Electron. Commun. Probab. **19** (2014), no. 63, 1–11.
DOI: 10.1214/ECP.v19-3502
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Concentration inequalities for Gibbs sampling under d_{l_2}-metric

Neng-Yi Wang*

Abstract

The aim of this paper is to investigate the Gibbs sampling that's used for computing the mean of observables with respect to some function f depending on a very small number of variables. For this type of observable, by using the d_{l_2}-metric one obtains the sharp concentration estimate for the empirical mean, which in particular yields the correct speed in the concentration for f depending on a single observable.

Keywords: Concentration inequality; Gibbs sampling; coupling method; Dobrushin's uniqueness condition; d_{l_2}-metric.
AMS MSC 2010: 60E15; 65C05.
Submitted to ECP on May 7, 2014, final version accepted on September 4, 2014.

1 Introduction

Let μ be a Gibbs probability measure on E^N with dimension N large, i.e.,

$$\mu(dx^1,\cdots,dx^N)=\frac{e^{-V(x^1,\cdots,x^N)}}{\int\cdots\int_{E^N} e^{-V(x^1,\cdots,x^N)}\pi(dx^1)\cdots\pi(dx^N)}\pi(dx^1)\cdots\pi(dx^N),$$

where π is some σ-finite reference measure on E. Our purpose is to study the Gibbs sampling–a Markov Chain Monte-Carlo method (MCMC in short) for approximating μ. Gibbs sampling is also called Glauber dynamics with systematic scan(see [6]).

Let $\mu_i(\cdot|x)$ $(x=(x^1,\cdots,x^N)\in E^N)$ be the regular conditional distribution of x^i knowing $(x^j, j\neq i)$ under μ, i.e.,

$$\mu_i(dx^i|x)=\frac{e^{-V(x^1,\cdots,x^N)}}{\int_E e^{-V(x^1,\cdots,x^N)}\pi(dx^i)}\pi(dx^i),$$

which is a one-dimensional measure, easy to simulate in practice.

By iterations of the one-dimensional conditional distributions $(\mu_i, i=1,\cdots,N)$, the Gibbs sampling is the time-homogeneous Markov chain $(Z_k, k=0,1,\cdots)$, where each Z_k is the random vector on E^N after the dynamics has been sequentially applied to all sites. (For details see Section 2.) In [6], Dyer, Goldberg and Jerrum study mixing time of Gibbs sampling on finite spin systems by Dobrushin uniqueness conditions. But we will study concentration inequalities for Gibbs sampling on the general space by Dobrushin conditions such as [17].

*School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China. E-mail: wangnengyi@hust.edu.cn

In [17], Wu and the author obtain some sharp concentration estimate for

$$\mathbb{P}(\frac{1}{n}\sum_{k=1}^{n} f(Z_k) - \mu(f) \geq t), t > 0, n \geq 1.$$

That result in particular yields the correct speed in the concentration of functions of type $f(x) = \frac{1}{N}\sum_{i=1}^{N} g(x^i)$, but not for f depending on a very small number of variables(for example $f(x) = g(x^1)$).

So the purpose of this paper is to solve the problem above, i.e., to establish a new sharp concentration estimate for $\mathbb{P}(\frac{1}{n}\sum_{k=1}^{n} f(Z_k) - \mu(f) \geq t), t > 0, n \geq 1$, where the function f depends on a small number of variables. Our method is to prove Talagrand's T_2-transport inequality with respect to (w.r.t. in short) the $d_{l_2^N}$-metric (see later the definition of (2.1)), which is much stronger than the T_1-transport inequality w.r.t. $d_{l_1^N}$-metric. The main new feature of our T_2-transport inequality is dimension free, now. As well known the T_2-transport inequality is much more difficult than the T_1-transport inequality (see [3, 7, 8]). Technically this obliges us to introduce a new type of Dobrushin interdependence coefficients and complicates much the process of tensorization.

This paper is organized as follows. The next section contains some preliminaries about transport inequality and Gibbs sampling. We present the main results in Section 3, and prove them in Section 4.

2 Some preliminaries

2.1 Transport inequality

Throughout the paper E is a Polish space with the Borel σ-field $\mathfrak{B}$, and d is a metric on E such that $d(x, y)$ is lower semi-continuous on E^2 (so d does not necessarily generate the topology of E). On the product space E^N, we consider the l_p^N $(p = 1, 2)$-metric

$$d_{l_p^N}(x, y) := \left(\sum_{i=1}^{N} d^p(x^i, y^i)\right)^{1/p}, \ x, y \in E^N. \tag{2.1}$$

Later sometimes d_{l_p} is short for $d_{l_p^N}$ (or $d_{l_p^n}$) when the index N (or n respectively) is obvious from the context. E^N is endowed with the $d_{l_p^N}$-metric unless otherwise stated.

Let $\mathcal{M}_1(E)$ be the space of Borel probability measures on E, and

$$\mathcal{M}_p^d(E) := \left\{\nu \in \mathcal{M}_1(E); \int_E d^p(x_0, x)\nu(dx) < \infty\right\}, p = 1, 2.$$

($x_0 \in E$ is some fixed point, but the definition above does not depend on x_0 by the triangle inequality). Given $\nu_1, \nu_2 \in \mathcal{M}_p^d(E)$, the L^p-Wasserstein distance between ν_1, ν_2 is given by

$$W_{p,d}(\nu_1, \nu_2) := \left(\inf_{\pi} \iint_{E\times E} d^p(x, y)\pi(dx, dy)\right)^{1/p}, \tag{2.2}$$

where the infimum is taken over all probability measures π on $E \times E$ such that its marginal distributions are respectively ν_1 and ν_2 (*called a coupling of* ν_1 *and* ν_2).

