Electronic Journal of Probability

Vol. 14 (2009), Paper no. 40, pages 1162–1180.

Journal URL
http://www.math.washington.edu/~ejpecp/

Concentration inequalities for Markov processes via coupling

Jean-René Chazottes[a], Frank Redig[b]
[a] Centre de Physique Théorique, CNRS, École Polytechnique
91128 Palaiseau, France
[b] Mathematisch Instituut Universiteit Leiden
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

Abstract

We obtain moment and Gaussian bounds for general coordinate-wise Lipschitz functions evaluated along the sample path of a Markov chain. We treat Markov chains on general (possibly unbounded) state spaces via a coupling method. If the first moment of the coupling time exists, then we obtain a variance inequality. If a moment of order $1+\epsilon$ of the coupling time exists, then depending on the behavior of the stationary distribution, we obtain higher moment bounds. This immediately implies polynomial concentration inequalities. In the case that a moment of order $1+\epsilon$ is finite uniformly in the starting point of the coupling, we obtain a Gaussian bound. We illustrate the general results with house of cards processes, in which both uniform and non-uniform behavior of moments of the coupling time can occur.

Key words: Gaussian bound, moment bounds, house of cards process, Hamming distance.

AMS 2000 Subject Classification: Primary 60E15.

Submitted to EJP on October 9, 2008, final version accepted April 22, 2009.

1 Introduction

In this paper we consider a stationary Markov chain $X_n, n \in \mathbb{Z}$, and want to obtain inequalities for the probability that a function $f(X_1, \ldots, X_n)$ deviates from its expectation. In the spirit of concentration inequalities, one can try to bound the exponential moment of $f - \mathbb{E}(f)$ in terms of the sum of squares of the Lipschitz constants of f, as can be done in the case of independent random variables by several methods [17].

In the present paper, we want to continue the line of thought developed in [7; 8] where concentration inequalities are obtained via a combination of martingale difference approach (telescoping $f - \mathbb{E}(f)$) and coupling of conditional distributions. In the case of an unbounded state space, we cannot expect to find a coupling of which the tail of the distribution of the coupling time can be controlled uniformly in the starting points. This non-uniform dependence is thus rather the rule than the exception and has to be dealt with if one wants to go beyond the finite (or compact) state space situation. Moreover, if the state space is continuous, then in general two copies of the process cannot be coupled such that they eventually coincide: we expect rather that in a coupling the distance between the two copies can be controlled and becomes small when we go further in time. We show that a control of the distance suffices to obtain concentration inequalities. This leads to a "generalized coupling time" which in discrete settings coincides with the ordinary coupling time (in the case of a successful coupling).

In order to situate our results in the existing literature, we want to stress that the main message of this paper is the connection between the behavior of the generalized coupling time and concentration inequalities. In order to illustrate the possibly non-uniform behavior of the coupling time, we concentrate on the simplest possible example of "house of cards" processes (Markov chains on the natural numbers). In this paper we restrict to the Gaussian concentration inequality and moment inequalities. In principle, moment inequalities with controll on the constants can be "summarized" in the form of Orlicz-norm inequalities, but we do not want to deal with this here.

The case of Markov chains was first considered by Marton [20; 21] : for uniformly contracting Markov chains, in particular for ergodic Markov chains with finite state space, Gaussian concentration inequalities are obtained. The method developed in that paper is based on transportation cost-information inequalities. With the same technique, more general processes were considered by her in [22]. Later, Samson [25] obtained Gaussian concentration inequalities for some classes of Markov chains and Φ-mixing processes, by following Marton's approach. Let us also mention the work by Djellout *et al.* [9] for further results in that direction. Chatterjee [6] introduced a version of Stein's method of exchangeable pairs to prove Gaussian as well as moment concentration inequalities. Using martingale differences, Gaussian concentration inequalities were obtained in [15; 24] for some classes of mixing processes. Markov contraction was used in [16] for "Markov-type" processes (e.g. hidden Markov chains).

Related work to ours is [10], [11], [12] where deviation or concentration inequalities ([10]) and speed of convergence to the stationary measure ([11], [12]) are obtained for subgeometric Markov chains, using a technique of regeneration times and Lyapounov functions. Concentration properties of suprema of additive functionals of Markov chains are studied in [1], using a technique of regeneration times.The example of the house of cards process, and in particular its speed of relaxation to the stationary measure is studied in [11], section 3.1. The speed of relaxation to the stationary measure is of course related to the coupling time, see e.g. [23] for a nice recent account. In fact, using an explicit coupling, we obtain concentration inequalities in the different regimes of relaxation studied in [11].

Our paper is organized as follows. We start by defining the context and introduce the telescoping procedure, combined with coupling. Here the notion of coupling matrix is introduced. In terms of this matrix we can (pointwise) bound the individual terms in the telescopic sum for $f - \mathbb{E}(f)$. We then turn to the Markov case, where there is a further simplification in the coupling matrix due to the Markov property of the coupling. In Section 5 we prove a variance bound under the assumption that the first moment of the (generalized) coupling time exists. In section 6 we turn to moment inequalities. In this case we require that a moment of order $1+\epsilon$ of the (generalized) coupling time exists. This moment $M_{x,y,1+\epsilon}$ depends on the starting point of the coupling. The moment inequality for moments of order $2p$ will then be valid if (roughly speaking) the $2p$-th moment of $M_{x,y,1+\epsilon}$ exists. In Section 7 we prove that if a moment of order $1+\epsilon$ of the coupling is finite, uniformly in the starting point, then we have a Gaussian concentration bound.

Finally, Section 8 contains examples. In particular, we illustrate our approach in the context of so-called house of cards processes, in which both the situation of uniform case (Gaussian bound), as well as the non-uniform case (all moments or moments up to a certain order) are met. We end with application of our moment bounds to measure concentration of Hamming neighborhoods and get non-Gaussian measure concentration bounds.

2 Setting

2.1 The process

The state space of our process is denoted by E. It is supposed to be a metric space with distance d. Elements of E are denoted by x, y, z. E is going to serve as state space of a double sided stationary process. Realizations of this process are thus elements of $E^{\mathbb{Z}}$ and are denoted by $\overline{x}, \overline{y}, \overline{z}$.

We denote by $(X_n)_{n\in\mathbb{Z}}$ a (two-sided) stationary process with values in E. The joint distribution of $(X_n)_{n\in\mathbb{Z}}$ is denoted by $\mathbb{P}$, and $\mathbb{E}$ denotes corresponding expectation.

