arXiv:math.PR/0503483 v1 23 Mar 2005

Deviation inequalities via coupling for stochastic processes and random fields

J.-R. Chazottes
Centre de Physique Théorique, CNRS UMR 7644
F-91128 Palaiseau Cedex, France
jeanrene@cpht.polytechnique.fr
P. Collet
Centre de Physique Théorique, CNRS UMR 7644
F-91128 Palaiseau Cedex, France
collet@cpht.polytechnique.fr
C. Külske
Department of Mathematics and Computing Sciences
University of Groningen, Blauwborgje 3
9747 AC Groningen, The Netherlands
kuelske@math.rug.nl
F. Redig
Faculteit Wiskunde en Informatica and Eurandom
Technische Universiteit Eindhoven, Postbus 513
5600 MB Eindhoven, The Netherlands
f.h.j.redig@tue.nl

Abstract

We present a new and simple approach to deviation inequalities for non-product measures, i.e., for dependent random variables. Our method is based on coupling. We illustrate our abstract results with chains with complete connections and Gibbsian random fields, both at high and low temperature.

Keywords and phrases: exponential deviation inequality, moment inequality, coupling matrix, Gibbsian random fields, chains with complete connections.

1 Introduction

By now, deviation and concentration inequalities for product measures have become a standard and powerful tool in many areas of probability and statistics, such as density estimation [5], geometric probability [20], etc. A recent monograph about this area is [11] where the reader can find much more information. Deviation inequalities for dependent, strongly mixing random variables were obtained for instance in [18, 17]. Later, in the context of dynamical systems Collet et al. [3] obtained an exponential deviation inequality using spectral analysis of the transfer operator. In [10], C. Külske obtained an exponential deviation inequality in the context of Gibbs random fields in the Dobrushin uniqueness regime. Therein the main input is Theorem 8.20 in [7] which allows to estimate uniformly the terms appearing in the martingale difference decomposition in terms of the Dobrushin matrix. Besides exponential deviation inequalities, moment inequalities have been obtained in, e.g., [2, 4, 5, 17]. In the dependent case, we also mention that K. Marton [13, 14, 15] obtained concentration inequalities based on "distance-divergence" inequalities and coupling (with a different approach than ours). In particular, she obtains in [14] results for a class of Gibbs random fields under a strong mixing condition close to Dobrushin-Shlosman condition. Let us notice that this method of "distance-divergence" inequalities inherently implies exponential deviation inequalities for Lipschitz functions (wrt to Hamming distance for instance).

In the present paper, we obtain abstract deviation inequalities using a coupling approach. We prove an upper bound for the probability of deviation from the mean for a general function of n variables, taking values in a finite alphabet, in term of a "coupling matrix" D. This matrix expresses how "well" one can couple in the far "future" if the "past" is given. If the coupling matrix can be uniformly controlled in the realization then an exponential deviation inequality follows. If the coupling matrix cannot be controlled uniformly in the realization then typically upper bounds for the moments are derived.

As a first application of our abstract inequalities, we obtain an exponential deviation inequality for Gibbsian random fields in a "high" temperature regime complementary to the Dobrushin uniqueness regime, and for chains with complete connections with a summable continuity rate. A second application is in the context of the low-temperature Ising model where we obtain upper bounds for the moments of a general local function. This is a typical situation where the coupling matrix cannot be controlled uniformly in the realization. Our deviation inequalities are new and yield various non-trivial applications which will the subject of a forthcoming paper.

The paper is organized as follows. In Section 2, we state and prove our abstract inequalities, first in the context of random processes, and next in the context of random fields. Section 3 deals with the examples of high-temperature Gibbs measures, chains with complete connections, and finally of the low temperature Ising model.

2 Main results

Let A be a finite set. Let $g : A^n \to \mathbb{R}$ be a function of n-variables. An element σ of the set $A^{\mathbb{N}}$ is an infinite sequence drawn from A, i.e., $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_i, \ldots)$ where $\sigma_i \in A$. With a slight abuse of notation, we also consider g as a function on $A^{\mathbb{N}}$ which does not depend on σ_k, for all $k > n$. The variation of g at site i is defined as

$$\delta_i g := \sup_{\sigma_j = \sigma'_j, \forall j \neq i} |g(\sigma) - g(\sigma')| \,.$$

A deviation inequality is an estimate for the probability of deviation of the function g from its expectation, i.e., an estimate for

$$\mathbb{P}\left\{g - \mathbb{E}g \geq t\right\} \tag{1}$$

for all $n \geq 1$ and all $t > 0$, within a certain class of probability measures $\mathbb{P}$. For example, an *exponential deviation inequality* is obtained by estimating the expectation

$$\mathbb{E}\left[e^{\lambda(g - \mathbb{E}g)}\right]$$

for any $\lambda \in \mathbb{R}$, and using the exponential Chebychev inequality.

However, there are natural examples where the exponential deviation inequality does not hold (see the example of the low-temperature Ising model below). In that case we are interested in bounding moments of the form

$$\mathbb{E}\left[(g - \mathbb{E}g)^{2p}\right]$$

to control the probability (1).

In this section, we use a combination of the classical martingale decomposition of $g - \mathbb{E}g$ and optimal coupling to perform a further telescoping which is adequate for the dependent case. This will lead us to a "coupling matrix" depending on the realization $\sigma \in A^{\mathbb{N}}$. This matrix quantifies how "good" future symbols can be coupled if past symbols are given according to σ. Typically, we have in mind applications to Gibbsian random fields. In that framework, the elements of the coupling matrix can be controlled uniformly in σ in the "high-temperature regime". This uniform control leads naturally to an exponential deviation inequality. At low temperature we can only control the coupling matrix for "good" configurations, but not uniformly. Therefore the exponential deviation fails and instead we will obtain Rosenthal-type inequalities for the moments of $g - \mathbb{E}g$; see e.g. [17] for the case of sums of random variables. Devroye inequality [4] is an example of such an inequality for the second moment (in the i.i.d. case).

2.1 The coupling matrix D

We now present our method. For $i = 1, 2, \ldots, n$, let $\mathcal{F}_i$ be the sigma-field generated by the random variables $\sigma_1, \ldots, \sigma_i$, and $\mathcal{F}_0$ be the trivial sigma-field $\{\emptyset, \Omega\}$. We write

$$g(\sigma_1, \ldots, \sigma_n) - \mathbb{E}g(\sigma_1, \ldots, \sigma_n) = \sum_{i=1}^{n} V_i(\sigma) \tag{2}$$

where

$$V_i(\sigma) := \mathbb{E}[g|\mathcal{F}_i](\sigma) - \mathbb{E}[g|\mathcal{F}_{i-1}](\sigma) =$$

$$\int \mathbb{P}(d\eta_{i+1}\cdots d\eta_n|\sigma_1,\ldots,\sigma_i)\ g(\sigma_1,\ldots,\sigma_i,\eta_{i+1},\ldots,\eta_n)$$

$$-\int \mathbb{P}(d\eta_i\cdots d\eta_n|\sigma_1,\ldots,\sigma_{i-1})\ g(\sigma_1,\ldots,\sigma_{i-1},\eta_i,\eta_{i+1},\ldots,\eta_n) =$$

