Combinatorics, Probability and Computing (2019), **28**, pp. 919–935
doi:10.1017/S0963548319000099

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Logarithmic Sobolev inequalities in discrete product spaces

Katalin Marton[†]

H-1364 POB 127, Budapest, Hungary
Email: marton@renyi.hu

(Received 1 October 2015; revised 30 December 2018; first published online 13 June 2019)

Abstract

The aim of this paper is to prove an inequality between relative entropy and the sum of average conditional relative entropies of the following form: for a fixed probability measure q on $\mathcal{X}^n$, ($\mathcal{X}$ is a finite set), and any probability measure $p = \mathcal{L}(Y)$ on $\mathcal{X}^n$,

$$D(p\|q) \leqslant C \cdot \sum_{i=1}^{n} \mathbb{E}_p D(p_i(\cdot\,|Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)\|q_i(\cdot\,|Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)), \qquad (*)$$

where $p_i(\cdot\,|y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ and $q_i(\cdot\,|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ denote the local specifications for p resp. q, that is, the conditional distributions of the ith coordinate, given the other coordinates. The constant C depends on (the local specifications of) q.

The inequality ($*$) is meaningful in product spaces, in both the discrete and the continuous case, and can be used to prove a logarithmic Sobolev inequality for q, provided uniform logarithmic Sobolev inequalities are available for $q_i(\cdot\,|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, for all fixed i and fixed $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. Inequality ($*$) directly implies that the Gibbs sampler associated with q is a contraction for relative entropy.

In this paper we derive inequality ($*$), and thereby a logarithmic Sobolev inequality, in discrete product spaces, by proving inequalities for an appropriate Wasserstein-like distance.

2010 MSC Codes: Primary 60J10; Secondary 52A40, 82C22

1. Introduction and statement of some results

Let $\mathcal{X}$ be a finite set, and $\mathcal{X}^n$ the set of n-length sequences from $\mathcal{X}$. Let $\mathcal{P}(\mathcal{X}^n)$ denote the space of probability measures on $\mathcal{X}^n$. For a sequence $x \in \mathcal{X}^n$ we let x_i denote the ith coordinate of x.

We consider a reference probability measure $q \in \mathcal{P}(\mathcal{X}^n)$ which will be fixed throughout Sections 1–3. In Section 4 we still consider a fixed reference measure denoted by q, with suitable subscripts.

The aim of this paper is to prove logarithmic Sobolev inequalities for measures on discrete product spaces, by proving inequalities for an appropriate Wasserstein-like distance. A logarithmic Sobolev inequality is, roughly speaking, a contractivity property of relative entropy with respect to some Markov semigroup. It is much easier to prove contractivity for a distance between measures than for relative entropy, since a distance satisfies the triangle inequality. Our

[†]This work was supported by grant OTKA K 105840 of the Hungarian Academy of Sciences and by National Research, Development and Innovation Office NKFIH K 120706.

method will be used to prove logarithmic Sobolev inequalities for measures satisfying a version of Dobrushin's uniqueness condition, as well as Gibbs measures satisfying a strong mixing condition.

To explain the results, we need some definitions and notation.

Notation 1.1. If r and s are two probability measures (on any measurable space) then we let $|r-s|$ denote their variational distance:

$$|r-s| = \sup_A |r(A) - s(A)|.$$

Definition 1.2 (**W_2 distance; see Theorem 8.2 of [1]**). For probability measures $r, s \in \mathcal{P}(\mathcal{X}^n)$, let Z and U represent r resp. s, that is, Z and U are random sequences with distributions $\mathcal{L}(Z) = r$ and $\mathcal{L}(U) = s$, respectively. We define

$$W_2(r,s) = \min_\pi \sqrt{\sum_{i=1}^n \mathbb{P}_\pi\{Z_i \neq U_i\}^2},$$

where the minimum is taken over all joint distributions $\pi = \mathcal{L}(Z, U)$ with marginals r and s.

Note that W_2 is a distance on $\mathcal{P}(\mathcal{X}^n)$, but it cannot be defined by taking the minimum expectation of some distance on $\mathcal{X}^n$.

Definition 1.3 (**relative entropy, conditional relative entropy**). For probability measures r and s defined on a finite set $\mathcal{Z}$, we let $D(r\|s)$ denote the *relative entropy* of r with respect to s:

$$D(r\|s) = \sum_{u \in \mathcal{Z}} r(u) \log \frac{r(u)}{s(u)}.$$

(We use the natural logarithm, with the convention $0 \log 0 = 0$ and $a \log 0 = \infty$ for $a > 0$.) If Z and U are random variables with values in $\mathcal{Z}$ and distributed according to $r = \mathcal{L}(Z)$ resp. $s = \mathcal{L}(U)$, then we also use the notation $D(Z\|U)$ for the relative entropy $D(r\|s)$. If, moreover, we are given a probability measure $\pi = \mathcal{L}(S)$ on another finite set $\mathcal{S}$, and conditional distributions $\mu(\cdot\,|s) = \mathcal{L}(Z|S=s)$, $\nu(\cdot\,|s) = \mathcal{L}(U|S=s)$, then we define the *conditional relative entropy*:

$$\mathbb{E}_\pi D(\mu(\cdot\,|S)\|\nu(\cdot\,|S)) \triangleq \sum_{s\in\mathcal{S}} \pi(s) D(\mu(\cdot\,|s)\|\nu(\cdot\,|s)).$$

For $\mathbb{E}_\pi D(\mu(\cdot\,|S)\|\nu(\cdot\,|S))$ we shall use any of the notations

$$\mathbb{E}D(\mu(\cdot\,|S)\|\nu(\cdot\,|S)), \quad \mathbb{E}D(\mu(\cdot\,|S)\|U|S),$$
$$\mathbb{E}D(Z|S)\|\nu(\cdot\,|S)), \quad \mathbb{E}D(Z|S)\|U|S)),$$

where expectation is taken with respect to $\pi = \mathcal{L}(S)$.

Notation 1.4. For $y = (y_1, y_2, \ldots, y_n) \in \mathcal{X}^n$ and $I \subset [1, n]$, we write

$$y_I = (y_k : k \in I) \quad \text{and} \quad \bar{y}_I = (y_k : k \notin I).$$

Moreover, if $p = \mathcal{L}(Y) \in \mathcal{P}(\mathcal{X}^n)$ then

$$p_I \triangleq \mathcal{L}(Y_I), \quad p_I(\cdot\,|\bar{y}_I) \triangleq \mathcal{L}(Y_I \mid \bar{Y}_I = \bar{y}_I), \quad \bar{p}_I \triangleq \mathcal{L}(\bar{Y}_I), \quad \bar{p}_I(\cdot\,|y_I) \triangleq \mathcal{L}(\bar{Y}_I|Y_I = y_I).$$

If $I = \{i\}$, then we write i instead of $\{i\}$.

Definition 1.5. The conditional distributions $q_i(\cdot\,|\bar{x}_i)$ are called the *local specifications* of the distribution q.

Theorem 1.6. *Set*

$$\alpha \triangleq \min q_i(x_i|\bar{x}_i), \tag{1.1}$$

where the minimum is taken over all $x \in \mathcal{X}^n$ satisfying $q(x) > 0$ and all $i \in [1, n]$. Assume that, for any $p \in \mathcal{P}(\mathcal{X}^n)$ satisfying

$$q(x) = 0 \implies p(x) = 0, \tag{1.2}$$

all the inequalities

$$W_2^2\Big(p_I(\cdot\,|\bar{y}_I), q_I(\cdot\,|\bar{y}_I)\Big) \leqslant C \cdot \mathbb{E}\left\{\sum_{i\in I} |p_i(\cdot\,|\bar{Y}_i) - q_i(\cdot\,|\bar{Y}_i)|^2 \,\middle|\, \bar{Y}_I = \bar{y}_I\right\} \tag{1.3}$$

are satisfied, where $I \subset [1, n]$, $\bar{y}_I \in \mathcal{X}^{[1,n]\setminus I}$ is a fixed sequence, and $\mathbb{E}\{\cdot|\cdot\}$ denotes conditional expectation with respect to the conditional distribution $\mathcal{L}(Y_I|\bar{Y}_I)$. Then, for all $p \in \mathcal{P}(\mathcal{X}^n)$ satisfying (1.2),

$$D(p\|q) \leqslant \frac{2C}{\alpha} \cdot \sum_{i=1}^n \mathbb{E}|p_i(\cdot\,|\bar{Y}_i) - q_i(\cdot\,|\bar{Y}_i)|^2 \leqslant \frac{C}{\alpha} \cdot \sum_{i=1}^n \mathbb{E}D\Big(Y_i|\bar{Y}_i\|q_i(\cdot\,|\bar{Y}_i)\Big). \tag{1.4}$$

Note that in (1.3)–(1.4) *expectation is taken with respect to $p = \mathcal{L}(Y)$.*

Condition (1.2) is necessary, since otherwise $D(p\|q)$ could be ∞, while the middle term is always finite. On the other hand, for the inequality between the first and last terms (1.2) is not necessary, since if $D(p\|q) = \infty$ then the last term is ∞ as well.

Remark. In [15] a bound analogous to the bound relating the first and last terms of (1.4) was proved for measures on Euclidean spaces (under reasonable conditions). That bound was used to derive a logarithmic Sobolev inequality, improving on an earlier result in [18]. In the present paper we shall deduce a logarithmic Sobolev inequality from the first inequality in (1.4) (Corollary 1.11 to Theorem 1.6).

