Symmetric Metropolis-within-Gibbs Algorithm for Lattice Gaussian Sampling

Zheng Wang and Cong Ling
Department of EEE
Imperial College London
London, SW7 2AZ, United Kingdom
Email: z.wang10, c.ling@imperial.ac.uk

***Abstract*—As a key sampling scheme in Markov chain Monte Carlo (MCMC) methods, Gibbs sampling is widely used in various research fields due to its elegant univariate conditional sampling, especially in tacking with multidimensional sampling systems. In this paper, a Gibbs-based sampler named as symmetric Metropolis-within-Gibbs (SMWG) algorithm is proposed for lattice Gaussian sampling. By adopting a symmetric Metropolis-Hastings (MH) step into the Gibbs update, we show the Markov chain arising from it is geometrically ergodic, which converges exponentially fast to the stationary distribution. Moreover, by optimizing its symmetric proposal distribution, the convergence efficiency can be further enhanced.**

Keywords: Lattice Gaussian sampling, MCMC methods, Gibbs sampling, Metropolis-Hastings algorithm.

I. Introduction

Sampling from the lattice Gaussian distribution is an important problem in coding and cryptography. In [1], lattice Gaussian distribution was employed to achieve the full shaping gain for lattice coding. On the other hand, both the capacity of the Gaussian channel and the secrecy capacity of the Gaussian wiretap channel can be obtained through it [2]. As for the field of cryptography, lattice Gaussian distribution has already become a central tool in the construction of many primitives such as lattice-based cryptosystems and fully-homomorphic encryption for cloud computing [3], [4]. Due to the central role of the lattice Gaussian distribution playing in these research areas, how to perform the sampling over it becomes the key. In fact, lattice Gaussian sampling itself essentially corresponds to the closest vector problem (CVP) via a polynomial-time dimension-preserving reduction, which allows to solve the lattice decoding problems with a suitable variance [5], [6].

However, in contrast to sampling from a continuous Gaussian distribution, it is by no means trivial to sample even from a low-dimensional discrete Gaussian distribution. Because of it, Markov chain Monte Carlo (MCMC) methods were introduced, which attempts to obtain the desired samples via Markov chains. In particular, the validity of Gibbs sampling in terms of convergence to the target lattice Gaussian distribution has been demonstrated in [7], and a flexible block Gibbs sampling was proposed to further improve the convergence, where the ergodicity of the chain is still preserved. However, ergodicity only implies an asymptotic convergence. Since the Markov chain associated with lattice Gaussian sampling corresponds to a countably infinite state space, how to specify the convergence of Gibbs sampling naturally becomes an open question of interest.

Besides Gibbs sampling, the traditional Metropolis-Hastings (MH) sampling from MCMC can also be used for lattice Gaussian sampling and some progress have been made. Specifically, the independent Metropolis-Hastings-Klein (IMHK) algorithm was proposed in [8], which is uniformly ergodic with an accessible convergence rate. Furthermore, by taking advantage of a symmetrical proposal distribution, a symmetric Metropolis-Klein (SMK) algorithm was given, which not only converges exponentially fast, but also takes the selection of the initial state into account. By definition from MCMC, such a convergence scheme is referred to as geometric ergodicity.

Inspired by the SMK algorithm, in this paper, a Gibbs-based sampler referred to as symmetric Metropolis-within-Gibbs algorithm is proposed. It not only makes use of the univariate sampling from Gibbs sampling, but also retains the acceptance-rejection rule from the MH sampling. We firstly show the underlying Markov chain formed by it is geometrically ergodic. Note that different from the SMK shown in [8], the proposed Gibbs-based algorithm performs the sampling over a 1-dimensional symmetric conditional distribution, which successfully gets rid off the need of Klein's algorithm. Then, to further enhance the convergence performance, an optimized scheme of the proposed algorithm is also given. With the optimized symmetric proposal distribution, it outperforms the original one by the convergence rate.

