}^{II}$. Then $\mathcal{P}^{M,N}$ converges weakly on $D(I, \mathcal{M}^{II})$ to δ_η, where η is the unique deterministic path in $C(I, \mathcal{M}^{II})$ satisfying $\eta_0 = \nu$ and solving the weak McKean-Vlasov equation*

$$\begin{aligned} \ll \eta_t, f_t \gg = \ll \eta_0, f_0 \gg &+ \int_0^t \ll \eta_s, \partial_s f_s \gg \, ds \\ &+ \int_0^t \ll \eta_s, \mathcal{G}_m(\eta_s) f_s \gg \, ds \qquad (f \in \mathcal{F}, 0 \le t \le T). \end{aligned}$$

Let $\mathcal{H}^{M,N}$ be the *Hamiltonian* associated to the generator $\mathcal{A}^{M,N}$ defined, for $\zeta \in \mathcal{M}^{II}$ and test functions $f \in \mathcal{F}$, as

$$\begin{aligned} \mathcal{H}^{M,N}(\zeta, f) &= e^{-\ll \zeta, f \gg} \mathcal{A}^{M,N} e^{\ll \zeta, f \gg} \\ &= MN \int \overline{m}(\zeta, d\mu, dx, dy) \Big\{ \exp\Big[\frac{1}{MN} D^N f(\mu, x, y)\Big] - 1 \Big\}. \end{aligned} \tag{2.1}$$

Define the *scaled Hamiltonian* $\mathcal{H}$ by

$$\begin{aligned} \mathcal{H}(\zeta, f) &= \lim_{M,N\to\infty} \frac{1}{MN} \mathcal{H}^{M,N}(\zeta, MNf) \\ &= \int \overline{m}(\zeta, d\mu, dx, dy) \Big\{ \exp[Df(\mu, x, y)] - 1 \Big\}. \end{aligned}$$

We introduce, for $0 \le t \le T$, the functionals J_t acting on $f \in \mathcal{F}$ and $\eta \in D(I, \mathcal{M}^{II})$:

$$J_t(\eta, f) = \ll \eta_t, f_t \gg - \ll \eta_0, f_0 \gg - \int_0^t \ll \eta_r, \partial_r f_r \gg - \mathcal{H}(\eta_r, f_r) \, dr. \tag{2.2}$$

Then, for any $\nu \in \mathcal{M}^{II}$ and $\eta \in D(I, \mathcal{M}^{II})$, let

$$S_\nu(\eta) = \begin{cases} \sup \Big\{ J_T(\eta, f) \;\Big|\; f \in \mathcal{F} \Big\} & \text{if } \eta_0 = \nu, \\ \infty & \text{otherwise.} \end{cases}$$

The functional S_ν is in fact the rate function governing the large deviation upper bound of our particle system:

Theorem I. *Suppose that Assumptions I and II are satisfied and that the starting points $\nu^{M,N}$ converge weakly to some $\nu \in \mathcal{M}^{II}$ as $N,\ M \to \infty$.*

Then

$$\limsup_{M,N\to\infty} \frac{1}{MN} \log \mathcal{P}^{M,N}(F) \le - \inf_{\eta \in F} S_\nu(\eta),$$

for each closed subset F of $D(I, \mathcal{M}^{II})$.

Our next goal is to derive additional representation formulae for the rate function S_ν.

3. Representations of the rate function and regularity of paths. We will now define a subset H_ν of admissible paths in the space $C(I,\mathcal{M}^{II})$ of continuous $\mathcal{M}^{II}$-valued mappings.

Definition I. *Fix $\nu \in \mathcal{M}^{II}$. Let H_ν denote the set of all paths $\eta \in C(I,\mathcal{M}^{II})$ such that $\eta_0 = \nu$ and which satisfy a weak McKean-Vlasov equation*

$$\ll \eta_t, f_t \gg = \ll \eta_0, f_0 \gg + \int_0^t \ll \eta_s, \partial_s f_s \gg \, ds \tag{2.3}$$
$$+ \int_0^t \ll \eta_s, \mathcal{G}_{k_s}(\eta_s) f_s \gg \, ds \qquad (f \in \mathcal{F}, 0 \le t \le T),$$

for kernels $k_s(x,\mu,\eta,dy)$ $(s \in I)$ which are absolutely continuous with respect to $m(x,\mu,\eta,dy)$ and who possess a Radon-Nikodym derivative

$$g_s(\xi,\mu,x,y) := \frac{k_s(\xi,\mu,x,dy)}{m(\xi,\mu,x,dy)}$$

which satisfies

$$\int_0^T dt \int d\overline{m}(\eta_t)\Big(g_t \log g_t - g_t + 1\Big) < \infty. \tag{2.4}$$

The next result gives a non-variational representation formula for S_ν. It relies essentially on work by Léonard [L1].

Theorem II. *Fix $\nu \in \mathcal{M}^{II}$ and $\eta \in D(I,\mathcal{M}^{II})$. Then*

$$S_\nu(\eta) < \infty \text{ if and only if } \eta \in H_\nu.$$

And in the latter case the following representation holds

$$S_\nu(\eta) = \int_0^T dt \int d\overline{m}(\eta_t)\Big(g_t \log g_t - g_t + 1\Big). \tag{2.5}$$

Remark. The fact that S_ν is concentrated on continuous paths implies that the large deviation upper bound of Theorem I extends even to subsets F that are closed with respect to the stronger *uniform topology* on $D(I,\mathcal{M}^{II})$ — provided that F is measurable with respect to all $\mathcal{P}^{N,M}$. The argument is the same as in Section 18 in Bilingsley [Bi].

