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.

1 Introduction

Let \( \mu \) be a Gibbs probability measure on \( E^N \) with dimension \( N \), i.e.,

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

where \( \pi \) is some \( \sigma \)-finite reference measure on \( E \).

Let \( \mu_i(x) \) be the regular conditional distribution of \( x^i \) knowing \( (x^j, j \neq i) \) under \( \mu \); and \( \bar{\mu}_i(dy|x) = \left( \prod_{j \neq i} \delta_{x^j}(dy^j) \right) \otimes \mu_i(dy^i|x) \) (product measure), where \( \delta \) is the Dirac measure at the point \( x \). 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 \( \mu \) 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 \( \mu \) 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 \( X_0 = (X^1_0, \cdots, X^N_0) \in E^N \), independently draw an index \( \sigma \) from the uniform distribution on the index set \( \{1, \cdots, N\} \), then draw \( X^\sigma_i \) from \( \mu_i(\cdot|x) \) 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
Convergence rates of the random scan Gibbs sampler

\[ X_i^1 = X_i^0, i \neq \sigma, \] completing one iteration of the scheme. After \( m \) such iterations, we obtain \( X_m = (X_1^1, \cdots, X_N^1) \). Thus this scan Gibbs sampler is exactly the time-homogeneous Markov chain \( \{X_m : m = 0, 1, \cdots \} \) with invariant distribution \( \mu \), whose one step transition probability \( P(x, dy) = \frac{1}{2} \sum_{i=1}^N \mu_i(dy|x) \).

Our objective is to study the convergence rate of the \( m \)-step transition probability \( P^m \) to \( \mu \) under Dobrushin’s uniqueness conditions as \( m \) tends to \( \infty \). 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 \( \sigma \) 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 \( \sigma \)-field \( \mathcal{B} \), 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

\[ d_{l_1^N}(x,y) := \sum_{i=1}^N d(x_i,y_i), \quad x = (x^1, \cdots, x^N), y = (y^1, \cdots, y^N) \in E^N. \]  

(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

\[ \mathcal{M}_1^1(E) := \left\{ \nu \in \mathcal{M}_1(E) : \int_E d(x_0,x)\nu(dx) < \infty \right\}. \]

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

\[ W_{1,d}(\nu_1, \nu_2) := \inf \int_{E \times E} d(x,y)\pi(dx,dy), \]  

(2.2)

where the infimum is taken over all probability measures \( \pi \) on \( E \times E \) such that its marginal distributions are respectively \( \nu_1 \) and \( \nu_2 \) (called a coupling of \( \nu_1 \) and \( \nu_2 \)).

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

\[ W_{1,d}(\nu_1, \nu_2) = \sup_{\|f\|_{L^1(d\mu)} \leq 1} \int_E f(d\nu_1 - d\nu_2), \quad \|f\|_{L^1(d\mu)} := \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}. \]

Throughout the paper we assume that \( \int_{E^N} d(y',x_0^j)\mu_i(y) < \infty, \mu_i(\cdot|x) \in \mathcal{M}_1^1(E) \) for all \( i = 1, \cdots, N \) and \( x \in E^N \), where \( x_0 \) is some fixed point of \( E^N \), and \( x \mapsto \mu_i(\cdot|x) \) is Lipschitzian from \( (E^N, d_{l_1^N}) \) to \( (\mathcal{M}_1^1(E), W_{1,d}) \). For \( x = (x^1, \cdots, x^N), y = (y', \cdots, y^N) \in E^N \), the expression

\[ x = y \text{ off} j \]

means that for \( k = 1, \cdots, N \),

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

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

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

(2.3)
Convergence rates of the random scan Gibbs sampler

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

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

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

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

or

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

Notice that $r_{\infty}$ (or $r_1$) coincides with the operator norm of the $N$ by $N$ matrix $C$ acting as an operator from $l^N_w$ (or $l^N_2$ respectively) to itself.

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

**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 $\nu_1, \nu_2$ on $E^N$,

$$|\nu_1 P^m f - \nu_2 P^m f| \leq \left(1 - \frac{1 - r_{\infty}}{N}\right)^m \max_{1 \leq i \leq N} \text{Ed}(X^i_0(1), X^i_0(2)) \sum_{i=1}^{N} \delta_i(f), \quad \forall m \geq 1,$$

where $(X_0(1), X_0(2))$ is a coupling of $(\nu_1, \nu_2)$, i.e., the law of $X_0(j)$ is $\nu_j$ for $j = 1, 2$.