II. Lattice Gaussian Sampling by Gibbs Sampler

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 generated by $\mathbf{B}$ is defined by

$$\Lambda = \{\mathbf{Bx} : \mathbf{x} \in \mathbb{Z}^n\}, \tag{1}$$

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

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

for all $\mathbf{z} \in \mathbb{R}^n$. When $\mathbf{c}$ or σ are not specified, we assume that they are $\mathbf{0}$ and 1 respectively. Then, the *discrete Gaussian*

distribution over Λ is defined as

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

for all $\mathbf{Bx} \in \Lambda$, 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.

On the MCMC front, lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$ can be viewed as a complex target distribution π lacking of direct sampling methods. Therefore, Gibbs sampling who makes use of the 1-dimensional conditional distribution as a tractable alternative to work with was introduced [7]. Specifically, at each Markov move of Gibbs sampling, each coordinate of $\mathbf{x}$ is sampled from the following 1-dimensional conditional distribution

$$\begin{aligned}\pi(x_i|\mathbf{x}_{[-i]}) &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\sum_{x_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}} \\ &= \frac{e^{-\frac{1}{2\sigma_i^2}|x_i-c|^2}}{\sum_{x_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}|x_i-c|^2}}. \end{aligned} \quad (4)$$

Here, $\sigma_i = \sigma/|b_i|$, b_i represents the scaling coefficient of x_i and c stands for the summation of the rest of $n-1$ components of $\mathbf{x}$ multiplied by their own scaling coefficients, $1 \le i \le n$ denotes the coordinate index of $\mathbf{x}$ and $\mathbf{x}_{[-i]} \triangleq [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]^T$. By repeating such a procedure with a randomly chosen coordinate i, an underlying Markov chain $\{\mathbf{X}_0, \mathbf{X}_1, \ldots\}$ is induced, whose transition probability between two adjacent states $\mathbf{x}$ and $\mathbf{y}$ is defined by the univariate conditional distribution [9],

$$P(\mathbf{x},\mathbf{y}) = P(x_i \to y_i|\mathbf{x}_{[-i]}) = \pi(y_i|\mathbf{x}_{[-i]}). \quad (5)$$

Clearly, $\mathbf{x}$ and $\mathbf{y}$ only differ from each other by at most one component while the sampling of y_i for the state $\mathbf{y}$ highly depends on the $n-1$ unchanged components of $\mathbf{x}$ rather than x_i. Without loss of generality, such a single-step move scheme of Gibbs sampling is considered in the context. To summarize, the Gibbs sampler for lattice Gaussian distribution is shown in Algorithm 1, where $\mathbf{r} = \{1/n, \ldots, 1/n\}$ stands for a selection probability set (nonuniform probability set is also possible, see more details in [10]).

Theorem 1 ([7])**.** *Given the target lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by Gibbs algorithm is ergodic:*

$$\lim_{t\to\infty} \|P^t(\mathbf{x};\cdot) - D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV} = 0 \quad (6)$$

for all states $\mathbf{x} \in \mathbb{Z}^n$, where $P^t(\mathbf{x};\cdot)$ denotes a row of the transition matrix $\mathbf{P}$ for t Markov moves and $\|\cdot\|_{TV}$ denotes the total variation distance.

III. Symmetric Metropolis-within-Gibbs Algorithm

A. Classical MH Algorithms

In [11], the original Metropolis algorithm was successfully extended to a more general scheme known as the Metropolis-Hastings (MH) algorithm. In particular, given the current state $\mathbf{x}$ for Markov chain $\mathbf{X}_t$, a state candidate $\mathbf{y}$ for the next Markov move $\mathbf{X}_{t+1}$ is generated from a proposal distribution $q(\mathbf{x},\mathbf{y})$. Then the acceptance ratio α is computed by

$$\alpha(\mathbf{x},\mathbf{y}) = \min\left\{1, \frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}\right\}, \quad (7)$$