It turns out that paths $\eta \in H_\nu$ enjoy certain differentiability properties. To this end, endow the space $\mathcal{F}_0$ with the norm

$$\|f\|_{\mathcal{F}_0} := \sup_{\mu} |f(\mu)| + \sup_{\mu,x,y} |Df(\mu,x,y)| \qquad (f \in \mathcal{F}_0).$$

The set of all finite signed measures on $\mathcal{M}$ then can be regarded as a subspace of the topological dual $\mathcal{F}_0'$ of $\mathcal{F}_0$ by taking the integral of a function in $\mathcal{F}_0$ against a measure on $\mathcal{M}$. We extend the notation $\ll \vartheta, f \gg$ to denote the duality relation between $\vartheta \in \mathcal{F}_0'$ and $f \in \mathcal{F}_0$. For $\vartheta \in \mathcal{F}_0'$ and $\zeta \in \mathcal{M}^{II}$, we denote with $\mathcal{L}(\vartheta,\zeta)$ the Legendre transform of $\mathcal{H}(\zeta,\cdot)$:

$$\mathcal{L}(\vartheta,\zeta) = \sup_{f \in \mathcal{F}_0} \big(\ll \vartheta, f \gg - \mathcal{H}(\zeta,f)\big).$$

Theorem III. *If $\eta \in H_\nu$,then the mapping $t \longmapsto \eta_t \in \mathcal{F}_0'$ is strongly differentiable Lebesgue-almost everywhere with derivative $\dot{\eta}_t$. Moreover, η can be represented as*

$$\eta_t = \nu + \int_0^t \dot{\eta}_s \, ds \qquad (0 \leq t \leq T),$$

where the integral is to be taken in the sense of Bochner. Moreover, the following Lagrangian representation of the rate function holds

$$S_\nu(\eta) = \int_0^T \mathcal{L}(\dot{\eta}_t, \eta_t) \, dt. \tag{2.6}$$

Remark. The same method we use here to prove Theorem III can be applied to the weak McKean-Vlasov equations relating to the one-level large deviations as considered in Léonard [L2] or Djehiche and Kaj [DK]. There, the operators corresponding to $\mathcal{G}_k$ are *bounded* linear operators on the space $C_b(\mathbb{E})$. This guarantees that the derivatives of the one-level paths exist even as finite signed measures and with respect to the total variation norm. Let us briefly give an example for this loss of regularity in the two-level case. If we assume that there is *no interaction between the colonies*, i.e. γ does not depend on ξ, Theorem 2.2 of Dawson and Gärtner [DG1] together with the main result in Djehiche and Kaj [DK] yield a large deviation principle with rate function

$$\overline{S}_\nu(\eta) = \inf \Big\{ \int R \, dQ \; \Big| \; Q \in \mathcal{M}\big(D(I, \mathcal{M})\big) \text{ with marginal flow } (\eta_t) \Big\}.$$

Here R denotes the one-level rate function of Djehiche and Kaj [DK]. Consequently, if $\mu(\cdot) \in D(I, \mathcal{M})$ is such that $R(\mu) < \infty$, the two-level path given by $\eta(t) = \delta_{\mu(t)}$, $(t \in I)$ satisfies $S_{\eta(0)}(\eta) \leq \overline{S}_{\eta(0)}(\eta) < \infty$. This simple example shows that one cannot expect much further regularity of the paths in H_ν than stated in Theorem II.

References

[Bi] P. Billingsley, Convergence of Probability Measures, John Wiley & Sons, Inc., New York, 1968.

[D1] D. A. Dawson, *Asymptotic analysis of multilevel stochastic systems*, Stochastic differ. systems (ed. M. Métivier and E. Pardoux). Lecture Notes in Control and Information Sciences, vol. 69, 1984, pp. 76-89.

[D2] D. A. Dawson, *Stochastic ensembles and hierarchies*, Stochastic Processes and their Applications (ed. K. Itô and T. Hida). Lecture Notes in Math., vol. 1203, 1985, pp. 20-37.

[DG1] D. A. Dawson, J. Gärtner, *Multilevel large deviations and interacting diffusions*, Probab. Theory Relat. Fields **98** (1994), 432-487.

[DG2] D. A. Dawson, J. Gärtner, *Large deviations from the McKean-Vlasov limit for weakly interacting diffusions*, Stochastics **20** (1987), 247-308.

[DK] B. Djehiche, I. Kaj, *The rate function for some measure-valued jump processes*, Annals of Probability **23** (1995), 1414-1438.

[DS] B. Djehiche, A. Schied, *Large deviations for hierachical systems of interacting jump processes. Revised and extended version.*, Preprint (1997).

[EK] S. Ethier, T. Kurtz, Markov Processes, Characterization and Convergence, Wiley-Interscience, New York, 1986.

[L1] C. Léonard, *On the large deviations for particle systems associated with spatially homogeneous Boltzmann type equations*, Probab. Theory Relat. Fields **101** (1995), 1-44.