$$\int \mathbb{P}(d\eta_{i+1}\cdots d\eta_n|\sigma_1,\ldots,\sigma_i)\ g(\sigma_1,\ldots,\sigma_i,\eta_{i+1},\ldots,\eta_n)$$

$$-\int \mathbb{P}(d\eta_i|\sigma_1\ldots,\sigma_{i-1})\int \mathbb{P}(d\eta_{i+1}\cdots d\eta_n|\sigma_1\ldots,\sigma_{i-1}\eta_i)\, g(\sigma_1,\ldots,\sigma_{i-1},\eta_i,\eta_{i+1},\ldots,\eta_n) \le$$

$$\max_{\alpha\in A}\int \mathbb{P}(d\eta_{i+1}\cdots d\eta_n|\sigma_1,\ldots,\sigma_i=\alpha)\ g(\sigma_1,\ldots,\sigma_{i-1},\alpha,\eta_{i+1},\ldots,\eta_n)$$

$$-\min_{\beta\in A}\int \mathbb{P}(d\eta_{i+1}\cdots d\eta_n|\sigma_1,\ldots,\sigma_i=\beta)\ g(\sigma_1,\ldots,\sigma_{i-1},\beta,\eta_{i+1},\ldots,\eta_n)\,.$$

$$=: Y_i(\sigma) - X_i(\sigma)\,. \tag{3}$$

Denote by $\mathbb{P}^{\sigma}_{i,\alpha,\beta} = \mathbb{P}^{\sigma_{<i}}_{i,\alpha,\beta}$ the optimal coupling of the conditional distributions $\mathbb{P}(d\eta_{\ge i+1}|\sigma_1\cdots\sigma_i=\alpha)$ and $\mathbb{P}(d\eta_{\ge i+1}|\sigma_1\cdots\sigma_i=\beta)$, and introduce the (infinite) upper-triangular matrix $D = D^{\sigma}$ defined for $i,j\in\mathbb{N}$ by

$$D_{ii} := 1$$

$$D_{i,i+j} = D^{\sigma}_{i,i+j} := \max_{\alpha,\beta\in A}\mathbb{P}^{\sigma}_{i,\alpha,\beta}\left\{\sigma^{(1)}_{i+j} \neq \sigma^{(2)}_{i+j}\right\}\,. \tag{4}$$

Notice that if the σ_i's are mutually independent, then D is the identity matrix because the conditional distributions $\mathbb{P}(d\eta_{\ge i+1}|\sigma_1\cdots\sigma_i=\alpha)$ and $\mathbb{P}(d\eta_{\ge i+1}|\sigma_1\cdots\sigma_i=\beta)$ are equal. Hence we have a perfect coupling in this case.

We proceed with the following simple telescoping identity:

$$g(\sigma_1,\ldots,\sigma_{i-1},\alpha,\sigma^{(1)}_{i+1},\ldots,\sigma^{(1)}_n) - g(\sigma_1,\ldots,\sigma_{i-1},\beta,\sigma^{(2)}_{i+1},\ldots,\sigma^{(2)}_n) =$$

$$[g(\sigma_1,\ldots,\sigma_{i-1},\alpha,\sigma^{(1)}_{i+1},\ldots,\sigma^{(1)}_n) - g(\sigma_1,\ldots,\sigma_{i-1},\beta,\sigma^{(1)}_{i+1},\ldots,\sigma^{(1)}_n)]+$$

$$[g(\sigma_1,\ldots,\sigma_{i-1},\beta,\sigma^{(1)}_{i+1},\ldots,\sigma^{(1)}_n) - g(\sigma_1,\ldots,\sigma_{i-1},\beta,\sigma^{(2)}_{i+1},\sigma^{(1)}_{i+2},\ldots,\sigma^{(1)}_n)]+$$

$$[g(\sigma_1,\ldots,\sigma_{i-1},\beta,\sigma^{(2)}_{i+1},\sigma^{(1)}_{i+2},\ldots,\sigma^{(1)}_n) - g(\sigma_1,\ldots,\sigma_{i-1},\beta,\sigma^{(2)}_{i+1},\sigma^{(2)}_{i+2},\sigma^{(1)}_{i+3},\ldots,\sigma^{(1)}_n)]$$

$$+\cdots+$$

$$[g(\sigma_1,\ldots,\sigma_{i-1},\beta,\sigma^{(2)}_{i+1},\sigma^{(2)}_{i+2},\ldots,\sigma^{(2)}_{n-1},\sigma^{(1)}_n) - g(\sigma_1,\ldots,\sigma_{i-1},\beta,\sigma^{(2)}_{i+1},\ldots,\sigma^{(2)}_n)]$$

$$=: \sum_{j=0}^{n-i}\nabla^{12}_{i,i+j}g\,.$$

We have the following implications:

$$\sigma^{(1)}_{i+j} = \sigma^{(2)}_{i+j} \Rightarrow \nabla^{12}_{i,i+j}g = 0$$

$$\sigma^{(1)}_{i+j} \neq \sigma^{(2)}_{i+j} \Rightarrow \nabla^{12}_{i,i+j} g \leq \delta_{i+j} g \, .$$

Therefore using (3) and (4) we arrive at the inequalities

$$V_i \leq Y_i - X_i \leq \sum_{j=0}^{n-i} D_{i,i+j} \delta_{i+j} g = (D\delta g)_i \tag{5}$$

where δg denotes the column vector with coordinates $\delta_j g$, for $j = 1, \ldots, n$, and 0 for $j > n$.

2.2 Uniform decay of D: exponential deviation inequality

Let $D^{\infty}_{i,j} := \sup_{\sigma \in A^{\mathbb{N}}} D^{\sigma}_{i,j}$. We assume that

$$\|D^{\infty}\|_2^2 := \sup_{x \in \ell_2(\mathbb{N}), \|x\|_2 = 1} \|D^{\infty} x\|_2^2 < \infty \, . \tag{6}$$

We then have the following theorem.

Theorem 1. *If (6) is valid, then for all $n \in \mathbb{N}$ and all functions $g : A^n \to \mathbb{R}$, we have the inequality*

$$\mathbb{P}\left\{ g - \mathbb{E}g \geq t \right\} \leq e^{-\frac{2t^2}{\|D^{\infty}\|_2^2 \ \|\delta g\|_2^2}} \, . \tag{7}$$

Proof. We have the following lemma which appears in [5].