Definition 1.7 (**Gibbs sampler**). For $i \in [1, n]$, let $\Gamma_i{:}\mathcal{P}(\mathcal{X}^n) \mapsto \mathcal{P}(\mathcal{X}^n)$ be the Markov kernel

$$\Gamma_i(z|y) = \delta(\bar{y}_i, \bar{z}_i) \cdot q_i(z_i|\bar{y}_i), \qquad y, z \in \mathcal{X}^n,$$

where δ denotes the Kronecker δ (*i.e.* Γ_i leaves all but the ith coordinates unchanged, and updates the ith coordinate according to $q_i(\cdot\,|\bar{y}_i)$). Finally, set

$$\Gamma = \Gamma_q = \frac{1}{n} \cdot \sum_{i=1}^n \Gamma_i.$$

That is, Γ selects an $i \in [1, n]$ at random, and applies Γ_i. It is easy to see that Γ preserves, and is reversible with respect to q. Γ is called the *Gibbs sampler* (or *Glauber dynamics*) governed by the local specifications of q.

Theorem 1.6 implies that the Gibbs sampler defined by the local specifications of q is a strict contraction for the relative entropy $D(p\|q)$, for any measure p satisfying (1.2).

Corollary 1.8. *If $q \in \mathcal{P}(\mathcal{X}^n)$ satisfies the conditions of Theorem 1.6, then, for all $p \in \mathcal{P}(\mathcal{X}^n)$ satisfying* (1.2),

$$D(p\Gamma\|q)\big(= D(p\Gamma_q\|q)\big) \leqslant \left(1 - \frac{\alpha}{nC}\right) \cdot D(p\|q). \tag{1.5}$$

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

Inequality (1.5) follows from Theorem 1.6 by the inequality

$$D(p\Gamma\|q) \leqslant \frac{1}{n}\sum_{i=1}^{n} D(p\Gamma_i\|q)$$

(a consequence of the convexity of relative entropy), together with the identity

$$D(p\|q) - D(p\Gamma_i\|q) = \mathbb{E}D\Big(p_i(\cdot\,|\bar{Y}_i)\|q_i(\cdot\,|\bar{Y}_i)\Big).$$

Theorem 1.6 also implies Gross's logarithmic Sobolev inequality, defined as follows.

Definition 1.9 (**Dirichlet form**). Let $(\mathcal{Z},\pi)$ be a finite probability space, and let $G:\mathcal{Z}\mapsto\mathcal{Z}$ be a Markov kernel with invariant measure π. The *Dirichlet form* associated with G is

$$\mathcal{E}_G(f,f) = \langle(\mathbb{I}-G)f,f\rangle_\pi,$$

where $f:\mathcal{Z}\mapsto\mathbb{R}_+$.

Definition 1.10 (**logarithmic Sobolev inequality for Markov kernels**). We say that G satisfies a *logarithmic Sobolev inequality* with logarithmic Sobolev constant c if, for every probability measure p on $\mathcal{Z}$, we have

$$D(p\|\pi) \leqslant c\cdot\mathcal{E}_G\big(\sqrt{f},\sqrt{f}\big),$$

where $f(z) = p(z)/\pi(z)$.

The property expressed by the logarithmic Sobolev inequality was defined by Leonard Gross [11] in 1975. For an introduction to logarithmic Sobolev inequalities and their manifold interpretations and uses, see [12] and [19].

A simple calculation shows that for any $p\in\mathcal{P}(\mathcal{X}^n)$

$$\mathcal{E}_\Gamma\left(\sqrt{\frac{p}{q}},\sqrt{\frac{p}{q}}\right) = \frac{1}{n}\cdot\mathbb{E}_p\sum_{i=1}^{n}\left(1-\left(\sum_{y_i\in\mathcal{X}}\sqrt{p_i(y_i|\bar{Y}_i)\cdot q_i(y_i|\bar{Y}_i)}\right)^2\right).$$

(Using the fact that, for fixed $\bar{y}_i$, the measure $p\Gamma_i$ does not depend on y_i, we just calculate the Dirichlet form for a matrix with identical rows.) For the Gibbs sampler

$$G = \Gamma_q = \frac{1}{n}\cdot\sum_{i=1}^{n}\Gamma_i$$

and $\mathbb{E}_q f^2 = 1$, the Dirichlet form can also be written in terms of the squared norm of the *discrete gradient* of f:

$$\mathcal{E}_\Gamma(f,f) = \frac{1}{n}\mathbb{E}_q\sum_{k=1}^{n}|\partial_k f|^2,$$

where $\partial_k f = (\mathbb{I}-\Gamma_k)f$. Applying this to

$$f(x) = \sqrt{\frac{p(x)}{q(x)}},$$

we obtain

$$\mathcal{E}_\Gamma\left(\sqrt{\frac{p}{q}},\sqrt{\frac{p}{q}}\right) = \frac{1}{n}\mathbb{E}_q\sum_{k=1}^{n}\left|\partial_k\sqrt{\frac{p}{q}}\right|^2.$$

(It is easy to verify that the above two expressions for $\mathcal{E}_\Gamma(\sqrt{p/q},\sqrt{p/q})$ are equal.)

Theorem 1.6 implies Gross's logarithmic Sobolev inequality for the Gibbs sampler Γ.

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

Corollary 1.11. *If q on $\mathcal{X}^n$ satisfies the conditions of Theorem 1.6 then, for any $p \in \mathcal{P}(\mathcal{X}^n)$,*

$$\begin{aligned}\frac{1}{n}\cdot D(p\|q) &\leqslant \frac{2C}{\alpha}\cdot \mathcal{E}_\Gamma\left(\sqrt{\frac{p}{q}},\sqrt{\frac{p}{q}}\right)\\ &= \frac{2C}{\alpha n}\cdot\sum_{i=1}^{n}\mathbb{E}_p\left(1-\left(\sum_{y_i\in\mathcal{X}}\sqrt{p_i\big(y_i|\bar{Y}_i\big)\cdot q_i\big(y_i|\bar{Y}_i\big)}\right)^2\right)\\ &= \frac{2C}{\alpha n}\mathbb{E}_q\sum_{k=1}^{n}\left|\partial_k\sqrt{\frac{p}{q}}\right|^2. \end{aligned}\tag{1.6}$$

This can be considered a dimension-free logarithmic Sobolev inequality, since Γ only updates one coordinate.

Remark. In [21] the 'Standard Logarithmic Sobolev Inequality', for the Gibbs sampler, is defined using the Dirichlet form defined in terms of the squared norm of discrete gradient (formula SLS).

Corollary 1.11 follows from the first inequality in (1.4) (Theorem 1.6) by the following lemma.

Lemma 1.12 (**Proposition 1 of [19]**). *Let r and s be two probability measures on $\mathcal{X}$. Then*

$$|r-s|^2 \leqslant 1-\left(\sum_{y\in\mathcal{X}}\sqrt{r(y)s(y)}\right)^2.$$

Remark. The inequality between the first and last term in (1.4) also implies an inequality of the form (1.6), but with a slightly worse constant. This follows from Theorem A.1 of [8].

Theorem 1.6 can be applied to distributions q satisfying the following version of *Dobrushin's uniqueness condition.*

Definition 1.13 (**Dobrushin's uniqueness condition**). We define the *coupling matrix* of $q \in \mathcal{P}(\mathcal{X}^n)$ as

$$A=(a_{k,i})_{k,i=1}^n,\quad a_{k,i}=\max_{z,s}|q_i(\cdot\,|\bar{z}_i)-q_i(\cdot\,|\bar{s}_i)|,\tag{1.7}$$

where the max is taken on sequences $z, s \in \mathcal{X}^n$, differing only in the kth coordinate. (Clearly $a_{k,k} = 0$ for all k.) We say that q satisfies (an $\mathbb{L}_2$-version of) Dobrushin's uniqueness condition if

$$\|A\|_2 < 1.$$

Here $\|A\|_2$ denotes the matrix norm corresponding to the Euclidean norm.

This differs from Dobrushin's original uniqueness condition where the maximum column sum of A is assumed to be < 1.

If sequences $z, s \in \mathcal{X}^n$ differ in several coordinates then, by a telescoping argument and the triangle inequality, (1.7) implies

$$|q_i(\cdot\,|\bar{z}_i)-q_i(\cdot\,|\bar{s}_i)| \leqslant \sum_{k\neq i}\delta(z_k,s_k)a_{k,i},\tag{1.7$'$}$$

where δ denotes the Kronecker δ.

Theorem 1.14. *Assume that the measure $q \in \mathcal{P}(\mathcal{X}^n)$ satisfies Dobrushin's uniqueness condition with coupling matrix A, $\|A\|_2 < 1$. Then the conditions of Theorem 1.6 are satisfied with $C = 1/(1-\|A\|)^2$. Thus, for any $p \in \mathcal{P}(\mathcal{X}^n)$ satisfying (1.2),*

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

$$D(p\|q) \leqslant \frac{2}{\alpha} \cdot \frac{1}{(1-\|A\|)^2} \cdot \sum_{i=1}^{n} \mathbb{E}|p_i(\cdot\,|\bar{Y}_i) - q_i(\cdot\,|\bar{Y}_i)|^2$$
$$\leqslant \frac{1}{\alpha} \cdot \frac{1}{(1-\|A\|)^2} \cdot \sum_{i=1}^{n} \mathbb{E}D\Big(Y_i|\bar{Y}_i\|q_i(\cdot\,|\bar{Y}_i)\Big),$$

and

$$D(p\Gamma\|q) \leqslant \left(1 - \frac{1}{n} \cdot \frac{\alpha}{2} \cdot (1-\|A\|)^2\right) \cdot D(p\|q). \tag{1.8}$$

Remark. In [24] a logarithmic Sobolev inequality is proved for discrete spin systems, where the title suggests that it uses Dobrushin's uniqueness condition. However, the condition used there is not identical to Dobrushin's uniqueness condition, just reminiscent of it. Moreover, an inequality of the form relating the first and last terms of (1.4) was proved in [2], assuming a condition only slightly reminiscent of Dobrushin's uniqueness condition.