$\mathscr{F}_{-\infty}^{i}$ denotes the sigma-fields generated by $\{X_k : k \leq i\}$,

$$\mathscr{F}_{-\infty} := \bigcap_i \mathscr{F}_{-\infty}^{i}$$

denotes the tail sigma-field, and

$$\mathscr{F} = \sigma\left(\bigcup_{i=-\infty}^{\infty} \mathscr{F}_{-\infty}^{i}\right).$$

We assume in the whole of this paper that $\mathbb{P}$ is tail trivial, i.e., for all sets $A \in \mathscr{F}_{-\infty}$, $\mathbb{P}(A) \in \{0, 1\}$.

For $i < j, i, j \in \mathbb{Z}$, we denote by X_i^j the vector $(X_i, X_{i+1}, \ldots, X_j)$, and similarly we have the notation $X_{-\infty}^{i}$, X_i^{∞}. Elements of $E^{\{i,i+1,\ldots,j\}}$ (i.e., realizations of X_i^j) are denoted by x_i^j, and similarly we have $x_{-\infty}^{i}$, x_i^{∞}.

2.2 Conditional distributions, Lipschitz functions

We denote by $\mathbb{P}_{x_{-\infty}^{i}}$ the joint distribution of $\{X_j : j \geq i+1\}$ given $X_{-\infty}^{i} = x_{-\infty}^{i}$. We assume that this object is defined for all $x_{-\infty}^{i}$, i.e., that there exists a specification with which $\mathbb{P}$ is consistent. This is automatically satisfied in our setting, see [14].

Further, $\hat{\mathbb{P}}_{x^i_{-\infty},y^i_{-\infty}}$ denotes a coupling of $\mathbb{P}_{x^i_{-\infty}}$ and $\mathbb{P}_{y^i_{-\infty}}$.

For $f : E^{\mathbb{Z}} \to \mathbb{R}$, we define the i-th Lipschitz constant

$$\delta_i(f) := \sup\left\{\frac{f(\overline{x}) - f(\overline{y})}{d(x_i, y_i)} : x_j = y_j,\ \forall j \neq i,\ x_i \neq y_i\right\}.$$

The function f is said to be Lipschitz in the i-th coordinate if $\delta_i(f) < \infty$, and Lipschitz in all coordinates if $\delta_i(f) < \infty$ for all i. We use the notation $\delta(f) = (\delta_i(f))_{i\in\mathbb{Z}}$. We denote by $\mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R})$ the set of all real-valued functions on $E^{\mathbb{Z}}$ which are Lipschitz in all coordinates.

3 Telescoping and the coupling matrix

We start with $f \in \mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R}) \cap L^1(\mathbb{P})$, and begin with the classical telescoping (martingale-difference) identity

$$f - \mathbb{E}(f) = \sum_{i=-\infty}^{\infty} \Delta_i$$

where

$$\Delta_i := \mathbb{E}(f \,|\, \mathscr{F}^i_{-\infty}) - \mathbb{E}(f \,|\, \mathscr{F}^{i-1}_{-\infty}).$$

We then write, using the notation of Section 2.1,

$$\Delta_i = \Delta_i(X^i_{-\infty}) = \int d\mathbb{P}_{X^{i-1}_{-\infty}}(z_i) \times$$

$$\int d\hat{\mathbb{P}}_{X^i_{-\infty}, X^{i-1}_{-\infty} z_i}(y^\infty_{i+1}, z^\infty_{i+1}) \left[f(X^i_{-\infty} y^\infty_{i+1}) - f(X^{i-1}_{-\infty} z^\infty_i) \right]. \tag{1}$$

For $f \in \mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R})$, we have the following obvious telescopic inequality

$$|f(\overline{x}) - f(\overline{y})| \leq \sum_{i\in\mathbb{Z}} \delta_i(f) d(x_i, y_i). \tag{2}$$

Combining (1) and (2) one obtains

$$|\Delta_i(X^i_{-\infty})| \leq \sum_{j=0}^{\infty} D^{X^i_{-\infty}}_{i,i+j} \delta_{i+j}(f) \tag{3}$$

where

$$D^{X^i_{-\infty}}_{i,i+j} := \int d\mathbb{P}_{X^{i-1}_{-\infty}}(z_i) \int d\hat{\mathbb{P}}_{X^i_{-\infty}, X^{i-1}_{-\infty} z_i}(y^\infty_{i+1}, z^\infty_{i+1})\, d(y_{i+j}, z_{i+j}). \tag{4}$$

This is an upper-triangular random matrix which we call the coupling matrix associated with the process (X_n) and $\hat{\mathbb{P}}$, the coupling of the conditional distributions. As we obtained before in [7], in the context of E a finite set, the decay properties of the matrix elements $D^{X^i_{-\infty}}_{i,i+j}$ (i.e., how these matrix elements become small when j becomes large) determine the concentration properties of

$f \in \mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R})$, via the control (3) on Δ_i, together with Burkholder's inequality [5, Theorem 3.1, p. 87], which relates the moments of $f - \mathbb{E}(f)$ with powers of the sum of squares of Δ_i. The non-uniformity (as a function of the realization of $X^i_{-\infty}$) of the decay of the matrix elements as a function of j (which we encountered e.g. in the low-temperature Ising model [7]) will be *typical* as soon as the state space E is unbounded. Indeed, if starting points in the coupling are further away, then it takes more time to get the copies close in the coupling .

Remark 3.1. *The same telescoping procedure can be obtained for "coordinate-wise Hölder" functions, i.e., functions such that for some* $0 < \alpha < 1$

$$\delta_i^\alpha(f) := \sup\left\{\frac{f(\overline{x}) - f(\overline{y})}{d^\alpha(x_i, y_i)} : x_j = y_j,\ \forall j \neq i,\ x_i \neq y_i\right\}.$$

is finite for all i. In (4)*, we then have to replace d by* d^α*.*

4 The Markov case

We now consider $(X_n)_{n\in\mathbb{Z}}$ to be a stationary and ergodic Markov chain. We denote by $p(x, dy) := \mathbb{P}(X_1 \in dy \big| X_0 = x)$ the transition kernel. We let ν be the unique stationary measure of the Markov chain. We denote by $\mathbb{P}_\nu$ the path space measure of the stationary process $(X_n)_{n\in\mathbb{Z}}$. By $\mathbb{P}_x$ we denote the distribution of (X_1^∞), for the Markov process conditioned on $X_0 = x$.

