Electron. Commun. Probab. **22** (2017), no. 56, 1–7.
DOI: 10.1214/17-ECP90
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Convergence rates of the random scan Gibbs sampler under the Dobrushin's uniqueness condition*

Neng-Yi Wang†

Abstract

In this paper, under the Dobrushin's uniqueness condition, we obtain explicit estimates of the geometrical convergence rate for the random scan Gibbs sampler in the Wasserstein metric.

Keywords: random scan Gibbs sampler; coupling method; Wasserstein metric; Dobrushin's uniqueness condition.
AMS MSC 2010: 60E15; 65C05.
Submitted to ECP on May 15, 2017, final version accepted on September 25, 2017.

1 Introduction

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

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; and $\bar{\mu}_i(dy|x) = \left(\prod_{j\neq i} \delta_{x^j}(dy^j)\right) \bigotimes \mu_i(dy^i|x)$(product measure), where $\delta_\cdot$ is the Dirac measure at the point $\cdot$. We see that

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

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

In order to approximate μ via iterations of the one-dimensional conditional distributions $\mu_i, i = 1, \cdots, N$, the various scan Gibbs samplers are often used (see [4]). In [6], Wu and the author studied systematic scan Gibbs sampler by Dobrushin's uniqueness conditions. In this paper, we will study the random scan Gibbs sampler.

The scheme of the random scan Gibbs sampler approximating μ is that, in each iteration, one randomly chooses one coordinate to update according to the one-dimensional conditional distributions $\mu_i, i = 1, \cdots, N$. It is described as follows. Given any initial value $X_0 = (X_0^1, \cdots, X_0^N) \in E^N$, independently draw an index σ from the uniform distribution on the index set $\{1, \cdots, N\}$, then draw X_1^σ from $\mu_\sigma(\cdot|x)$ and take

*Supported by National Natural Science Foundation of China(No.11601170), and Fundamental Research Funds for the Central Universities, HUST, China, under Grants(No.2014QNRC015, No.2015QN127).

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

$X_1^i = X_0^i, i \neq \sigma$, completing one iteration of the scheme. After m such iterations, we obtain $X_m = (X_m^1, \cdots, X_m^N)$. Thus this scan Gibbs sampler is exactly the time-homogeneous Markov chain $\{X_m : m = 0, 1, \cdots\}$ with invariant distribution μ, whose one step transition probability $P(x, dy) = \frac{1}{N}\sum_{i=1}^N \bar{\mu}_i(dy|x)$.

Our objective is to study the convergence rate of the $m-$step transition probability P^m to μ under Dobrushin's uniqueness conditions as m tends to ∞. To this end, by coupling, our main idea is to establish some contractive properties in the sense of the maximum or sum distance (respectively, see the two lemmas in Section 3). Although the coupling is similar to [6], to prove the contractive properties is very different because of this scan Gibbs sampler with the random index σ instead of the systematic scan in [6].

This paper is organized as follows. We present the main results in Section 2, and then prove them in Section 3.

2 Main results

Throughout the paper E is a Polish space with the Borel σ-field $\mathfrak{B}$, and d is a metric on E such that $d(\cdot,\cdot)$ is lower semi-continuous on E^2. On the product space E^N, we consider the l_1^N-metric

$$d_{l_1^N}(x,y) := \sum_{i=1}^N d(x^i, y^i),\ x = (x^1, \cdots, x^N), y = (y^1, \cdots, y^N) \in E^N. \tag{2.1}$$

The product space E^N is always endowed with the $d_{l_1^N}$-metric unless otherwise stated. Let $\mathcal{M}_1(E)$ be the space of Borel probability measures on E, and

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

(Here $x_0 \in E$ is some fixed point, but the definition above does not depend on x_0 by the triangle inequality). Given $\nu_1, \nu_2 \in \mathcal{M}_1^d(E)$, the L^1-Wasserstein distance between ν_1, ν_2 is given by

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

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).