[L2] C. Léonard, *Large deviations for long range interacting particle systems with jumps*, Ann. Inst. Henri Poincaré **B31** (1995), 289-323.

[Pa] K. R. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, New York, 1967.

Boualem Djehiche
Division of Mathematical Statistics
The Royal Institute of Technology
S-100 44 Stockholm, Sweden
boualem@math.kth.se

Alexander Schied
Institut für Mathematik
Humboldt-Universität
Unter den Linden 6
D-10099 Berlin, Germany
schied@mathematik.hu-berlin.de

LARGE AND MODERATE DEVIATIONS OF PRODUCTS OF EMPIRICAL MEASURES AND U-EMPIRICAL MEASURES IN STRONG TOPOLOGIES*

PETER EICHELSBACHER AND UWE SCHMOCK

ABSTRACT. We present large and moderate deviation principles (LDP, MDP) for m-fold products of empirical measures and for U-empirical measures, where the underlying i.i.d. random variables take values in a measurable (not necessarily Polish) space $(S, \mathcal{S})$. The LDP can be formulated on a suitable subset $\mathcal{M}_1^\Phi(S^m)$ of all probability measures on $(S^m, \mathcal{S}^{\otimes m})$, for the MDP we consider a corresponding subset $\mathcal{M}^\Phi(S^m)$ of all signed measures on $(S^m, \mathcal{S}^{\otimes m})$. We endow both spaces with topologies, which are stronger than the usual τ-topology and which make the map $\nu \mapsto \int_{S^m} \varphi \, d\nu$ continuous even for certain unbounded φ taking values in a separable Banach space. A special feature of the LDP is the non-convexity of the rate function for $m \geq 2$.

1. LARGE DEVIATION PRINCIPLES

Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables taking values in a measurable space $(S, \mathcal{S})$. For simplicity we assume that they are defined on the product space $(\Omega, \mathcal{A}, \mathbb{P}) = (S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}})$, where μ is the distribution of X_1. For every $n \in \mathbb{N}$ the empirical measure is defined by $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. The large deviations of $\{L_n\}_{n\in\mathbb{N}}$ have been studied in several papers. Sanov [15] considered the problem when $S = \mathbb{R}$ and the space of probability measures $\mathcal{M}_1(S)$ is endowed with the weak topology. Donsker and Varadhan [10] and Bahadur and Zabell [4] (see also Azencott [3]) proved – with different approaches – a large deviation principle for $\{L_n\}_{n\in\mathbb{N}}$ when S is a Polish space with Borel σ-algebra $\mathcal{S}$, again in the weak topology of $\mathcal{M}_1(S)$. Groeneboom, Oosterhoff and Ruymgaart [13] obtained a stronger result in which S is a Hausdorff topological space with Borel σ-algebra $\mathcal{S}$ and $\mathcal{M}_1(S)$ is endowed with the $\tau_1(\mathbb{R})$-topology, that is, the coarsest topology which makes the maps $\mathcal{M}_1(S) \ni \nu \mapsto \int_S \varphi \, d\nu$ continuous for all φ in the space $B(S, \mathbb{R})$ of bounded, real-valued, $\mathcal{S}$-measurable functions on S. De Acosta [1] proved the large deviation principle for $\{L_n\}_{n\in\mathbb{N}}$ in the $\tau_1(\mathbb{R})$-topology setting when $(S, \mathcal{S})$ is a measurable space.

An extension in another direction was obtained by Wu [16, Section 2.4.4 and Section 4.1]. He considers a Polish space S with Borel σ-algebra $\mathcal{S}$ and a collection Φ consisting of all $\mathcal{S}$-measurable functions $f\colon S \to \mathbb{R}$, which are dominated in absolute value by some multiple of a single measurable reference function $\psi\colon S \to [1, \infty]$ with $\int_S \exp(\alpha\psi) \, d\mu < \infty$ for all $\alpha > 0$. On a suitable subset of $\mathcal{M}_1(S)$ he defines a

1991 *Mathematics Subject Classification.* 60F10 (primary); 62H10 and 28A35 (secondary).

Key words and phrases. Large deviations, moderate deviations, empirical measures, Sanov's theorem, strong topology, U-statistics, von Mises statistics.

*Research partially supported by the Swiss National Foundation, Contract No. 21-298333.90.

topology such that $\nu \mapsto \int_S \varphi \, d\nu$ is continuous for every $\varphi \in \Phi$, and proves a large deviation principle for $\{L_n\}_{n\in\mathbb{N}}$ with respect to this topology.

In this section we present large deviations results which include de Acosta's and Wu's results for general measurable spaces (not just Polish spaces), and extend them in several directions: Instead of real valued functions we consider functions with values in a separable real Banach space; we consider more general collections Φ, which satisfy appropriate exponential moment conditions; and – most important in our opinion – we consider products of empirical measures and U-empirical measures. These extensions enable us to derive large deviation principles for Banach-space valued U- and V-statistics as a particular application. These results improve the large deviation principle for $\mathbb{R}^d$-valued U- and V-statistics obtained in [11], because we use weaker moment conditions and make no additional assumptions on the state space S. As a further application of our extension to U-empirical measures, we can improve the Gibbs conditioning principle for interacting ensembles of particles, see [12] for details.