Lemma 1. *Suppose $\mathcal{F}$ is a sigma-field and Z_1, Z_2, V are random variables such that*

1. $Z_1 \leq V \leq Z_2$
2. $\mathbb{E}(V|\mathcal{F}) = 0$
3. *Z_1 and Z_2 are $\mathcal{F}$-measurable*

Then, for all $\lambda \in \mathbb{R}$, we have

$$\mathbb{E}(e^{\lambda V}|\mathcal{F}) \leq e^{\lambda^2 (Z_2 - Z_1)^2/8} \, . \tag{8}$$

We apply this lemma with $V = V_i$, $\mathcal{F} = \mathcal{F}_{i-1}$, $Z_1 = X_i - \mathbb{E}[g|\mathcal{F}_{i-1}]$, $Z_2 = Y_i - \mathbb{E}[g|\mathcal{F}_{i-1}]$. Remember the inequality

$$Y_i - X_i \leq (D\delta g)_i \, .$$

We obtain

$$\mathbb{E}(e^{\lambda V_i}|\mathcal{F}_{i-1}) \leq e^{\lambda^2 (D\delta g)_i^2/8} \, . \tag{9}$$

Therefore, by successive conditioning, and the exponential Chebychev inequality,

$$\begin{aligned} & \mathbb{P}\left\{ g - \mathbb{E}g \geq t \right\} \leq e^{-\lambda t} \mathbb{E}\left(e^{\lambda \sum_{i=1}^{n} V_i} \right) \\ \leq \quad & e^{-\lambda t} \mathbb{E}\left(\mathbb{E}(e^{\lambda V_n}|\mathcal{F}_{n-1}) e^{\lambda \sum_{i=1}^{n-1} V_i} \right) \\ \leq \quad & \cdots \leq e^{-\lambda t} e^{\lambda^2 \|D^{\infty} \delta g\|_2^2/8} \leq e^{-\lambda t} e^{\lambda^2 \|D^{\infty}\|_2^2 \ \|\delta g\|_2^2/8} \, . \end{aligned} \tag{10}$$

Now choose the optimal $\lambda = 4t/(\|D^\infty\|_2^2 \; \|\delta g\|_2^2)$ to obtain the result. $\square$

From (7) we deduce

$$\mathbb{P}\{|g - \mathbb{E}g| \geq t\} \leq 2e^{-\frac{2t^2}{\|D^\infty\|_2^2 \|\delta g\|_2^2}}.$$

2.3 Non-uniform decay of D: moment inequalities

If the dependence on σ of the elements of the coupling matrix cannot be controlled uniformly, then in many cases we can still control the moments of the coupling matrix. To this aim, we introduce the (non-random, i.e., not depending on σ) matrices

$$\mathcal{D}_{i,j}^{(p)} := \mathbb{E}(D_{i,j}^p)^{1/p} \tag{11}$$

for all $p \in \mathbb{N}$.

A typical example of non-uniformity which we will encounter, for instance in the low-temperature Ising model, is an estimate of the following form:

$$D_{i,i+j}^\sigma \leq \mathbb{1}\{\ell_i(\sigma) \geq j\} + \psi_{i,i+j} \tag{12}$$

where $\psi_{i,i+j}$ does not depend on σ, and where ℓ_i are unbounded functions of σ with a distribution independent of i. The idea is that the matrix elements $D_{i,i+j}$ "start to decay" when $j \geq \ell_i(\sigma)$. The "good" configurations σ are those for which $\ell_i(\sigma)$ is "small".

In the particular case when (12) holds, in principle one still *can* have an exponential deviation inequality provided one is able to bound

$$\mathbb{E}\left(e^{\lambda \sum_{i=1}^n \ell_i^2}\right).$$

However, in the examples given below, the tail of the ℓ_i will be exponential or stretched exponential. Henceforth, we cannot deduce an exponential deviation inequality from these estimates.

We now prove an inequality for the variance of g which is a generalization of an inequality derived in [4] in the i.i.d. case.

Theorem 2. *For all $n \in \mathbb{N}$, all functions $g : A^n \to \mathbb{R}$ we have the inequality*

$$\mathbb{E}\left[(g - \mathbb{E}g)^2\right] \leq \|\mathcal{D}^{(2)}\|_2^2 \; \|\delta g\|_2^2. \tag{13}$$

Proof. We start again from the decomposition (2). Recall the fact that $\mathbb{E}[V_i|\mathcal{F}_j] = 0$ for all $i > j$, from which it follows that $\mathbb{E}[V_i V_j] = 0$ for $i \neq j$.

Using (5) and Cauchy-Schwarz inequality we obtain

$$
\begin{aligned}
\mathbb{E}\left[(g-\mathbb{E}g)^2\right] &= \mathbb{E}\sum_{i=1}^n V_i^2 \\
&\leq \mathbb{E}\left(\sum_i^n (D\delta g)_i^2\right) \\
&= \sum_{i=1}^n\sum_{k=1}^n\sum_{l=1}^n \mathbb{E}\left(D_{i,k}D_{i,l}\right)\delta_k g\delta_l g \\
&\leq \sum_{i=1}^n\sum_{k=1}^n\sum_{l=1}^n \mathbb{E}\left(D_{i,k}^2\right)^{\frac{1}{2}}\mathbb{E}\left(D_{i,l}^2\right)^{\frac{1}{2}}\delta_k g\delta_l g \\
&= \|\mathcal{D}^{(2)}\delta g\|_2^2 \\
&\leq \|\mathcal{D}^{(2)}\|_2^2\,\|\delta g\|_2^2\,.
\end{aligned}
$$

□

Remark 1. *In the i.i.d. case, the coupling matrix D is the identity matrix. Hence inequality* (13) *reduces to*

$$\mathbb{E}\left[(g-\mathbb{E}g)^2\right] \leq \|\delta g\|_2^2$$

which the analogue of Devroye inequality [4].

In case (12) holds, we have the following proposition.

Proposition 1. *Assume that there exists $\epsilon > 0$ such that $\mathbb{E}(\ell_0^{2+\epsilon}) < \infty$, and assume moreover that $\|\psi\|_2 < \infty$. Then $\|\mathcal{D}^{(2)}\|_2 < \infty$.*

Proof. Let $P_{i,i+j} := \mathbb{E}(\mathbb{1}\{\ell_i \geq j\})^{\frac{1}{2}} = \mathbb{P}(\ell_0 \geq j)^{\frac{1}{2}}$, where we used that the distribution of ℓ_i is independent of i by assumption. It suffices to prove that $\|P\|_2 < \infty$. Since

$$\|P\|_2^2 \leq \|P\|_1\|P\|_\infty$$

it suffices to prove that $\|P\|_1, \|P\|_\infty < \infty$. We have

$$
\begin{aligned}
\|P\|_\infty &= \sup_i \sum_j j^{\frac{1+\epsilon}{2}}\mathbb{P}(\ell_i \geq j)^{\frac{1}{2}} j^{-\frac{1+\epsilon}{2}} \qquad (14)\\
&\leq \left(\sum_j j^{1+\epsilon}\mathbb{P}(\ell_0 \geq j)\right)^{\frac{1}{2}}\left(\sum_j j^{-(1+\epsilon)}\right)^{\frac{1}{2}} = C_\epsilon\,\mathbb{E}(\ell_0^{2+\epsilon})
\end{aligned}
$$

where $C_\epsilon > 0$. We have for the other norm:

$$
\begin{aligned}
\|P\|_1 &= \sup_j \sum_i |P_{i,j}| = \sup_j \sum_{i\leq j} |P_{i,j}| \qquad (15)\\
&= \sup_j \sum_{i\leq j} \mathbb{P}(\ell_i \geq j-i)^{\frac{1}{2}} = \sup_j \sum_{i\leq j} \mathbb{P}(\ell_0 \geq j-i)^{\frac{1}{2}} \\
&\leq C_\epsilon\mathbb{E}(\ell_0^{2+\epsilon})\,.
\end{aligned}
$$

□

Notice that this proposition can also be proved using Young's inequality since P is a convolution operator.

