On the Geometric Ergodicity of Gibbs Algorithm for Lattice Gaussian Sampling

Zheng Wang
College of Electronic and Information Engineering
Nanjing University of Aeronautics and Astronautics
Nanjing, 210000, China
Email: z.wang@ieee.org

Cong Ling
Department of EEE
Imperial College London
London, SW7 2AZ, United Kingdom
Email: cling@ieee.org

***Abstract*—Sampling from the lattice Gaussian distribution is emerging as an important problem in coding and cryptography. In this paper, the conventional Gibbs sampling algorithm is demonstrated to be geometrically ergodic in tackling with lattice Gaussian sampling, which means its induced Markov chain converges exponentially fast to the stationary distribution. Moreover, as the exponential convergence rate is dominated by the spectral radius of the forward operator of the Markov chain, a comprehensive analysis is given and we show that the convergence performance can be further enhanced by usages of blocked sampling strategy and choices of selection probabilities.**

Keywords: Lattice Gaussian sampling, Markov chain Monte Carlo, lattice coding and decoding.

I. Introduction

Nowadays, lattice Gaussian distribution has drawn a lot of attentions in various research fields. In mathematics, Banaszczyk firstly applied it to prove the transference theorems for lattices [1]. In coding, lattice Gaussian distribution was employed to obtain the full shaping gain for lattice coding [2], and to achieve the capacity of the Gaussian channel and the secrecy capacity of the Gaussian wiretap channel, respectively [3]. Meanwhile, lattice Gaussian distribution is also applied to relay network under the compute-and-forward strategy for the physical layer security [4]. In cryptography, the lattice Gaussian distribution has already become a central tool in the construction of many primitives. Specifically, Micciancio and Regev used it to propose lattice-based cryptosystems based on the worst-case hardness assumptions [5]. Meanwhile, it also has underpinned the fully-homomorphic encryption for cloud computing [6]. Algorithmically, lattice Gaussian sampling with a suitable variance allows to solve the shortest vector problem (SVP) and the closest vector problem (CVP) [7]; for example, it has led to efficient lattice decoding for multi-input multi-output (MIMO) systems [8].

Because of the central role of lattice Gaussian distribution playing in these fields, its sampling algorithms become an important computational problem. However, different from the case of continuous Gaussian density, sampling from the lattice Gaussian distribution is not straightforward at all even for a low-dimensional system. One feasible way is proposed in [7] but besides the exponential time complexity, it also requires exponential space complexity. Another sampling algorithm for lattices is due to Klein, originally proposed for bounded-distance decoding (BDD) [9]. However, Klein's algorithm is only valid when the standard deviation of lattice Gaussian distribution is sufficiently large (i.e., $\sigma \geq \omega(\sqrt{\log n}) \cdot \max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$), rendering Klein's algorithm inapplicable to many cases of interest.

To this end, Markov chain Monte Carlo (MCMC) methods are introduced as an alternative way for lattice Gaussian sampling, which attempts to sample from the target distribution by building a Markov chain [10], [11]. After a burn-in time, the Markov chain will step into a stationary distribution, where samples from the target distribution can be obtained thereafter. In [10], Gibbs algorithm was introduced into lattice Gaussian sampling by showing its ergodicity, which employs univariate conditional sampling to build the Markov chain. Although an artificially designed symmetric Metropolis-within-Gibbs sampling algorithm has proved to be geometrical ergodicity [12], the convergence behaviour of classic Gibbs sampler is still unknown.

In this paper, a comprehensive analysis is conducted to prove the geometric ergocidity of Gibbs sampler for lattice Gaussian distribution, which means the underlying Markov chain converges exponentially fast to the lattice Gaussian distribution. Furthermore, by showing the derived spectral radius of the forward operator (or the lag-1 maximal correlation between two consecutive states) exactly characterizes the convergence rate of the Markov chain, we give a convergence analysis aiming to further enhance the convergence performance, where blocked sampling and reasonable choice of selection probabilities are shown to be beneficial to the rapid Markov mixing.