When μ, ν are probability measures, the Kullback information (or relative entropy) of ν with respect to μ is defined as

$$H(\nu|\mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} d\nu, & \text{if } \nu \ll \mu, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.3}$$

For $p = 1, 2$, we say that the probability measure μ satisfies the L^p-*transport-entropy inequality* on (E, d) with some constant $c > 0$, if

$$W_{p,d}(\mu,\nu) \le \sqrt{2cH(\nu|\mu)}, \text{for every } \nu \in \mathcal{M}_1(E). \tag{2.4}$$

To be short, we write $\mu \in T_p(c)$ (or $T_{p,d}(c)$) for this relation. This inequality, related to the phenomenon of measure concentration, was introduced and studied by Marton [11, 12], developed subsequently by Talagrand [15], Bobkov-Götze [1], Otto-Villani [14], Djellout *et al.* [3] and amply explored by Ledoux [10, 9], Villani [16] and Gozlan-Léonard [8].

2.2 Gibbs sampling

Let $\mu_i(dx^i|x)(x = (x^1, \cdots, x^N) \in E^N)$ be the given regular conditional distribution of x^i knowing $(x^j, j \ne i)$ under μ, and $\bar\mu_i(dy|x)$ be the lift of μ_i to E^N.

Gibbs sampling is described as follows. Given a initial point $x_0 = (x_0^1, \cdots, x_0^N) \in E^N$, let $(X_n, n \ge 0)$ be a non-homogeneous Markov chain starting from x_0 defined on some probability space $(\Omega, , \mathfrak{F}, \mathbb{P}_{x_0})$, and given $X_{kN+i-1} = x = (x^1, \cdots, x^N) \in E^N, (k \in \mathbb{N}, 1 \le i \le N)$, then $X^j_{kN+i} = x^j$ for $j \ne i$ and the conditional law of X^i_{kN+i} is $\mu_i(\cdot|x)$.

In other words, the transition probability at step $kN + i$ is $\mathbb{P}(X_{kN+i} \in dy|X_{kN+i-1} = x) = \bar\mu_i(dy|x)$. Therefore for $\forall k \ge 1$,

$$\begin{aligned}\mathbb{P}(X_{kN} \in dy|X_{(k-1)N} = x) &= \int_{E^N} \bar\mu_1(dx_1|x) \cdots \int_{E^N} \bar\mu_{N-1}(dx_{N-1}|x_{N-2})\bar\mu_N(dy|x_{N-1}) \\ &=: P(x, dy),\end{aligned}$$

and the Gibbs sampling is the time-homogeneous Markov chain $(Z_k = X_{kN}, k = 0, 1, \cdots)$, whose transition probability is P.

3 Main results

Throughout the paper we assume that $\int_{E^N} d^2(y^i, x_0^i)d\mu(y) < \infty$, $\mu_i(\cdot|x) \in \mathcal{M}_2^d(E)$ for all $i = 1, \cdots, N$ and $x \in E^N$, where x_0 is some fixed point of E^N, and $x \to \mu_i(\cdot|x)$ is Lipschitzian from $(E^N, d_{l_2^N})$ to $(\mathcal{M}_2^d(E), W_{2,d})$.

For $p = 1, 2$, define the matrix of the d-Dobrushin interdependence coefficients $C^{(p)} := (c_{ij}^{(p)})_{i,j=1,\cdots,N}$ as

$$c_{ij}^{(p)} := \sup_{x=y \text{ off} j} \frac{W_{p,d}(\mu_i(\cdot|x), \mu_i(\cdot|y))}{d(x^j, y^j)}, i, j = 1, \cdots, N. \tag{3.1}$$

Obviously $c_{ii}^{(p)} = 0$. Denote by $\|A\|_p$ the operator norm of a general N by N matrix A acting as an operator from l_p^N to itself. Then the well known Dobrushin uniqueness condition (see [4, 5]) is

$$\|C^{(1)}\|_1 = \max_{1\le j\le N} \sum_{i=1}^N c_{ij}^{(1)} < 1.$$

So the generalization of Dobrushin uniqueness condition is read as

$$(H1) \quad C^{(2)} : l_2^N \to l_2^N \text{ with } \|C^{(2)}\|_2 < 1.$$

Let $r_\infty := \|C^{(2)}\|_\infty = \max_{1\le i\le N} \sum_{j=1}^N c_{ij}^{(2)}$ and $r_1 := \|C^{(2)}\|_1 = \max_{1\le j\le N} \sum_{i=1}^N c_{ij}^{(2)}$.