Theorem 1.6 is proved in Section 2, and Theorem 1.14 in Section 3. In Section 4 we are going to deduce a logarithmic Sobolev inequality from a strong mixing condition, for measures q on $\mathcal{X}^{\mathbb{Z}}$ (under the additional condition that the local specifications $q_k(x_k|x_i, i \neq k)$, if not equal to 0, are bounded from below). The strong mixing condition we use is the same as Dobrushin and Shlosman's strong mixing conditions, but we do not assume that q is a Markov field. Our strong mixing condition can also be considered as a generalization of Φ-mixing for (stationary) probability measures on $\mathcal{X}^{\mathbb{Z}}$. For non-Markov stationary probability measures on $\mathcal{X}^{\mathbb{Z}}$, it is more restrictive than usual strong mixing.

The first proof for the implication that Dobrushin and Shlosman's strong mixing conditions imply a logarithmic Sobolev inequality for Markov fields was given by Stroock and Zegarlinski [22, 23] in 1992. The arguments in [23] are quite hard to follow. In 2001, Cesi [3] proved that Dobrushin and Shlosman's strong mixing conditions imply a logarithmic Sobolev inequality; his approach is quite different from the previous ones, and much simpler.

We feel that there is still room for alternative and perhaps simpler proofs in this important topic. Moreover, our proof is valid without the Markovity assumption. (It may be, however, that the proofs in [23] and [3] can also be generalized for the non-Markovian case, but it has not been tried.)

We believe that the separate parts of our proof (Theorem 1.6 and its applicability) are comprehensible in themselves, thus making the whole proof easier to follow.

There is another approach to strong mixing, for measures q on $\mathcal{X}^{\mathbb{Z}}$ with finite range of interaction, developed by Olivieri, Picco and Martinelli: see [13]. Their aim was to replace condition (4.2) of strong mixing (see below) with a milder one, requiring (4.2) only for 'non-pathological' sets Λ, *i.e.* for sets with boundary much smaller than volume. Martinelli and Olivieri [14] proved a logarithmic Sobolev inequality under this modified condition, for measures q with finite range of interaction. In Appendix B we briefly sketch the Olivieri–Picco–Martinelli approach, and show how to modify Theorem 1.6 and Lemma 4.4 (below), to get logarithmic Sobolev inequalities under this weaker assumption.

2. Proof of Theorem 1.6

We need the following lemma.

Lemma 2.1 **(see [21]).** *Let r and s be two probability measures on $\mathcal{X}$. Set*

$$\alpha_s = \min_{s(x)\neq 0} s(x).$$

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

If $D(r\|s)<\infty$, *then*

$$D(r\|s)\leqslant\frac{2}{\alpha_s}\cdot|r-s|^2. \tag{2.1}$$

Remark. Inequality (2.1) can be considered as a converse to the Pinsker–Csiszár–Kullback inequality, which says that

$$|r-s|^2\leqslant\frac{1}{2}D(r\|s).$$

However, there is no uniform converse: the reverse inequality must depend on the measure s. Note also that in [21] the following stronger inequality is proved:

$$D(r\|s)\leqslant\log\left(1+\frac{2}{\alpha_s}\cdot|r-s|^2\right). \tag{2.1′}$$

However, this logarithmic improvement does not yield any improvement in Theorem 1.6.

We proceed to the proof of Theorem 1.6. Let $\pi=\mathcal{L}(Y,X)$ be a coupling of $p=\mathcal{L}(Y)$ and $q=\mathcal{L}(X)$ that achieves $W_2(p,q)$.

We apply induction on n. Assume that the theorem holds for $n-1$.

By the expansion formula for relative entropy, we have

$$D(p\|q)=\frac{1}{n}\cdot\sum_{i=1}^{n}D(Y_i\|X_i)+\frac{1}{n}\cdot\sum_{i=1}^{n}\mathbb{E}D\big(\bar{Y}_i|Y_i\|\bar{q}_i|Y_i\big). \tag{2.2}$$

For each fixed y_i, the measure $\bar{q}_i(\cdot\,|y_i)$ satisfies the conditions of the theorem. By the induction hypothesis,

$$\begin{aligned}\frac{1}{n}\cdot\sum_{i=1}^{n}\mathbb{E}D\big(\bar{Y}_i|Y_i\|\bar{q}_i(\cdot\,|Y_i)\big)&\leqslant\frac{1}{n}\cdot\frac{2C}{\alpha}\cdot\sum_{i=1}^{n}\sum_{j\neq i}|p_j\big(\cdot\,|\bar{Y}_j\big)-q_j\big(\cdot\,|\bar{Y}_j\big)|^2\\&=\left(1-\frac{1}{n}\right)\cdot\frac{2C}{\alpha}\cdot\sum_{j=1}^{n}|p_j\big(\cdot\,|\bar{Y}_j\big)-q_j\big(\cdot\,|\bar{Y}_j\big)|^2.\end{aligned} \tag{2.3}$$

To estimate the first term on the right-hand side of (2.2), observe that by the definition of α, $\mathbb{P}\{X_i=x\}\geqslant\alpha$ for any $i\in[1,n]$ and $x\in\mathcal{X}$. Thus, by Lemma 2.1,

$$D(Y_i\|X_i)\leqslant\frac{2}{\alpha}\cdot|\mathcal{L}(Y_i)-\mathcal{L}(X_i)|^2. \tag{2.4}$$

Further, condition (1.3) implies

$$\begin{aligned}\sum_{i=1}^{n}|\mathcal{L}(Y_i)-\mathcal{L}(X_i)|^2&\leqslant\sum_{i=1}^{n}\mathbb{P}_\pi\{Y_i\neq X_i\}^2=W_2^2(p,q)\\&\leqslant C\cdot\mathbb{E}\sum_{i=1}^{n}|p_i\big(\cdot\,|\bar{Y}_i\big)-q_i\big(\cdot\,|\bar{Y}_i\big)|^2.\end{aligned} \tag{2.5}$$

Putting together (2.4) and (2.5), it follows that the first term on the right-hand side of (2.2) can be bounded as follows:

$$\frac{1}{n}\cdot\sum_{i=1}^{n}D(Y_i\|X_i)\leqslant\frac{1}{n}\cdot\frac{2C}{\alpha}\cdot\sum_{i=1}^{n}\mathbb{E}|p_i\big(\cdot\,|\bar{Y}_i\big)-q_i\big(\cdot\,|\bar{Y}_i\big)|^2. \tag{2.6}$$

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

Substituting (2.3) and (2.6) into (2.2), we get the first inequality in (1.4). The second inequality follows from the Pinsker–Csiszár–Kullback inequality. □

3. Proof of Theorem 1.14

Let both $p, q \in \mathcal{P}(\mathcal{X}^n)$ be fixed. We want to show that (1.3) holds with $C = 1/(1 - \|A\|_2)^2$, where A is the coupling matrix for q. It is enough to prove this for $I = [1, n]$, since for any $I \subset [1, n]$ and $\bar{y}_I$ the conditional distribution $q_I(\cdot\,|\bar{y}_I)$ satisfies Dobrushin's uniqueness condition with a minor of A as its coupling matrix. (The idea of the proof for $I = [1, n]$ goes back to Dobrushin's papers [4, 5], although he worked with another matrix norm.)

We will prove that Dobrushin's uniqueness condition implies that the Gibbs sampler Γ is a contraction with respect to the W_2-distance with rate $1 - 1/n \cdot (1 - \|A\|_2)$.

To achieve this, let r and s be two probability measures on $\mathcal{X}^n$, and let (U, Z) be a random pair of sequences, with marginals r and s, and achieving $W_2(r, s)$.