We further suppose that the coupling $\hat{\mathbb{P}}$ of Section 2.2 is Markovian, and denote by $\hat{\mathbb{P}}_{x,y}$ the coupling started from x, y, and corresponding expectation by $\hat{\mathbb{E}}_{x,y}$. More precisely, by the Markov property of the coupling we then have that

$$\hat{\mathbb{P}}_{x^i_{-\infty}, y^i_{-\infty}} = \hat{\mathbb{P}}_{x_i, y_i}$$

is a Markovian coupling $((\hat{X}_n^{(1)}, \hat{X}_n^{(2)}))_{n\in\mathbb{N}}$ of the Markov chains $(X_n)_{n\geq 0}$ starting from $X_0 = x_i$, resp. $Y_0 = y_i$. In this case the expression (4) of the coupling matrix simplifies to

$$\Psi_{X_{i-1},X_i}(j) := D^{X_{i-1},X_i}_{i,i+j} = \int p(X_{i-1}, dy) \int d\hat{\mathbb{P}}_{X_i,y}(u_0^\infty, v_0^\infty)\, d(u_j, v_j) \tag{5}$$

With this notation, (3) reads

$$|\Delta_i(X_{i-1}, X_i)| \leq \sum_{j\geq 0} \Psi_{X_{i-1},X_i}(j)\delta_{i+j} f. \tag{6}$$

We define the "generalized coupling time"

$$\mathcal{T}(u_0^\infty, v_0^\infty) := \sum_{j=0}^{\infty} d(u_j, v_j). \tag{7}$$

In the case E is a discrete (finite or countable) alphabet, the "classical" coupling time is defined as usual

$$T(u_0^\infty, v_0^\infty) := \inf\{k \geq 0 : \forall j \geq k : u_j = v_j\}.$$

If we use the trivial distance $d(x, y) = 1$ if $x \neq y$ and $d(x, y) = 0$ if $x = y$, for $x, y \in E$, then we have

$$d(u_j, v_j) \leq \mathbb{1}\{T(u_0^\infty, v_0^\infty) \geq j\} \tag{8}$$

and hence

$$\mathcal{T}(u_0^\infty, v_0^\infty) \le T(u_0^\infty, v_0^\infty).$$

Of course, the same inequality remains true if E is a bounded metric space with $d(x,y) \le 1$ for $x, y \in E$. However a "successful coupling" (i.e., a coupling with $T < \infty$) is not expected to exist in general in the case of a non-discrete state space. It can however exist, see e.g. [13] for a successful coupling in the context of Zhang's model of self-organized criticality. Let us also mention that the "generalized coupling time" unavoidably appears in the context of dynamical systems [8].

In the discrete case, using (6) and (8), we obtain the following inequality:

$$\Psi_{x,y}(j) \le \sum_z p(x,z)\hat{\mathbb{E}}_{z,y}\left(\mathbb{1}\{T(u_0^\infty, v_0^\infty) \ge j\}\right) \tag{9}$$

whereas in the general (not necessarily discrete) case we have, by (7), and monotone convergence,

$$\sum_{j\ge 0}\Psi_{x,y}(j) = \int p(x,dz)\hat{\mathbb{E}}_{y,z}(\mathcal{T}) \le \int p(x,dz)\hat{\mathbb{E}}_{y,z}(T). \tag{10}$$

Remark 4.1. *So far, we made a telescoping of $f - \mathbb{E}(f)$ using an increasing family of sigma-fields. One can as well consider a decreasing family of sigma-fields, such as $\mathscr{F}_i^\infty$, defined to be the sigma-fields generated by $\{X_k : k \ge i\}$. We then have, mutatis mutandis, the same inequalities using "backward telescoping"*

$$f - \mathbb{E}(f) = \sum_{i=-\infty}^{\infty} \Delta_i^*,$$

where

$$\Delta_i^* := \mathbb{E}(f \,|\, \mathscr{F}_i^\infty) - \mathbb{E}(f \,|\, \mathscr{F}_{i+1}^\infty).$$

and estimating Δ_i^* in a completely parallel way, by introducing a lower-triangular analogue of the coupling matrix matrix.

Backward telescoping is natural in the context of dynamical systems where the forward process is deterministic, hence cannot be coupled (as defined above) with two different initial conditions such that the copies become closer and closer. However, backwards in time, such processes are non-trivial Markov chains for which a coupling can be possible with good decay properties of the coupling matrix. See [8] for a concrete example with piecewise expanding maps of the interval.

5 Variance inequality

For a real-valued sequence $(a_i)_{i\in\mathbb{Z}}$, we denote the usual ℓ_p norm

$$\|a\|_p = \left(\sum_{i\in\mathbb{Z}} |a_i|^p\right)^{1/p}.$$

Our first result concerns the variance of a $f \in \mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R})$.

Theorem 5.1. *Let $f \in \mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R}) \cap L^2(\mathbb{P}_\nu)$. Then*

$$\mathrm{Var}(f) \leq C \|\delta(f)\|_2^2 \tag{11}$$

where

$$C = \int \nu(dx) \times$$

$$\int p(x, dz) \int p(x, dy) \int p(x, du) \hat{\mathbb{E}}_{z,y}(\mathcal{T}) \hat{\mathbb{E}}_{z,u}(\mathcal{T}). \tag{12}$$

As a consequence, we have the concentration inequality

$$\forall t > 0, \quad \mathbb{P}(|f - \mathbb{E}(f)| \geq t) \leq C \frac{\|\delta(f)\|_2^2}{t^2}. \tag{13}$$

Proof. We estimate, using (6) and stationarity

$$\mathbb{E}(\Delta_i^2) \leq \mathbb{E}\left(\sum_{j \geq 0} \Psi_{X_0, X_1}(j) \delta_{i+j}(f)\right)^2 = \mathbb{E}\left(\left(\Psi_{X_0,X_1} * \delta(f)\right)_i^2\right).$$

where $*$ denotes convolution, and where we extended Ψ to $\mathbb{Z}$ by putting it equal to zero for negative integers. Since

$$\mathrm{Var}(f) = \sum_{i=-\infty}^{\infty} \mathbb{E}\left((\Delta_i)^2\right)$$

Using Young's inequality, we then obtain,

$$\mathrm{Var}(f) \leq \mathbb{E}\left(\left\|\Psi_{X_0,X_1} * \delta(f)\right\|_2^2\right) \leq \mathbb{E}\left(\left\|\Psi_{X_0,X_1}\right\|_1^2\right) \left\|\delta(f)\right\|_2^2.$$

Now, using the equality in (10)

$$\begin{aligned}
& \mathbb{E}\left(\left\|\Psi_{X_0,X_1}\right\|_1^2\right) \\
= \; & \mathbb{E}\left(\int p(X_0, dy) \hat{\mathbb{E}}_{X_1,y}(\mathcal{T})\right)^2 \\
= \; & \int \nu(dx) p(x, dz) \left(\int p(x, dy) \hat{\mathbb{E}}_{z,y}(\mathcal{T})\right)^2 \\
= \; & \int \nu(dx) \int p(x, dz) \int p(x, dy) \int p(x, du) \hat{\mathbb{E}}_{z,y}(\mathcal{T}) \hat{\mathbb{E}}_{z,u}(\mathcal{T}),
\end{aligned}$$

which is (11). Inequality (13) follows from Chebychev's inequality. □

The expectation in (12) can be interpreted as follows. We start from a point x drawn from the stationary distribution and generate three independent copies y, u, z from the Markov chain at time $t = 1$ started from x. With these initial points we start the coupling in couples (y, z) and (u, z), and compute the expected coupling time.