II. Gibbs Sampler for Lattice Gaussian

Let $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n] \subset \mathbb{R}^n$ consist of n linearly independent vectors. The n-dimensional lattice Λ based on $\mathbf{B}$ is defined by

$$\Lambda = \mathcal{L}(\mathbf{B}) = \{\mathbf{B}\mathbf{x} : \mathbf{x} \in \mathbb{Z}^n\}, \tag{1}$$

where $\mathbf{B}$ is known as the lattice basis. We define the Gaussian function centered at $\mathbf{c} \in \mathbb{R}^n$ for standard deviation $\sigma > 0$ as

$$\rho_{\sigma,\mathbf{c}}(\mathbf{z}) = e^{-\frac{\|\mathbf{z}-\mathbf{c}\|^2}{2\sigma^2}}, \tag{2}$$

978-1-5090-3097-2/17/$31.00 ©2017 IEEE

for all $\mathbf{z} \in \mathbb{R}^n$. Then, the discrete Gaussian distribution over Λ is defined as

$$D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x}) = \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{Bx})}{\rho_{\sigma,\mathbf{c}}(\Lambda)} = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}} \tag{3}$$

for all $\mathbf{x} \in \mathbb{Z}^n$, where $\rho_{\sigma,\mathbf{c}}(\Lambda) \triangleq \sum_{\mathbf{Bx}\in\Lambda} \rho_{\sigma,\mathbf{c}}(\mathbf{Bx})$ is just a scaling to make a probability distribution. We claim that such a definition differs slightly from the one in [5], where σ is scaled by a constant factor $\sqrt{2\pi}$ (i.e., $s = \sqrt{2\pi}\sigma$). It has been demonstrated in [13] Klein's algorithm is capable to sample from $D_{\Lambda,\sigma,\mathbf{c}}$ within a negligible statistical distance if

$$\sigma \geq \omega(\sqrt{\log n}) \cdot \max\nolimits_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|. \tag{4}$$

However, Gaussian sampling algorithms are lacking for the range $\sigma < \omega(\sqrt{\log n}) \cdot \max_i\|\widehat{\mathbf{b}}_i\|$.

From the perceptive of MCMC, lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$ with $\sigma < \omega(\sqrt{\log n}) \cdot \max_i\|\widehat{\mathbf{b}}_i\|$ can be seen as a complex target distribution lacking direct sampling methods. Therefore, Gibbs sampler that makes use of the 1-dimensional conditional distribution becomes a tractable alternative to work with [14]. More specifically, each coordinate of $\mathbf{x}$ is sampled from the following 1-dimensional conditional distribution

$$P_i(x_i|\mathbf{x}_{[-i]}) = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\sum_{x_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}, \tag{5}$$

where $1 \leq i \leq n$ denotes the coordinate index of $\mathbf{x}$, $\mathbf{x}_{[-i]} \triangleq [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]^T$. During the univariate sampling, the other $n-1$ variables contained in $\mathbf{x}_{[-i]}$ are leaving unchanged. Note that since x_i is a variable, it will not be completely determined by $\mathbf{x}_{[-i]}$ while all the candidates of x_i in the sampling space are possible to be sampled with certain probabilities, that is,

$$\text{var}(x_i|\mathbf{x}_{[-i]}) > 0. \tag{6}$$

If time permits to reach the stationary distribution, the proposed Gibbs sampler will draw samples from $D_{\Lambda,\sigma,\mathbf{c}}$ no matter what value σ takes, which means the obstacle encountered by Klein's algorithm is overcome.

Theoretically, there are various scan schemes to proceed the component updating in Gibbs sampler. Among them, random scan is the basic one and will be considered throughput the context. Typically, in random scan Gibbs sampler, the coordinate index i is randomly chosen based on the selection probabilities $[\alpha_1, \ldots, \alpha_n]$, where $\sum_{i=1}^n \alpha_i = 1$ and $\alpha_i > 0$.

In particular, the transition probability of the Markov chain in Gibbs sampler is a weighted sum of the full conditional probabilities as

$$P(\mathbf{X}^t, \mathbf{X}^{t+1}) = \sum_{i=1}^{n} \alpha_i P_i(x_i|\mathbf{x}_{[-i]}), \tag{7}$$

where t is the time index of the Markov chain. To summarize, Algorithm 1 illustrates the operation of Gibbs sampler for lattice Gaussian distribution. The initial Markov state $\mathbf{x}^0$ can be chosen from the state space $\Omega = \mathbb{Z}^n$ arbitrarily or from the output of a suboptimal algorithm, while $t_{\text{mix}}(\epsilon)$ denotes the mixing time of the Markov chain.