We now turn to higher moment estimates. We have the following theorem.

Theorem 3. *There exists a constant $C > 0$ such that for all $n \in \mathbb{N}$, all functions $g : A^n \to \mathbb{R}$, for any $p \in \mathbb{N}$, we have*

$$\mathbb{E}\left[(g - \mathbb{E}g)^{2p}\right] \leq C^p \; p^{2p} \; \|\mathcal{D}^{(2p)}\|_2^{2p} \|\delta g\|_2^{2p} \, .$$

Proof. We start from (2) and get

$$\mathbb{E}\left[(g - \mathbb{E}g)^{2p}\right] = \sum_{i_1} \cdots \sum_{i_{2p}} \mathbb{E}\left(V_{i_1} \cdots V_{i_{2p}}\right) \, .$$

This sum can be estimated by applying the martingale version of the Marcinkiewicz-Zygmund inequality [19, Theorem 3.3.6] since $\mathbb{E}[V_i|\mathcal{F}_j] = 0$ for all $i > j$. This gives

$$\mathbb{E}\left[(g - \mathbb{E}g)^{2p}\right] \leq C^p \; p^{2p} \; \mathbb{E}\big[(\sum_i V_i^2)^p\big]$$

where the constant $C^p p^{2p}$ can be deduced from the proof of Theorem 3.3.6 in [19].

We now estimate the rhs by using (5):

$$\begin{aligned}
& \mathbb{E}\;\big[(\sum_i V_i^2)^p\big] && (16)\\
= \;& \sum_{i_1} \cdots \sum_{i_p} \mathbb{E}\left(V_{i_1}^2 \cdots V_{i_p}^2\right) && (17)\\
\leq \;& \sum_{i_1} \cdots \sum_{i_p} \mathbb{E}\left[(D\delta g)_{i_1}^2 \cdots (D\delta g)_{i_p}^2\right] \\
= \;& \sum_{i_1 \cdots i_p} \sum_{j_1 \cdots j_p} \sum_{k_1 \cdots k_p} \mathbb{E}\left(\prod_{r=1}^{p} D_{i_r, j_r} D_{i_r, k_r}\right) \left(\prod_{r=1}^{p} \delta_{j_r} g \; \delta_{k_r} g\right) \\
\leq \;& \sum_{i_1 \cdots i_p} \sum_{j_1 \cdots j_p} \sum_{k_1 \cdots k_p} \prod_{r=1}^{p} \left(\mathcal{D}^{(2p)}_{i_r, j_r} \mathcal{D}^{(2p)}_{i_r, k_r} \delta_{j_r} g \; \delta_{k_r} g\right) \\
= \;& \|\mathcal{D}^{(2p)} \delta g\|_{2p}^{p} \leq \|\mathcal{D}^{(2p)}\|_2^{2p} \; \|\delta g\|_2^{2p}
\end{aligned}$$

where in the fourth step we used the inequality

$$\mathbb{E}(f_1 \cdots f_{2p}) \leq \prod_{i=1}^{2p} (\mathbb{E}(f_i^{2p}))^{\frac{1}{2p}}$$

which follows from Hölder inequality. □

In the situation when (12) holds, we have the following proposition.

Proposition 2. *Let $p \in \mathbb{N}$. Assume that there exists $\epsilon > 0$ such that $\mathbb{E}(\ell_0^{2p+\epsilon}) < \infty$, and that there exist a constant $c > 0$ and $0 < \alpha \leq 1$ such that $\psi_{i,i+j} < e^{-cj^\alpha}$ for all $i \in \mathbb{N}$. Then $\|\mathcal{D}^{(p)}\|_2 < \infty$.*

Proof. The proof follows the line of the proof of Proposition 1. Now let $P_{i,i+j} = \mathbb{P}(\ell_i \geq j)^{\frac{1}{2p}} = \mathbb{P}(\ell_0 \geq j)^{\frac{1}{2p}}$. It suffices to show that $\|P\|_2 < \infty$. In turn, it is sufficient to prove that $\|P\|_\infty < \infty$ and $\|P\|_1 < \infty$. Let $\epsilon' > 0$ to be fixed later on. We have

$$\begin{aligned} \|P\|_\infty &= \sum_j j^{\frac{2p-1}{2p}(1+\epsilon')}\, \mathbb{P}(\ell_0 \geq j)^{\frac{1}{2p}}\, j^{-\frac{2p-1}{2p}(1+\epsilon')} \qquad (18) \\ &\leq \left(\sum_j j^{(2p-1)(1+\epsilon')}\, \mathbb{P}(\ell_0 \geq j)\right)^{\frac{1}{2p}} \left(\sum_j j^{-1-\epsilon'}\right)^{\frac{2p-1}{2p}} = C'_{\epsilon,p}\, \mathbb{E}(\ell_0^{2p+\epsilon})^{\frac{1}{2p}} \end{aligned}$$

where in the last step we have chosen $\epsilon' = \epsilon/(2p-1)$ and where $C'_{\epsilon,p} > 0$. $\square$

Using Theorem 3, Proposition 2 and Chebychev inequality, we immediately obtain the following deviation inequality:

$$\mathbb{P}\{|g - \mathbb{E}g| > t\} \leq K_p \frac{\|\delta g\|_2^{2p}}{t^{2p}}$$

for all $t > 0$, where $K_p := 2(p(2p-1))^p \|\mathcal{D}^{(2p)}\|_2^{2p}$.

Remark 2. *The assumption on ψ in the proposition is far from being optimal. However, it will be satisfied in all examples below.*

2.4 Random fields

We now present the extension of our previous results to random fields. This requires mainly notational changes. We work with lattice spin systems. The configuration space is $\Omega = \{-,+\}^{\mathbb{Z}^d}$, endowed with the product topology. We could of course take any finite set A instead of $\{-,+\}$. For $\Lambda \subset \mathbb{Z}^d$ and $\sigma, \eta \in \Omega$ we denote $\sigma_\Lambda \eta_{\Lambda^c}$ the configuration coinciding with σ (resp. η) on Λ (resp. Λ^c). For $\sigma \in \Omega$ and $x \in \mathbb{Z}^d$, σ^x denotes the configuration obtained from σ by flipping the spin at x. A local function $g : \Omega \to \mathbb{R}$ is such that there exists a finite subset $\Lambda \subset \mathbb{Z}^d$ such that for all σ, η, ω, $g(\sigma_\Lambda \omega_{\Lambda^c}) = g(\sigma_\Lambda \eta_{\Lambda^c})$.

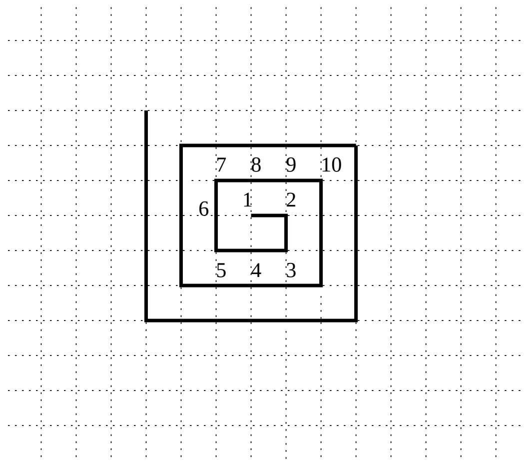
We denote $\delta_x g = \sup_\sigma |g(\sigma_x) - g(\sigma)|$ the variation of g at x. δg denotes the map $\mathbb{Z}^d \to \mathbb{R} : x \mapsto \delta_x g$.