For any function $f : E^N \to \mathbb{R}$, let $\|f\|_{Lip(d_{l_p^N})} := \sup_{x\ne y} \frac{|f(x)-f(y)|}{d_{l_p^N}(x,y)}, p = 1, 2.$

Our main results are the following:

Theorem 3.1. *Assume*

$$r_\infty r_1 < \frac{1}{2},$$

and for some constant $c > 0$,

$$(H2) \quad \forall i = 1, \cdots N,\ \forall x \in E^N,\ \mu_i(\cdot|x) \in T_2(c).$$

Then for any Lipschitzian function f on E^N with $\|f\|_{Lip(d_{l_2^N})} \le \alpha$, for any starting point of the chain $x = (x^1, \cdots, x^N) \in E^N$, we have

(a)

$$P(x, \cdot) \in T_{2,d_{l_2^N}}\left(\frac{c}{(1 - \|C^{(2)}\|_2)^2}\right); \tag{3.2}$$

(b) *for* $\forall t > 0, n \ge 1$,

$$\begin{aligned} &\mathbb{P}_x\left(\tfrac{1}{n}\sum_{k=1}^n f(Z_k) - \tfrac{1}{n}\sum_{k=1}^n \mathbb{E}_x(f(Z_k)) \ge t\right) \\ \le\quad & \exp\left\{-\tfrac{nt^2(1-\sqrt{r_\infty r_1/(1-r_\infty r_1)})^2(1-\|C^{(2)}\|_2)^2}{2c\alpha^2}\right\}; \end{aligned} \tag{3.3}$$

(c) *for* $\forall t > 0, n \ge 1$,

$$\begin{aligned} &\mathbb{P}_x\left(\tfrac{1}{n}\sum_{k=1}^n f(Z_k) - \mu(f) \ge \tfrac{\alpha M_x}{n} + t\right) \\ \le\quad & \exp\left\{-\tfrac{nt^2(1-\sqrt{r_\infty r_1/(1-r_\infty r_1)})^2(1-\|C^{(2)}\|_2)^2}{2c\alpha^2}\right\}, \end{aligned} \tag{3.4}$$

where

$$M_x = \frac{\sqrt{r_\infty r_1}}{\sqrt{1 - r_\infty r_1} - \sqrt{r_\infty r_1}} \sqrt{\int_{E^N} \sum_{i=1}^N d(x^i, y^i)^2 \mu(dy)}.$$

Remark 3.2. Under the assumption of $r_\infty r_1 < 1$, by the Riesz interpolation inequality, $\|C^{(2)}\|_2 \le \sqrt{r_\infty r_1} < \sqrt{\frac{1}{2}}$, which implies $(H1)$ holds.

Remark 3.3. Recall some results from [17, Lemma 3.4 and Theorem 2.7]: assume that $\|C^{(1)}\|_1 < \frac{1}{2}$, and for some constant $c > 0$,

$$\forall i = 1, \cdots N,\ \forall x \in E^N,\ \mu_i(\cdot|x) \in T_1(c),$$

Then for any Lipschitzian function f on E^N with $\|f\|_{Lip(d_{l_1^N})} \le \alpha$, one has

(a) (from [17, Lemma 3.4])

$$P(x, \cdot) \in T_{1,d_{l_1^N}}\left(\frac{Nc}{(1 - \|C^{(1)}\|_1)^2}\right), \forall x = (x^1, \cdots, x^N) \in E^N; \tag{3.5}$$

(b) (from [17, Theorem 2.7])

$$\begin{aligned} &\mathbb{P}_x\left(\tfrac{1}{n}\sum_{k=1}^n f(Z_k) - \tfrac{1}{n}\sum_{k=1}^n \mathbb{E}_x f(Z_k) \ge t\right) \\ \le\quad & \exp\left\{-\tfrac{n}{N} \cdot \tfrac{t^2(1-2\|C^{(1)}\|_1)^2}{2c\alpha^2}\right\},\ \forall t > 0, n \ge 1. \end{aligned} \tag{3.6}$$

So in Theorem 3.1, the present transport inequality (3.2) (and concentration inequality (3.3)), contrary to previous results (3.5) (and (3.6) respectively), are dimension-free in the sense that N does not explicitly appear in the quantitative estimates. But also note that the quantity $L := \sqrt{\int_{E^N} \sum_{i=1}^N d(x^i, y^i)^2 \mu(dy)} = W_{2,d_{l_2^N}}(\delta_x, \mu)$ appearing in the bias M_x in (3.4) is dimension-dependent.

Remark 3.4. By a result of Gozlan (for details see [8]): $T_2(c)$ is equivalent to dimension-free concentration on product spaces, which is one main difference between T_2 and T_1.

Remark 3.5. It's easy to show that the concentration inequality (3.3) is sharp: in fact, take $E^N = \mathbb{R}^N$, $\mu = \gamma^{\bigotimes N}$ where γ is the Gaussian law $\mathcal{N}(0,1)$, then $(Z_k, k \geq 1)$ is an independent identically distributed sequence, $C^{(2)} = 0$, $c = 1$, $r_1 = r_\infty = 0$, and so the inequality (3.3) becomes sharp for $f(x) = x^1$: in this case, the inequality (3.3) is read as:

$$\mathbb{P}_x\left(\frac{1}{n}\sum_{k=1}^n Z_k^1 - \frac{1}{n}\sum_{k=1}^n \mathbb{E}_x Z_k^1 \geq t\right) \leq \exp\left\{-\frac{nt^2}{2}\right\};$$

however, it's well known that $\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}_x\left(\frac{1}{n}\sum_{k=1}^n Z_k^1 - \frac{1}{n}\sum_{k=1}^n \mathbb{E}_x Z_k^1 \geq t\right) = -\frac{t^2}{2}$.