6 Moment inequalities

In order to control higher moments of $(f - \mathbb{E}(f))$, we have to tackle higher moments of the sum $\sum_i \Delta_i^2$ and for these we cannot use the simple stationarity argument used in the estimation of the variance.

Instead, we start again from (6) and let $\lambda^2(j) := (j+1)^{1+\epsilon}$ where $\epsilon > 0$.

We then obtain, using Cauchy-Schwarz inequality:

$$\begin{aligned} |\Delta_i| &\leq \sum_{j\geq 0} \lambda(j)\Psi_{X_{i-1},X_i}(j)\frac{\delta_{i+j}(f)}{\lambda(j)} \\ &\leq \left(\sum_{j\geq 0}\lambda(j)^2\left(\Psi_{X_{i-1},X_i}(j)\right)^2\sum_{k\geq 0}\left(\frac{\delta_{i+k}(f)}{\lambda(k)}\right)^2\right)^{1/2}. \end{aligned}$$

Hence

$$\Delta_i^2 \leq \Psi_\epsilon^2(X_{i-1},X_i)\left((\delta(f))^2 * \frac{1}{\lambda^2}\right)_i \tag{14}$$

where $\delta(f)^2$ denotes the sequence with components $\left(\delta_i(f)\right)^2$, and where

$$\Psi_\epsilon^2(X_{i-1},X_i) = \sum_{j\geq 0}(j+1)^{1+\epsilon}\left(\Psi_{X_{i-1},X_i}(j)\right)^2. \tag{15}$$

Moment inequalities will now be expressed in terms of moments of Ψ_ϵ^2.

6.1 Moment inequalities in the discrete case

We first deal with a discrete state space E. Recall (8).

Lemma 6.1. *In the discrete case, i.e., if E is a countable set with the discrete metric, then, for all $\epsilon > 0$, we have the estimate*

$$\begin{aligned} &\Psi_\epsilon^2(X_{i-1},X_i) \\ &\leq \frac{1}{2}\left(\sum_z p(X_{i-1},z)\hat{\mathbb{E}}_{X_i,z}((T+1)^{1+\frac{\epsilon}{2}})\right)^2. \end{aligned} \tag{16}$$

Proof. Start with

$$\begin{aligned} \Psi_\epsilon^2 &= \sum_{j\geq 0}(j+1)^{1+\epsilon}\left(\Psi_{X_{i-1},X_i}(j)\right)^2 \\ &\leq \sum_{z,u}\sum_{j\geq 0}(j+1)^{1+\epsilon}p(X_{i-1},z)p(X_{i-1},u)\hat{\mathbb{P}}_{X_i,z}(T\geq j)\hat{\mathbb{P}}_{X_i,u}(T\geq j). \end{aligned}$$

Proceed now with

$$
\begin{aligned}
&\sum_{j\geq 0}(j+1)^{1+\epsilon}p(X_{i-1},z)p(X_{i-1},u)\hat{\mathbb{P}}_{X_i,z}(T\geq j)\hat{\mathbb{P}}_{X_i,u}(T\geq j)\\
=\ &\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{j=0}^{l\wedge k}(j+1)^{1+\epsilon}\ \hat{\mathbb{P}}(T_1=k,T_2=l)\\
\leq\ &\frac{1}{2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(l\wedge k+1)^{2+\epsilon}\ \hat{\mathbb{P}}(T_1=k,T_2=l)\\
=\ &\hat{\mathbb{E}}\left(((T_1+1)\wedge(T_2+1))^{2+\epsilon}\right),
\end{aligned}
$$

where we denoted by T_1 and T_2 two independent coupling times corresponding to two independent copies of the coupling started from (X_i,z), resp. (X_i,u).

Now use that for two independent non-negative real-valued random variables we have

$$
\mathbb{E}\big((X\wedge Y)^{2+\epsilon}\big)\leq\mathbb{E}(X^{1+\frac{\epsilon}{2}})\mathbb{E}(Y^{1+\frac{\epsilon}{2}}).
$$

The lemma is proved. □

In order to arrive at moment estimates, we want an estimate for $\mathbb{E}\big(\sum_i\Delta_i^2\big)^p$. This is the content of the next lemma. We denote, as usual, $\zeta(s)=\sum_{n=1}^{\infty}(1/n)^s$.

Lemma 6.2. *For all $\epsilon>0$ and integers $p>0$ we have*

$$
\begin{aligned}
\mathbb{E}\big(\sum_i\Delta_i^2\big)^p\ \leq\ &\left(\frac{\zeta(1+\epsilon)}{2}\right)^p\|\delta(f)\|_2^{2p}\sum_{x,y}\nu(x)p(x,y)\times\\
&\left(\sum_z p(x,z)\hat{\mathbb{E}}_{y,z}((T+1)^{1+\frac{\epsilon}{2}})\right)^{2p}.
\end{aligned}
\tag{17}
$$

Proof. We start from

$$
\begin{aligned}
&\mathbb{E}\big(\sum_i\Delta_i^2\big)^p\\
&\leq\sum_{i_1,\ldots,i_p}\mathbb{E}\left(\prod_{l=1}^{p}\Psi_\epsilon(X_{i_l-1},X_{i_l})^2\right)\prod_{l=1}^{p}\left((\delta(f))^2*\frac{1}{\lambda^2}\right)_{i_l}.
\end{aligned}
$$

Then use Hölder's inequality and stationarity, to obtain

$$
\begin{aligned}
\mathbb{E}\big(\sum_i\Delta_i^2\big)^p\ &\leq\ \mathbb{E}(\Psi_\epsilon^{2p}(X_0,X_1))\times\left\|(\delta(f))^2*\frac{1}{\lambda^2}\right\|_1^p\\
&\leq\ \mathbb{E}(\Psi_\epsilon^{2p}(X_0,X_1))\times\left\|\frac{1}{\lambda^2}\right\|_1^p\|(\delta(f))^2\|_1^p\\
&=\ \mathbb{E}(\Psi_\epsilon^{2p}(X_0,X_1))\times\left\|\frac{1}{\lambda^2}\right\|_1^p\|(\delta(f))\|_2^{2p}
\end{aligned}
$$

where in the second inequality we used Young's inequality. The lemma now follows from (16). □

We can now formulate our moment estimates in the discrete case.