Select an index $\nu \in [1, n]$ at random, independently of (U, Z), and define

$$U_k{}' = U_k, \quad Z_k{}' = Z_k \quad \text{for } k \neq \nu.$$

Then define $\mathcal{L}(U_i{}', Z_i{}' | \nu = i, \bar{U}_i = \bar{u}_i, \bar{Z}_i = \bar{z}_i)$ as that coupling of $q_i(\cdot\,|\bar{u}_i)$ and $q_i(\cdot\,|\bar{z}_i)$ that achieves $|q_i(\cdot\,|\bar{u}_i) - q_i(\cdot\,|\bar{z}_i)|$. It is clear that $\mathcal{L}(U') = r\Gamma$, and $\mathcal{L}(Z') = s\Gamma$. Moreover,

$$\mathbb{P}\{U_k{}' \neq Z_k{}' \mid \nu = i\} = \mathbb{P}\{U_k \neq Z_k\} \quad \text{for } k \neq i,$$

and, by the definition of the coupling matrix, more precisely by (1.7′):

$$\mathbb{P}\{U_i{}' \neq Z_i{}' \mid \nu = i\} = \sum_{\bar{u}_i, \bar{z}_i} \mathbb{P}\{\bar{U}_i = \bar{u}_i, \bar{Z}_i = \bar{z}_i\} \cdot |q_i(\cdot\,|\bar{u}_i) - q_i(\cdot\,|\bar{z}_i)|$$

$$\leqslant \sum_{\bar{u}_i, \bar{z}_i} \mathbb{P}\{\bar{U}_i = \bar{u}_i, \bar{Z}_i = \bar{z}_i\} \cdot \sum_{k \neq i} a_{k,i} \cdot \delta(u_k, z_k) = \sum_{k \neq i} a_{k,i} \cdot \mathbb{P}\{U_k \neq Z_k\}.$$

Thus

$$\mathbb{P}\{U_i{}' \neq Z_i{}'\} \leqslant (1 - 1/n) \cdot \mathbb{P}\{U_i \neq Z_i\} + 1/n \cdot \sum_{k \neq i} a_{k,i} \cdot \mathbb{P}\{U_k \neq Z_k\}.$$

It follows that

$$\sqrt{\sum_{i=1}^{n} \mathbb{P}\{U_i{}' \neq Z_i{}'\}^2} \leqslant \|B\|_2 \cdot \sqrt{\sum_{i=1}^{n} \mathbb{P}\{U_i \neq Z_i\}^2},$$

where

$$B = (1 - 1/n) \cdot \mathbb{I}_n + 1/n \cdot A$$

($\mathbb{I}_n$ is the identity matrix). Thus

$$\sqrt{\sum_{i=1}^{n} \mathbb{P}\{U_i{}' \neq Z_i{}'\}^2} \leqslant \left(1 - \frac{1}{n} \cdot (1 - \|A\|_2)\right) \cdot \sqrt{\sum_{i=1}^{n} \mathbb{P}\{U_i \neq Z_i\}^2}.$$

This proves the contractivity of Γ with rate $1 - 1/n \cdot (1 - \|A\|_2)$.

By the triangle inequality,

$$W_2(p, q) \leqslant W_2(p, p\Gamma) + W_2(p\Gamma, q).$$

By contractivity of Γ, and since q is invariant with respect to Γ, it follows that

$$W_2(p, q) \leqslant W_2(p, p\Gamma) + (1 - 1/n \cdot (1 - \|A\|_2)) \cdot W_2(p, q),$$

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

that is,

$$W_2(p,q) \leqslant \frac{n}{1-\|A\|_2} \cdot W_2(p, p\Gamma).$$

But it is easy to see that

$$W_2(p, p\Gamma) = \frac{1}{n} \cdot \sqrt{\mathbb{E} \sum_{i=1}^{n} |p_i(\cdot\,|\bar{Y}_i) - q_i(\cdot\,|\bar{Y}_i)|^2}.$$

By the last two inequalities, (1.3) (for $I = [1, n]$), and hence Theorem 1.14, is proved. □

4. Gibbs measures with the strong mixing property

4.1 Definitions, notation and statement of Theorem 4.2

In this section we work with measures on $\mathcal{X}^\Lambda$, where Λ is a subset of the d-dimensional cubic lattice $\mathbb{Z}$. Most of the time Λ will be finite. The notation $\Lambda \subset\subset \mathbb{Z}$ expresses that $\Lambda \subset \mathbb{Z}$ is finite.

The lattice points in $\mathbb{Z}$ will be called *sites*. We use the following distance on $\mathbb{Z}$:

$$\rho(k,i) = \max_\nu |k_\nu - i_\nu|, \quad \text{where } k = (k_1, k_2, \ldots, k_d),\ i = (i_1, i_2, \ldots, i_d).$$

The elements of $\mathcal{X}$ are called *spins*, and the elements of the set $\mathcal{X}^\Lambda$ ($\Lambda \subset \mathbb{Z}$, possibly infinite) are called *spin configurations*, or just *configurations*, over Λ.

We consider an *ensemble of conditional distributions* $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ where $\Lambda \subset\subset \mathbb{Z}$, and $\bar{\Lambda}$ is the complement of Λ. (We prefer to write $\bar{x}_\Lambda$ in place of $x_{\bar{\Lambda}}$, and, accordingly, use the notation $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$.) The measure $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ is considered as the conditional distribution of a random spin configuration over Λ, given the *outside configuration* $\bar{x}_\Lambda$ (*i.e.* the configuration outside Λ). For a site $i \in \mathbb{Z}$ we use the notation $q_i(\cdot\,|\bar{x}_i)$.

The conditional distribution $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ ($\Lambda \subset\subset \mathbb{Z}$) naturally defines the conditional distributions $q_M(\cdot\,|\bar{x}_M)$ for any $M \subset \Lambda$. We assume that the ensemble of the conditional distributions $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ satisfies the natural compatibility conditions. The conditional distribution $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ also defines the conditional distributions $q_M(\cdot\,|\bar{x}_\Lambda)$ for all subsets $M \subset \Lambda$.

Under the compatibility conditions there exists at least one probability measure $q = \mathcal{L}(X)$ on the space of configuration s $\mathcal{X}^{\mathbb{Z}}$, compatible with the local specifications $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$:

$$\mathcal{L}(X_\Lambda | \bar{X}_\Lambda = \bar{x}_\Lambda) = q_\Lambda(\cdot\,|\bar{x}_\Lambda).$$

Here X_Λ denotes the marginal of the random configuration X for the sites in Λ, and $\bar{x}_\Lambda$ is called an *outside configuration* for Λ. The conditional distributions $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ are called the *local specifications* of q, and q is called a *Gibbs measure* compatible with the local specifications $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$.

We say that the ensemble of local specifications $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ has *finite range of interaction* R (or is Markov of order R) if $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ only depends on those coordinates x_k, $k \in \bar{\Lambda}$, that are in the R-neighbourhood of Λ.

In general, the local specifications do not uniquely determine the Gibbs measure. The question of uniqueness has been extensively studied in the case of local specifications with finite range of interaction, and a sufficient condition for uniqueness was given by Dobrushin and Shlosman [6]. But the general question of uniqueness is open, even for measures with finite range of interaction.

A property stronger than uniqueness is strong mixing.

In their celebrated paper [7] in 1987, Dobrushin and Shlosman gave a characterization of complete analyticity of Markov Gibbs measures over $\mathbb{Z}$. Their characterization was formulated in twelve conditions which were proved to be equivalent, and are referred to as *Dobrushin and Shlosman's strong mixing conditions*. The following definition is the same as one of these twelve

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

(III C), except that we do not assume Markovity, and replace the function $K \cdot \exp(-\gamma r)$ with a more general function $\varphi(r)$. In the Markov case $\varphi(r)$ necessarily has the form $K \cdot \exp(-\gamma r)$.

In order to define strong mixing, let $\varphi:\mathbb{Z}_+ \mapsto \mathbb{R}_+$ be a function satisfying

$$\sum_{i \in \mathbb{Z}} \varphi(\rho(0, i)) < \infty. \tag{4.1}$$

Definition 4.1 (**strong mixing**). The Gibbs measure q is called *strongly mixing with coupling function* φ if for any sets $M \subset \Lambda \subset\subset \mathbb{Z}$ and any two outside configurations $\bar{y}_\Lambda$ and $\bar{z}_\Lambda$ differing only at one single site $k \notin \Lambda$:

$$|q_M(\cdot\,|\bar{y}_\Lambda) - q_M(\cdot\,|\bar{z}_\Lambda)| \leqslant \varphi(\rho(k, M)). \tag{4.2}$$

If two outside configurations $\bar{y}_\Lambda$ and $\bar{z}_\Lambda$ differ at several sites then a telescoping argument, together with the triangle inequality, shows that

$$|q_M(\cdot\,|\bar{y}_\Lambda) - q_M(\cdot\,|\bar{z}_\Lambda)| \leqslant \sum_{k \notin \Lambda} \delta(y_k, z_k) \cdot \varphi(\rho(k, M)). \tag{4.2$'$}$$

For stationary probability measures on $\mathcal{X}^{\mathbb{Z}}$, this definition is more restrictive than usual strong mixing, and is equivalent to Φ-mixing. On $\mathbb{Z}$ the term strong mixing has only been used for Markov fields; for simplicity we extend its use without adding any qualification.

Our aim in this section is to prove the following theorem.

Theorem 4.2. *Assume that the ensemble* $\{q_\Lambda(\cdot\,|\bar{x}_\Lambda):\Lambda \subset\subset \mathbb{Z}\}$ *satisfies the strong mixing condition with coupling function* φ*. Moreover, assume that*

$$\alpha \triangleq \inf q_i(x_i|\bar{x}_i) > 0,$$

where the infimum is taken for all $x \in \mathcal{X}^{\mathbb{Z}}$ *and* $i \in \mathbb{Z}$ *such that* $q_i(x_i|\bar{x}_i) > 0$*. Fix a* $\Lambda \subset\subset \mathbb{Z}$ *together with an outside configuration* $\bar{y}_\Lambda$*. Then the measure* $q_\Lambda(\cdot\,|\bar{y}_\Lambda)$ *satisfies condition* (1.3) *of Theorem 1.6, with a constant C, independent of* Λ *and* $\bar{y}_\Lambda$*. Moreover, it is enough to assume* (4.2) *for sets* Λ *of diameter at most* m_0*, where* m_0 *depends on the dimension d and the function* φ*. The constant C depends on the dimension d, the function* φ *and on* α*.*

Remark. If q has finite range of interaction then Theorem 4.2 implies that condition (4.2) is constructive, in the sense of Dobrushin and Shlosman.