Next we emphasize differences and improvements of this theorem compared with [17, Theorem 2.7](see (3.6)). Take $f(x) = \sum_{i=1}^l g(x^i), x = (x^1, \cdots, x^N), 1 \leq l \leq N$, where $g : (E, d) \to \mathbb{R}$ is d-Lipschitzian with $\|g\|_{Lip(d)} = \alpha$. Since $\|f\|_{Lip(d_{l_2^N})} \leq \alpha\sqrt{l}$, the inequality (3.3) implies for all $t > 0, n \geq 1$,

$$\mathbb{P}_x\left(\frac{1}{n}\sum_{k=1}^n\sum_{i=1}^l g(Z_k^i) - \mathbb{E}_x\frac{1}{n}\sum_{k=1}^n\sum_{i=1}^l g(Z_k^i) \geq t\right)$$

$$\leq \exp\left\{-\frac{n}{l}\cdot\frac{t^2(1-\sqrt{r_\infty r_1/(1-r_\infty r_1)})^2(1-\|C^{(2)}\|_2)^2}{2c\alpha^2}\right\}, \tag{3.7}$$

which is of speed n/l. However if one applies [17, Theorem 2.7](see (3.6)), one obtains only the concentration inequality of speed n/N. In other words, for functions depending on a small number of variables (in particular, the case $l = 1$), this current theorem improves essentially those in [17, Theorem 2.7].

Remark 3.6. Let E is a Riemannian manifold. The assumption $(H2)$ can be verified by Bakry-Emery's Γ_2-criterion ([10, Theorem 5.2]) or the more general criterion of F.Y. Wang (see [9]) for the log-Sobolev inequality (which is stronger than $T_2(c)$ by Otto-Villani [14]), or the very general sufficient condition of Lyapunov function method for $T_2(c)$ by Cattiaux *et al.* [2].

4 Proofs of the main results

4.1 The construction of the coupling.

Given any two initial distributions ν_1 *and* ν_2 *on* E^N, we begin by constructing our coupled non-homogeneous Markov chain $(X_i, Y_i)_{i\geq 0}$, which is similar to the coupling in [17] or [13].

Let (X_0, Y_0) be a coupling of (ν_1, ν_2). And given

$$(X_{kN+i-1}, Y_{kN+i-1}) = (x, y) \in E^N \times E^N, k \in \mathbb{N}, 1 \leq i \leq N,$$

then

$$X_{kN+i}^j = x^j, Y_{kN+i}^j = y^j, j \neq i,$$

and

$$\mathbb{P}((X_{kN+i}^i, Y_{kN+i}^i) \in \cdot | (X_{kN+i-1}, Y_{kN+i-1}) = (x, y)) = \pi(\cdot|x, y),$$

where $\pi(\cdot|x, y)$ is an optimal coupling of $\mu_i(\cdot|x)$ and $\mu_i(\cdot|y)$ such that

$$\left(\iint_{E^2} d^2(\tilde{x}, \tilde{y})\pi(d\tilde{x}, d\tilde{y}|x, y)\right)^{1/2} = W_{2,d}(\mu_i(\cdot|x), \mu_i(\cdot|y)).$$

(For the existence of such a coupling, refer to Villani [16].) Define the partial order on $\mathbb{R}^N$ by $a \le b$ if and only if $a^i \le b^i, i = 1, \cdots, N$. Then, by the triangle inequality for the metric $W_{2,d}$,

$$W_{2,d}(\mu_i(\cdot|x), \mu_i(\cdot|y)) \le \sum_{j=1}^N c_{ij}^{(2)} d(x^j, y^j),\ 1 \le i \le N. \tag{4.1}$$

We have for $\forall k \in \mathbb{N},\ 1 \le i \le N$,

$$\mathbb{E}[d^2(X_i^i, Y_i^i)|(X_{i-1}, Y_{i-1})] \le \left[\sum_{j=1}^N c_{ij}^{(2)} d(X_{i-1}^j, Y_{i-1}^j)\right]^2 \le \left(\sum_{j=1}^N c_{ij}^{(2)}\right)\left(\sum_{j=1}^N c_{ij}^{(2)} d^2(X_{i-1}^j, Y_{i-1}^j)\right),$$

and so

$$\begin{pmatrix} \mathbb{E}[d^2(X_{kN+i}^1, Y_{kN+i}^1)|X_{kN+i-1}, Y_{kN+i-1}] \\ \vdots \\ \vdots \\ \mathbb{E}[d^2(X_{kN+i}^N, Y_{kN+i}^N)|X_{kN+i-1}, Y_{kN+i-1}] \end{pmatrix} \le A_i \begin{pmatrix} d^2(X_{kN+i-1}^1, Y_{kN+i-1}^1) \\ \vdots \\ \vdots \\ d^2(X_{kN+i-1}^N, Y_{kN+i-1}^N) \end{pmatrix},$$

where

$$A_i = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ s_i c_{i1}^{(2)} & s_i c_{i2}^{(2)} & \cdots & \cdots & \cdots & s_i c_{iN}^{(2)} \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} (\textit{the blank in the matrix means } 0),$$

and

$$s_i := \sum_{j=1}^N c_{ij}^{(2)}.$$

Therefore by iterations, we have

$$\begin{pmatrix} \mathbb{E}d^2(X_N^1, Y_N^1) \\ \vdots \\ \vdots \\ \mathbb{E}d^2(X_N^N, Y_N^N) \end{pmatrix} \le A_N A_{N-1} \cdots A_1 \begin{pmatrix} \mathbb{E}d^2(X_0^1, Y_0^1) \\ \vdots \\ \vdots \\ \mathbb{E}d^2(X_0^N, Y_0^N) \end{pmatrix}. \tag{4.2}$$