To formulate our results, we need some additional notations. Let $(E, \|\cdot\|_E)$ be a separable real Banach space with Borel σ-algebra $\mathcal{E}$. We always exclude the case $E = \{0\}$. Given $m \in \mathbb{N}$ and a measurable space $(S, \mathcal{S})$ as before, we consider the set $\mathcal{M}_1(S^m)$ of probability measures on the product space S^m, equipped with the product σ-algebra $\mathcal{S}^{\otimes m}$. Let Φ be a collection of $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable functions $\varphi\colon S^m \to E$ containing the set $B(S^m, E)$ of all bounded measurable ones. Define the Φ-restricted set of probability measures on S^m by

$$\mathcal{M}_1^{\Phi}(S^m) = \left\{ \nu \in \mathcal{M}_1(S^m) \;\middle|\; \int_{S^m} \|\varphi\|_E \, d\nu < \infty \text{ for every } \varphi \in \Phi \right\}.$$

Then the Bochner integral $\int_{S^m} \varphi \, d\nu$ is defined for every $\varphi \in \Phi$ and $\nu \in \mathcal{M}_1^{\Phi}(S^m)$. Let $\tau_1^{\Phi}(E)$ denote the coarsest topology on $\mathcal{M}_1^{\Phi}(S^m)$ such that the map $\mathcal{M}_1^{\Phi}(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ is continuous for every $\varphi \in \Phi$. If $\Phi = B(S^m, E)$, then $\mathcal{M}_1^{\Phi}(S^m) = \mathcal{M}_1(S^m)$ and we write $\tau_1(E)$ instead of $\tau_1^{\Phi}(E)$. If in addition $E = \mathbb{R}$, then the topology $\tau_1^{\Phi}(E)$ coincides with the usual $\tau_1(\mathbb{R})$-topology introduced above. The σ-algebra on $\mathcal{M}_1(S^m)$ is defined to be the smallest one such that $\mathcal{M}_1^{\Phi}(S^m)$ and all the maps $\mathcal{M}_1(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ with $\varphi \in B(S^m, E)$ are measurable.

We consider the following three exponential moment conditions for Φ:

Condition 1.1 (Weak Cramér condition). *For every function $\varphi \in \Phi$ there exists at least one $\alpha_\varphi > 0$ such that $\int_{S^m} \exp(\alpha_\varphi \|\varphi\|_E) \, d\mu^{\otimes m} < \infty$.*

Condition 1.2 (Strong Cramér condition). *For every $\varphi \in \Phi$ and every $\alpha > 0$, $\int_{S^m} \exp(\alpha \|\varphi\|_E) \, d\mu^{\otimes m} < \infty$.*

Condition 1.3. *For every $\varphi \in \Phi$ there exists at least one $\alpha_\varphi > 0$ such that*

$$\int_{S^m} \exp(\alpha_\varphi \|\varphi \circ \pi_\tau\|_E) \, d\mu^{\otimes m} < \infty$$

for every map $\tau\colon \{1, \ldots, m\} \to \{1, \ldots, m\}$, where $\pi_\tau\colon S^m \to S^m$ is defined by $\pi_\tau(s) = (s_{\tau(1)}, \ldots, s_{\tau(m)})$ for every $s = (s_1, \ldots, s_m) \in S^m$.

Note that all three conditions are satisfied in the case $\Phi = B(S^m, E)$. Conditions 1.2 and 1.3 are needed to prove a full large deviation principle in the $\tau_1^{\Phi}(E)$-topology for the m-fold products $L_n^{\otimes m}\colon \Omega \to \mathcal{M}_1(S^m)$ of the empirical measures, expressible as

$$L_n^{\otimes m} = \frac{1}{n^m} \sum_{i_1,\ldots,i_m=1}^{n} \delta_{(X_{i_1},\ldots,X_{i_m})}$$

for $n \in \mathbb{N}$. For applications to U-statistics, it is also of interest to prove a large deviation principle for the measures $L_n^m \colon \Omega \to \mathcal{M}_1(S^m)$ with $n \geq m$, which are defined by

$$L_n^m = \frac{1}{n_{(m)}} \sum_{(i_1,\ldots,i_m)\in I(m,n)} \delta_{(X_{i_1},\ldots,X_{i_m})}, \tag{1.4}$$

where $n_{(m)} \equiv \prod_{k=0}^{m-1}(n-k)$ and where $I(m,n) \subset \{1,\ldots,n\}^m$ contains all m-tuples with pairwise different components. Due to their immediate statistical application, we call the measures $\{L_n^m\}_{n\geq m}$ the U-empirical measures of order m. These U-empirical measures seem to be the proper generalization of the usual empirical measures $\{L_n\}_{n\in\mathbb{N}}$ because there is no dependence within the individual terms of (1.4). Using the U-empirical measures, it is possible to model a weak but long-range interaction between the i. i. d. random variables $\{X_i\}_{i\in\mathbb{N}}$. This also means that we leave the realm of independence for $m \geq 2$. Our large deviations result for the U-empirical measures only requires Condition 1.2.