Proof of Theorem 4.2. Consider the infinite symmetric matrix

$$\Phi = \big(\varphi(\rho(k, i))\big)_{k,i \in \mathbb{Z}}.$$

Since the entries are non-negative, and the row-sums equal, $\|\Phi\|_2$ equals the row-sum:

$$\|\Phi\|_2 = \sum_{i \in \mathbb{Z}} \varphi(\rho(0, i)) < \infty.$$

Fix a measure $p_\Lambda \in \mathcal{P}(\mathcal{X}^\Lambda)$ satisfying

$$q_\Lambda(x_\Lambda|\bar{y}_\Lambda) = 0 \implies p_\Lambda(x_\Lambda) = 0 \quad \text{for } x_\Lambda \in \mathcal{X}^\Lambda.$$

It is enough to prove that

$$W_2^2\Big(p_\Lambda, q_\Lambda\big(\cdot\,|\bar{y}_\Lambda\big)\Big) \leqslant C \cdot \mathbb{E} \sum_{i \in \Lambda} W_2^2\Big(p_i\big(\cdot\,|\bar{Y}_i\big), q_i\big(\cdot\,|\bar{Y}_i\big)\Big) \tag{4.3}$$

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

(with C independent of Λ and $\bar{y}_\Lambda$), since for any $M \subset \Lambda$ and any fixed $y_{\Lambda\setminus M}$, the conditional distribution $q_M(\cdot\,|\bar{y}_M)$ (where $\bar{y}_M = (y_{\Lambda\setminus M}, \bar{y}_\Lambda)$) satisfies the strong mixing condition with the same function φ. □

We start with a weaker version of (4.3), namely Lemma 4.4 below.

Notation 4.3. Let $\mathcal{I}_m = \mathcal{I}_m(\Lambda)$ denote the set of m-sided cubes in $\mathbb{Z}$ that intersect Λ. Set

$$\Theta_m \triangleq \min_R \left[\|\Phi\|_2 \cdot \frac{d \cdot R}{m} + 2d \cdot \sum_{r=R}^{\infty} (2r+1)^{d-1} \varphi(r) \right]. \tag{4.4}$$

Note that we can achieve

$$\Theta_m < 1 \tag{4.5}$$

by selecting R large enough to make the second term in (4.4) small, and then selecting m.

Lemma 4.4. *If m is so large that $\Theta_m < 1$, then*

$$\begin{aligned} W_2^2\Big(p_\Lambda, q_\Lambda(\cdot\,|\bar{y}_\Lambda)\Big) &\leqslant \frac{1}{m^d} \cdot \frac{1}{(1-\Theta_m)^2} \cdot \sum_{I \in \mathcal{I}_m} \mathbb{E} W_2^2(p_{I\cap\Lambda}(\cdot\,|\bar{Y}_{I\cap\Lambda}), q_{I\cap\Lambda}(\cdot\,|\bar{Y}_{I\cap\Lambda})) \\ &\leqslant \frac{1}{(1-\Theta_m)^2} \cdot \sum_{I \in \mathcal{I}_m} \mathbb{E}|p_{I\cap\Lambda}\big(\cdot\,|\bar{Y}_{I\cap\Lambda}\big) - q_{I\cap\Lambda}\big(\cdot\,|\bar{Y}_{I\cap\Lambda}\big)|^2. \end{aligned} \tag{4.6}$$

If the ensemble $\{q_\Lambda(\cdot\,|\bar{x}_\Lambda)\}$ has finite range of interaction R then Lemma 4.4 holds with $\|\Phi\|_2 \cdot (d \cdot R)/m$ in place of Θ_m.

The second inequality in (4.6) follows from the first one by the trivial inequality

$$W_2^2(r, s) \leqslant m^d \cdot |r - s|^2 \quad \text{for measures } r, s \text{ on } \mathcal{X}^{I\cap\Lambda}, I \in \mathcal{I}_m.$$

The proof of Lemma 4.4 follows that of Theorem 1.14, but we use a more general Gibbs sampler, updating (the intersection of Λ with) an m-sided cube at a time, not just one site. (Here we follow [7], where Gibbs samplers updating large sets of sites at a time were used.) Let us extend the definition of p_Λ so that on $\bar{\Lambda}$ it is concentrated on the fixed $\bar{y}_\Lambda$.

Definition 4.5. For $I \in \mathcal{I}_m$, let $\Gamma_I : \mathcal{P}(\mathcal{X}^\Lambda) \mapsto \mathcal{P}(\mathcal{X}^\Lambda)$ be the Markov kernel:

$$\Gamma_I(z_\Lambda | y_\Lambda) = \delta_{y_{\Lambda\setminus I}, z_{\Lambda\setminus I}} \cdot q_{I\cap\Lambda}(z_{I\cap\Lambda} | \bar{y}_{I\cap\Lambda})$$

(for $k \in \bar{\Lambda}$, y_k is defined by the fixed $\bar{y}_\Lambda$). Then set

$$\Gamma_{\mathcal{I}_m} = \frac{1}{|\mathcal{I}_m|} \cdot \sum_{I \in \mathcal{I}_m} \Gamma_I.$$

Then $\Gamma_{\mathcal{I}_m}$ preserves, and is reversible with respect to $q_\Lambda(\cdot\,|\bar{y}_\Lambda)$. We call $\Gamma_{\mathcal{I}_m}$ the Gibbs sampler for measure $q_\Lambda(\cdot\,|\bar{y}_\Lambda)$, defined by the local specifications $q_{I\cap\Lambda}(\cdot\,|\bar{y}_{I\cap\Lambda})$, $I \in \mathcal{I}_m$.

Proof of Lemma 4.4. To estimate $W_2^2(p_\Lambda, q_\Lambda(\cdot\,|\bar{y}_\Lambda))$, we are going to prove that if $\Theta_m < 1$ then the Gibbs sampler $\Gamma_{\mathcal{I}_m}$ is a contraction with respect to the W_2-distance, with rate $1 - m^d/|\mathcal{I}_m| \cdot (1 - \Theta_m)$.

To achieve this, let r and s be two probability measures on $\mathcal{X}^\Lambda$, and let $\mathcal{L}(Y, Z)$ be a joint distribution with marginals r and s, and achieving $W_2(r, s)$. We extend the definition of $\mathcal{L}(Y, Z)$, letting $\bar{Y}_\Lambda = \bar{Z}_\Lambda = \bar{y}_\Lambda$, where $\bar{y}_\Lambda$ is the fixed outside configuration. Let Y' and Z' be random variables representing $r\Gamma_{\mathcal{I}_m}$ and $s\Gamma_{\mathcal{I}_m}$.

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

When carrying out one step of the Gibbs sampler $\Gamma_{\mathcal{I}_m}$, we select a random element ν from $\mathcal{I}_m$, independently of (Y, Z). Then we can assume that

$$\mathbb{P}\{Y_i{}' \neq Z_i{}' \mid \nu = I\} = \mathbb{P}\{Y_i \neq Z_i\} \quad \text{for all } i \in \Lambda \setminus I.$$

Moreover, we can define $\mathcal{L}(Y_{I\cap\Lambda}{}', Z_{I\cap\Lambda}{}' \mid \nu = I)$ as that coupling of $q_{I\cap\Lambda}(\cdot\,|\bar{y}_{I\cap\Lambda})$ and $q_{I\cap\Lambda}(\cdot\,|\bar{z}_{I\cap\Lambda})$ that achieves $W_2^2(q_{I\cap\Lambda}(\cdot\,|\bar{y}_{I\cap\Lambda}), q_{I\cap\Lambda}(\cdot\,|\bar{z}_{I\cap\Lambda}))$.

At this point we need the following lemma.