Let

$$B := A_N A_{N-1} \cdots A_1. \tag{4.3}$$

By (4.2) above, Markov property and iterations,

$$\begin{pmatrix} \mathbb{E}d^2(X_{kN}^1, Y_{kN}^1) \\ \vdots \\ \vdots \\ \mathbb{E}d^2(X_{kN}^N, Y_{kN}^N) \end{pmatrix} \le B^k \begin{pmatrix} \mathbb{E}d^2(X_0^1, Y_0^1) \\ \vdots \\ \vdots \\ \mathbb{E}d^2(X_0^N, Y_0^N) \end{pmatrix}. \tag{4.4}$$

4.2 Proof of the main results

For prove Theorem 3.1, we first need to prove a few lemmas.

Let $\widetilde{c}_{ij} = s_i c_{ij}^{(2)}, i, j = 1, \cdots, N$, $\widetilde{C} = (\widetilde{c}_{ij})_{N\times N}$, then

$$\begin{aligned} \|\widetilde{C}\|_1 &:= \max_{1\le j\le N} \sum_{i=1}^N \widetilde{c}_{ij} = \max_{1\le j\le N} \sum_{i=1}^N s_i c_{ij}^{(2)} = \max_{1\le j\le N} \sum_{i=1}^N \sum_{k=1}^N c_{ik}^{(2)} c_{ij}^{(2)} \\ &= \max_{1\le j\le N} \sum_{k=1}^N ((C^{(2)})^T C^{(2)})_{kj} = \|(C^{(2)})^T C^{(2)}\|_1 \\ &\le \|(C^{(2)})^T\|_1 \|C^{(2)}\|_1 = r_\infty r_1, \end{aligned}$$

where $(C^{(2)})^T$ denotes the transposition of the matrix $(C^{(2)})$.

Lemma 4.1. *Assume $r_\infty r_1 < 1$, then for the matrix B given in (4.3),*

$$\|B\|_1 := \max_{1\le j\le N} \sum_{k=1}^N B_{kj} \le \frac{r_\infty r_1}{1 - r_\infty r_1}. \tag{4.5}$$

In particular

$$W_{2,d_{l_2^N}}(P(x,\cdot), P(y,\cdot)) \le \sqrt{\frac{r_\infty r_1}{1 - r_\infty r_1}} d_{l_2^N}(x,y), \ \forall x, y \in E^N. \tag{4.6}$$

Proof. The last conclusion (4.6) follows from (4.5) and (4.2). We only need to show (4.5). Just take the matrix $\widetilde{C}$ in place of the matrix C in [17, Lemma 3.2](i.e., $\widetilde{c}_{ij}$ takes place of c_{ij}, $i, j = 1, \cdots, N$). Here we give a sketch of the proof (for details refer to the proof of [17, Lemma 3.2]): first we obtain for $1 \le k \le N$,

$$B_{kj} = \begin{cases} 0, & \text{if } j = 1, \\ \sum_{h=1}^{j-1} \left(\sum_{l=1}^{k-1} \sum_{k>i_l>\cdots>i_2>i_1=h} \widetilde{c}_{k,i_l} \widetilde{c}_{i_l,i_{l-1}} \cdots \widetilde{c}_{i_2,i_1=h} \widetilde{c}_{h,j} + \widetilde{c}_{h,j} 1_{h=k} \right), & \text{if } 2 \le j \le N. \end{cases} \tag{4.7}$$

And then for fixed $j : 2 \le j \le N$, when $l : 1 \le l \le N-1$ and $h : 1 \le h \le j-1$,

$$\sum_{k=1}^N \sum_{k>i_l>\cdots>i_2>i_1=h} \widetilde{c}_{k,i_l} \widetilde{c}_{i_l,i_{l-1}} \cdots \widetilde{c}_{i_2,i_1=h} \le \sum_{k=1}^N (\widetilde{C}^l)_{kh} \le \|\widetilde{C}^l\|_1 \le (r_\infty r_1)^l,$$

thus by calculation we can show for $2 \le j \le N$, $\sum_{k=1}^N B_{kj} \le r_\infty r_1 + \cdots + (r_\infty r_1)^N \le \frac{r_\infty r_1}{1 - r_\infty r_1}$.

□

Lemma 4.2. *Assume (H1) and (H2), then*

$$P(x_0, \cdot) \in T_{2,d_{l_2^N}} \left(\frac{c}{(1 - \|C^{(2)}\|_2)^2} \right), \forall x_0 = (x_0^1, \cdots, x_0^N) \in E^N.$$

Proof. The proof is similar to the one used by [3, Theorem 2.5] or [17, Lemma 3.4]. First for simplicity denote $P(x_0, \cdot)$ by P and note that for $1 \le i \le N$,

$$X_N^1 = X_1^1, \cdots, X_N^i = X_i^i,$$

$$P(X_N^i \in \cdot | X_N^1, \cdots, X_N^{i-1}) = \mu_i(\cdot | X_N^1, \cdots, X_N^{i-1}, x_0^{i+1}, \cdots, x_0^N),$$

and thus

$$W_{2,d}(P(X_N^i \in \cdot|X_N^1 = x^1, \cdots, X_N^{i-1} = x^{i-1}), P(X_N^i \in \cdot|X_N^1 = y^1, \cdots, X_N^{i-1} = y^{i-1})) \\ \le \sum_{j=1}^{i-1} c_{ij}^{(2)} d(x^j, y^j).$$