Algorithm 1 Gibbs sampler for lattice Gaussian distribution

```
Input: $\mathbf{B}, \sigma, \mathbf{c}, \mathbf{x}^0, t_{\text{mix}}(\epsilon)$
Output: $\mathbf{x} \sim D_{\Lambda,\sigma,\mathbf{c}}$
for $t$ =1,2, ... do
    randomly choose the index $i$ based on $[\alpha_1, \ldots, \alpha_n]$
    sample $x_i$ from $P_i(x_i|\mathbf{x}_{[-i]})$ shown in (5)
    update $\mathbf{x}$ with the sampled $x_i$ and let $\mathbf{X}_t = \mathbf{x}$
    if $t \geq t_{\text{mix}}(\epsilon)$ then
        output the state of $\mathbf{X}_t$
    end if
end for
```

Theorem 1 ([10]). *Given the invariant distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by the Gibbs sampler converges to the stationary distribution in the total variation (TV) distance as $t \rightarrow \infty$:*

$$\lim_{t\rightarrow\infty} \|P^t(\mathbf{x};\cdot) - D_{\Lambda,\sigma,\mathbf{c}}\|_{TV} = 0. \tag{8}$$

III. Geometric Ergodicity of Gibbs Sampler

Although *ergodicity* implies asymptotic convergence to stationarity, it does not entail the way of the convergence, resulting in an intractable Markov chain [15]. Among the kinds of ergodicity in literature [16], *geometric ergodicity* which converges exponentially is defined as:

Definition 1. *A Markov chain with stationary distribution $\pi(\cdot)$ is geometrically ergodic if there exists $0 < \rho < 1$ and $M(\mathbf{x}) < \infty$ such that for all $\mathbf{x} \in \Omega$*

$$\|P^t(\mathbf{x},\cdot) - \pi(\cdot)\|_{TV} \leq M(\mathbf{x})\rho^t, \tag{9}$$

where Ω represents the state space of the Markov chain and $M(\mathbf{x})$ is parameterized by the initial state $\mathbf{x}$.

Here, the notion of spectral gap γ of the Markov chain is induced. Then, according to the following Theorem from [17], the geometric ergodicity can be simply verified by $\gamma > 0$ for a reversible, irreducible and aperiodic Markov chain.

Theorem 2 ([17]). *A reversible, irreducible and aperiodic chain with the spectral gap $\gamma = 1 - spec(\mathbf{F}) > 0$ in L_0^2 converges exponentially to the stationary distribution π*

$$\|P^t(\mathbf{x},\cdot) - \pi(\cdot)\|_{TV} \leq M(\mathbf{x})(1-\gamma)^t, \tag{10}$$

where $\text{spec}(\cdot)$ *denotes the spectral radius and* $\mathbf{F}$ *represents the forward operator of the Markov chain.*

Particularly, given the transition probability $P(\mathbf{X}^t, \mathbf{X}^{t+1})$, the forward operator $\mathbf{F}$ of the Markov chain is defined as [14]

$$\mathbf{F}h(\mathbf{x}) \triangleq \sum_{\mathbf{y}\in\Omega} h(\mathbf{y})P(\mathbf{x},\mathbf{y}) = E[h(\mathbf{y})|\mathbf{x}] \tag{11}$$

with induced operator norm

$$\|\mathbf{F}\| = \sup_{h\in L_0^2(\pi),\mathrm{var}(h)=1} \|\mathbf{F}h\| = \sup_{h\in L_0^2(\pi),\mathrm{var}(h)=1} \langle h(\mathbf{X}^t), h(\mathbf{X}^{t+1})\rangle. \quad (12)$$