Theorem 6.1. *Suppose E is a countable set with discrete metric. Let $p \geq 1$ be an integer and $f \in \mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R}) \cap L^{2p}(\mathbb{P})$. Then for all $\epsilon > 0$ we have the estimate*

$$\mathbb{E}(f - \mathbb{E}(f))^{2p} \leq C_p \|\delta(f)\|_2^{2p} \tag{18}$$

where

$$\begin{aligned} C_p &= (2p-1)^{2p} \left(\frac{\zeta(1+\epsilon)}{2} \right)^p \times \\ & \sum_{x,y} \nu(x) p(x,y) \left(\sum_z p(x,z) \hat{\mathbb{E}}_{y,z} \left((T+1)^{1+\frac{\epsilon}{2}} \right) \right)^{2p}. \end{aligned} \tag{19}$$

As a consequence we have the concentration inequalities

$$\forall t > 0, \quad \mathbb{P}(|f - \mathbb{E}(f)| \geq t) \leq C_p \frac{\|\delta(f)\|_2^{2p}}{t^{2p}}. \tag{20}$$

Proof. By Burkholder's inequality [5, Theorem 3.1, p. 87], one gets

$$\mathbb{E}\left((f - \mathbb{E}(f))^{2p} \right) \leq (2p-1)^{2p} \, \mathbb{E}\Big((\sum_i \Delta_i^2)^p \Big)$$

and (19) then follows from (17), whereas (20) follows from (19) by Markov's inequality. □

Remark 6.1. *Theorem 6.1 for $p = 1$ is weaker than Theorem 5.1: indeed, for* (11) *to hold we only need to have the first moment of the coupling time to be finite.*

Remark 6.2. *A typical behavior (see the examples below) of the coupling time is as follows:*

$$\hat{\mathbb{P}}_{x,y}(T \geq j) \leq C(x,y)\phi(j)$$

Here $C(x,y)$ is a constant that depends, in general in an unbounded way, on the starting points (x,y) in the coupling and where $\phi(j)$, determining the tail of the coupling time does not depend on the starting points. Therefore, for the finiteness of the constant C_p in (19) *we need that the tail-estimate $\phi(j)$ decays fast enough so that $\sum_j j^\epsilon \phi(j) < \infty$ (that does not depend on p), and next the $2p$-th power of the constant $C(x,y)$ has to be integrable (this depends on p).*

6.2 The general state space case

In order to formulate the general state space version of these results, we introduce the expectation

$$\tilde{\mathbb{E}}_{x,y}(F(\overline{u}, \overline{v})) = \int p(x, dz) \int \hat{\mathbb{P}}_{y,z}(d\overline{u}, d\overline{v}) F(\overline{u}, \overline{v}).$$

We can then rewrite

$$\Psi_\epsilon^2(x,y) = \sum_{j \geq 0} (j+1)^{1+\epsilon} \left(\tilde{\mathbb{E}}_{x,y} d(u_j, v_j) \right)^2.$$

We introduce

$$\alpha_j^{x,y} = \left(\tilde{\mathbb{E}}_{x,y}(d(u_j, v_j)) - \tilde{\mathbb{E}}_{x,y}(d(u_{j+1}, v_{j+1}))\right).$$

This quantity is the analogue of $\hat{\mathbb{P}}(T = j)$ of the discrete case. We then define

$$M_r^{x,y} = \sum_{j\geq 0}(j+1)^r \alpha_j^{x,y} \tag{21}$$

which is the analogue of the r-th moment of the coupling time. The analogue of Theorem 6.1 then becomes the following.

Theorem 6.2. *Let $p \geq 1$ be an integer and $f \in \mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R}) \cap L^{2p}(\mathbb{P})$. Then for all $\epsilon > 0$ we have the estimate*

$$\mathbb{E}(f - \mathbb{E}(f))^{2p} \leq C_p \|\delta(f)\|_2^{2p}$$

where

$$C_p = (2p-1)^{2p}\left(\frac{\zeta(1+\epsilon)}{2}\right)^p \int \nu(dx)p(x,dy)\left(M_{1+\frac{\epsilon}{2}}^{x,y}\right)^{2p}.$$

7 Gaussian concentration bound

If one has a uniform estimate of the quantity (15), we obtain a corresponding uniform estimate for Δ_i^2, and via Hoeffding's inequality, a Gaussian bound for $(f - \mathbb{E}(f))$. This is formulated in the following theorem.

Theorem 7.1. *Let E be a countable set with the discrete metric. Let $f \in \mathrm{Lip}(E^{\mathbb{Z}}, \mathbb{R})$ such that $\exp(f) \in L^1(\mathbb{P})$. Then for all $\epsilon > 0$ we have*

$$\mathbb{E}\left(e^{f - \mathbb{E}(f)}\right) \leq e^{C\|\delta(f)\|_2^2/16} \tag{22}$$

where

$$C = \zeta(1+\epsilon)\left(\sup_{u,v}\hat{\mathbb{E}}_{u,v}(T^{1+\epsilon})\right)^2. \tag{23}$$

In particular, we get the concentration inequality

$$\forall t \geq 0, \quad \mathbb{P}(|f - \mathbb{E}(f)| \geq t) \leq 2\exp\left(\frac{-4Ct^2}{\|\delta(f)\|_2^2}\right). \tag{24}$$

The general state space analogue of these bounds is obtained by replacing $\hat{\mathbb{E}}_{u,v}(T^{1+\epsilon})$ by $M_{1+\epsilon}^{u,v}$ in (23) (where $M_r^{x,y}$ is defined in (21)).