For any probability measure Q on E^N such that $H(Q|P) < \infty$, let $Q_i(\cdot|x^{[1,i-1]})$ be the regular conditional law of x^i knowing $x^{[1,i-1]}$, where $i \ge 2, x^{[1,i-1]} = (x^1, \cdots, x^{i-1})$, and $Q_i(\cdot|x^{[1,i-1]})$ the law of x^1 for $i = 1$, all under law Q. Define $P_i(\cdot|x^{[1,i-1]})$ similarly but under P. We shall use the Kullback information between conditional distributions,

$$H_i(y^{[1,i-1]}) = H(Q_i(\cdot|y^{[1,i-1]})|P_i(\cdot|y^{[1,i-1]})),$$

and exploit the following important identity:

$$H(Q|P) = \sum_{i=1}^{N} \int_{E^N} H_i(y^{[1,i-1]}) dQ(y).$$

The key is to construct an appropriate coupling of Q and P, that is, two random sequences $Y^{[1,N]}$ and $X^{[1,N]}$ taking values on E^N distributed according to Q and P, respectively, on some probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. We define a joint distribution $\mathfrak{L}(Y^{[1,N]}, X^{[1,N]})$ by induction as follows (the Marton coupling).

At first the law of (Y^1, X^1) is the optimal coupling of $Q(x^1 \in \cdot)$ and $P(x^1 \in \cdot)$ $(= \mu_1(\cdot|x_0))$ such that

$$\mathbb{E}(d^2(Y^1, X^1)) = W_{2,d}^2(Q(x^1 \in \cdot), P(x^1 \in \cdot)).$$

Assume that for some $i, 2 \le i \le N, (Y^{[1,i-1]}, X^{[1,i-1]}) = (y^{[1,i-1]}, x^{[1,i-1]})$ is given. Then the joint conditional distribution $\mathfrak{L}(Y^i, X^i|Y^{[1,i-1]} = y^{[1,i-1]}, X^{[1,i-1]} = x^{[1,i-1]})$ is the optimal coupling of $Q_i(\cdot|y^{[1,i-1]})$ and $P_i(\cdot|x^{[1,i-1]})$, that is

$$\mathbb{E}(d^2(Y^i, X^i)|Y^{[1,i-1]} = y^{[1,i-1]}, X^{[1,i-1]} = x^{[1,i-1]}) = W_{2,d}^2(Q_i(\cdot|y^{[1,i-1]}), P_i(\cdot|x^{[1,i-1]})).$$

Obviously, $Y^{[1,N]}, X^{[1,N]}$ are of law Q, P respectively. By the triangle inequality for the $W_{2,d}$ distance,

$$\mathbb{E}(d^2(Y^i, X^i)|Y^{[1,i-1]} = y^{[1,i-1]}, X^{[1,i-1]} = x^{[1,i-1]}) \\ \le \left[W_{2,d}(Q_i(\cdot|y^{[1,i-1]}), P_i(\cdot|y^{[1,i-1]})) + W_{2,d}(P_i(\cdot|y^{[1,i-1]}), P_i(\cdot|x^{[1,i-1]}))\right]^2 \\ \le \left[W_{2,d}(Q_i(\cdot|y^{[1,i-1]}), P_i(\cdot|y^{[1,i-1]})) + \sum_{j=1}^{i-1} c_{ij}^{(2)} d(x^j, y^j)\right]^2.$$

The above inequality gives us

$$\sum_{i=1}^{N} \mathbb{E} d^2(Y^i, X^i) \le \sum_{i=1}^{N} \mathbb{E} \left[W_{2,d}(Q_i(\cdot|Y^{[1,i-1]}), P_i(\cdot|Y^{[1,i-1]})) + \sum_{j=1}^{i-1} c_{ij}^{(2)} d(X^j, Y^j)\right]^2.$$

Let $\xi = \left(d(Y^i, X^i)\right)_{i=1,\cdots,N}$, $\eta = \left(W_{2,d}(Q_i(\cdot|Y^{[1,i-1]}), P_i(\cdot|Y^{[1,i-1]}))\right)_{i=1,\cdots,N}$, and note that the norm of a general random vector $a = (a^i)_{i=1,\cdots,N}$ is defined to be $\sqrt{\sum_{i=1}^N \mathbb{E}(a^i)^2}(=: \|a\|_2)$, then

$$\|\xi\|_2 \le \|\eta + C^{(2)}\xi\|_2 \le \|\eta\|_2 + \|C^{(2)}\|_2 \|\xi\|_2.$$

So

$$\|\xi\|_2 \le \frac{\|\eta\|_2}{1-\|C^{(2)}\|_2} \le \frac{\sqrt{\sum_{i=1}^N \mathbb{E}[2cH_i(Y^{[1,i-1]})]}}{1-\|C^{(2)}\|_2} = \frac{\sqrt{2cH(Q|P)}}{1-\|C^{(2)}\|_2}.$$

Hence

$$W_{2,d_{l_2^N}}(Q,P) \le \sqrt{\frac{2cH(Q|P)}{(1-\|C^{(2)}\|_2)^2}},$$

i.e., $P = P(x_0,\cdot) \in T_{2,d_{l_2^N}}\left(\frac{c}{(1-\|C^{(2)}\|_2)^2}\right)$. □

In order to prove Theorem 3.1, we need the following dependent tensorization of T_2(from the result of Djellout-Guillin-Wu [3, Theorem 2.5]).