Note that the set $\mathcal{M}_1^\Phi(S^m)$ is convex and contains the Dirac measure δ_x for every $x \in S^m$, hence $L_n^{\otimes m}$ and L_n^m take values in $\mathcal{M}_1^\Phi(S^m)$ and both mappings turn out to be measurable. Let us recall the definition of the relative entropy $H_m(\nu\,|\,\tilde\nu)$ of $\nu \in \mathcal{M}_1(S^m)$ with respect to $\tilde\nu \in \mathcal{M}_1(S^m)$:

$$H_m(\nu\,|\,\tilde\nu) = \begin{cases} \int_{S^m} f \log f \, d\tilde\nu, & \text{if } \nu \ll \tilde\nu \text{ and } f = \frac{d\nu}{d\tilde\nu}, \\ \infty & \text{otherwise.} \end{cases}$$

The rate function $J_m \colon \mathcal{M}_1(S^m) \to [0,\infty]$ is defined by

$$J_m(\nu) = \begin{cases} \frac{1}{m} H_m(\nu\,|\,\mu^{\otimes m}), & \text{if } \nu = \nu_1^{\otimes m}, \\ \infty & \text{otherwise,} \end{cases}$$

where ν_1 denotes the first marginal of ν. For every $B \subset \mathcal{M}_1(S^m)$ define $J_m(B) = \inf_{\nu\in B} J_m(\nu)$. Note that, if there exists $A \in \mathcal{S}$ satisfying $0 < \mu(A) < 1$, then the rate function J_m is non-convex for all $m \geq 2$. To see this, define the conditional probability measures $\nu = \mu(\,\cdot\,|A)$ and $\tilde\nu = \mu(\,\cdot\,|A^c)$ and show that $\frac{1}{2}(\nu^{\otimes m} + \tilde\nu^{\otimes m})$ is not a product measure. For every level $l \in [0,\infty)$ define the level set

$$K(J_m, l) = \{\, \nu \in \mathcal{M}_1(S^m) \mid J_m(\nu) \leq l \,\}.$$

Our main large deviations results in [12] are the following two theorems:

Theorem 1.5 (Large deviations of U-empirical measures of order m)**.**

(a) *For every measurable $B \subset \mathcal{M}_1(S^m)$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in B) \geq -J_m(\operatorname{int}_{\tau_1^\Phi(E)}(B)), \tag{1.6}$$

where $\operatorname{int}_{\tau_1^\Phi(E)}(B)$ denotes the interior of the set $B \cap \mathcal{M}_1^\Phi(S^m)$ with respect to the $\tau_1^\Phi(E)$-topology.

(b) *If Condition* 1.1 *holds, then $K(J_m,l) \subset \mathcal{M}_1^\Phi(S^m)$ for every level $l \in [0,\infty)$.*

(c) *If Condition* 1.2 *holds, then* $K(J_m, l)$ *is* $\tau_1^{\Phi}(E)$*-compact and sequentially* $\tau_1^{\Phi}(E)$*-compact for every* $l \in [0, \infty)$.

(d) *If Condition* 1.2 *holds, then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in B) \le -J_m(\mathrm{cl}_{\tau_1^{\Phi}(E)}(B)) \tag{1.7}$$

for every measurable $B \subset \mathcal{M}_1(S^m)$*, where* $\mathrm{cl}_{\tau_1^{\Phi}(E)}(B)$ *denotes the closure of the set* $B \cap \mathcal{M}_1^{\Phi}(S^m)$ *with respect to the* $\tau_1^{\Phi}(E)$*-topology.*

Theorem 1.8 (Large deviations of m-fold products of empirical measures)**.**

(a) *If Condition* 1.3 *holds, then* (1.6) *is true with* $L_n^{\otimes m}$ *in place of* L_n^m.

(b) *If Conditions* 1.2 *and* 1.3 *hold, then* (1.7) *is true with* $L_n^{\otimes m}$ *in place of* L_n^m.

Theorem 1.5, specialized to $m = 1$ and $E = \mathbb{R}$, already combines and extents the aforementioned large deviations results of de Acosta and Wu for the empirical measures $\{L_n\}_{n\in\mathbb{N}}$ to the $\tau_1^{\Phi}(\mathbb{R})$-topology. Note that the lower bound in Theorem 1.5(a) does not make use of Condition 1.1. But if Condition 1.1 does not hold, then the lower bound might loose its strength because measures $\nu \in B$ with $J_m(\nu) < \infty$ might not be contained in $\mathcal{M}_1^{\Phi}(S^m)$.

For the proofs of the lower bounds of the above theorems, we adopt the method outlined in [9, Exercise 3.2.23(ii)] and combine it with the law of large numbers for U-statistics and V-statistics, respectively. To prove the upper bounds, we basically use de Acosta's projective limit approach contained in [1] combined with a Banach-space version of Cramér's theorem, which is due to Donsker and Varadhan. (Choosing $m = 1$, $S = E$ and $\Phi = B(S, E) \cup \{\mathrm{id}_E\}$, their result can be recovered from Theorem 1.5 by using the contraction principle [8, Theorem 4.2.1] and identifying the rate function via [10, Theorem 5.2(iv)].) Of course, we have to cope with the unbounded, Banach-space valued functions in Φ, which is one reason for the $\tau_1^{\Phi}(E)$-topology to be finer than the usual $\tau_1(\mathbb{R})$-topology.