Proof. From (16) we get

$$\begin{aligned}
\Psi_\epsilon^2(x,y) &\leq \frac{1}{2}\sum_{z,u} p(x,z)p(x,u)\hat{\mathbb{E}}_{y,z}(T^{1+\frac{\epsilon}{2}})\hat{\mathbb{E}}_{y,u}(T^{1+\frac{\epsilon}{2}}) \\
&\leq \frac{1}{2}\left(\sup_{u,v}\hat{\mathbb{E}}_{u,v}(T^{1+\frac{\epsilon}{2}})\right)^2.
\end{aligned}$$

Now start from the classical Azuma-Hoeffding inequality [19]

$$\mathbb{E}\left(e^{f-\mathbb{E}(f)}\right) \le e^{\frac{1}{8}\sum_i \|\Delta_i\|_\infty^2}.$$

Therefore, we estimate, using (14) and Young's inequality

$$\sum_i \|\Delta_i\|_\infty^2 \le \frac{1}{2}\zeta(1+\epsilon)\left(\sup_{u,v}\hat{\mathbb{E}}_{u,v}(T^{1+\frac{\epsilon}{2}})\right)^2 \|\delta(f)\|_2^2$$

which establishes (22). Inequality (24) follows from (22) by the optimized exponential Chebychev inequality. □

Remark 7.1. *The assumption that a moment of order* $1+\epsilon$ *of the coupling time exists, which is uniformly bounded in the starting point, can be weakened to the same property for the first moment, if we have some form of monotonicity. More precisely, we say that a coupling has the monotonicity property, if there exist "worse case starting points"* x_u, x_l*, which have the property that*

$$\sup_{x,y}\hat{\mathbb{P}}_{x,y}(T\ge j) \le \hat{\mathbb{P}}_{x_u,x_l}(T\ge j) \tag{25}$$

for all $j \ge 0$*. In that case, using* (9)*, we can start from* (6) *and obtain, in the discrete case, the uniform bound*

$$\Delta_i \le \sum_{j\ge 0}\hat{\mathbb{P}}_{x_u,x_l}(T\ge j)\delta_{i+j}f \tag{26}$$

and via Azuma-Hoeffding inequality, combined with Young's inequality, we then obtain the Gaussian bound (22) *with*

$$C = \frac{1}{2}\hat{\mathbb{E}}_{x_u,x_l}(T).$$

Finally, it can happen (especially if the state space is unbounded) that the coupling has no worst case starting points, but there is a sequence x_u^n, x_l^n *of elements of the state space such that* $\hat{\mathbb{P}}_{x_u^n,x_l^n}(T\ge j)$ *is a non-decreasing sequence in n for every fixed j and*

$$\hat{\mathbb{P}}_{x,y}(T\ge j) \le \lim_{n\to\infty}\hat{\mathbb{P}}_{x_u^n,x_l^n}(T\ge j)$$

(E.g. in the case of the state space $\mathbb{Z}$ *we can think of the sequence* $x_u^n \to \infty$ *and* $x_l^n \to -\infty$*.) In that case, from monotone convergence we have the Gaussian concentration bound with*

$$C = \lim_{n\to\infty}\frac{1}{2}\hat{\mathbb{E}}_{x_u^n,x_l^n}(T).$$

8 Examples

8.1 Finite-state Markov chains

As we mentioned in the introduction, this case was already considered by K. Marton (and others), but it illustrates our method in the most simple setting, and gives also an alternative proof in this setting.

Indeed, if the chain is aperiodic and irreducible, then it is well-known [26],

$$\sup_{u,v\in E}\hat{\mathbb{P}}_{u,v}(T\ge j) \le c\,\rho^j$$

for all $j \ge 1$ and some $c > 0$. Hence the Gaussian bound (22) holds.

8.2 House of cards processes

These are Markov chains on the set of natural numbers which are useful in the construction of couplings for processes with long-range memory, and dynamical systems, see e.g. [4].

More precisely, a house of cards process is a Markov chain on the natural numbers with transition probabilities

$$\mathbb{P}(X_{k+1} = n+1|X_k = n) = 1 - q_n = 1 - \mathbb{P}(X_{k+1} = 0|X_k = n),$$

for $n = 0, 1, 2, \ldots$, i.e., the chain can go "up" with one unit or go "down" to zero. Here, $0 < q_n < 1$.

In the present paper, house of card chains serve as a nice class of examples where we can have moment inequalities up to a certain order, depending on the decay of q_n, and even Gaussian inequalities. Given a sequence of independent uniformly distributed random variables (U_k) on $[0,1]$, we can view the process X_k generated via the recursion

$$X_{k+1} = (X_k + 1)\mathbb{1}\{U_{k+1} \geq q_{X_k}\}. \tag{27}$$

This representation also yields a coupling of the process for different initial conditions. The coupling has the property that when the coupled chains meet, they stay together forever. In particular, they will stay together forever after they hit together zero. For this coupling, we have the following estimate.

Lemma 8.1. *Consider the coupling defined via* (27)*, started from initial condition* (k,m) *with* $k \geq m$*. Then we have*

$$\hat{\mathbb{P}}_{k,m}(T \geq t) \leq \prod_{j=0}^{t-1}(1 - q^*_{k+j}) \tag{28}$$

where

$$q^*_n = \inf_{s\leq n} q_s.$$

Proof. Call Y^k_t the process defined by (27) started from k, and define Z^k_t, a process started from k defined via the recursion

$$Z_{t+1} = (Z_t + 1)\mathbb{1}\left\{U_{t+1} \geq q^*_{Z_t}\right\},$$

where U_t is the same sequence of independent uniformly distributed random variables as in (27). We claim that, for all $t \geq 0$,

$$Y^k_t \leq Z^k_t, Y^m_t \leq Z^k_t.$$

Indeed, the inequalities hold at time zero. Suppose they hold at time t, then, since q^*_n is non-increasing as a function of n,

$$q_{Y^k_t} \geq q^*_{Y^k_t} \geq q^*_{Z^k_t} \quad \text{and} \quad q_{Y^m_t} \geq q^*_{Y^m_t} \geq q^*_{Z^k_t}$$

whence

$$\mathbb{1}\left\{U_{t+1} \geq q^*_{Z^k_t}\right\} \geq \mathbb{1}\left\{U_{t+1} \geq q_{Y^k_t}\right\} \text{ and } \mathbb{1}\left\{U_{t+1} \geq q^*_{Z^k_t}\right\} \geq \mathbb{1}\left\{U_{t+1} \geq q_{Y^m_t}\right\}.$$

Therefore, in this coupling, if $Z_t^k = 0$, then $Y_t^m = Y_t^k = 0$, and hence the coupling time is dominated by the first visit of Z_t^k to zero, which gives

$$\hat{\mathbb{P}}_{k,m}(T \geq t) \leq \mathbb{P}(Z_n^k \neq 0, n = 1, \dots, t-1) = \prod_{j=1}^{t-1}(1 - q_{k+j}^*).$$

□

The behavior (28) of the coupling time shows the typical non-uniformity as a function of the initial condition. More precisely, the estimate in the rhs of (28) becomes bad for large k. We now look at three more concrete cases.