and $\mathbf{y}$ will be accepted as the new state by $\mathbf{X}_{t+1}$ with probability α. Otherwise, $\mathbf{x}$ will be retained by $\mathbf{X}_{t+1}$. In this way, a Markov chain $\{\mathbf{X}_0, \mathbf{X}_1, \ldots\}$ is established with the transition probability $P(\mathbf{x},\mathbf{y})$ as follows:

$$P(\mathbf{x},\mathbf{y}) = \begin{cases} q(\mathbf{x},\mathbf{y})\alpha(\mathbf{x},\mathbf{y}) & \text{if } \mathbf{y} \ne \mathbf{x}, \\ 1 - \sum_{\mathbf{z}\ne\mathbf{x}} q(\mathbf{x},\mathbf{z})\alpha(\mathbf{x},\mathbf{z}) & \text{if } \mathbf{y} = \mathbf{x}. \end{cases} \quad (8)$$

It is interesting that in MH algorithms, $q(\mathbf{x},\mathbf{y})$ can be any fixed distribution from which we can conveniently draw samples. However, the MH algorithm also pays a price for its flexibility. If the proposal distribution is poorly chosen, either the acceptance rate is low, or the Markov chain converges slowly.

Algorithm 1 Gibbs algorithm for lattice Gaussian sampling

Input: $\mathbf{B}, \sigma, \mathbf{c}, \mathbf{r}, \mathbf{X}_0$,
Output: samples from the target distribution $\pi = D_{\Lambda,\sigma,\mathbf{c}}$

```
for t =1,2, ... do
    randomly choose coordinate index i from set r
    let x denote the state of X_{t-1}
    generate y by sampling y_i from π(y_i|x_{[-i]})
    if Markov chain goes to steady then
        output y as the state X_t
    end if
end for
```

B. The Proposed Sampling Algorithm

In principle, Gibbs sampling is a special case of MH sampling that tackles with multi-dimensional problems through the univariate conditional sampling. More precisely, by letting $q(\mathbf{x},\mathbf{y}) = \pi(y_i|\mathbf{x}_{[-i]})$, then the acceptance ratio α is always 1 by definition, namely,

$$\frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})} = \frac{\pi(y_i|\mathbf{x}_{[-i]})\pi(\mathbf{x}_{[-i]})\pi(x_i|\mathbf{x}_{[-i]})}{\pi(x_i|\mathbf{x}_{[-i]})\pi(\mathbf{x}_{[-i]})\pi(y_i|\mathbf{x}_{[-i]})} = 1, \quad (9)$$

thus resulting in the classic Gibbs sampling. Inspired by this, the Gibbs update is often replaced by a Metropolis-Hastings step, yielding the Metropolis-within-Gibbs algorithm [10].

Now, we propose the symmetric Metropolis-within-Gibbs sampling algorithm, where $q(\mathbf{x},\mathbf{y})$ is designed as a 1-dimensional conditional symmetric Gaussian distribution:

$$\begin{aligned} q(\mathbf{x},\mathbf{y}) &= Q(x_i \to y_i|\mathbf{x}_{[-i]}) = \frac{e^{-\frac{1}{2\sigma^2}|y_i-x_i|^2}}{\sum_{y_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}|y_i-x_i|^2}} \\ &= \frac{e^{-\frac{1}{2\sigma^2}|y_i-x_i|^2}}{\sum_{z_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}|z_i|^2}} \\ &= q(\mathbf{y},\mathbf{x}). \end{aligned} \quad (10)$$

Clearly, by doing this, the generation of the state candidate $\mathbf{y}$ is completely independent of the other $n-1$ unchanged components, but heavily depends on the i-th component, i.e., x_i, in the previous state $\mathbf{x}$. Meanwhile, since the chain is symmetric, the calculation of the acceptance ratio α is also greatly simplified by such an inhere elegance:

$$\alpha = \min\left\{1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right\} = \min\left\{1, \frac{\pi(y_i|\mathbf{x}_{[-i]}) \cdot \pi(\mathbf{x}_{[-i]})}{\pi(x_i|\mathbf{x}_{[-i]}) \cdot \pi(\mathbf{x}_{[-i]})}\right\} = \min\left\{1, \frac{e^{-\frac{1}{2\sigma_i^2}|y_i - c|^2}}{e^{-\frac{1}{2\sigma_i^2}|x_i - c|^2}}\right\}. \tag{11}$$