Lemma 4.6. *(For a proof, see Appendix A.) Let us fix the set $M \subset\subset \mathbb{Z}$, together with two outside configurations $\bar{y}_M$ and $\bar{z}_M$, differing only at site $k \notin M$. Let Y and Z be random variables realizing $q_M(\cdot\,|\bar{y}_M)$ and $q_M(\cdot\,|\bar{z}_M)$. Define*

$$J_i = J_{k,M,i} = \{j \in M{:}\rho(k,j) \geqslant \rho(k,i)\}. \tag{4.7}$$

Then there exists a coupling $\mathcal{L}(Y, Z|\bar{y}_M, \bar{z}_M)$ of $\mathcal{L}(Y)$ and $\mathcal{L}(Z)$, satisfying

$$\mathbb{P}\{Y_i \neq Z_i\} \leqslant |q_{J_i}(\cdot\,|\bar{y}_M) - q_{J_i}(\cdot\,|\bar{z}_M)|, \quad i \in M.$$

If q satisfies the strong mixing condition with function φ then, for this coupling,

$$\mathbb{P}\{Y_i \neq Z_i\} \leqslant \varphi(\rho(k,i)) \quad \text{for all } i \in M.$$

By Lemma 4.6 and inequality (4.2′), for fixed I, $\bar{y}_{I\cap\Lambda}$ and $\bar{z}_{I\cap\Lambda}$, we can define a coupling

$$\mathcal{L}\big(Y_{I\cap\Lambda}{}', Z_{I\cap\Lambda}{}' \mid \nu = I, \bar{Y}_{I\cap\Lambda} = \bar{y}_{I\cap\Lambda}, \bar{Z}_{I\cap\Lambda} = \bar{z}_{I\cap\Lambda}\big),$$

satisfying

$$\mathbb{P}\{Y_i{}' \neq Z_i{}' \mid \nu = I, \bar{Y}_{I\cap\Lambda} = \bar{y}_{I\cap\Lambda}, \bar{Z}_{I\cap\Lambda} = \bar{z}_{I\cap\Lambda}\} \leqslant \sum_{k\in\Lambda\setminus I} \delta(y_k, z_k)\cdot\varphi(\rho(k,i)), \quad \text{for all } i \in I\cap\Lambda.$$

Thus

$$\mathbb{P}\{Y_i{}' \neq Z_i{}' \mid \nu = I\} \leqslant \sum_{k\in\Lambda\setminus I} \mathbb{P}\{Y_k \neq Z_k\}\cdot\varphi(\rho(k,i)) \quad \text{for all } i \in I\cap\Lambda. \tag{4.8}$$

We calculate $\mathbb{P}\{Y_i{}' \neq Z_i{}'\}$ by averaging for $I \in \mathcal{I}_m$. Set $N = |\mathcal{I}_m|$. Since each $i \in \Lambda$ is covered by exactly m^d cubes from $\mathcal{I}_m$, (4.8) implies

$$\mathbb{P}\{Y_i{}' \neq Z_i{}'\} \leqslant \left(1 - \frac{m^d}{N}\right)\cdot\mathbb{P}\{Y_i \neq Z_i\} + \frac{1}{N}\cdot\sum_{I\ni i}\sum_{k\in\Lambda\setminus I} \mathbb{P}\{Y_k \neq Z_k\}\cdot\varphi(\rho(k,i)). \tag{4.9}$$

Consider the vectors

$$u = \big(\mathbb{P}\{Y_k \neq Z_k\}\big)_{k\in\Lambda} \quad \text{and} \quad v = \big(\mathbb{P}\{Y_i{}' \neq Z_i{}'\}\big)_{i\in\Lambda},$$

and let D denote the matrix with entries

$$d_{k,i} = \varphi(\rho(k,i))\cdot\sum_{I\ni i,\Lambda\setminus I\ni k} 1, \quad k, i \in \Lambda.$$

With this notation, (4.9) means that

$$v \leqslant \left(\left(1 - \frac{m^d}{N}\right)\cdot\mathbb{I} + \frac{1}{N}\cdot D\right)\cdot u$$

coordinatewise, thus

$$\|v\|_2 \leqslant \left(\left(1 - \frac{m^d}{N}\right) + \frac{1}{N}\|D\|_2\right)\cdot\|u\|_2. \tag{4.10}$$

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

We claim that, for fixed k and i,

$$\sum_{I:k\notin I, I\ni i} 1 \leqslant m^{d-1} \cdot \min\{d \cdot \rho(k,i),\ \ m\}. \tag{4.11}$$

It is clear that $\sum_{I:I\ni i} 1 \leqslant m^d$. To prove the rest of (4.11), assume for simplicity that k and i differ in all coordinates. Then we can assume that $k = (0, 0, \ldots, 0)$, and $i_j > 0$ for all j. Let H_j denote the half-space $\{z^d \in \mathbb{Z}^d : z_j > 0\}$. Assume that for an I, $i \in I \in \mathcal{I}_m$, I intersects $\bar{H}_j$, that is, there is a $z^{(j)} \in I$ whose jth coordinate is non-positive. Then the projection of I on the coordinate axis $\{z_j = 0\}$ will contain k. It follows that if $I \cap \bar{H}_j \neq \emptyset$ for every j then $k \in I$. Consequently, if $i \in I \in \mathcal{I}_m$ and $k \notin I$ then there is at least one j such that $I \subset H_j$. There are $m^{d-1} \cdot |k_j - i_j| \leqslant m^{d-1} \cdot \rho(k,i)$ possible positions for an m-sided cube containing i and contained in H_j. This proves that $\sum_{I:k\notin I, I\ni i} 1 \leqslant d \cdot m^{d-1} \cdot \rho(k,i)$, and thereby (4.11). If some coordinates of k and i are equal, then the same argument works in a subspace.

Since the right-hand side of (4.11) is symmetric in k and i, we have

$$\|D\|_2 \leqslant m^d \cdot \sum_i \varphi(\rho(k,i)) \cdot \min\left\{\frac{d \cdot \rho(k,i)}{m}, 1\right\}. \tag{4.12}$$

Now fix an R, and divide the sum in (4.12) into two parts, for i satisfying $\rho(k,i) \leqslant R$ and $\rho(k,i) > R$, respectively. We see that

$$\|D\|_2 \leqslant m^d \cdot \left(\frac{\|\Phi\|_2 \cdot d \cdot R}{m} + \sum_{i:\rho(k,i)>R} \varphi(\rho(k,i))\right).$$

Taking the minimum in R, we get

$$\|D\|_2 \leqslant m^d \cdot \Theta_m. \tag{4.13}$$

By (4.10) and the definition of the vectors u and v, (4.13) implies that

$$\sqrt{\sum_{i\in\Lambda} \mathbb{P}\{Y_i' \neq Z_i'\}^2} \leqslant \left(1 - \frac{m^d}{N} \cdot (1 - \Theta_m)\right) \cdot \sqrt{\sum_{k\in\Lambda} \mathbb{P}\{Y_k \neq Z_k\}^2},$$

that is,

$$W_2(r\Gamma_{\mathcal{I}_m}, s\Gamma_{\mathcal{I}_m}) \leqslant \left(1 - \frac{m^d}{N} \cdot (1 - \Theta_m)\right) \cdot W_2(r,s). \tag{4.14}$$

The stated contractivity is proved.

By the triangle inequality it follows that

$$\begin{aligned} W_2\big(p_\Lambda, q_\Lambda(\cdot\,|\bar{y}_\Lambda)\big) &\leqslant W_2\big(p_\Lambda, p_\Lambda\Gamma_{\mathcal{I}_m}\big) + W_2\big(p_\Lambda\Gamma_{\mathcal{I}_m}, q_\Lambda(\cdot\,|\bar{y}_\Lambda)\big) \\ &\leqslant W_2\big(p_\Lambda, p_\Lambda\Gamma_{\mathcal{I}_m}\big) + \left(1 - \frac{m^d}{N} \cdot (1 - \Theta_m)\right) \cdot W_2\Big(p_\Lambda, q_\Lambda\big(\cdot\,|\bar{y}_\Lambda\big)\Big), \end{aligned}$$

whence

$$W_2\Big(p_\Lambda, q_\Lambda\big(\cdot\,|\bar{y}_\Lambda\big)\Big) \leqslant \frac{N}{m^d} \cdot \frac{1}{(1 - \Theta_m)} \cdot W_2\big(p_\Lambda, p_\Lambda\Gamma_{\mathcal{I}_m}\big). \tag{4.15}$$

To complete the proof of Lemma 4.4, we have to estimate $W_2(p_\Lambda, p_\Lambda\Gamma_{\mathcal{I}_m})$ in terms of the quantities

$$\mathbb{E}W_2^2\Big(p_{I\cap\Lambda}\big(\cdot\,|\bar{Y}_{I\cap\Lambda}\big), q_{I\cap\Lambda}\big(\cdot\,|\bar{Y}_{I\cap\Lambda}\big)\Big).$$

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

To do this, fix an $I \in \mathcal{I}_m$, together with a configuration $y_{\Lambda\setminus I} \in \mathcal{X}^{\Lambda\setminus I}$, and define a coupling $\pi_{I\cap\Lambda}(\cdot\,|y_{\Lambda\setminus I})$ of $p_{I\cap\Lambda}(\cdot\,|\bar{y}_{I\cap\Lambda})$ and $q_{I\cap\Lambda}(\cdot\,|\bar{y}_{I\cap\Lambda})$ that achieves W_2-distance. We extend $\pi_{I\cap\Lambda}(\cdot\,|y_{\Lambda\setminus I})$ to a measure on $\mathcal{X}^\Lambda \times \mathcal{X}^\Lambda$ concentrated on the diagonal $(y_{\Lambda\setminus I}, y_{\Lambda\setminus I})$, for coordinates outside I. Finally, we define the coupling π of p_Λ and $p_\Lambda \Gamma_{\mathcal{I}_m}$ by averaging the distributions $\pi_{I\cap\Lambda}(\cdot\,|y_{\Lambda\setminus I})$ with respect to I and $y_{\Lambda\setminus I}$.

Using this construction, an easy computation (using the Cauchy–Schwarz inequality) shows that

$$W_2^2(p_\Lambda, p_\Lambda\Gamma_{\mathcal{I}_m}) \leqslant \frac{m^d}{N^2} \sum_{I\in\mathcal{I}_m} \mathbb{E} W_2^2\Big(p_{I\cap\Lambda}\big(\cdot\,|\bar{Y}_{I\cap\Lambda}\big), q_{I\cap\Lambda}\big(\cdot\,|\bar{Y}_{I\cap\Lambda}\big)\Big). \tag{4.16}$$

Substituting (4.16) into (4.15), we get the first inequality in (4.6). Understanding the proof, one easily sees that the statement for Gibbs measures with finite range of interaction holds true. Lemma 4.4 is proved. □

To complete the proof of Theorem 4.2 we have to deduce (4.3) from Lemma 4.4. To do this we need the following lemma.