1. **Case 1:**
$$q_n = \frac{1}{n^\alpha}\ ,\ n \geq 2,\ 0 < \alpha < 1.$$
Then it is easy to deduce from (28) that
$$\hat{\mathbb{P}}_{k,m}(T \geq t) \leq C \exp\left(-\frac{1}{1-\alpha}\left((t+k)^{1-\alpha} - k^{1-\alpha}\right)\right). \tag{29}$$
The stationary (probability) measure is given by:
$$\nu(k) = \pi_0 c_k \tag{30}$$
with
$$c_k = \prod_{j=0}^{k}(1 - q_j) \tag{31}$$
which is bounded from above by
$$c_k \leq C' \exp\left(-\frac{1}{1-\alpha}k^{1-\alpha}\right). \tag{32}$$
From (29), combined with (30), (32), it is then easy to see that the constant C_p of (19) is finite for all $p \in \mathbb{N}$. Therefore, in that case the moment inequalities (18) hold, for all $p \geq 1$.

2. **Case 2:**
$$q_n = \frac{\gamma}{n} \quad (\gamma > 0)$$
for $n \geq \gamma + 1$, and other values q_i are arbitrary. In this case we obtain from (28) the estimate
$$\hat{\mathbb{P}}_{k,m}(T \geq t) \leq C_\gamma \frac{(k+1)^\gamma}{(k+t)^\gamma}$$
and for the stationary measure we have (31) with
$$c_k \leq C'_\gamma k^{-\gamma}.$$
The constant C_p of (19) is therefore bounded by
$$C_p \leq C_\gamma C'_\gamma C_1 C_2$$

where $C_1 = (2p-1)^{2p}(\zeta(1+\epsilon)/2)^p$ is finite independent of γ, and where

$$C_2 = C_2(p) \le \sum_{k\ge 1} k^{-\gamma}\left(\hat{\mathbb{E}}_{k+1,0}(T+1)^{1+\frac{\epsilon}{2}}\right)^{2p}$$

so we estimate

$$\hat{\mathbb{E}}_{k+1,0}(T+1)^{1+\frac{\epsilon}{2}} \le (1+\delta)\sum_{t=0}^{\infty}\frac{(k+1)^\gamma(t+1)^\delta}{(t+k)^\gamma},$$

where $\delta := \epsilon/2$. To see when $C_2 < \infty$, we first look at the behavior of

$$I(a,b,k) := \sum_{t=1}^{\infty}\frac{t^a}{(t+k)^b}.$$

The sum in the rhs is convergent for $b-a>1$, in which case it behaves as k^{1+a-b} for k large, which gives for our case $a=\delta$, $b=\gamma$, $\gamma>1+\delta$. In that case, we find that $C_2(p)$ is finite as soon as

$$\sum_{k\ge 1} k^{-\gamma+2p\gamma}k^{2p(\delta-\gamma+1)} < \infty$$

which gives

$$\gamma > 1 + 2p\delta + 2p.$$

Hence, in this case, for $\gamma > 1+\delta$, we obtain the moment estimates (6.1) up to order $p < (\gamma-1)/2(\delta+1)$.

3. **Case 3:**

$$q := \inf\{q_n : n \in \mathbb{N}\} > 0$$

then we have the uniform estimate

$$\sup_{k,m}\hat{\mathbb{P}}_{k,m}(T \ge t) \le (1-q)^t \tag{33}$$

which gives the Gaussian concentration bound (22) with $C = \frac{1}{2(1-q)}$.

8.3 Ergodic interacting particle systems

As a final example, we consider spin-flip dynamics in the so-called $M<\epsilon$ regime. These are Markov processes on the space $E = \{0,1\}^S$, with S a countable set. This is a metric space with distance

$$d(\eta,\xi) = \sum_{n=1}^{\infty} 2^{-n}|\eta_{i_n} - \xi_{i_n}|$$

where $n \mapsto i_n$ is a bijection from $\mathbb{N}$ to S.

The space E is interpreted as set of configurations of "spins" η_i which can be up (1) or down (0) and are defined on the set S (usually taken to be a lattice such as $\mathbb{Z}^d$). The spin at site $i \in S$ flips at a configuration dependent rate $c(i,\eta)$. The process is then defined via its generator on local functions defined by

$$Lf(\eta) = \sum_{i\in S} c(i,\eta)(f(\eta^i) - f(\eta))$$

where η^i is the configuration η obtained from η by flipping at site i. See [18] for more details about existence and ergodicity of such processes.

We assume here that we are in the so-called "$M < \epsilon$ regime", where we have the existence of a coupling (the so-called basic coupling) for which we have the estimate

$$\hat{\mathbb{P}}_{\eta^j,\eta}(\eta_i(t) \neq \zeta_i(t)) < e^{-\epsilon t} e^{\Gamma(i,j)t} \tag{34}$$

with $\Gamma(i,j)$ a matrix indexed by S with finite l_1-norm $M < \epsilon$. As a consequence, from any initial configuration, the system evolves exponentially fast to its unique equilibrium measure which we denote μ. The stationary Markov chain is then defined as $X_n = \eta_{n\delta}$ where $\delta > 0$, and $\eta_0 = X_0$ is distributed according to μ.

In the basic coupling, from (34), we obtain the uniform estimate

$$\tilde{\mathbb{E}}_{\eta,\zeta}(d(\eta(k),\zeta(k)) \leq e^{-(M-\epsilon)k\delta}.$$

As a consequence, the quantity $M_r^{\eta,\zeta}$ of (21) is finite uniformly in η,ζ, for every $r > 0$. Therefore, we have the Gaussian bound (22) with

$$C \leq \sum_{k\geq 1} k^{1+\epsilon} e^{-(M-\epsilon)k\delta} < \infty.$$

8.4 Measure concentration of Hamming neighborhoods

We apply Theorem 6.1 to measure concentration of Hamming neighborhoods. The case of contracting Markov chains was already (and first) obtained in [20] as a consequence of an information divergence inequality. We can easily obtain such Gaussian measure concentration from (22). But, by a well-known result of Bobkov and Götze [2], (22) and that information divergence inequality are in fact equivalent. The interesting situation is when (22) does not hold but only have moment bounds.