There are two non-trivial problems in transferring de Acosta's approach to the case $m \ge 2$: The dependence of the different terms in (1.4) and the non-convexity of the rate function J_m. The first problem is treated by using the existence of almost regular partitions of complete hypergraphs, see Baranyai [5]. This result allows us to conveniently decompose the sum in (1.4) into an n-dependent number of partial sums, each of which consists of independent terms. To circumvent the non-convexity of J_m, we use the convexity of the relative entropy $H_m(\,\cdot\,|\,\mu^{\otimes m})$ on $\mathcal{M}_1(S^m)$ and the fact that $L_n^{\otimes m}$ is a product measure as well as a good approximation for L_n^m. Finally we have to show that the infimum of $H_m(\,\cdot\,|\,\mu^{\otimes m})$ over all "η-approximate product measures" of a $\tau_1^{\Phi}(E)$-closed set C tends to the infimum over all product measures in C as $\eta \downarrow 0$.

2. Moderate deviation principles

As before, let $(S, \mathcal{S})$ be a measurable space and $m \in \mathbb{N}$. Let $\mathcal{M}(S^m)$ denote the set of all signed measures on $(S^m, \mathcal{S}^{\otimes m})$ with finite total variation. Given a sequence $\{b_n\}_{n\in\mathbb{N}} \subset (0, \infty)$ satisfying $\lim_{n\to\infty} b_n/n = 0$ and $\lim_{n\to\infty} n/b_n^2 = 0$, we define the moderate U-emprirical measure $M_n^m\colon \Omega \to \mathcal{M}(S^m)$ by

$$M_n^m = \frac{n}{b_n}(L_n^m - \mu^{\otimes m})$$

for every $n \geq m$. If $m = 1$, then we also write M_n instead of M_n^1. Similarly to the previous section, let $(E, \|\cdot\|_E)$ be a separable real Banach space with Borel σ-algebra $\mathcal{E}$, and let Φ be a set of $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable and Bochner $\mu^{\otimes m}$-integrable functions $\varphi\colon S^m \to E$. We always assume that $B(S^m, E) \subset \Phi$. We define the Φ-restricted set of signed measures by

$$\mathcal{M}^\Phi(S^m) = \Big\{ \nu \in \mathcal{M}(S^m) \;\Big|\; \int_{S^m} \|\varphi\|_E \, d|\nu| < \infty \text{ for every } \varphi \in \Phi \Big\},$$

where $|\nu|$ denotes the total variation measure corresponding to ν. Let $\tau^\Phi(E)$ be the coarsest topology on $\mathcal{M}^\Phi(S^m)$ such that $\mathcal{M}^\Phi(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ is continuous for every $\varphi \in \Phi$. If $\Phi = B(S^m, E)$, then we write $\tau(E)$ instead of $\tau^\Phi(E)$.

The large deviations of $\{M_n\}_{n\in\mathbb{N}}$ on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ have been studied by Borovkov and Mogulskii [6, Section 3] in the $\tau(\mathbb{R})$-topology on the space $\mathcal{M}(S)$ when S is a Hausdorff topological space with Borel σ-algebra $\mathcal{S}$. They considered special subsets of $\mathcal{M}(S)$ for the lower and the upper bound. De Acosta [2, Section 3] generalized this result to a full large deviation principle on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ when $(S, \mathcal{S})$ is a general measurable space. In [12] we prove large deviations of the moderate U-empirical measures $\{M_n^m\}_{n\geq m}$ in stronger topologies, again generated by a collection Φ of functions $\varphi\colon S^m \to E$ satisfying appropriate moments conditions. We obtain a moderate deviation principle for Banach-space valued U-statistics as a particular application.

The σ-algebra on $\mathcal{M}(S^m)$ is defined to be the smallest one containing $\mathcal{M}^\Phi(S^m)$ such that all the maps $\mathcal{M}(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ with $\varphi \in B(S^m, E)$ are measurable. Since $\Phi \subset L_1(\mu^{\otimes m}, E)$ by assumption, the measure-valued maps $\{M_n^m\}_{n\geq m}$ actually take values in $\mathcal{M}^\Phi(S^m)$ and are therefore measurable.

For the moderate deviations upper bound in Theorem 2.3(d) below, we have to formulate an additional growth restriction. This in turn requires – at least for the unbounded $\varphi \in \Phi$ – a decomposition, which for symmetric φ is closely related to the Hoeffding decomposition [7, (1.1.9)] for the corresponding U-statistic. (Although it might seem tempting, we do not want to consider collections Φ consisting only of symmetric functions, because for $\mathcal{S} \neq \{\varnothing, S\}$ and $m \geq 2$, we would not be able to separate the zero measure from the measure $\delta_x^{\otimes m-1} \otimes \delta_y - \delta_y \otimes \delta_x^{\otimes m-1}$ with $x \in A \in \mathcal{S}$ and $y \in S \setminus A$ for example; hence the topology $\tau^\Phi(E)$ would loose the Hausdorff property.) Given $\varphi \in L_1(\mu^{\otimes m}, E)$ and a nonempty subset A of $\{1, \ldots, m\}$, define $\varphi_A \in L_1(\mu^{\otimes |A|}, E)$ by μ-integrating $\varphi(s_1, \ldots, s_m)$ with respect to every s_i with $i \in \{1, \ldots, m\} \setminus A$. By convention, $\varphi_\varnothing \equiv \int_{S^m} \varphi \, d\mu^{\otimes m} \in E$. Furthermore, define $\tilde\varphi_A \in L_1(\mu^{\otimes |A|}, E)$ by