To summarize, different from Gibbs sampling who always accepts the new sampling candidate determinately, uncertainty for the sample acceptance in the proposed sampling algorithm is retained in an Metropolis way [12]. Then, based on (10) and (11), the transition probability $P(\mathbf{x}, \mathbf{y})$ of the symmetric Metropolis-within-Gibbs algorithm are given by

$$P(\mathbf{x},\mathbf{y}) = \begin{cases} \min\left\{q(\mathbf{x},\mathbf{y}), \frac{\pi(\mathbf{y})q(\mathbf{x},\mathbf{y})}{\pi(\mathbf{x})}\right\} & \text{if } \mathbf{y} \neq \mathbf{x}, \\ q(\mathbf{x},\mathbf{x}) + \sum\limits_{\mathbf{z}\neq\mathbf{x}} \max\left\{0, q(\mathbf{x},\mathbf{z}) - \frac{\pi(\mathbf{z})q(\mathbf{x},\mathbf{z})}{\pi(\mathbf{x})}\right\} & \text{if } \mathbf{y} = \mathbf{x}. \end{cases} \tag{12}$$

Then, we can easily arrive at the following Theorem. The proof is straightforward but omitted here. For more details, readers are referred to [7].

Theorem 2. *Given the target lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by the proposed symmetric Metropolis-within-Gibbs algorithm is ergodic:*

$$\lim_{t\to\infty} \|P^t(\mathbf{x};\cdot) - D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV} = 0 \tag{13}$$

for all states $\mathbf{x} \in \mathbb{Z}^n$.

IV. Geometric Ergodicity

In this section, the proof of the geometric ergodicity for the proposed algorithm is presented. It should be noticed that at each step, the proposed Gibbs-based algorithm focuses on the sampling over a 1-dimensional symmetric distribution rather than a full dimension distribution. Therefore, compared to the SMK algorithm, it is more suitable for multidimensional target distributions.

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

$$\|P^t(\mathbf{x},\cdot) - \pi(\cdot)\|_{TV} \leq M(\mathbf{x})(1-\delta)^t. \tag{14}$$

Note that the selection of the initial state also matters, which is the main difference between *geometric ergodicity* and *uniform ergodicity* [13].

In MCMC, the *drift condition* is the well-known straightforward way to prove the geometric ergodicity [14], and its definition with respect to a discrete state space Ω is shown below [13].

Algorithm 2 Symmetric Metropolis-within-Gibbs Algorithm for Lattice Gaussian Sampling

Input: $\mathbf{B}, \sigma, \mathbf{c}, \mathbf{X}_0$
Output: samples from the target distribution $\pi = D_{\Lambda,\sigma,\mathbf{c}}$

```
for t = 1,2, ... do
    randomly choose coordinate index i from set r
    let x denote the state of X_{t-1}
    generate y by the proposal distribution q(x, y) in (10)
    calculate the acceptance ratio α(x, y) in (11)
    generate a sample u from the uniform density U[0, 1]
    if u ≤ α(x, y) then
        let X_t = y
    else
        X_t = x
    end if
    if Markov chain goes to steady then
        output the state of X_t
    end if
end for
```

Definition 2. *A Markov chain with discrete state space Ω satisfies the drift condition if there are constants $0 < \lambda < 1$ and $b < \infty$, and a function $V : \Omega \to [1, \infty]$, such that*

$$\sum_{\Omega} P(\mathbf{x},\mathbf{y})V(\mathbf{y}) \leq \lambda V(\mathbf{x}) + b\mathbf{1}_C(\mathbf{x}) \tag{15}$$

for all $\mathbf{x} \in \Omega$, where C is a small set, $\mathbf{1}_C(\mathbf{x})$ equals to 1 when $\mathbf{x} \in C$ and 0 otherwise.