We introduce the spiraling enumeration $\Gamma : \mathbb{Z}^d \to \mathbb{N}$ illustrated in the figure for the case $d = 2$.

We will use the abbreviation $(\leq x) = \{y \in \mathbb{Z}^d : \Gamma(y) \leq \Gamma(x)\}$ and similarly we introduce the abbreviations $(< x)$. By definition $\mathcal{F}_{\leq x}$ denotes the sigma-field generated by $\sigma(y)$, $y \leq x$ and $\mathcal{F}_{<0}$ denotes the trivial sigma-field.

For any local function $g : \Omega \to \mathbb{R}$, we have the analog decomposition as in (2):

$$g - \mathbb{E}(g) = \sum_{x \in \mathbb{Z}^d} V_x \qquad (19)$$



where

$$V_x := \mathbb{E}\left[g|\mathcal{F}_{\leq x}\right] - \mathbb{E}[g|\mathcal{F}_{<x}] \ .$$

The analog of the coupling matrix is the following matrix indexed by lattice sites $x, y \in \mathbb{Z}^d$

$$D_{x,y}(\sigma) := \hat{\mathbb{P}}^{\sigma}_{x,+,-}\left\{X_1(y) \neq X_2(y)\right\} \tag{20}$$

where $\hat{\mathbb{P}}^{\sigma}_{x,+,-}$ denotes the optimal coupling between the conditional measures $\mathbb{P}(\cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\cdot|\sigma_{<x,-_x})$.

We first consider the case of uniform decay of D. In that case, the exponential deviation inequality of Theorem 1 holds with the norm of $\ell_2(\mathbb{Z}^d)$, i.e., $\|\delta g\|_2^2 = \sum_{x\in\mathbb{Z}^d}(\delta_x g)^2$.

Theorem 1' *Assume that*

$$D^{\infty}_{x,y} := \sup_{\sigma} D_{x,y}(\sigma) \tag{21}$$

is a bounded operator in $\ell_2(\mathbb{Z}^d)$. Then for all local functions g we have the following inequality

$$\mathbb{P}\left\{g - \mathbb{E}g \geq t\right\} \leq e^{-\frac{2t^2}{\|D^{\infty}\|_2^2\ \|\delta g\|_2^2}} \ . \tag{22}$$

In the non-uniform case, the moment inequalities of Theorems 2 and 3 extend immediately as follows. The analog of (12) is

$$D_{x,y}(\sigma) \leq 1\!\!1\{\ \ell_x(\sigma) \geq |y-x|\} + \psi(|y-x|) \tag{23}$$

where $\psi(n)$ decays at least as a stretched exponential, i.e., there exist $C, c > 0$ and $0 < \alpha \leq 1$, such that $\psi(n) \leq Ce^{-cn^{\alpha}}$ for all $n \geq 1$. We assume that the distribution of ℓ_x is independent of x. We extend the matrix $\mathcal{D}$ defined in (11) by putting

$$\mathcal{D}^{(p)}_{x,y} := \mathbb{E}(D^p_{x,y})^{1/p}$$

for $x, y \in \mathbb{Z}^d$.

Theorem 2' *For all local functions g, for any $p \in \mathbb{N}$, we have*

$$\mathbb{E}\left[(g - \mathbb{E}g)^{2p}\right] \leq (p(2p-1))^p \|\mathcal{D}^{(2p)}\|_2^{2p} \|\delta g\|_2^{2p}.$$

The analog of Propositions 1 and 2 is the following:

Proposition 2' *Let $p \in \mathbb{N}$. Assume (23) is satisfied and that there exists $\epsilon > 0$ such that $\mathbb{E}(\ell_0^{2dp+\epsilon}) < \infty$. Then $\|\mathcal{D}^{(2p)}\|_2 < \infty$.*

Remark 3. *It is immediate to extend the previous inequalities to integrable functions g belonging to the closure of the set of local functions with the norm $\|\!|g|\!\| := \|\delta g\|_2$. Notice that Theorem 1' implies that such functions are $L^p(\mathbb{P})$ for any $p \in \mathbb{N}$.*

2.5 Existence of the coupling by bounding the variation

We continue with random fields and state a proposition which says that if we have an estimate of the form

$$V_x \leq (D\delta g)_x$$

for some matrix D, then there exists a coupling with coupling matrix $\hat{D}$ such that its matrix elements decay at least as fast as the matrix elements of D. We formulate the proposition more abstractly:

Proposition 3. *Suppose that $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on Ω and $g : \Omega \to \mathbb{R}$ such that we have the estimate*

$$|\mathbb{E}_{\mathbb{P}}[g] - \mathbb{E}_{\mathbb{Q}}[g]| \leq \sum_{x \in \mathbb{Z}^d} \rho(x) \delta_x g \tag{24}$$

for some "weights" $\rho : \mathbb{Z}^d \to \mathbb{R}^+$. Suppose $\varphi : \mathbb{Z}^d \to \mathbb{R}^+$ is such that

$$\sum_{x \in \mathbb{Z}^d} \rho(x)\varphi(x) < \infty.$$

Then there exists a coupling $\hat{\mu}$ of $\mathbb{P}$ and $\mathbb{Q}$ such that

$$\sum_{x \in \mathbb{Z}^d} \hat{\mu}\left\{X_1(x) \neq X_2(x)\right\} \varphi(x) \leq \sum_{x \in \mathbb{Z}^d} \varphi(x)\rho(x) < \infty.$$

Proof. Let $B_n := [-n, n]^d \cap \mathbb{Z}^d$. Define the "cost" function

$$C_n^{\varphi}(\sigma, \sigma') := \sum_{x \in B_n} |\sigma_x - \sigma'_x| \, \varphi(x).$$

Denote by $\mathbb{P}_n$, resp. $\mathbb{Q}_n$, the joint distribution of $\{\sigma_x, x \in B_n\}$ under $\mathbb{P}$, resp. $\mathbb{Q}$. Consider the class of functions

$$\mathcal{G}_{C_n^{\varphi}} := \{g | \; g \in \mathcal{F}_{B_n}, \; |g(\sigma) - g(\sigma')| \leq \sum_{x \in \mathbb{Z}^d} \varphi(x) 1\!\!1\{\sigma_x \neq \sigma'_x\}, \forall \sigma, \sigma' \in \Omega\}.$$

It is obvious from the definition that $g \in \mathcal{G}_{C_n^\varphi}$, if, and only if, g is $\mathcal{F}_{B_n}$-measurable and

$$(\delta_x g)(\sigma) \leq \varphi(x) \quad \forall x \in B_n, \forall \sigma \in \Omega\,.$$