$$\tilde\varphi_A\big((s_i)_{i\in A}\big) = \sum_{B\subset A} (-1)^{|A\setminus B|} \varphi_B\big((s_i)_{i\in B}\big),$$

for every nonempty $A \subset \{1, \ldots, m\}$, and let $\tilde\varphi_\varnothing \equiv \varphi_\varnothing$. According to the inclusion–exclusion principle or the Möbius inversion formula,

$$\varphi(s_1, \ldots, s_m) = \sum_{A\subset\{1,\ldots,m\}} \tilde\varphi_A\big((s_i)_{i\in A}\big)$$

for $\mu^{\otimes m}$-almost all $(s_1, \ldots, s_m) \in S^m$. Hence, for every $n \geq m$,

$$\int_{S^m} \varphi \, dL_n^m = \tilde\varphi_0 + \sum_{a=1}^{m} \int_{S^a} \tilde\varphi_a \, dL_n^a \qquad \mathbb{P}\text{-almost surely,} \tag{2.1}$$

where, for every $a \in \{0,1,\dots,m\}$,

$$\tilde\varphi_a \equiv \sum_{\substack{A\subset\{1,\dots,m\}\\ |A|=a}} \tilde\varphi_A.$$

Note that every $\tilde\varphi_A$ with nonempty $A \subset \{1,\dots,m\}$ is completely μ-degenerate in the sense that $\mu^{\otimes|A|-1}$-almost surely $\int_S \tilde\varphi_A((s_i)_{i\in A})\,\mu(ds_j) = 0$ for every $j \in A$.

We can now state the additional growth restriction on $\varphi \in \Phi$ in terms of the norms of the corresponding completely μ-degenerate functions $\tilde\varphi_2,\dots,\tilde\varphi_m$:

Condition 2.2. *For every $\varphi \in \Phi$ and every $a \in \{2,\dots,m\}$,*

$$\biggl(\int_{S^a} \|\tilde\varphi_a\|_E^{b_n^2/n}\,d\mu^{\otimes a}\biggr)^{n/b_n^2} = o\biggl(\Bigl(\frac{n}{b_n}\Bigr)^{a-1}\biggr) \qquad \text{as } n\to\infty.$$

This condition makes sure that the terms with $a \geq 2$ in (2.1) have no influence on the scale of moderate deviations. If $\Phi = B(S^m,E)$, then all corresponding completely μ-degenerate functions are bounded and Condition 2.2 certainly holds.

We define the moderate-deviations rate function $I_m\colon \mathcal{M}(S^m) \to [0,\infty]$ by

$$I_m(\nu) = \frac12 \int_S \Bigl(\frac{d\nu_1}{d\mu}\Bigr)^2 d\mu$$

if the first marginal $\nu_1 \in \mathcal{M}(S)$ of ν satisfies $\nu_1(S) = 0$ and $\nu_1 \ll \mu$ such that $\nu = \sum_{i=1}^m \mu^{\otimes i-1}\otimes\nu_1\otimes\mu^{\otimes m-i}$; and we define $I_m(\nu)=\infty$ otherwise. Using the the convexity of $\mathbb{R} \ni x \mapsto x^2$, it follows that I_m is convex. For every $l \in [0,\infty)$ we define the corresponding level set by $K(I_m,l) = \{\,\nu\in\mathcal{M}(S^m) \mid I_m(\nu)\le l\,\}$.

Theorem 2.3 (Moderate deviations of U-empirical measures)**.**

(a) *If $\Phi \subset L_2(\mu^{\otimes m},E)$, then $K(I_m,l) \subset \mathcal{M}^\Phi(S^m)$ for every $l\in[0,\infty)$.*

(b) *If $\Phi \subset L_2(\mu^{\otimes m},E)$, then the level set $K(I_m,l)$ is $\tau^\Phi(E)$-compact and sequentially $\tau^\Phi(E)$-compact for every $l\in[0,\infty)$.*

(c) *Assume that there exists a $p\in(1,2]$ such that the Banach space $(E,\|\cdot\|_E)$ is of type p and $\lim_{n\to\infty} n/b_n^p = 0$. If $\Phi\subset L_p(\mu^{\otimes m},E)$ in the case $m=1$, or if $\Phi\subset L_2(\mu^{\otimes m},E)$ in the case $m\ge 2$, then*

$$\liminf_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}(M_n^m\in B) \ge -I_m(\operatorname{int}_{\tau^\Phi(E)}(B)) \tag{2.4}$$

for every measurable $B\subset\mathcal{M}(S^m)$, where $\operatorname{int}_{\tau^\Phi(E)}(B)$ denotes the interior of the set $B\cap\mathcal{M}^\Phi(S^m)$ with respect to the $\tau^\Phi(E)$-topology.

(d) *If Conditions* 1.2 *and* 2.2 *hold and if the Banach space E is of type* 2*, then*

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}(M_n^m\in B) \le -I_m(\operatorname{cl}_{\tau^\Phi(E)}(B))$$

for every measurable $B\subset\mathcal{M}(S^m)$, where $\operatorname{cl}_{\tau^\Phi(E)}(B)$ denotes the closure of the set $B\cap\mathcal{M}^\Phi(S^m)$ with respect to the $\tau^\Phi(E)$-topology.