(b) In particular for any initial distribution $\nu$ on $E^N$,

$$W_{1,d_{N}}(\nu^{P^m}, \nu) \leq N \left(1 - \frac{1 - r_{\infty}}{N}\right)^m \max_{1 \leq i \leq N} \text{Ed}(X^i_0(1), X^i_0(2)), \quad \forall m \geq 1,$$

where $(X_0(1), X_0(2))$ is a coupling of $(\nu, \nu)$.

**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 $\nu_1, \nu_2$ on $E^N$,

$$|\nu_1 P^m f - \nu_2 P^m f| \leq \left(1 - \frac{1 - r_1}{N}\right)^m \|f\|_{\text{Lip}(d_{N})} \text{Ed}_{l_{N}}(X_0(1), X_0(2)), \quad \forall m \geq 1,$$  

where $(X_0(1), X_0(2))$ is a coupling of $(\nu_1, \nu_2)$.

(b) In particular for any initial distribution $\nu$ on $E^N$,

$$W_{1,d_{l_{N}}}(\nu^{P^m}, \nu) \leq \left(1 - \frac{1 - r_1}{N}\right)^m \text{Ed}_{l_{N}}(X_0(1), X_0(2)), \quad \forall m \geq 1,$$

where $(X_0(1), X_0(2))$ is a coupling of $(\nu, \nu)$.

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

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

i.e.,

Lemma 3.1. 

ECP 22 (2017), paper 56.

3 Proofs of the main results

Given any two initial distributions ν1 and ν2 on EN, we construct our coupled homoge-

neous Markov chain (Xm,Ym)m\geq0, which is quite close to the coupling by K. Marton

[3] (see also [6]).

Let (X0,Y0) be a coupling of (ν1,ν2). And given

(Xm−1,Ym−1) = (x,y) ∈ EN × EN, σ = k,

then

Xm

i

\overset{\mathrm{d}}{=} x^i, Ym

i

= y^i, i \neq k,

and

\mathbb{P}( (Xm

k

,Ym

k

) ∈ ⋃(Xm−1,Ym−1) = (x,y), σ = k ) = π(|x,y),

where π(|x,y) is an optimal coupling of μk(|x) and μk(|y) such that

\int \int_{E^2} d(\tilde{x},y)\pi(\tilde{x},\tilde{y}|x,y) = W_{1,d}(\mu_k(|x),\mu_k(|y)),

i.e.,

\mathbb{E}[d(Xm

k

,Ym

k

)|(Xm−1,Ym−1) = (x,y), σ = k] = W_{1,d}(\mu_k(|x),\mu_k(|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(Xm

i

,Ym

i

)] \leq \left(1 - \frac{r}{N}\right)^m \max_{1 \leq i \leq N} \mathbb{E}d(X_0

i

,Y_0

i

).

Proof. For ∀i \geq 1, m \geq 1,

\mathbb{E}[d(Xm

i

,Ym

i

)|(Xm−1,Ym−1) = (x,y)]

= \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}[d(Xm

i

,Ym

i

)|(Xm−1,Ym−1) = (x,y), σ = k]

= \frac{1}{N} \left[ W_{1,d}(\mu_i(|x),\mu_i(|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],

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

\mathbb{E}[d(Xm

i

,Ym

i

)|X_{m-1},Y_{m-1}] \leq \frac{1}{N} \left[ \sum_{k=1}^{N} c_{ik} d(Xm_{k-1},Ym_{k-1}) + (N-1)d(Xm_{i-1},Ym_{i-1}) \right],
Convergence rates of the random scan Gibbs sampler

\[
E[d(X^i_m, Y^i_m)] \leq \frac{1}{N} \left[ \sum_{k=1}^{N} c_{ik} E[d(X^k_{m-1}, Y^k_{m-1})] + (N - 1) E[d(X^i_{m-1}, Y^i_{m-1})] \right]
\]

where \(c_{ik} = \max_{1 \leq k \leq N} E[d(X^k_{m-1}, Y^k_{m-1})] \leq \frac{r_{\infty} + N - 1}{N} \max_{1 \leq k \leq N} E[d(X^k_{m-1}, Y^k_{m-1})] \)

\[
= \left(1 - \frac{1 - r_{\infty}}{N}\right) \max_{1 \leq k \leq N} E[d(X^k_{m-1}, Y^k_{m-1})].
\]

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

\[
\max_{1 \leq i \leq N} E[d(X^i_m, Y^i_m)] \leq \left(1 - \frac{1 - r_{\infty}}{N}\right)^m \max_{1 \leq i \leq N} E[d(X^i_0, Y^i_0)].
\]