Therefore, if (24) holds, then for all $g \in \mathcal{G}_{C_n^\varphi}$,

$$|\mathbb{E}_{\mathbb{P}}[g] - \mathbb{E}_{\mathbb{Q}}[g]| \leq \sum_{x\in\mathbb{Z}^d} \rho(x)\delta_x g \leq \sum_{x\in\mathbb{Z}^d} \rho(x)\varphi(x)\,.$$

Hence, by the Kantorovich-Rubinstein duality theorem [16], there exists a coupling $\hat{\mu}_n$ of $\mathbb{P}_n$ and $\mathbb{Q}_n$ such that

$$\hat{\mu}_n\left(C_n^\varphi(\sigma,\sigma')\right) = \hat{\mu}_n\left(\sum_{x\in B_n} \varphi(x)1\!\!1\{X_1(x) \neq X_2(x)\}\right) \leq \sum_{x\in\mathbb{Z}^d} \varphi(x)\rho(x)\,.$$

By compactness (in the weak topology), there exists a subsequence along which $\hat{\mu}_n$ converges weakly to some probability measure $\hat{\mu}$. For any $k \leq n$, we have

$$\hat{\mu}_n\left(\sum_{x\in B_k} \varphi(x)1\!\!1\{X_1(x) \neq X_2(x)\}\right) \leq$$

$$\hat{\mu}_n\left(\sum_{x\in B_n} \varphi(x)1\!\!1\{X_1(x) \neq X_2(x)\}\right) \leq \sum_{x\in\mathbb{Z}^d} \varphi(x)\rho(x)\,.$$

Therefore, taking the limit $n \to \infty$ along the above subsequence yields

$$\hat{\mu}\left(\sum_{x\in B_k} \varphi(x)1\!\!1\{X_1(x) \neq X_2(x)\}\right) \leq \sum_{x\in\mathbb{Z}^d} \varphi(x)\rho(x)\,.$$

We now take the limit $k \to \infty$ and use monotonicity to conclude that

$$\hat{\mu}\left(\sum_{x\in\mathbb{Z}^d} \varphi(x)1\!\!1\{X_1(x) \neq X_2(x)\}\right) \leq \sum_{x\in\mathbb{Z}^d} \varphi(x)\rho(x)\,.$$

$\square$

We shall illustrate below this proposition with the example of Gibbs random fields at high-temperature under the Dobrushin uniqueness condition.

3 Examples

3.1 High-temperature Gibbs measures

For the sake of convenience, we briefly recall a few facts about Gibbs measures. We refer to [7] for details.

A finite range potential (with range R) is a family of functions $U(A,\sigma)$ indexed by finite subsets A of $\mathbb{Z}^d$ such that the value of $U(A,\sigma)$ depends only on σ_A and such that $U(A,\sigma)=0$ if $\mathrm{diam}(A)>R$. If $R=1$ then the potential is nearest-neighbor.

The associated finite volume Hamiltonian with boundary condition η is then given by

$$H^\eta_\Lambda(\sigma)=\sum_{A\cap\Lambda\neq\emptyset} U(A,\sigma_\Lambda\eta_{\Lambda^c})\,.$$

The specification is then defined as

$$\gamma_\Lambda(\sigma|\eta)=\frac{e^{-H^\eta_\Lambda(\sigma)}}{Z^\eta_\Lambda}\,.$$

We then say that $\mathbb{P}$ is Gibbs measure with potential U if $\gamma_\Lambda(\sigma|\cdot)$ is a version of the conditional probability $\mathbb{P}(\sigma_\Lambda|\mathcal{F}_{\Lambda^c})$.

Before we state our result, we need some notions from [6]. What we mean by "high temperature" will be an estimate on the variation of single-site conditional probabilities, which will imply a uniform estimate for disagreement percolation. Let

$$p_x := 2\sup_{\sigma,\sigma'}\left|\mathbb{P}(\sigma_x=+|\sigma_{\mathbb{Z}^d\setminus x})-\mathbb{P}(\sigma'_x=+|\sigma'_{\mathbb{Z}^d\setminus x})\right|\,.$$

From [6, Theorem 7.1] it follows that there exists a coupling $\mathbb{P}^\sigma_{x,+,-}$ of the conditional distributions $\mathbb{P}(\cdot|\sigma_{<x},+_x)$ and $\mathbb{P}(\cdot|\sigma_{<x},-_x)$ such that under this coupling

1. For $x>y$, the event $X_1(y)\neq X_2(y)$ coincides with the event that there exists a path $\gamma\subset\mathbb{Z}^d\setminus(<x)$ from x to y such that, for all $z\in\gamma$, $X_1(z)\neq X_2(z)$. We denote this event by "$x \not\leftrightarrow y$".

2. The distribution of $\mathbb{1}\{X_1(y)\neq X_2(y)\}$ for $y\in\mathbb{Z}^d\setminus(\leq x)$ under $\mathbb{P}^\sigma_{x,+,-}$ is dominated by the product measure

$$\prod_{y\in\mathbb{Z}^d\setminus(\leq x)}\nu_{p_y}\,.$$

Let $p_c=p_c(d)$ be the critical percolation threshold for site-percolation on $\mathbb{Z}^d$. It then follows from statements 1 & 2 above that, if

$$\sup\{p_y : y\in\mathbb{Z}^d\}<p_c \tag{25}$$

then we have the uniform estimate

$$\mathbb{P}^\sigma_{x,+,-}\{X_1(y)\neq X_2(y)\}\leq\prod_{y\in\mathbb{Z}^d\setminus(\leq x)}\nu_{p_y}(x \not\leftrightarrow y)\leq e^{-c|x-y|}\,. \tag{26}$$

Then we can apply Theorem 1' to obtain

Theorem 4. *Let U be a nearest-neighbor potential such that* (25) *holds. Then for the coupling matrix* (20) *we have the uniform estimate*

$$D_{x,y}(\sigma)\leq e^{-C|x-y|}$$

for some $C>0$. Hence we have the following exponential deviation inequality: for any local function g and for all $t>0$

$$\mathbb{P}\left\{g-\mathbb{E}g\geq t\right\}\leq e^{-\frac{2t^2}{\|D^\infty\|_2^2\;\|\delta g\|_2^2}}\,.$$

Remark 4. *Theorem 4 can easily be extended to any finite range potential.*

Theorem 4 was obtained in [10] in the Dobrushin uniqueness regime using a different approach. The high-temperature condition which we use here is sometimes less restrictive than Dobrushin uniqueness condition, but sometimes it is more restrictive. However, Dobrushin uniqueness condition is not limited to finite range potentials. We now apply Proposition 3 to show that in the Dobrushin uniqueness regime, there does exist a coupling of $\mathbb{P}(\cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\cdot|\sigma_{<x,-_x})$ such that the elements of the associated coupling matrix decay at least as fast as the elements of the Dobrushin matrix. The Dobrushin uniqueness condition is based on the matrix

$$C_{x,y}:=2\sup_{\sigma,\sigma':\sigma_{\mathbb{Z}^d\setminus y}=\sigma'_{\mathbb{Z}^d\setminus y}}\left|\mathbb{P}(\sigma_x=+|\sigma_{\mathbb{Z}^d\setminus x})-\mathbb{P}(\sigma_x=+|\sigma'_{\mathbb{Z}^d\setminus x})\right|\,.$$