The following example, derived from [14, Example 7.11], shows that Theorem 2.3(c) does not hold without an assumption on the type of the separable Banach space E, even in the case $\Phi = B(S^m,E)$ with $m=1$, $S=E$, and a set B determined by a bounded continuous function.

Example 2.5. Let $(c_0,\|\cdot\|_\infty)$ denote the separable Banach space of all real sequences tending to zero. Let $\{\xi_{i,k}\}_{i,k\in\mathbb{N}}$ be a collection of independent random variables with $\mathbb{P}(\xi_{i,k} = 1) = \mathbb{P}(\xi_{i,k} = -1) = 1/2$ for all $i, k \in \mathbb{N}$. Consider $\{b_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ such that $b_n/n \downarrow 0$ as $n \to \infty$. Define $\alpha_k = \sqrt{b_n/n}$ whenever $(n-1)2^{n-1} < k \le n2^n$. Then $X_i \equiv (\alpha_k \xi_{i,k})_{k\in\mathbb{N}}$ for $i \in \mathbb{N}$ gives an i.i.d. sequence of symmetric, c_0-valued random variables with $\|X_i\|_\infty \le \bar\alpha \equiv \sup_{k\in\mathbb{N}} \alpha_k$ for all $i \in \mathbb{N}$. Define $S = E = c_0$, let μ denote the distribution of X_1, and define $\varphi\colon S \to E$ by $\varphi(s) = s$ for $\|s\|_\infty \le \bar\alpha$ and $\varphi(s) = \bar\alpha s/\|s\|_\infty$ otherwise. Note that $\varphi \in C_{\mathrm{b}}(S,E) \subset \Phi \equiv B(S,E)$, hence $B \equiv \{\, \nu \in \mathcal{M}(S) \mid \|\int_S \varphi\, d\nu\|_\infty < 1 \,\}$ is $\tau(c_0)$-open. Furthermore, B contains the zero measure, which satisfies $I_1(0) = 0$.

To evaluate the left-hand side of (2.4), first note that

$$\{M_n \in B\} = \left\{ \Big\| \sum_{i=1}^n X_i \Big\|_\infty < b_n \right\} \subset \bigcap_{k=1}^{n2^n} \left\{ \Big| \sum_{i=1}^n \xi_{i,k} \Big| < \sqrt{b_n n} \right\},$$

where we used that $\alpha_k \ge \sqrt{b_n/n}$ for all $k \le n2^n$. Using $\mathbb{P}(\sum_{i=1}^n \xi_{i,k} = n) = 2^{-n}$, it follows for all $n \in \mathbb{N}$ with $b_n \le n$ that $\mathbb{P}(M_n \in B) \le (1 - 2^{-n})^{n2^n} \le e^{-n}$. Hence, the left-hand side of (2.4) equals minus infinity.

References

[1] A. de Acosta, *On large deviations of empirical measures in the τ-topology*, J. Appl. Probab. **31A** (1994), 41–47, (special volume).

[2] ———, *Projective systems in large deviation theory II: some applications*, Probability in Banach Spaces, 9 (J. Hoffman-Jørgensen, J. Kuelbs, and M. B. Marcus, eds.), Progress in Probability, vol. 35, Birkhäuser, Boston, 1994, pp. 241–250.

[3] R. Azencott, *Grandes Déviations et Applications*, Lectures Notes in Mathematics, vol. 774, Springer-Verlag, Berlin, 1980.

[4] R. R. Bahadur and S. L. Zabell, *Large deviations of the sample mean in general vector spaces*, Ann. Probab. **7** (1979), 587–621.

[5] Z. Baranyai, *On the factorisation of the complete uniform hypergraph*, Infinite and Finite Sets (A. Hajnal, T. Rado, and V. T. Sós, eds.), North-Holland, Amsterdam, 1975, pp. 91–108.

[6] A. A. Borovkov and A. A. Mogulskii, *Probabilities of large deviations in topological spaces II*, Siberian Math. J. **21** (1980), 653–664.

[7] Yu. V. Borovskikh, *U-Statistics in Banach Spaces*, VSP, Utrecht, 1996.

[8] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston, MA, 1993.

[9] J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Pure and Applied Mathematics, vol. 137, Academic Press, San Diego, 1989.

[10] M. D. Donsker and S. R. S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time III*, Comm. Pure Appl. Math. **29** (1976), 389–461.

[11] P. Eichelsbacher and M. Löwe, *Large deviation principle for m-variate von Mises-statistics and U-statistics*, J. Theoret. Probab. **8** (1995), 807–824.

[12] P. Eichelsbacher and U. Schmock, *Large and moderate deviations of products of empirical measures and U-empirical measures in strong topologies*, preprint, 1996.

[13] P. Groeneboom, J. Oosterhoff, and F. H. Ruymgaart, *Large deviation theorems for empirical probability measures*, Ann. Probab. **7** (1979), 553–586.

[14] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, Berlin, 1991.

[15] I. N. Sanov, *On the probability of large deviations of random variables*, Selected Transl. Math. Statist. and Prob. **I** (1961), 214–244.

[16] L. Wu, *Some general methods of large deviations and applications*, preprint, 1993.

(P. Eichelsbacher) Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany

E-mail address: peter@mathematik.uni-bielefeld.de

(U. Schmock) Departement Mathematik, ETH Zentrum, CH-8092 Zürich, Switzerland

E-mail address: schmock@math.ethz.ch