Here, the *small set* C means that there exist $k > 0$, $1 > \delta > 0$ and a probability measure v on Ω such that

$$P^k(\mathbf{x}, \mathcal{B}) \geq \delta v(\mathcal{B}), \quad \forall \mathbf{x} \in C \tag{16}$$

for all measurable subsets $\mathcal{B} \subseteq \Omega$ (also known as *minorisation condition* in literature) [15]. Then, following the footstep from [8], we try to prove the Markov chain arising from the proposed algorithm is geometrically ergodic by satisfying the drift condition.

Theorem 3. *Given the invariant lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain established by the symmetric Metropolis-within-Gibbs algorithm satisfies the drift condition. Therefore, it is geometrically ergodic.*

Proof. By definition, it is easy to verify that any nonempty bounded set $C \subseteq \mathbb{Z}$ corresponds to a small set. In order to specify a small set C, we define

$$C = \{x_i \in \mathbb{Z} : \pi(x_i|\mathbf{x}_{[-i]}) \geq \frac{1}{d^2}\}, \tag{17}$$

where $d > 1$ is a constant set initially.

At each Markov move, given the acceptance ratio α shown in (11), the acceptance region A_{x_i} and the potential rejection region R_{x_i} for the chain started from $\mathbf{x}$ are defined as

$$A_{x_i} = \{y_i \in \mathbb{Z} | \pi(y_i|\mathbf{x}_{[-i]}) \geq \pi(x_i|\mathbf{x}_{[-i]})\}; \tag{18}$$

$$R_{x_i} = \{y_i \in \mathbb{Z} | \pi(y_i|\mathbf{x}_{[-i]}) < \pi(x_i|\mathbf{x}_{[-i]})\}. \tag{19}$$

In other words, state candidate $\mathbf{y}$ will be accepted by $\mathbf{X}_{t+1}$ without uncertainty if $y_i \in A_{x_i}$. On the other hand, $\mathbf{y}$ has a certain risk to be rejected if $y_i \in R_{x_i}$. Then, divided by A_{x_i} and R_{x_i}, the discrete term $\sum_{y_i\in\mathbb{Z}} P(\mathbf{x},\mathbf{y})V(\mathbf{y})$ can be expressed as

$$\sum_{y_i\in\mathbb{Z}} P(\mathbf{x},\mathbf{y})V(\mathbf{y}) = \sum_{y_i\in A_{x_i}} P(\mathbf{x},\mathbf{y})V(\mathbf{y}) + \sum_{y_i\in R_{x_i}} P(\mathbf{x},\mathbf{y})V(\mathbf{y})$$
$$= \sum_{y_i\in A_{x_i}} q(\mathbf{x},\mathbf{y})V(\mathbf{y}) + \sum_{y_i\in R_{x_i}} q(\mathbf{x},\mathbf{y})\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}V(\mathbf{y}) + \sum_{y_i\in R_{x_i}} q(\mathbf{x},\mathbf{y})\left[1-\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right]V(\mathbf{x}).$$

Let $V(\mathbf{x}) = \pi(\mathbf{x})^{-\frac{1}{2}}$, for the consideration of the indicator function $\mathbf{1}_C(\mathbf{x})$, we will discuss the two cases of $x_i \notin C$ and $x_i \in C$ respectively to reveal that the drift condition is satisfied.