This condition is defined by requiring that

$$\sup_{x\in\mathbb{Z}^d}\sum_{y\in\mathbb{Z}^d}C_{x,y}<1$$

and the Dobrushin matrix is then defined as

$$\Delta_{x,y}:=\sum_{n\geq 0}C^n_{x,y}\,.$$

We now have the following proposition:

Proposition 4. *Assume that the Dobrushin uniqueness condition holds. For any $\varphi:\mathbb{Z}^d\to\mathbb{R}^+$ such that for any $x\in\mathbb{Z}^d$,*

$$\sum_{y\in\mathbb{Z}^d}\varphi(y)\Delta_{y,x}<\infty$$

then there exists a coupling $\hat{\mathbb{P}}^\sigma_{x,+,-}$ of $\mathbb{P}(\cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\cdot|\sigma_{<x,-_x})$ such that

$$\sum_{y\in\mathbb{Z}^d}\varphi(y)\;\hat{\mathbb{P}}^{\sigma_{<x}}_{x,+,-}\left\{X_1(y)\neq X_2(y)\right\}<\infty\,.$$

Proof. From [10, Lemma 1], we have the estimate

$$\left|\int\mathbb{P}(d\eta|\sigma_{<x,+_x})g(\eta)-\int\mathbb{P}(d\eta|\sigma_{<x,-_x})g(\eta)\right|\leq\sum_{y\in\mathbb{Z}^d}(\mathbb{1}_{x,y}+\Delta_{y,x})\delta_y g$$

(where $\mathbb{1}_{x,y}$ denotes the Kronecker symbol).

We can apply Proposition 3 to conclude the proof. $\square$

As an example we mention that if the potential is finite-range and translation-invariant and satisfies the Dobrushin uniqueness condition, we have for large enough $|x-y|$

$$\Delta_{y,x} \leq e^{-c|x-y|}$$

and hence there exists a coupling $\hat{\mathbb{P}}^{\sigma_{<x}}_{x,+,-}$ such that

$$\hat{\mathbb{P}}^{\sigma_{<x}}_{x,+,-}\left\{X_1(y) \neq X_2(y)\right\} \leq e^{-c'|x-y|}$$

for all $c' < c$ and large enough $|x-y|$.

Unhappily, we are not able to construct explicitly such a coupling.

3.2 Chains with complete connections

Here we deal with a class of chains with complete connections and use a coupling estimate proved in [9]. Let A be a finite set (the alphabet). A chain with complete connections $(\sigma_j)_{j\in\mathbb{Z}}$, $\sigma_j \in A$, distributed according to $\mathbb{P}$ has the property that the sequence defined as

$$\varepsilon_n := \sup_{k\in\mathbb{Z}} \sup_{\sigma,\xi\in A^{\mathbb{Z}}} \left| \mathbb{P}_k(\sigma_k|\sigma_{-\infty}^{k-1}) - \mathbb{P}_k(\sigma_k|\sigma_{k-n}^{k-1}\xi_{-\infty}^{k-n-1}) \right|$$

converges to 0 as n tend to ∞. This sequence is called the continuity rate of the chain. We further assume that the continuity rate is summable, i.e.

$$\sum_n \varepsilon_n < \infty\,. \tag{27}$$

We also assume the following non-nullness condition to hold:

$$\inf_{k\in\mathbb{Z}} \inf_{\sigma\in A^{\mathbb{Z}}} \mathbb{P}_k(\sigma_k|\sigma_{-\infty}^{k-1}) =: \vartheta > 0\,. \tag{28}$$

For this class of chain with complete connections we have the following exponential deviation inequality.

Theorem 5. *Assume that $(\sigma_j)_{j\in\mathbb{Z}}$ is a chain with complete connections such that (27) and (28) hold. Then there exists a constant $C>0$ such that for all $n\in\mathbb{N}$, all functions $g : A^n \to \mathbb{R}$ and all $t>0$, we have the estimate*

$$\mathbb{P}\left\{g - \mathbb{E}g \geq t\right\} \leq e^{-\frac{Ct^2}{\|\delta g\|_2^2}}\,.$$

Proof. The theorem will be proved if the assumption of Theorem 1, i.e. (6), is satisfied by our class of chains with complete connections. But the proof of the main theorem in [9] contains an estimate which immediately implies that

$$\sup_{i\in\mathbb{N}} D_{i,i+j}^{\infty} \leq (1-\vartheta)^j$$

which obviously implies (6) since $\vartheta > 0$. $\square$

3.3 The low-temperature Ising model

It is clear that for the Ising model in the phase coexistence region, no exponential deviation inequalities can hold. Indeed, this would contradict the surface-order large deviations for the magnetization in that regime (see e.g. [8] and reference therein). Nevertheless, we shall show that we can control the moments of all local functions.

We consider the low-temperature plus phase of the Ising model on $\mathbb{Z}^d$, $d \geq 2$. This is a probability measure $\mathbb{P}_\beta^+$ on lattice spin configurations $\sigma \in \Omega$, defined as the weak limit as $\Lambda \uparrow \mathbb{Z}^d$ of the following finite volume measures:

$$\mathbb{P}^+_{\Lambda,\beta}(\sigma_\Lambda) = \exp\left(\beta \sum_{<xy>\in\Lambda} \sigma_x\sigma_y + \beta \sum_{<xy>, x\in\partial\Lambda,\ y\notin\Lambda} \sigma_x\right) \Big/ Z^+_{\Lambda,\beta} \tag{29}$$

where $\beta \in \mathbb{R}^+$, and $Z^+_{\Lambda,\beta}$ is the partition function. In (29) $< xy >$ denotes nearest neighbor bonds and $\partial\Lambda$ the inner boundary, i.e. the set of those $x \in \Lambda$ having at least one neighbor $y \notin \Lambda$. The existence of the limit $\Lambda \uparrow \mathbb{Z}^d$ of $\mathbb{P}^+_{\Lambda,\beta}$ is by a standard and well-known monotonicity argument, see e.g. [7].

For any $\eta \in \Omega$, $\Lambda \subset \mathbb{Z}^d$ we denote by $\mathbb{P}^\eta_{\Lambda,\beta}$ the corresponding finite volume measure with boundary condition η:

$$\mathbb{P}^\eta_{\Lambda,\beta}(\sigma_\Lambda) = \exp\left(\beta \sum_{<xy>\in\Lambda} \sigma_x\sigma_y + \beta \sum_{x\in\Lambda,\ y\notin\Lambda} \sigma_x\eta_x\right) \Big/ Z^\eta_{\Lambda,\beta}\,.$$

Later on we will have to choose β large enough.

We can now formulate our result on moments of arbitrary local functions. We shall show that we can apply Theorem 2' and Proposition 2'.