Recall the Kantorovich-Rubinstein duality relation ([5])

$$W_{1,d}(\nu_1, \nu_2) = \sup_{\|f\|_{Lip(d)} \le 1} \int_E f d(\nu_1 - \nu_2), \quad \|f\|_{Lip(d)} := \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}.$$

Throughout the paper we assume that $\int_{E^N} d(y^i, x_0^i) d\mu(y) < \infty$, $\mu_i(\cdot|x) \in \mathcal{M}_1^d(E)$ 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 $(E^N, d_{l_1^N})$ to $(\mathcal{M}_1^d(E), W_{1,d})$. For $x = (x^1, \cdots, x^N), y = (y^1, \cdots, y^N) \in E^N$, the expression

$$x = y \text{ off} j$$

means that for $k = 1, \cdots, N$,

$$\begin{cases} x^k \neq y^k, & \text{if} \quad k = j, \\ x^k = y^k, & \text{if} \quad k \neq j. \end{cases}$$

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

$$c_{ij} := \sup_{x=y \text{ off} j} \frac{W_{1,d}(\mu_i(\cdot|x), \mu_i(\cdot|y))}{d(x^j, y^j)}, i, j = 1, \cdots, N. \tag{2.3}$$

Obviously $c_{ii} = 0$. By the triangle inequality of the metric $W_{1,d}$, for any $x, y \in E^N$,

$$W_{1,d}(\mu_i(\cdot|x), \mu_i(\cdot|y)) \le \sum_{j=1}^N c_{ij} d(x^j, y^j),\ 1 \le i \le N. \tag{2.4}$$

Then the well known Dobrushin uniqueness condition (see [1, 2] or [6]) is read as

$$(H1) \quad r_\infty := \max_{1\le i\le N} \sum_{j=1}^N c_{ij} < 1$$

or

$$(H2) \quad r_1 := \max_{1\le j\le N} \sum_{i=1}^N c_{ij} < 1.$$

Notice that r_∞ (or r_1) coincides with the operator norm of the N by N matrix C acting as an operator from l_∞^N (or l_1^N respectively) to itself.

Recall that for any function $f : E^N \to \mathbb{R}$, $\|f\|_{Lip(d_{l_1^N})} = \max_{1\le i\le N} \delta_i(f)$, where $\delta_i(f) := \sup_{x=y \text{ off } i} \frac{|f(x)-f(y)|}{d(x^i,y^i)}$.

Theorem 2.1. (Convergence Rate 1) *Under the Dobrushin uniqueness condition $(H1)$, we have:*

(a) *For any Lipschitzian function f on E^N and two initial distributions ν_1, ν_2 on E^N,*

$$|\nu_1 P^m f - \nu_2 P^m f| \le \left(1 - \frac{1-r_\infty}{N}\right)^m \max_{1\le i\le N} \mathbb{E} d(X_0^i(1), X_0^i(2)) \sum_{i=1}^N \delta_i(f),\ \forall m \ge 1, \tag{2.5}$$

where $(X_0(1), X_0(2))$ is a coupling of (ν_1, ν_2), i.e., the law of $X_0(j)$ is ν_j for $j = 1, 2$.

(b) *In particular for any initial distribution ν on E^N,*

$$W_{1,d_{l_1^N}}(\nu P^m, \mu) \le N \left(1 - \frac{1-r_\infty}{N}\right)^m \max_{1\le i\le N} \mathbb{E} d(X_0^i(1), X_0^i(2)),\ \forall m \ge 1,$$

where $(X_0(1), X_0(2))$ is a coupling of (ν, μ).

Theorem 2.2. (Convergence Rate 2) *Under the Dobrushin uniqueness condition $(H2)$, we have:*

(a) *For any Lipschitzian function f on E^N and two initial distributions ν_1, ν_2 on E^N,*

$$|\nu_1 P^m f - \nu_2 P^m f| \le \left(1 - \frac{1-r_1}{N}\right)^m \|f\|_{Lip(d_{l_1^N})} \mathbb{E} d_{l_1^N}(X_0(1), X_0(2)),\ \forall m \ge 1, \tag{2.6}$$

where $(X_0(1), X_0(2))$ is a coupling of (ν_1, ν_2).