(i). In the case of $x_i \notin C$, as $\mathbf{1}_C(\mathbf{x}) = 0$, λ can be expressed directly as

$$\lambda = \frac{\sum\limits_{y_i\in\mathbb{Z}} P(\mathbf{x},\mathbf{y})V(\mathbf{y})}{V(\mathbf{x})} \tag{20}$$

Then, by substitution, it follows that

$$\lambda = 1 - \sum_{y_i\in A_{x_i}} q(\mathbf{x},\mathbf{y})\left[1-\frac{V(\mathbf{y})}{V(\mathbf{x})}\right] + \sum_{y_i\in R_{x_i}} q(\mathbf{x},\mathbf{y})\left[\frac{V(\mathbf{x})}{V(\mathbf{y})} - \frac{V(\mathbf{x})^2}{V(\mathbf{y})^2}\right].$$

Insight into the term $V(\mathbf{x})/V(\mathbf{y})$, we have

$$\frac{V(\mathbf{x})}{V(\mathbf{y})} = \left(\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right)^{\frac{1}{2}} = \left(\frac{e^{-\frac{1}{2\sigma_i^2}|y_i-c|^2}}{e^{-\frac{1}{2\sigma_i^2}|x_i-c|^2}}\right)^{\frac{1}{2}}. \tag{21}$$

For further investigation, define a function

$$\omega(x) = e^{-\frac{1}{2\sigma_i^2}|x-c|^2} \tag{22}$$

and it is straightforward to verify that

$$\lim_{|x|\to\infty} l(x)\cdot\nabla\log\,\omega(x) = -\infty, \tag{23}$$

where $l(x) = x/|x|$ and ∇ represents the gradient. This condition implies that for any arbitrarily large $\gamma > 0$, there exists $R > 0$ such that

$$\frac{\omega(x + a\cdot l(x))}{\omega(x)} \le e^{-a\cdot\gamma}, \tag{24}$$

where $|x| \ge R, a \ge 0$. Put it another way, as $|x|$ goes to infinity, ω is at least exponentially decaying with a rate γ tending to infinity. Hence, once $|x|$ is large enough, even a minimum discrete integer increment, namely, $|\Delta| = 1$, can make $\omega(x+\Delta)$ become extremely smaller or larger than $\omega(x)$.

Now, suppose $y_i^1 = x_i + \Delta \in R_x$, with large enough $|x_i|$, the ratio of $\omega(y_i^1)/\omega(x_i)$ could be arbitrarily small, that is

$$\frac{\omega(y_i^1)}{\omega(x_i)} \to 0 \text{ for } |x_i| \to \infty. \tag{25}$$

As y_i^1 is the closest candidate to x_i in set R_x, then the following relationship holds due to the arbitrarily large exponential decay rate of ω

$$\omega(y_i^2) \ll \omega(y_i^1) \text{ for } |x_i| \to \infty, \tag{26}$$

where $y_i^2 \in R_x$, $y_i^2 \ne y_i^1$. Therefore, we have

$$\frac{\omega(y_i^2)}{\omega(x_i)} \ll \frac{\omega(y_i^1)}{\omega(x_i)} \to 0 \text{ for } |x_i| \to \infty, \tag{27}$$

implying the summation term about $V(\mathbf{x})/V(\mathbf{y})$ for $y_i \in R_x$ will tend to be 0 as $|x_i|$ goes to infinity

$$\sum_{y_i\in R_x} \frac{V(\mathbf{y})}{V(\mathbf{x})} = \sum_{y_i\in R_x} \left(\frac{\omega(x_i)}{\omega(y_i)}\right)^{\frac{1}{2}} \to 0 \text{ for } |x_i| \to \infty. \tag{28}$$

Similarly, the same thing happens to the summation term about $V(\mathbf{y})/V(\mathbf{x})$ for $y_i \in A_x$, namely,

$$\sum_{y_i\in A_x} \frac{V(\mathbf{x})}{V(\mathbf{y})} = \sum_{y_i\in A_x} \left(\frac{\omega(y_i)}{\omega(x_i)}\right)^{\frac{1}{2}} \to 0 \text{ for } |x_i| \to \infty. \tag{29}$$