Theorem 6. *Let $\mathbb{P} = \mathbb{P}^+_\beta$ be the plus phase of the low-temperature Ising model defined above. There exists $\beta_0 > \beta_c$, such that for all $\beta > \beta_0$, for al $p \in \mathbb{N}$, there exists a constant $C_p \in (0,\infty)$ such that for all local functions g, we have*

$$\mathbb{E}\left((g-\mathbb{E}g)^{2p}\right) \leq C_p\ \|\delta g\|_2^{2p}\,.$$

Consequently, for all $t > 0$, we have the deviation inequalities

$$\mathbb{P}\left[|g-\mathbb{E}g| > t\right] \leq C_p\ \frac{\|\delta g\|_2^{2p}}{t^{2p}}\,.$$

Proof. The theorem follows from Theorem 2' and Proposition 2' if we obtain the bound (23) with good decay properties for the tail of the distribution of ℓ_x. This is the content of the following proposition. □

Proposition 5. *Let $\mathbb{P} = \mathbb{P}^+_\beta$ be the plus phase of the low-temperature Ising model. There exists $\beta_0 > \beta_c$ such that for all $\beta > \beta_0$, the inequality (23) holds together with the estimate*

$$\psi(n) \leq Ce^{-cn}$$

for all $n \in \mathbb{N}$ and

$$\mathbb{P}\{\ell_0 \geq n\} \leq C' e^{-c' n^\alpha}$$

for some $c, c', C, C' > 0$ and $0 < \alpha \leq 1$.

Proof. We shall make a coupling of the conditional measures $\mathbb{P}(\cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\cdot|\sigma_{<x,-_x})$. This coupling already appeared in [1] (see also [6]). Both conditional measures are a distribution of a random field ω_y, $y \notin (\leq x)$. We start with the first site $y_1 > x$ according to the order induced by Γ. We generate $X_1(y_1)$ and $X_2(y_1)$ as a realization of the optimal coupling between $\mathbb{P}(\sigma_{y_1} = \cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\sigma_{y_1} = \cdot|\sigma_{<x,-_x})$. Given that we have generated $X_1(y), X_2(y), \ldots, X_1(y_n), X_2(y_n)$ for $y = y_1, \ldots, y_n$, we generate $X_1(y_{n+1})$, $X_2(y_{n+1})$ for the smallest $y_{n+1} > y_n$ as a realization of the optimal coupling between

$$\mathbb{P}(\sigma_{y_{n+1}} = \cdot|X_1(y_1)\cdots X_1(y_n)\sigma_{<x,+_x}) \text{ and } \mathbb{P}(\sigma_{y_{n+1}} = \cdot|X_2(y_1)\cdots X_2(y_n)\sigma_{<x,-_x})\,.$$

By the Markov property of $\mathbb{P}$ we have the following: if there exists a contour separating y from x such that for all sites z belonging to that contour we have $X_1(z) = X_2(z)$, then $X_1(y) = X_2(y)$. The complement of this event (of having such a contour) is contained in the event that there exists a path of disagreement from x to y, i.e., a path $\gamma \subset \mathbb{Z}^d \setminus (< x)$ such that for all $z \in \gamma$, $X_1(z) \neq X_2(z)$. Denote that event by E_{xy}. Clearly its probability is bounded from above by the probability of the same event in the product coupling. In turn the event E_{xy} is contained in the event E^+_{xy} that there exists a path γ from x to y in $\mathbb{Z}^d \setminus (< x)$ such that for all $z \in \gamma$, $(X_1(z), X_2(z)) \neq (+,+)$. In [12] the probability of that event in the product coupling is precisely estimated from above by

$$C e^{-c|x-y|} + \mathbb{1}\{\ell_x(\sigma) \geq |x-y|\} \tag{30}$$

for some $C, c > 0$, where $\ell_x(\sigma)$ is an unbounded function of σ with tail estimate

$$\mathbb{P}(\ell_x(\sigma) \geq n) = \mathbb{P}(\ell_0(\sigma) \geq n) \leq C' e^{-c' n^\alpha}$$

for some $C', c' > 0$ and $0 < \alpha < 1$. For the reader's convenience, we briefly comment on these estimates. The ideas is that the conditional measure $\mathbb{P}(\cdot|\xi_{\leq x})$ resembles the original unconditioned plus phase (in $\mathbb{Z}^d \setminus (\leq x)$) provided ξ contains "enough" pluses. "Containing enough pluses" is exactly quantified by the random variable $\ell_x(\xi)$: $(\ell_x(\xi) \leq n)$ is the event that for all self-avoiding path γ of length at least n, the magnetization along γ,

$$m_\gamma(\xi) := \frac{1}{|\gamma|} \sum_{z \in \gamma} \xi_z$$

is close "enough to one". If this is the case then under the conditional measure we still have a Peierls' estimate, which produces the exponential term in (30). We refer to [12] for more details. $\square$

References

[1] J. van den Berg, C. Maes, *Disagreement percolation in the study of Markov fields*, Ann. Probab. **22**, 749–763 (1994).

[2] J.-R. Chazottes, P. Collet, B. Schmitt, *Devroye inequality for a class of non-uniformly hyperbolic dynamical systems*, preprint, 2005.

[3] P. Collet, S. Martínez, B. Schmitt, *Exponential inequalities for dynamical measures of expanding maps of the interval*, Probab. Theor. Rel. Fields **123**, 301–322 (2002).

[4] L. Devroye, *Exponential inequalities in nonparametric estimation*, Nato ASI series C, Math. Phys. Sci 335, p. 31-44, Kluwer Academic Publishers (1991).

[5] L. Devroye and G. Lugosi, *Combinatorial methods in density estimation*, Springer Series in Statistics, Springer, New York (2001).

[6] H.O. Georgii, O. Häggström and C. Maes, *The Random Geometry of Equilibrium Phases* in Phase transitions and critical phenomena, Vol. 18, Eds. C. Domb and J.L. Lebowitz, 1–142, Academic Press London (2001)

[7] H.-O. Georgii. *Gibbs Measures and Phase Transitions*. Walter de Gruyter & Co., Berlin, 1988.

[8] D. Ioffe, *Exact large deviation bounds up to T_c for the Ising model in two dimensions*, Probab. Theory Related Fields **102** (1995), no. 3, 313–330.

[9] M. Iosifescu, *A coupling method in the theory of dependence with complete connections according to Doeblin*, Rev. Roumaine Math. Pures Appl. **37** (1992), no.1, 59–66.

[10] C. Külske, *Concentration inequalities for functions of Gibbs fields with applications to diffraction and random Gibbs measures*, Comm. Math. Phys. **239**, 29–51 (2003).

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

[12] C. Maes, F. Redig, S. Shlosman and A. Van Moffaert, *Percolation, Path large deviations and weak Gibbsianity*, Comm. Math. Phys. **209**, 517–545 (2000).

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

[14] K. Marton, *Measure concentration and strong mixing*, Studia Sci. Math. Hungar. **40** (2003), no. 1-2, 95–113.

[15] K. Marton, *Measure concentration for Euclidean distance in the case of dependent random variables*, Ann. Probab. **32** (2004), no. 3B, 2526–2544.

[16] S.T. Rachev, *Probability metrics and the stability of stochastic models.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley and Sons, Ltd., Chicheter, 1991.

[17] E. Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*, Mathématiques & Applications **31**, Springer, 2000.

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

[19] W. Stout, *Almost sure convergence*, Probability and Mathematical Statistics **24**, Academic Press, New York-London, 1974.

[20] J. E. Yukich, *Probability theory of classical Euclidean optimization problems*, Lecture Notes in Mathematics **1675**, Springer-Verlag, Berlin, 1998.