(b) *In particular for any initial distribution ν on E^N,*

$$W_{1,d_{l_1^N}}(\nu P^m, \mu) \le \left(1 - \frac{1-r_1}{N}\right)^m \mathbb{E} d_{l_1^N}(X_0(1), X_0(2)),\ \forall m \ge 1,$$

where $(X_0(1), X_0(2))$ is a coupling of (ν, μ).

Remark 2.3. These results above show geometric convergences of the distribution of X_m to the invariant distribution μ, which is useful in practice for the rapid convergence of this sampler.

Remark 2.4. As indicated by a referee, it is especially relevant for statistical applications to investigate exponential concentration inequalities for the convergence of empirical means $\frac{1}{n}\sum_{i=1}^n f(X_i)$ to $\int_{E^N} f d\mu$. But regretfully unlike the systematic scan sampler of [6], because the current sampler has random index σ, we don't succeed in establishing those concentration inequalities under the Dobrushin's uniqueness condition.

3 Proofs of the main results

Given any two initial distributions ν_1 *and* ν_2 *on* E^N, we construct our coupled homogeneous Markov chain $(X_m, Y_m)_{m\geq 0}$, which is quite close to the coupling by K. Marton [3] (see also [6]).

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

$$(X_{m-1}, Y_{m-1}) = (x, y) \in E^N \times E^N, \quad \sigma = k,$$

then

$$X_m^i = x^i, Y_m^i = y^i, i \neq k,$$

and

$$\mathbb{P}((X_m^k, Y_m^k) \in \cdot | (X_{m-1}, Y_{m-1}) = (x, y), \sigma = k) = \pi(\cdot | x, y),$$

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

$$\iint_{E^2} d(\tilde{x}, \tilde{y}) \pi(d\tilde{x}, d\tilde{y}|x, y) = W_{1,d}(\mu_k(\cdot|x), \mu_k(\cdot|y)),$$

i.e.,

$$\mathbb{E}[d(X_m^k, Y_m^k)|(X_{m-1}, Y_{m-1}) = (x, y), \sigma = k] = W_{1,d}(\mu_k(\cdot|x), \mu_k(\cdot|y)).$$

Then we have:

Lemma 3.1. *Under the Dobrushin uniqueness condition* $(H1)$*, for any* $m \geq 1$*,*

$$\max_{1\leq i\leq N} \mathbb{E}[d(X_m^i, Y_m^i)] \leq \left(1 - \frac{1 - r_\infty}{N}\right)^m \max_{1\leq i\leq N} \mathbb{E} d(X_0^i, Y_0^i).$$

Proof. For $\forall i \geq 1, m \geq 1$,

$$\begin{aligned}
&\mathbb{E}[d(X_m^i, Y_m^i)|(X_{m-1}, Y_{m-1}) = (x, y)] \\
&= \frac{1}{N}\sum_{k=1}^N \mathbb{E}[d(X_m^i, Y_m^i)|(X_{m-1}, Y_{m-1}) = (x, y), \sigma = k] \\
&= \frac{1}{N}\left[W_{1,d}(\mu_i(\cdot|x), \mu_i(\cdot|y)) + \sum_{k\neq i} d(x^i, y^i)\right] \\
&\leq \frac{1}{N}\left[\sum_{k=1}^N c_{ik} d(x^k, y^k) + (N-1) d(x^i, y^i)\right],
\end{aligned}$$

where the last inequality above holds because of (2.4). And thus

$$\mathbb{E}[d(X_m^i, Y_m^i)|X_{m-1}, Y_{m-1}] \leq \frac{1}{N}\left[\sum_{k=1}^N c_{ik} d(X_{m-1}^k, Y_{m-1}^k) + (N-1) d(X_{m-1}^i, Y_{m-1}^i)\right],$$