Here, $\mathbf{x}$ and $\mathbf{y}$ denote the Markov states of $\mathbf{X}^t$ and $\mathbf{X}^{t+1}$ respectively. $L^2(\pi)$ is the Hilbert space of square integrable functions with respect to π so that $L_0^2(\pi) \triangleq \{h(\mathbf{x}) : E[h(\mathbf{x})] = 0, \mathrm{var}[h(\mathbf{x})] < \infty\}$ denotes the subspace of $L^2(\pi)$ consisting of functions with zero mean relative to π. More precisely, for $h(\cdot), g(\cdot) \in L_0^2(\pi)$, the inner product defined by the space is

$$\langle h(\mathbf{x}), g(\mathbf{x})\rangle = E[h(\mathbf{x})g(\mathbf{x})] \quad (13)$$

with variance

$$\mathrm{var}_\pi[h(\mathbf{x})] = \langle h(\mathbf{x}), h(\mathbf{x})\rangle = \|h(\mathbf{x})\|^2. \quad (14)$$

Clearly, from Theorem 2, the convergence rate of the Markov chain is exactly characterized by the spectral radius of $\mathbf{F}$, i.e., $\rho = \mathrm{spec}(\mathbf{F})$. Based on it, we then arrive at the following Corollary to demonstrate the geometric ergodicity of Gibbs sampler for lattice Gaussian distribution.

Corollary 1. *Given the invariant lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by Gibbs sampler is geometrically ergodic as*

$$\|P^t(\mathbf{x},\cdot) - D_{\Lambda,\sigma,\mathbf{c}}\|_{TV} \le M(\mathbf{x})\mathrm{spec}^t(\mathbf{F}). \quad (15)$$

Proof: To start with, it is easy to verify that the Markov chain induced by Gibbs sampler is irreducible, aperiodic and reversible. Then, according to Theorem 2, in order to show the geometric ergodicity, we only need to prove $\gamma > 0$ by $0 < \mathrm{spec}(\mathbf{F}) < 1$.

Typically, $\mathrm{spec}(\mathbf{F})$ is closely related with the norm of $\mathbf{F}$ as [17], [18]

$$\mathrm{spec}(\mathbf{F}) = \lim_{t\to\infty} \|\mathbf{F}^t\|^{1/t}. \quad (16)$$

In theory, reversibility corresponds to a self-adjoint operator $\mathbf{F}$ with [19]

$$\|\mathbf{F}^t\| = \|\mathbf{F}\|^t, \quad (17)$$

then we have

$$\mathrm{spec}(\mathbf{F}) = \|\mathbf{F}\|. \quad (18)$$

Subsequently, according to (12) and (13), the spectral radius of $\mathbf{F}$ can be further expressed as

$$\begin{aligned}\mathrm{spec}(\mathbf{F}) &= \sup_{h\in L_0^2(\pi),\mathrm{var}(h)=1} \langle h(\mathbf{X}^t), h(\mathbf{X}^{t+1})\rangle. \\ &= \sup_{h\in L_0^2(\pi),\mathrm{var}(h)=1} E[h(\mathbf{X}^t)h(\mathbf{X}^{t+1})]. \quad (19)\end{aligned}$$

Because only one component of $\mathbf{x}$ (i.e., x_i) over $\mathbf{X}^t$ and $\mathbf{X}^{t+1}$ is different and because the coordinate index i is determined randomly, $\mathbf{X}^t_{[-\mathbf{i}]}$ can be viewed as a joint function made up by $\mathbf{X}^t$ and $\mathbf{i}$, i.e., $\mathbf{X}_{[-\mathbf{i}]} = f(\mathbf{X}, \mathbf{i})$, where the random variable $\mathbf{i}$ serves as the update index i at each move. Hence, given $\mathbf{X}^t_{[-i]}$, $\mathbf{X}^t$ and $\mathbf{X}^{t+1}$ are conditionally independent on sampling x_i, which is also referred to as interleaving Markov property in data augmentation literatures [14]. Then, we have

$$\begin{aligned}E[h(\mathbf{X}^t)h(\mathbf{X}^{t+1})] &= E[E[h(\mathbf{X}^t)h(\mathbf{X}^{t+1})|\mathbf{X}^t_{[-\mathbf{i}]}]] \\ &= E\left[\sum_{i=1}^n \alpha_i E[h(\mathbf{X}^t)h(\mathbf{X}^{t+1})|\mathbf{X}^t_{[-i]}]\right] \\ &= E\left[\sum_{i=1}^n \alpha_i E[h(\mathbf{X}^t)|\mathbf{X}^t_{[-i]}]E[h(\mathbf{X}^{t+1})|\mathbf{X}^t_{[-i]}]\right] \\ &\overset{(a)}{=} E\left[\sum_{i=1}^n \alpha_i E^2[h(\mathbf{x})|\mathbf{x}_{[-i]}]\right] \\ &\overset{(b)}{=} \sum_{i=1}^n \alpha_i E[h^2(\mathbf{x})] - \sum_{i=1}^n \alpha_i E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]] \\ &\overset{(c)}{=} \mathrm{var}[h(\mathbf{x})] - \sum_{i=1}^n \alpha_i E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]. \quad (20)\end{aligned}$$