**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,

\[
|\nu_1 P^m f - \nu_2 P^m f| = |E[f(X_m(1))] - E[f(X_m(2))]| \leq \sum_{i=1}^{N} \delta_i(f) E[d(X^i_m(1), X^i_m(2))]
\]

\[
\leq \max_{1 \leq i \leq N} E[d(X^i_m(1), X^i_m(2))] \sum_{i=1}^{N} \delta_i(f)
\]

\[
\leq \left(1 - \frac{1 - r_{\infty}}{N}\right)^m \max_{1 \leq i \leq N} E[d(X^i_0(1), X^i_0(2))] \sum_{i=1}^{N} \delta_i(f).
\]

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

\[
W_{1,d^N_1}(\nu P^m, \mu) = W_{1,d^N_1}(\nu P^m, \mu P^m) \leq E[d^N_1(X_m(1), X_m(2))] = \sum_{i=1}^{N} E[d(X^i_m(1), X^i_m(2))]
\]

\[
\leq N \max_{1 \leq i \leq N} E[d(X^i_m(1), X^i_m(2))]
\]

\[
\leq N \left(1 - \frac{1 - r_{\infty}}{N}\right)^m \max_{1 \leq i \leq N} E[d(X^i_0(1), X^i_0(2))].
\]

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

\[
E[d^N_1(X_m, Y_m)] \leq \left(1 - \frac{1 - r_{\infty}}{N}\right)^m E[d^N_1(X_0, Y_0)].
\]

ECP 22 (2017), paper 56. http://www.imstat.org/ecp/
Proof.

\[
\mathbb{E}[d_{lN}^N(X_m, Y_m)|X_{m-1}, Y_{m-1}) = (x, y)]
\]

\[
= \sum_{i=1}^{N} \mathbb{E}[d(X_m^i, Y_m^i)|(X_{m-1}, Y_{m-1}) = (x, y)]
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} \mathbb{E}[d(X_m^i, Y_m^i)|(X_{m-1}, Y_{m-1}) = (x, y), \sigma = k]
\]

\[
= \frac{1}{N} \sum_{k=1}^{N} \left\{ W_1.d(\mu_k(\cdot|x), \mu_k(\cdot|y)) + \sum_{i \neq k} d(x_i, y_i) \right\}
\]

\[
\leq \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{N} c_k d(x_i, y_i) + \frac{1}{N} \sum_{k=1}^{N} \sum_{i \neq k} d(x_i, y_i)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} d(x_i, y_i) \sum_{k=1}^{N} c_k + \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{N} d(x_i, y_i) - \frac{1}{N} \sum_{k=1}^{N} d(x_k, y_k)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} d(x_i, y_i) \left( \frac{1}{N} \sum_{k=1}^{N} c_k + \frac{N-1}{N} \right)
\]

\[
\leq \frac{r_1 + N - 1}{N} d_{lN}^N(x, y) \leq \left(1 - \frac{1 - r_1}{N}\right) d_{lN}^N(x, y),
\]

i.e.,

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

and thus,

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

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

**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,

\[
|\nu_1 P^m f - \nu_2 P^m f| = |\mathbb{E}f(X_m(1)) - \mathbb{E}f(X_m(2))|
\]

\[
\leq \|f\|_{\text{Lip}(d_{lN}^N)} \mathbb{E}d_{lN}^N(X_m(1), X_m(2))
\]

\[
\leq \left(1 - \frac{1 - r_1}{N}\right)^m \|f\|_{\text{Lip}(d_{lN}^N)} \mathbb{E}d_{lN}^N(X_0(1), X_0(2)).
\]

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

\[
W_{1,d_{lN}^N}(\nu P^m, \mu) = W_{1,d_{lN}^N}(\nu P^m, \mu P^m) \leq \mathbb{E}d_{lN}^N(X_m(1), X_m(2))
\]

\[
\leq \left(1 - \frac{1 - r_1}{N}\right)^m \mathbb{E}d_{lN}^N(X_0(1), X_0(2)). \qed
\]
Convergence rates of the random scan Gibbs sampler

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.
Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS$^1$)
- Easy interface (EJMS$^2$)

Economical model of EJP-ECP

- Non profit, sponsored by IMS$^3$, BS$^4$, ProjectEuclid$^5$
- Purely electronic

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

---

$^1$LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
$^2$EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
$^3$IMS: Institute of Mathematical Statistics http://www.imstat.org/
$^4$BS: Bernoulli Society http://www.bernoulli-society.org/
$^5$Project Euclid: https://projecteuclid.org/
$^6$IMS Open Access Fund: http://www.imstat.org/publications/open.htm