Lemma 4.7. *(The proof is in Appendix A.) Let $p = \mathcal{L}(Y)$ and q be two measures on $\mathcal{X}^n$. Let α be defined by (1.1). Then*

$$|p-q|^2 \leqslant \left(\frac{2}{(|\mathcal{X}|\cdot\alpha)^2}\right)^{n+\log_2 n} \cdot \sum_{i=1}^n \mathbb{E}|p_i\big(\cdot\,|\bar{Y}_i\big) - q_i\big(\cdot\,|\bar{Y}_i\big)|^2.$$

Using Lemma 4.7, we estimate the terms in the last sum in (4.6). We get

$$W_2^2\Big(p_\Lambda, q_\Lambda\big(\cdot\,|\bar{y}_\Lambda\big)\Big) \leqslant \frac{m^d}{(1-\Theta_m)^2} \cdot \left(\frac{2}{\big(|\mathcal{X}|\cdot\alpha\big)^2}\right)^{m^d+\log_2(m^d)} \cdot \mathbb{E}\sum_{i\in\Lambda} |p_i\big(\cdot\,|Y_{\Lambda\setminus i}\big) - q_i\big(\cdot\,|Y_{\Lambda\setminus i}, \bar{y}_\Lambda\big)|^2.$$

Thus (4.3) is fulfilled with

$$C = \frac{m^d}{(1-\Theta_m)^2} \cdot \left(\frac{2}{(|\mathcal{X}|\cdot\alpha)^2}\right)^{m^d+\log_2(m^d)},$$

as soon as m is large enough for $\Theta_m < 1$.

We used the strong mixing condition (4.2) in proving Lemma 4.6, and Lemma 4.6 was used for subsets of m-sided cubes. It was enough to consider m-sided cubes with m so large that $\Theta_m < 1$ holds, a condition depending on d and φ. This proves the last two statements of Theorem 4.2. □

Remark. An argument similar to Lemma 4.7 was also present in [23].

Acknowledgement

The author thanks I. Sason for calling her attention to the papers [20] and [9].

References

[1] Boucheron, S., Lugosi, G. and Massart, P. (2013) *Concentration Inequalities*, Oxford University Press.
[2] Caputo, P., Menz, G. and Tetali, P. (2015) Approximate tensorization of entropy at high temperature. *Ann Fac Sci Toulouse Math Sér 6* **24** 691–716.
[3] Cesi, F. (2001) Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probab. Theory Rel. Fields* **120** 569–584.
[4] Dobrushin, R. L. (1968) The description of a random field by means of conditional probabilities and condition of its regularity (in Russian). *Theory Probab. Appl.* **13** 197–224.
[5] Dobrushin, R. L. (1970) Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl.* **15** 458–486.

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

[6] Dobrushin, R. L. and Shlosman, S. B. (1985) Constructive criterion for the uniqueness of Gibbs field. In *Statistical Physics and Dynamical Systems* (J. Fritz, A. Jaffe and D. Szász, eds), Springer, pp. 371–403.

[7] Dobrushin, R. L. and Shlosman, S. B. (1987) Completely analytical interactions: Constructive description. *J. Statist. Phys.* **46** 983–1014.

[8] Diaconis, P. and Saloff-Coste, L. (1996) Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750.

[9] Gibbs, A. L. and Su, F. E. (2002) On choosing and bounding probability metrics. *Internat. Statist. Rev.* **70** 419–435.

[10] Goldstein, S. (1979) *Maximal coupling. Z. Wahrscheinlichkeitstheor. verw. Geb.* **46**, 193–204.

[11] Gross, L. (1975) Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** 1061–1083.

[12] Ledoux, M. (1999) Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités XXXIII*, Vol. 1709 of Lecture Notes in Mathematics, Springer, pp. 120–216.

[13] Martinelli, F. and Olivieri, E. (1994) Approach to equilibrium of Glauber dynamics in the one phase region, I: The attractive case. *Commun. Math. Phys.* **161** 447–486.

[14] Martinelli, F. and Olivieri, E. (1994) Approach to equilibrium of Glauber dynamics in the one phase region, II: The general case. *Commun. Math. Phys.* **161** 487–514.

[15] Marton, K. (2013) An inequality for relative entropy and logarithmic Sobolev inequalities in Euclidean spaces. *J. Funct. Anal.* **264** 34–61.

[16] Olivieri, E. (1988) On a cluster expansion for lattice spin systems: A finite size condition for the convergence. *J. Statist. Phys.* **50** 1179–1200.

[17] Olivieri, E. and Picco, P. (1990) Clustering for D-dimensional lattice systems and finite volume factorization properties. *J. Statist. Phys.* **59** 221–256.

[18] Otto, F. and Reznikoff, M. (2011) A new criterion for the logarithmic Sobolev inequality and two applications. *J. Funct. Anal.* **243** 121–157.

[19] Royer, G. (1999) *Une Initiation aux Inegalités de Sobolev Logarithmiques*, Société Mathématique de France.

[20] Sason, I. (2015) Tight bounds for symmetric divergence measures and a refined bound for lossless source coding. *IEEE Trans. Inform. Theory* **61** 701–707.

[21] Sason, I. (2015) On reverse Pinsker inequalities. arXiv:1503.07118v4

[22] Stroock, D. W. and Zegarlinski, B. (1992) The equivalence of the logarithmic Sobolev inequality and the Dobrushin–Shlosman mixing condition. *Commun. Math. Phys.* **144** 303–323.

[23] Stroock, D. W. and Zegarlinski, B. (1992) The logarithmic Sobolev inequality for discrete spin systems on the lattice. *Comm. Math. Phys.* **149** 175–193.

[24] Zegarlinski, B. (1992) Dobrushin uniqueness theorem and logarithmic Sobolev inequalities. *J. Funct. Anal.* **105** 77–111.

Appendix A

Proof of Lemma 4.6. Order the elements of Λ so that

$$\rho(k, i_1) \leqslant \rho(k, i_2) \leqslant \cdots \leqslant \rho(k, i_{|\Lambda|}),$$

that is, the sequence of sets $J_i = J_{k,M,i}$ (see (4.7)) is decreasing in i. Let Y_{J_i} and Z_{J_i} denote the marginals of Y and Z, respectively, for the sites in J_i. Then $(Y_{J_1}, \ldots, Y_{J_{|M|}})$ and $(Z_{J_1}, \ldots, Z_{J_{|M|}})$ are Markov chains (in fact $Y_{J_{i+1}}$ is a function of Y_{J_i}). Therefore, by a theorem of Goldstein [10], there exists a coupling $\pi = \mathcal{L}(Y, Z|\bar{y}_M, \bar{z}_M)$ of $\mathcal{L}(Y)$ and $\mathcal{L}(Z)$, satisfying

$$\mathbb{P}_\pi\{Y_{J_i} \neq Z_{J_i}\} = |\mathcal{L}(Y_{J_i}) - \mathcal{L}(Z_{J_i})| = |q_{J_i}(\cdot\,|\bar{y}_M) - q_{J_i}(\cdot\,|\bar{z}_M)|.$$

Since $i \in J_i$, and $\rho(k, i) = \rho(k, J_i)$, the statement of Lemma 4.6 follows. □

Proof of Lemma 4.7. Note first that if r and s are probability measures on $\mathcal{X}$, and $r(x), s(x) \geqslant \alpha$, then

$$|r - s| \leqslant 1 - |\mathcal{X}| \cdot \alpha.$$

Let $\mathcal{Y}$ and $\mathcal{Z}$ be finite sets, and consider two probability measures $r = \mathcal{L}(Y_1, Z_2)$ and s on $\mathcal{Y} \times \mathcal{Z}$. Denote $s_1(\cdot \mid \cdot)$ the conditional distribution of the first coordinate given the second one, defined by s, and define $s_2(\cdot \mid \cdot)$ similarly. Assume that $s_1(y_1|z_2) \geqslant \alpha_1$ and $s_2(z_2|y_1) \geqslant \alpha_2$ for all $y_1, z_2 \in \mathcal{Y} \times \mathcal{Z}$. Then

$$|s_2(\cdot\,|y_1) - s_2(\cdot\,|y_1{}')| \leqslant 1 - |\mathcal{Z}| \cdot \alpha_2 \quad \text{and} \quad |s_1(\cdot\,|z_2) - s_1(\cdot\,|z_2{}')| \leqslant 1 - |\mathcal{Y}| \cdot \alpha_1$$

for all $y_1, y_1{}' \in \mathcal{Y}$ and $z_2, z_2{}' \in \mathcal{Z}$.

https://doi.org/10.1017/S0963548319000099 Published online by Cambridge University Press

Thus in this case Dobrushin's uniqueness condition is satisfied with a 2×2 coupling matrix, with entries $1 - |\mathcal{Y}| \cdot \alpha_1$ and $1 - |\mathcal{Z}| \cdot \alpha_2$ outside the diagonal. (It does not matter that $\mathcal{Y}$ and $\mathcal{Z}$ may be different.) The coupling matrix has norm

$$\leqslant \max\{1 - |\mathcal{Y}| \cdot \alpha_1, 1 - |\mathcal{Z}| \cdot \alpha_2\}.$$