Here, equality (a) holds due to the fact that given $\mathbf{X}^t_{[-i]}$, $\mathbf{X}^t$ and $\mathbf{X}^{t+1}$ are independent and identically distributed with respect to x_i, (b) and (c) respectively come from the properties of random variable in statistics shown below

$$E^2[\mathbf{A}] = E[\mathbf{A}^2] - \mathrm{var}[\mathbf{A}] \quad (21)$$

and

$$E[E[\mathbf{A}|\mathbf{B}]] = E[\mathbf{A}]. \quad (22)$$

Then, by substitution, it follows that

$$\begin{aligned}\mathrm{spec}(\mathbf{F}) &= \sup_{h\in L_0^2(\pi),\mathrm{var}(h)=1}\left\{\mathrm{var}[h(\mathbf{x})] - \sum_{i=1}^n \alpha_i E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]\right\} \\ &= 1 - \inf_{h\in L_0^2(\pi),\mathrm{var}(h)=1}\left\{\sum_{i=1}^n \alpha_i E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]\right\} \\ &= 1 - \inf_{h\in L_0^2(\pi),\mathrm{var}(h)=1}\left\{\sum_{i=1}^n \alpha_i \sum_{\mathbf{x}_{[-i]}} \mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]P(\mathbf{x}_{[-i]})\right\} \quad (23)\end{aligned}$$

Clearly, as $h(\cdot)$ in the infimum is required to satisfy $\mathrm{var}(h) = 1$, the value of $\inf_{h\in L_0^2(\pi),\mathrm{var}(h)=1}\{\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]\}$ is actually determined by the Pearson correlation coefficient between x_i and $\mathbf{x}_{[-i]}$, i.e., $\mathrm{corr}(x_i, \mathbf{x}_{[-i]})$. Typically, it equals to 1 when x_i is independent of $\mathbf{x}_{[-i]}$, i.e., $|\mathrm{corr}(x_i, \mathbf{x}_{[-i]})| = 0$. On the other hand, it turns out to be 0 if and only if x_i completely depends on $\mathbf{x}_{[-i]}$, i.e., $|\mathrm{corr}(x_i, \mathbf{x}_{[-i]})| = 1$, which obviously contradicts the sampling setup shown in (6) since x_i is a random variable and any candidate of x_i is possible to be sampled. Therefore, it follows that

$$\inf_{h\in L_0^2(\pi),\mathrm{var}(h)=1}\{\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]\} > 0 \quad (24)$$

and so as to the summation

$$\inf_{h\in L_0^2(\pi),\mathrm{var}(h)=1}\left\{\sum_{i=1}^n \alpha_i \sum_{\mathbf{x}_{[-i]}} \mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]P(\mathbf{x}_{[-i]})\right\} > 0. \quad (25)$$

Meanwhile, because of $\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}] < \mathrm{var}[h(\mathbf{x})]$, we can

978-1-5090-3097-2/17/$31.00 ©2017 IEEE

immediately arrive at

$$\inf_{h\in L_0^2(\pi),\mathrm{var}(h)=1}\left\{\sum_{i=1}^{n}\alpha_i E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]\right\} \quad < \quad 1 \tag{26}$$

and

$$0<\mathrm{spec}(\mathbf{F})<1, \tag{27}$$

thus completing the proof by invoking Theorem 2 with $0<\gamma=1-\mathrm{spec}(\mathbf{F})<1$. ■

To summarize, the Markov chain converges exponentially fast to the lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, where the exponential convergence rate ρ is characterized by $\mathrm{spec}(\mathbf{F})$. In fact, this can be interpreted from maximal correlation point of view since $\mathrm{spec}(\mathbf{F})$ shown in (19) essentially represents the lag-1 maximal correlation between two consecutive Markov states $\mathbf{X}^t$ and $\mathbf{X}^{t+1}$ [20], which means Markov states further apart in the chain turns out to be gradually uncorrelated in an exponential way. Additionally, note that in geometric ergodicity, the initial state $\mathbf{x}\in\mathbb{Z}^n$ can be carefully selected for a smaller coefficient $M(\mathbf{x})<\infty$.