Let $A,B \subset E^n$ be two sets and denote by $\overline{d}(A,B)$ their normalized Hamming distance, $\overline{d}(A,B) = \inf\{\overline{d}(x_1^n,y_1^n) : x_1^n \in A, y_1^n \in B\}$, where

$$\overline{d}(x_1^n,y_1^n) = \frac{1}{n}\sum_{i=1}^n d(x_i,y_i),$$

$d(x_i,y_i) = 1$ if $x_i \neq y_i$, and 0 otherwise. The ε-neighborhood of A is then

$$[A]_\varepsilon = \{y_1^n : \inf_{x_1^n\in A} \overline{d}(x_1^n,y_1^n) \leq \varepsilon\}.$$

Theorem 8.1. *Take any $n \in \mathbb{N}$ and let $A \subset E^n$ a measurable set with $\mathbb{P}(A) > 0$. Then, under the assumptions of Theorem 6.1, we have, for all $p \geq 1$,*

$$\mathbb{P}([A]_\varepsilon) \geq 1 - \frac{1}{n^p}\,\frac{1}{\left(\frac{\varepsilon}{C_p^{1/2p}} - \frac{1}{\sqrt{n}(\mathbb{P}(A))^{1/2p}}\right)^{2p}} \tag{35}$$

for all $\varepsilon > \frac{C_p^{1/2p}}{\sqrt{n}(\mathbb{P}(A))^{1/2p}}$.

Proof. We apply Theorem 6.1 to $f = \overline{d}(\cdot, A)$, which is a function defined on E^n. It is easy to check that $\delta_i(f) \leq 1/n$, $i = 1, \ldots, n$. We first estimate $\mathbb{E}(f)$ by using (18), which gives (using the fact that $f_{|A} = 0$)

$$\mathbb{E}(f) \leq \frac{C_p^{1/2p}}{\sqrt{n}(\mathbb{P}(A))^{1/2p}}. \tag{36}$$

Now we apply (20) with $t = \varepsilon$,

$$\mathbb{P}(f > \varepsilon) \leq \frac{C_p}{n^p} \frac{1}{\left(\varepsilon - \frac{C_p^{1/2p}}{\sqrt{n}(\mathbb{P}(A))^{1/2p}}\right)^{2p}}.$$

The result then easily follows. □

REMARK 8.1. *For $p = 1$, the theorem holds under the assumption of Theorem 5.1; see Remark 6.1.*

As we saw in Section 8.2, we cannot have Gaussian bounds for certain house of cards processes, but only moment estimates up to a critical order. In particular, this means that we cannot have a Gaussian measure concentration of Hamming neighborhoods. But in that case we can apply the previous theorem and get polynomial measure concentration.

Acknowledgment. The authors thank E. Verbitskiy for useful discussions on the house of cards process, and an anonymous referee for useful remarks.

References

[1] R. Adamczak, A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. Electron. J. Prob. **13**, 1000-1034, (2008). MR2424985

[2] S. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163** (1999), 1–28. MR1682772

[3] S. Boucheron, O. Bousquet, G. Lugosi and P. Massart, Moment inequalities for functions of independent random variables. Ann. Prob. **33**, 514-560, (2005). MR2123200

[4] X. Bressaud, R. Fernández, and A. Galves. Decay of correlations for non-Hölderian dynamics. A coupling approach. Electron. J. Probab. **4** (1999), no. 3 (19 pp.). MR1675304

[5] D.L. Burkholder. Sharp inequalities for martingales and stochastic integrals. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987). Astérisque No. 157-158 (1988), 75–94. MR0976214

[6] S. Chatterjee. Stein's method for concentration inequalities. Probab. Theory Related Fields **138** (2007), no. 1-2, 305–321. MR2288072

[7] J.-R. Chazottes, P. Collet, C. Külske and F. Redig. Concentration inequalities for random fields via coupling. Probab. Theory Related Fields **137** (2007), no. 1-2, 201–225. MR2278456

[8] P. Collet. Variance and exponential estimates via coupling. Bull. Braz. Math. Soc. **37** (2006), no. 4, 461–475. MR2284882

[9] H. Djellout, A. Guillin and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. Ann. Probab. **32** (2004), no. 3B, 2702–2732. MR2078555

[10] R. Douc, A. Guillin and E. Moulines. Bounds on Regeneration Times and Limit Theorems for Subgeometric Markov Chains. Annales Inst. H. Poincaré, to appear. MR2446322

[11] R. Douc, G. Fort, E. Moulines and P. Soulier, Practical drift conditions for subgeometric rates of convergence. Ann. Appl. Prob. **14**, 1353-1377, (2004). MR2071426

[12] R. Douc, E. Moulines and P. Soulier, Computable convergence rates for sub-geometric ergodic Markov chains. Bernoulli **13**, 831-848 (2007). MR2348753

[13] A. Fey-den Boer, R. Meester, Ronald, C. Quant, and F. Redig. A probabilistic approach to Zhang's sandpile model. Comm. Math. Phys. **280**, 351–388, (2008). MR2395474

[14] S. Goldstein. A note on specifications. Z. Wahrsch. Verw. Gebiete, **46**, 45-51 (1978/79). MR0512331

[15] L. Kontorovich and K. Ramanan. Concentration Inequalities for Dependent Random Variables via the Martingale Method, prepint (2007), to appear in Ann. Probab. MR2478678

[16] L. Kontorovich. Obtaining Measure Concentration from Markov Contraction, preprint, 2007 (arXiv:0711.0987).

[17] M. Ledoux. *The concentration of measure phenomenon*, Mathematical Surveys and Monographs **89**. American Mathematical Society, Providence R.I., 2001. MR1849347

[18] T.M. Liggett, *Interacting particle systems.* Reprint of the 1985 original. Classics in Mathematics. Springer-Verlag, Berlin, 2005. MR2108619

[19] C. McDiarmid. On the method of bounded differences, in *Surveys in Combinatorics* 1989, Cambridge University Press, Cambridge (1989) 148–188. MR1036755

[20] K. Marton. Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration. Ann. Probab. **24** (1996), no. 2, 857–866. MR1404531

[21] K. Marton. A measure concentration inequality for contracting Markov chains. Geom. Funct. Anal. **6** (1996), no. 3, 556–571. [Erratum: Geom. Funct. Anal. 7 (1997), no. 3, 609–613.] MR1392329

[22] K. Marton. Measure concentration for a class of random processes. Probab. Theory Related Fields **110** (1998), no. 3, 427–439. MR1616492

[23] Y.H. Mao, Convergence rates in strong ergodicity for Markov processes. Stochastic Process. Appl. **116**, no. 12, 1964–1976, (2006). MR2307067

[24] E. Rio. Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. [Hoeffding inequalities for Lipschitz functions of dependent sequences] C. R. Acad. Sci. Paris Sér. I Math. **330** (2000), no. 10, 905–908. MR1771956

[25] P.-M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. Ann. Probab. **28** (2000), no. 1, 416–461. MR1756011

[26] H. Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000. MR1741181