By the argument proving Theorem 1.14, it follows that

$$W_2(r, s) \leqslant \max\left\{\frac{1}{(|\mathcal{Y}| \cdot \alpha_1)^2}, \frac{1}{(|\mathcal{Z}| \cdot \alpha_2)^2}\right\} \cdot \mathbb{E}(|r_1(\cdot\,|Y_2) - s_1(\cdot\,|Y_2)|^2 + |r_2(\cdot\,|Y_1) - s_2(\cdot\,|Y_1)|^2),$$

and, consequently,

$$|r - s|^2 \leqslant \max\left\{\frac{2}{(|\mathcal{Y}| \cdot \alpha_1)^2}, \frac{2}{(|\mathcal{Z}| \cdot \alpha_2)^2}\right\} \cdot \mathbb{E}(|r_1(\cdot\,|Y_2) - s_1(\cdot\,|Y_2)|^2 + |r_2(\cdot\,|Y_1) - s_2(\cdot\,|Y_1)|^2). \tag{A.1}$$

Lemma 4.7 follows from (A.1) by a recursive argument, dividing the index set into two possibly equal parts of size $\lceil n/2 \rceil$ and $\lfloor n/2 \rfloor$, and applying (A.1) to each part. Then

$$\max\left\{\frac{2}{(|\mathcal{Y}| \cdot \alpha_1)^2}, \frac{2}{(|\mathcal{Z}| \cdot \alpha_2)^2}\right\}$$

will be replaced by

$$\left(\frac{2}{(|\mathcal{X}| \cdot \alpha)^2}\right)^{\lceil n/2 \rceil}.$$

Repeating this step about $\log_2 n$ times we get the statement of the lemma. □

Appendix B

In this section we still work in the lattice $\mathbb{Z}$, but think of sites as placed in the centre of the lattice cubes.

Let $\mathbb{Z}/l$ ($l \geqslant 1$ integer) denote the sublattice in $\mathbb{Z}$, consisting of points whose coordinates are all multiples of l, and let $\mathcal{C}_l$ denote the set of finite unions of l-sided cubes with vertices in $\mathbb{Z}/l$. The said l-sided cubes contain l^d sites each, and the set of sites in distinct cubes are disjoint. We shall identify the sets in $\mathcal{C}_l$ with the set of sites contained in them.

The approach by Olivieri and Picco is based on the following definition of strong mixing.

Definition B.1 (**Olivieri and Picco**). The Gibbs measure q on $\mathcal{X}^{Z^d}$ with finite range of interaction is called *strongly mixing over $\mathcal{C}_l$ in the sense of Olivieri–Picco–Martinelli*, if there exist numbers $\gamma > 0$, $K > 0$ such that, for any sets $\Lambda \in \mathcal{C}_l$, $M \subset \Lambda$ and any two outside configurations $\bar{y}_\Lambda$ and $\bar{z}_\Lambda$ differing only at a single site $k \notin \Lambda$, we have

$$|q_M(\cdot\,|\bar{y}_\Lambda) - q_M(\cdot\,|\bar{z}_\Lambda)| \leqslant K \cdot \exp\,(-\gamma \cdot \rho(k, M)). \tag{B.1}$$

Given the following theorem, if l is sufficiently large then it is enough to require (B.1) just for cubes in $\mathcal{C}_l$, to get (B.1) for all $\Lambda \in \mathcal{C}_l$, albeit with different γ and K.

Olivieri and Picco's Effectivity Theorem [17, 13]. Assume that the Gibbs measure q on $\mathcal{X}^{Z^d}$ has finite range of interaction. For any $\gamma, K > 0$ there exists an l_0 such that if, for some $l \geqslant l_0$, (B.1) holds for all l-sided *cubes* $\Lambda \in \mathcal{C}_l$, all $M \subset \Lambda$ and all $k \notin \Lambda$, then (B.1) also holds for all $\Lambda \in \mathcal{C}_l$, and M and k as above, with different γ and K.

We use a slightly more general definition, although we cannot justify it with an analogue of the above Effectivity Theorem.

Definition B.2 (**strong mixing over** $\mathcal{C}_l$). Let $\varphi:\mathbb{Z}_+ \mapsto \mathbb{R}_+$ be a function satisfying (4.1). Fix an integer $l \geqslant 1$. The ensemble of conditional distributions $q_\Lambda(\cdot\,|\bar{x}_\lambda)$ on $\mathcal{X}^{Z^d}$ is called *strongly mixing over* $\mathcal{C}_l$, with coupling function φ if, for any sets $\Lambda \in \mathcal{C}_l$, $M \subset \Lambda$, and any two outside configurations $\bar{y}_\Lambda$ and $\bar{z}_\Lambda$ differing only at the single site k, (4.2) holds. (We do not assume finite range of interaction.)

For measures strongly mixing over $\mathcal{C}_l$ one can prove a logarithmic Sobolev inequality by means of the following modifications of Theorem 1.6 and Lemma 4.4.

Theorem 1.6′. *Consider a measure q^Λ on $\mathcal{X}^\Lambda = \prod_{j=1}^n \mathcal{X}^{\Lambda_j}$, where*

$$\Lambda = \cup_{j=1}^n \Lambda_j, \quad \Lambda_j \cap \Lambda_k = \emptyset \quad \textit{for } j \neq k, \quad |\Lambda_j| = m.$$

Set

$$\alpha = \min\{q_i(x_i|\bar{x}_i): \quad q_\Lambda(x_\Lambda) > 0, i \in \Lambda\}.$$

Fix a $p_\Lambda = \mathcal{L}(Y_\Lambda)$ on $\mathcal{X}^\Lambda$ satisfying

$$q_\Lambda(x_\Lambda) = 0 \Longrightarrow p_\Lambda(x_\Lambda) = 0.$$

Assume that q_Λ satisfies all the inequalities

$$W_2^2(p_I(\cdot\,|\bar{y}_I), q_I(\cdot\,|\bar{y}_I)) \leqslant C \cdot \mathbb{E}\left\{ \sum_{\Lambda_j \subset I} W^2(p_{\Lambda_j}(\cdot\,|\bar{Y}_{\Lambda_j}), q_{\Lambda_j}(\cdot\,|\bar{Y}_{\Lambda_j})) \,\middle|\, \bar{Y}_I = \bar{y}_I \right\},$$

where $I \subset \Lambda$ is the union of some of the sets Λ_j, and $\bar{y}_I \in \mathcal{X}^{\Lambda \setminus I}$ is a fixed sequence. Then

$$D(p_\Lambda \| q_\Lambda) \leqslant \frac{4Cm}{\alpha^m} \cdot \sum_{j=1}^n \mathbb{E} W^2\Big(p_{\Lambda_j}(\cdot\,|\bar{Y}_{\Lambda_j}), q_{\Lambda_j}(\cdot\,|\bar{Y}_{\Lambda_j})\Big).$$

This can be proved by the same argument as Theorem 1.6, using Lemma 1.12, the inequalities

$$|p_{\Lambda_j}(\cdot\,|\bar{Y}_{\Lambda_j}) - q_{\Lambda_j}(\cdot\,|\bar{Y}_{\Lambda_j})|^2 \leqslant m \cdot W^2(p_{\Lambda_j}(\cdot\,|\bar{Y}_{\Lambda_j}), q_{\Lambda_j}(\cdot\,|\bar{Y}_{\Lambda_j})),$$

and, in each induction step, fixing a whole new block Y_{Λ_j}.

Lemma 4.4′ (**for measures strongly mixing over** $\mathcal{C}_l$)**.** *Fix an integer l, and assume that the ensemble of conditional distributions $q_\Lambda(\cdot\,|\bar{x}_\Lambda)$ on $\mathcal{X}^{Z^d}$ satisfies the strong mixing condition over $\mathcal{C}_l$, with coupling function φ. Let $\Lambda \in \mathcal{C}_l$, and fix an outside configuration $\bar{y}_\Lambda$. For fixed m let $\mathcal{I}_{ml}$ denote the set of $m \cdot l$-sided cubes from $\mathcal{C}_l$ intersecting Λ. Then, for large enough m and any measure p_Λ on $\mathcal{X}^\Lambda$,*

$$\begin{aligned} W_2^2\Big(p_\Lambda, q_\Lambda(\cdot\,|\bar{y}_\Lambda)\Big) &\leqslant C \cdot \sum_{I \in \mathcal{I}_{ml}} \mathbb{E} W_2^2\Big(p_{I\cap\Lambda}(\cdot\,|\bar{Y}_{I\cap\Lambda}), q_{I\cap\Lambda}(\cdot\,|\bar{Y}_{I\cap\Lambda})\Big) \\ &\leqslant C \cdot m^d \cdot \sum_{I \in \mathcal{I}_{ml}} \mathbb{E}|p_{I\cap\Lambda}(\cdot\,|\bar{Y}_{I\cap\Lambda}) - q_{I\cap\Lambda}(\cdot\,|\bar{Y}_{I\cap\Lambda})|^2, \end{aligned}$$

where C and m depend on the dimension d and the function φ.

The proof uses a Gibbs sampler, updating (intersections with Λ of) randomly chosen cubes of side $m \cdot l$ from $\mathcal{C}_l$ (for an appropriate m).

Cite this article: Marton K (2019). Logarithmic sobolev inequalities in discrete product spaces. *Combinatorics, Probability and Computing* **28**, 919–935. https://doi.org/10.1017/S0963548319000099