IV. Convergence Analysis

It is not surprising that the lag-1 maximal correlation govern the rapid mixing since the lower correlation, the greater the amount of information contained in a given number of draws from the posterior, making it a useful diagnostics on the convergence rate of the MCMC sampler. However, from (23), it is difficult to calculate $\mathrm{spec}(\mathbf{F})$ explicitly, except in some rare cases. Nevertheless, comprehensive convergence analysis still can be carried out, which targets at a smaller convergence rate.

A. Blocked Sampling

Specifically, although Gibbs sampler will converge to the stationary distribution eventually, the way it functions by individually sampling only one component of $\mathbf{x}$ at each time tends to yield a slow convergence. Especially when components of $\mathbf{x}$ are highly correlated with each other, the Markov chain will most likely be trapped by some local minima for a long time. With the increase of the dimension, such a flaw will get worse, which severely diminishes the convergence.

From the regression point of view one should avoid introducing unnecessary components into the sampler, however removing one component by integrating it out of $\mathbf{x}$ turns out to be difficult in most cases of interest, rendering blocked sampling as a feasible alternative to improve the convergence. Because of this, sampling over multiple components of $\mathbf{x}$ is worthwhile to be considered. In what follows we show that such a mechanism motivated by blocked sampling is able to achieve a faster convergence than the standard Gibbs sampler by a smaller $\mathrm{spec}(\mathbf{F})$.

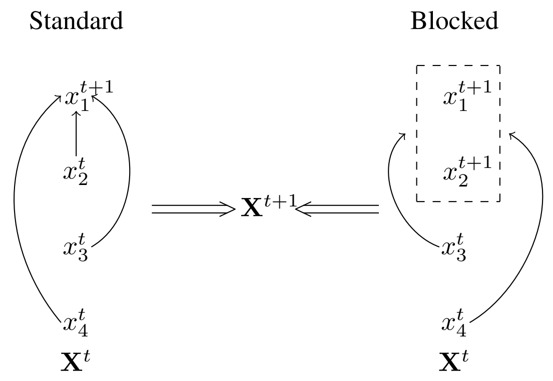

Fig. 1. Illustration of standard Gibbs sampler and blocked Gibbs sampler. Components within the dashed block are sampled as a whole by blocked Gibbs sampler.

Proposition 1. *Given the selection probabilities α_i's, the blocked version of Gibbs sampler achieves a faster convergence rate by a smaller convergence rate $\rho=spec(\mathbf{F})$, i.e.,*

$$\mathrm{spec}(\mathbf{F})^{\mathrm{block}}\leq\mathrm{spec}(\mathbf{F}). \tag{28}$$

Proof: First of all, according to the *law of total variance* shown below

$$\mathrm{var}(\mathbf{A})=E[\mathrm{var}(\mathbf{A}|\mathbf{B})]+\mathrm{var}[E(\mathbf{A}|\mathbf{B})], \tag{29}$$

the term shown in (20) can be derived as

$$E[h(\mathbf{X}^t)h(\mathbf{X}^{t+1})]=\sum_{i=1}^{n}\alpha_i\mathrm{var}[E[h(\mathbf{x})|\mathbf{x}_{[-i]}]] \tag{30}$$

and subsequently, we have

$$\mathrm{spec}(\mathbf{F})=\sup_{h\in L_0^2(\pi),\mathrm{var}(h)=1}\sum_{i=1}^{n}\alpha_i\mathrm{var}[E[h(\mathbf{x})|\mathbf{x}_{[-i]}]]. \tag{31}$$