Lemma 4.3. ([3, Theorem 2.5]) *Let $\mathbb{P}$ be a probability measure on the product space $(E^n, \mathfrak{B}^n), n \ge 2$. For any $x = (x_1, \cdots, x_n) \in E^n, x_{[1,i]} := (x_1, \cdots, x_i)$. Let $\mathbb{P}_i(\cdot|x_{[1,i-1]})$ denote the regular conditional law of x_i given $x_{[1,i-1]}$ under $\mathbb{P}$ for $2 \le i \le n$, and $\mathbb{P}_i(\cdot|x_{[1,i-1]})$ be the distribution of x_1 for $i = 1$ where $x_{[1,0]}$ denotes some fixed point x_0 on E.*

Assume that

(1) For some metric d on E, there is a constant $\kappa > 0$ such that $\mathbb{P}_i(\cdot|x_{[1,i-1]}) \in T_2(\kappa)$ on (E,d) for all $i \ge 1, x_{[1,i-1]}$ in $E^{i-1}(E^0 := \{x_0\})$;

(2) there exist $a_j \ge 0$ with $r^2 := \sum_{j=1}^\infty a_j^2 < 1$ such that

$$[W_{2,d}(\mathbb{P}_i(\cdot|x_{[1,i-1]}), \mathbb{P}_i(\cdot|\tilde{x}_{[1,i-1]}))]^2 \le \sum_{j=1}^{i-1} (a_j)^2 d^2(x_{i-j}, \tilde{x}_{i-j}), \tag{4.8}$$

for all $i \ge 1, x_{[1,i-1]}, \tilde{x}_{[1,i-1]}$ in E^{i-1}.

Then for any probability measure $\mathbb{Q}$ on E^n,

$$W_{2,d_{l_2^n}}(\mathbb{Q},\mathbb{P}) \le \frac{\sqrt{2\kappa H(\mathbb{Q}|\mathbb{P})}}{1-r}.$$

By Lemma 4.3 above, we can obtain the following key lemma, which can be considered as the main theoretical result of this paper.

Lemma 4.4. *On the path space $(E^N)^n$, consider the following $(d_{l_2})_{l_2}$-metric*

$$(d_{l_2})_{l_2}(\omega,\tilde{\omega}) := \left(\sum_{k=1}^n \sum_{j=1}^N d^2(\omega_k^j, \tilde{\omega}_k^j)\right)^{1/2}, \ \omega, \tilde{\omega} \in (E^N)^n.$$

Let $\mathbb{P}_x$ be the distribution of our Gibbs sampling $(Z_1, \cdots, Z_n)$ on $(E^N)^n$ equipped with the Borel-σ algebra, where the starting point $x \in E^N$ is arbitrary. Assume $r_\infty r_1 < \frac{1}{2}$ and $(H2)$. Then for any probability measure Q on $((E^N)^n, (d_{l_2})_{l_2})$, we have

$$W_{2,(d_{l_2})_{l_2}}(\mathbb{Q},\mathbb{P}_x) \le \frac{\sqrt{2cH(\mathbb{Q}|\mathbb{P}_x)/(1-\|C^{(2)}\|_2)^2}}{1-\sqrt{r_\infty r_1/(1-r_\infty r_1)}},$$

In other words

$$\mathbb{P}_x \in T_{2,(d_{l_2})_{l_2}}\left(\frac{c}{[1-\sqrt{r_\infty r_1/(1-r_\infty r_1)}]^2(1-\|C^{(2)}\|_2)^2}\right).$$

Proof. We will apply Lemma 4.3 with (E,d) (and $\mathbb{P}_i(\cdot|x_{[1,i-1]})$) being $(E^N, d_{l_2^N})$ (and $(P(x,\cdot))$ respectively. By the Remark 3.2, $(H1)$ holds. By $(H2)$ and Lemma 4.2, $P(x,\cdot)$ satisfies Talagrand's T_2-transport inequality uniformly on $x \in E^N$,i.e., the first assumption of Lemma 4.3 holds. Since $r_\infty r_1 < \frac{1}{2}$, by (4.6) of Lemma 4.1, the contraction constant r in Lemma 4.3 satisfies $r \le \sqrt{\frac{r_\infty r_1}{1-r_\infty r_1}}$ (<1). So Lemma 4.3 yields the desired result. □

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. By the Remark 3.2, $r_\infty r_1 < \frac{1}{2}$ implies $(H1)$, and because of assumption $(H2)$, Lemma 4.2 yields part (a) of Theorem 3.1.

Let $F(Z_1,\cdots,Z_n) = \frac{1}{n}\sum_{k=1}^n f(Z_k)$, then the Lipschitzian norm $\|F\|_{Lip}$ of F with respect to the metric $(d_{l_2})_{l_2}$ is not greater than $\|f\|_{Lip(d_{l_2^N})}/\sqrt{n} \le \alpha/\sqrt{n}$. Let $\mathbb{P}_x$ be the law of $(Z_1,\cdots,Z_n)$ on $(E^N)^n$, then by Lemma 4.4 and the famous Bobkov-Götze's criterion (see [17, Lemma 2.1]), we obtain the desired part (b) in Theorem 3.1.