hence

$$\begin{aligned}\mathbb{E}[d(X_m^i, Y_m^i)] &\leq \frac{1}{N}\left[\sum_{k=1}^{N} c_{ik}\mathbb{E}d(X_{m-1}^k, Y_{m-1}^k) + (N-1)\mathbb{E}d(X_{m-1}^i, Y_{m-1}^i)\right] \\ &\leq \frac{1}{N}\left(\sum_{k=1}^{N} c_{ik} + N - 1\right)\max_{1\leq k\leq N}\mathbb{E}d(X_{m-1}^k, Y_{m-1}^k) \\ &\leq \frac{r_\infty + N - 1}{N}\max_{1\leq k\leq N}\mathbb{E}d(X_{m-1}^k, Y_{m-1}^k) \\ &= \left(1 - \frac{1-r_\infty}{N}\right)\max_{1\leq k\leq N}\mathbb{E}d(X_{m-1}^k, Y_{m-1}^k).\end{aligned}$$

Because the inequalities hold for any $i \geq 1, m \geq 1$, and by induction,

$$\max_{1\leq i\leq N}\mathbb{E}[d(X_m^i, Y_m^i)] \leq \left(1 - \frac{1-r_\infty}{N}\right)^m \max_{1\leq i\leq N}\mathbb{E}d(X_0^i, Y_0^i).$$ □

Proof of Theorem 2.1. Let $X_m(1) = X_m, X_m(2) = Y_m$.

(a) For any Lipschitzian function $f : E^N \to \mathbb{R}$, by Lemma 3.1,

$$\begin{aligned}|\nu_1 P^m f - \nu_2 P^m f| = |\mathbb{E}f(X_m(1)) - \mathbb{E}f(X_m(2))| &\leq \sum_{i=1}^{N}\delta_i(f)\mathbb{E}d(X_m^i(1), X_m^i(2)) \\ &\leq \max_{1\leq i\leq N}\mathbb{E}d(X_m^i(1), X_m^i(2))\sum_{i=1}^{N}\delta_i(f) \\ &\leq \left(1 - \frac{1-r_\infty}{N}\right)^m \max_{1\leq i\leq N}\mathbb{E}d(X_0^i(1), X_0^i(2))\sum_{i=1}^{N}\delta_i(f).\end{aligned}$$

(b) For $\nu_1 = \nu$, $\nu_2 = \mu$, since $\mu P = \mu$, we have:

$$\begin{aligned}W_{1,d_{l_1^N}}(\nu P^m, \mu) = W_{1,d_{l_1^N}}(\nu P^m, \mu P^m) &\leq \mathbb{E}d_{l_1^N}(X_m(1), X_m(2)) = \sum_{i=1}^{N}\mathbb{E}d(X_m^i(1), X_m^i(2)) \\ &\leq N\max_{1\leq i\leq N}\mathbb{E}d(X_m^i(1), X_m^i(2)) \\ &\leq N\left(1 - \frac{1-r_\infty}{N}\right)^m \max_{1\leq i\leq N}\mathbb{E}d(X_0^i(1), X_0^i(2)).\end{aligned}$$

□

Lemma 3.2. *Under the Dobrushin uniqueness condition* $(H2)$, *for any* $m \geq 1$,

$$\mathbb{E}d_{l_1^N}(X_m, Y_m) \leq \left(1 - \frac{1-r_1}{N}\right)^m \mathbb{E}d_{l_1^N}(X_0, Y_0).$$

Proof.