For ease of presentation, a two-component blocked sampling scenario depicted in Fig. 1 is firstly considered. Typically, suppose x_i and x_j of $\mathbf{x}$ can be sampled together, then considering the fact that

$$E[h(\mathbf{x})|\mathbf{x}_{[-i,-j]}]=E[E[h(\mathbf{x})|\mathbf{x}_{[-i]}]|\mathbf{x}_{[-j]}], \tag{32}$$

we can immediately get

$$\mathrm{var}(E[h(\mathbf{x})|\mathbf{x}_{[-i,-j]}])\leq\mathrm{var}(E[h(\mathbf{x})|\mathbf{x}_{[-i]}]]) \tag{33}$$

and

$$\mathrm{var}(E[h(\mathbf{x})|\mathbf{x}_{[-i,-j]}])\leq\mathrm{var}(E[h(\mathbf{x})|\mathbf{x}_{[-j]}]]). \tag{34}$$

Obviously, this two-component blocked sampling can be easily extended to any larger size blocked sampling. To conclude, it follows that

$$\mathrm{var}(E[h(\mathbf{x})|\mathbf{x}_{[-\mathrm{block}]}])\leq\mathrm{var}(E[h(\mathbf{x})|\mathbf{x}_{[-i]}]]), \tag{35}$$

where the block contains the component x_i.

Consequently, according to (31) and (35), we can arrive at

$$\mathrm{spec}(\mathbf{F})^{\mathrm{block}}\leq\mathrm{spec}(\mathbf{F}), \tag{36}$$

completing the proof. ■

Compared to conventional univariate sampling, by sampling multiple components together, the slow, componentwise moves will be replaced by the fast moves incorporating the information about dependence between components. Moreover, it is straightforward to see that the convergence performance also improves gradually with grouping more elements into the block

$$\mathrm{var}(E[h(\mathbf{x})|\mathbf{x}_{[-\mathrm{block},-j]}]) \leq \mathrm{var}(E[h(\mathbf{x})|\mathbf{x}_{[-\mathrm{block}]}]), \quad (37)$$

where the insight behind this is intuitive to understand since larger blocks allow moves in more general directions. To be more specific, if all the components forming a single block could be sampled directly, there would be no any need for MCMC sampling. In this regard, blocked technique is strongly recommended if sampling over multi-component can be efficiently carried out. For more details on low complexity implementation, readers are referred to [10] for Gibbs-Klein sampling algorithm.

B. Selection Probabilities α_i's

Since every two adjacent states differ from each other by only one coordinate of $\mathbf{x}$, Gibbs sampler is naturally characterized by the selection probabilities α_i's, which determine the percentage of visits to a specific element of $\mathbf{x}$. As shown in (7), there is a great flexibility in choice of α_i's while extensions to other scan schemes can be easily constructed based on it.

As can be seen clearly from (31), spec$(\mathbf{F})$ is a function of α_i's and optimal allocation of visiting percentage can be therefore designed to minimize it. In particular, according to (23), the following relationship can be revealed

$$\mathrm{spec}(\mathbf{F}) \propto \frac{1}{\sum_{i=1}^{n}\alpha_i E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]}. \quad (38)$$

Intuitively, in order to reduce spec$(\mathbf{F})$, a small α_i is preferred when $E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]$ turns out to be small, which can be interpreted to make fewer visits to a component that is less variable. Therefore, compared to equal selection probabilities, i.e., $\alpha_i = 1/n$, considerable convergence potential can be exploited by a sophisticated design of α_i's. For this reason, a heuristic allocation method of α_i's is introduced to enhance the convergence.

In particular, since $h(\cdot)$ belonging to the Hilbert space $L_0^2(\pi)$ is hard to track, the variance of x_i during the univariate conditional sampling $P_i(x_i|\mathbf{x}_{[-i]})$ is applied as an alternative metric to approximately evaluate $E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]$ in a sense, that is,

$$\mathrm{var}[P_i(x_i|\mathbf{x}_{[-i]})] \propto E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]. \quad (39)$$

Then, based on such a relationship, α_i's can be optimized according to

$$\alpha_i = \frac{\mathrm{var}[P_i(x_i|\mathbf{x}_{[-i]})]}{\sum_{i=1}^{n}\mathrm{var}[P_i(x_i|\mathbf{x}_{[-i]})]}, \quad (40)$$

where a large $E[\mathrm{var}[h(\mathbf{x})|\mathbf{x}_{[-i]}]]$ corresponds to a relatively large α_i. By doing this, the component of $\mathbf{x}$ with large sampling variance is desired to be sampled more frequently for a better convergence performance.