For any $x \in E^N$, let $(Z_k = X_{kN}, Z_k' = Y_{kN})$ be the coupled Markov chain with initial condition $(X_0 = x, Y_0)$ as a coupling of (δ_x, μ), constructed at the beginning of this section. we have

$$\begin{aligned}|\frac{1}{n}\sum_{k=1}^n \mathbb{E}_x(f(Z_k)) - \mu(f)| &\le \frac{1}{n}\sum_{k=1}^n |\mathbb{E}_x(f(Z_k)) - \mu(f)| \le \frac{1}{n}\sum_{k=1}^n \mathbb{E}|f(Z_k) - f(Z_k^{'})| \\ &\le \frac{1}{n}\sum_{k=1}^n \|f\|_{Lip(d_{l_2^N})}\mathbb{E}d_{l_2^N}(Z_k, Z_k^{'}) \le \frac{1}{n}\sum_{k=1}^n \|f\|_{Lip(d_{l_2^N})}\sqrt{\mathbb{E}d^2_{l_2^N}(Z_k, Z_k^{'})},\end{aligned}$$

the last inequality holds because of Jensen's inequality.

By (4.4) and Lemma 4.1, the last term is bounded from above by

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^n \|f\|_{Lip(d_{l_2^N})}\sqrt{\|B^k\|_1 \mathbb{E}d^2_{l_2^N}(X_0,Y_0)} &\le \frac{1}{n}\sum_{k=1}^n \|f\|_{Lip(d_{l_2^N})}\sqrt{(\frac{r_\infty r_1}{1-r_\infty r_1})^k \mathbb{E}d^2_{l_2^N}(X_0,Y_0)} \\ &\le \frac{\alpha}{n}\frac{\sqrt{r_\infty r_1}}{\sqrt{1-r_\infty r_1} - \sqrt{r_\infty r_1}}\sqrt{\int_{E^N}\sum_{i=1}^N d(x^i,y^i)^2\mu(dy)}.\end{aligned}$$

Thus we obtain part (c) in Theorem 3.1 from its part (b). □

References

[1] Bobkov, S. and Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal., 163(1): 1-28,* 1999. MR-1682772

[2] Cattiaux, P.; Guillin, A. and Wu, L.: A note on Talagrand's transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Related Fields* 148, no. 1-2, 285-304, 2010. MR-2653230

[3] Djellout, H.; Guillin, A. and Wu, L.: Transportation cost-information inequalities and application to random dynamical systems and diffusions. *Ann. Probab.,* 32(3B): 2702-2732, 2004. MR-2078555

[4] Dobrushin, R. L.: The description of a random field by means of conditional probabilities and condition of its regularity. *Theory Probab. Appl.,* 13(2): 197-224, 1968. MR-0231434

[5] Dobrushin, R. L.: Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl.,* 15(3): 458-486, 1970.

[6] Dyer, M.; Goldberg, L. A. and Jerrum, M.: Dobrushin conditions and Systematic scan. *Combin. Probab. Comput,* 17(6): 761-779, 2008. MR-2463409

[7] Gozlan, N. and Léonard, C.: A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* 139(1-2): 235-283, 2007. MR-2322697

[8] Gozlan, N. and Léonard, C.: Transport inequalities. A survey. Markov Process. Related Fields 16(4): 635-736, 2010. MR-2895086

[9] Ledoux, M.: Concentration of measure and logarithmic Sobolev inequalities. *Séminaire de Probabilités XXXVI.* Lecture Notes in Math. 1709, 120-216. Springer-Verlag, Berlin, 1999. MR-1767995

[10] Ledoux, M.: The concentration of measure phenomenon. *Amer. Math. Soc.*, Providence, RI 2001. MR-1849347

[11] Marton, K.: Bounding $\bar{d}$-distance by information divergence : A method to prove measure concentration. *Ann. Probab.*, 24(2): 857-866, 1996. MR-1404531

[12] Marton, K.: A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3): 556-571, 1997. MR-1392329

[13] Marton, K.: Measure concentration for Euclidean distance in the case of dependent random variables. *Ann. Probab.*, 32(3B): 2526-2544, 2004. MR-2078549

[14] Otto, F. and Villani, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* 173, 361-400, 2000. MR-1760620

[15] Talagrand, M: Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3): 587-600, 1996. MR-1392331

[16] Villani, C: *Topics in Optimal Transportation*. Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003. MR-1964483

[17] Wang, N.-Y. and Wu, L.: Convergence rate and concentration inequalities for Gibbs sampling in high dimension. To appear in Bernoulli.

Acknowledgments. I am grateful to the referees for their conscientious and helpful comments which significantly improved the paper.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)

Economical model of EJP-ECP

- Low cost, based on free software (OJS[1])
- Non profit, sponsored by IMS[2], BS[3], PKP[4]
- Purely electronic and secure (LOCKSS[5])

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1] OJS: Open Journal Systems http://pkp.sfu.ca/ojs/
[2] IMS: Institute of Mathematical Statistics http://www.imstat.org/
[3] BS: Bernoulli Society http://www.bernoulli-society.org/
[4] PK: Public Knowledge Project http://pkp.sfu.ca/
[5] LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[6] IMS Open Access Fund: http://www.imstat.org/publications/open.htm