$$\begin{aligned}
&\mathbb{E}[d_{l_1^N}(X_m, Y_m)|(X_{m-1}, Y_{m-1}) = (x, y)] \\
&= \sum_{i=1}^N \mathbb{E}[d(X_m^i, Y_m^i)|(X_{m-1}, Y_{m-1}) = (x, y)] \\
&= \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^N \mathbb{E}[d(X_m^i, Y_m^i)|(X_{m-1}, Y_{m-1}) = (x, y), \sigma = k] \\
&= \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^N \mathbb{E}[d(X_m^i, Y_m^i)|(X_{m-1}, Y_{m-1}) = (x, y), \sigma = k] \\
&= \frac{1}{N} \sum_{k=1}^N \left\{ W_{1,d}(\mu_k(\cdot|x), \mu_k(\cdot|y)) + \sum_{i \neq k} d(x^i, y^i) \right\} \\
&= \frac{1}{N} \sum_{k=1}^N W_{1,d}(\mu_k(\cdot|x), \mu_k(\cdot|y)) + \frac{1}{N} \sum_{k=1}^N \sum_{i \neq k} d(x^i, y^i) \\
&\leq \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^N c_{ki} d(x^i, y^i) + \frac{1}{N} \sum_{k=1}^N \sum_{i \neq k} d(x^i, y^i) \\
&= \frac{1}{N} \sum_{i=1}^N d(x^i, y^i) \sum_{k=1}^N c_{ki} + \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^N d(x^i, y^i) - \frac{1}{N} \sum_{k=1}^N d(x^k, y^k) \\
&= \frac{1}{N} \sum_{i=1}^N d(x^i, y^i) \sum_{k=1}^N c_{ki} + \frac{N-1}{N} \sum_{i=1}^N d(x^i, y^i) \\
&= \sum_{i=1}^N d(x^i, y^i) \left(\frac{1}{N} \sum_{k=1}^N c_{ki} + \frac{N-1}{N} \right) \\
&\leq \frac{r_1 + N - 1}{N} d_{l_1^N}(x, y) \leq \left(1 - \frac{1 - r_1}{N}\right) d_{l_1^N}(x, y),
\end{aligned}$$

i.e.,

$$\mathbb{E}[d_{l_1^N}(X_m, Y_m)|X_{m-1}, Y_{m-1}] \leq \left(1 - \frac{1 - r_1}{N}\right) d_{l_1^N}(X_{m-1}, Y_{m-1}),$$

and thus,

$$\mathbb{E}[d_{l_1^N}(X_m, Y_m)] \leq \left(1 - \frac{1 - r}{N}\right) \mathbb{E}d_{l_1^N}(X_{m-1}, Y_{m-1}),$$

and by induction, $\mathbb{E}d_{l_1^N}(X_m, Y_m) \leq \left(1 - \frac{1-r_1}{N}\right)^m \mathbb{E}d_{l_1^N}(X_0, Y_0)$. □

Proof of Theorem 2.2. Let $X_m(1) = X_m, X_m(2) = Y_m$.

(a) For any Lipschitzian function $f : E^N \to \mathbb{R}$, by Lemma 3.2,

$$\begin{aligned}
|\nu_1 P^m f - \nu_2 P^m f| &= |\mathbb{E}f(X_m(1)) - \mathbb{E}f(X_m(2))| \\
&\leq \|f\|_{Lip(d_{l_1^N})} \mathbb{E}d_{l_1^N}(X_m(1), X_m(2)) \\
&\leq \left(1 - \frac{1 - r_1}{N}\right)^m \|f\|_{Lip(d_{l_1^N})} \mathbb{E}d_{l_1^N}(X_0(1), X_0(2)).
\end{aligned}$$

(b) For $\nu_1 = \nu$, $\nu_2 = \mu$, since $\mu P = \mu$, we have:

$$\begin{aligned}
W_{1,d_{l_1^N}}(\nu P^m, \mu) = W_{1,d_{l_1^N}}(\nu P^m, \mu P^m) &\leq \mathbb{E}d_{l_1^N}(X_m(1), X_m(2)) \\
&\leq \left(1 - \frac{1 - r_1}{N}\right)^m \mathbb{E}d_{l_1^N}(X_0(1), X_0(2)).
\end{aligned}$$ □

References

[1] 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

[2] Dobrushin, R. L.: Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl.*, 15(3): 458–486, 1970. MR-0298716

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

[4] Roberts, Gareth O. and Rosenthal, Jeffrey S.: General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1: 20–71, 2004. MR-2095565

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

[6] Wang, N.-Y. and Wu, L.: Convergence rate and concentration inequalities for Gibbs sampling in high dimension. *Bernoulli*, 20(4): 1698–1716, 2014. MR-3263086

Acknowledgments. The author would like to thank the anonymous referees for their helpful suggestions which improved the presentation of this paper.