V. Conclusion

In this paper, we demonstrated that the classic Gibbs sampling from MCMC methods converges exponentially fast to the lattice Gaussian distribution. An explicit expression of the convergence rate was presented, which is characterized by the spectral radius of the forward operator of the Markov chain. In order to improve the convergence performance, effective operations by blocked sampling technique and reasonable allocation of selection probabilities were presented, which are proven to be beneficial for the Markov mixing.

References

[1] W. Banaszczyk, "New bounds in some transference theorems in the geometry of numbers," *Math. Ann.*, vol. 296, pp. 625–635, 1993.

[2] G. Forney and L.-F. Wei, "Multidimensional constellations–Part II: Voronoi constellations," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 941–958, Aug. 1989.

[3] C. Ling and J.-C. Belfiore, "Achieiving the AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inform. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.

[4] S. Vatedka, N. Kashyap, and A. Thangaraj, "Secure compute-and-forward in a bidirectional relay," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2531–2556, May 2015.

[5] D. Micciancio and O. Regev, "Worst-case to average-case reductions based on Gaussian measures," in *Proc. Ann. Symp. Found. Computer Science*, Rome, Italy, Oct. 2004, pp. 372–381.

[6] C. Gentry, A. Sahai, and B. Waters, "Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based," in *CRYPTO*, Springer, Heidelberg, pp. 75-92, 2013.

[7] D. Aggarwal, D. Dadush, O. Regev, and N. Stephens-Davidowitz, "Solving the shortest vector problem in 2^n time via discrete Gaussian sampling," *STOC*, 2015.

[8] Z. Wang, S. Liu, and C. Ling, "Decoding by sampling - Part II: Derandomization and soft-output decoding," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4630–4639, Nov. 2013.

[9] P. Klein, "Finding the closest lattice vector when it is unusually close," in *ACM-SIAM Symp. Discr. Algorithms*, 2000, pp. 937–941.

[10] Z. Wang, C. Ling, and G. Hanrot, "Markov chain Monte Carlo algorithms for lattice Gaussian sampling," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Honolulu, USA, Jun. 2014, pp. 1489–1493.

[11] Z. Wang and C. Ling, "On the geometric ergodicity of Metropolis-Hastings algorithms for lattice Gaussian sampling," *accepted by IEEE Transactions on Information Theory, to appear soon.*, [Online] Available:http://arxiv.org/pdf/1501.05757v2.pdf.

[12] ——, "Symmetric Metropolis-within-Gibbs algorithm for lattice Gaussian sampling," in *Proc. IEEE Information Theory Workshop (ITW)*, Cambridge, United Kingdom, Sept. 2016, pp. 394–398.

[13] C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapdoors for hard lattices and new cryptographic constructions," in *Proc. 40th Ann. ACM Symp. Theory of Comput.*, Victoria, Canada, 2008, pp. 197–206.

[14] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, New York: Springer-Verlag, 2001.

[15] L. Tierney, "Markov chains for exploring posterior distributions (with discussion)," in *Proc Ann. Stat.*, vol. 22, 1994, pp. 1701–1762.

[16] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. UK, Cambridge University Press, 2009.

[17] I. Kontoyannis and S. P. Meyn, "Geometric ergodicity and spectral gap of non-reversible real valued Markov chains," in *Proc. Probab. Theory and related Fields*, vol. 154, 2012, pp. 327–339.

[18] J. A. Fill, "Eigenvalue bounds on convergence to stationary for nonreversible Markov chains, with application to the exclusion process." in *Proc. Annals of Applied Probability*, vol. 1, 1991, pp. 62–87.

[19] J. S. Liu, W. H. Wong, and A. Kong, "Covariance structure and convergence rate of the Gibbs sampler with various scans," *J. Roy. Statist. Soc. Series B*, **57**(1): 157-169, 1995.

[20] J. S. Liu, "Fraction of missing information and convergence rate of data augmentation," in *Computationally Intensive Statistical Methods: Proceedings of the 26th symposium on the Interface*, North Carolina, 1994, pp. 490–497.