Consequently, based on (28) and (29), as $|x_i|$ goes to infinity, the following derivation holds

$$\begin{aligned}
\lambda &= \limsup_{|x_i|\to\infty} \frac{\sum\limits_{y_i\in\mathbb{Z}} P(\mathbf{x},\mathbf{y})V(\mathbf{y})}{V(\mathbf{x})} \\
&= 1 - \liminf_{|x_i|\to\infty} \sum_{y_i\in A_{x_i}, y_i\ne x_i} q(\mathbf{x},\mathbf{y}) \\
&= 1 - \liminf_{|x_i|\to\infty} \sum_{y_i\in A_{x_i}, y_i\ne x_i} \frac{e^{-\frac{1}{2\sigma^2}|y_i-x_i|^2}}{\sum_{y_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}|y_i-x_i|^2}} \\
&< 1,
\end{aligned} \tag{30}$$

which means the drift condition shown in (15) is satisfied in the case of $x_i \notin C$.

(ii). On the other hand, if $x_i \in C$, i.e., $\mathbf{1}_C(\mathbf{x}) = 1$, then for $y_i \in A_{x_i}$, we have

$$\begin{aligned}
V(\mathbf{y}) = \pi(\mathbf{y})^{-\frac{1}{2}} &= \left(\pi(y_i|\mathbf{x}_{[-i]})\cdot\pi(\mathbf{x}_{[-i]})\right)^{-\frac{1}{2}} \\
&\le \left(\pi(x_i|\mathbf{x}_{[-i]})\cdot\pi(\mathbf{x}_{[-i]})\right)^{-\frac{1}{2}} \\
&\overset{(a)}{\le} d\cdot\pi(\mathbf{x}_{[-i]})^{-\frac{1}{2}},
\end{aligned} \tag{31}$$

where inequality (a) holds by the definition of small set given in (17). As the rest of $n-1$ components $\mathbf{x}_{[-i]}$ keep unchanged, the probability $\pi(\mathbf{x}_{[-i]})$ is a constant here. Therefore, the term $\sum_{y_i\in A_{x_i}} q(\mathbf{x},\mathbf{y})V(\mathbf{y})$ can always be upper bounded by a constant $b > 0$, namely,

$$\sum_{y_i\in A_{x_i}} q(\mathbf{x},\mathbf{y})V(\mathbf{y}) \le b < \infty. \tag{32}$$

Hence, as $|x|$ goes to infinity, it follows that

$$\begin{aligned}
\limsup_{|x_i|\to\infty} \sum_{y_i\in\mathbb{Z}} P(\mathbf{x},\mathbf{y})V(\mathbf{y}) &\le b + \limsup_{|x|\to\infty} \sum_{y_i\in R_{x_i}} q(\mathbf{x},\mathbf{y})V(\mathbf{x}) \\
&= b + \lambda V(\mathbf{x}),
\end{aligned} \tag{33}$$

and based on (28), similarly, it is easy to verify that

$$\lambda = \lim_{|x_i|\to\infty} \sup \sum_{y_i \in R_{x_i}} q(\mathbf{x}, \mathbf{y}) < 1, \tag{34}$$

completing the proof. □

Overall, the exponential convergence rate can be interpreted in two folds. On one hand, when $x_i \notin C$, the Markov chain shrinks geometrically towards the small set C as $\lambda < 1$. On the other hand, if $x_i \in C$, the Markov chain will converge to the stationary distribution exponentially fast. Obviously, there is a trade-off between them depending on the set size of C. However, the problem lies on the fact that C is determined artificially, resulting in δ and λ flexible and sensitive to any slight change.

V. Algorithm Optimization

To further enhance the convergence performance, we now provide an optimized scheme for the proposed symmetric Metropolis-within-Gibbs algorithm. In particular, the proposal distribution is slightly changed as

$$q_{\text{opt}}(\mathbf{x}, \mathbf{y}) = Q_{\text{opt}}(x_i \to y_i) = \frac{Q(x_i \to y_i)}{1 - Q(x_i \to x_i)}, \tag{35}$$

which corresponds to eliminate x_i itself from the sampling list of y_i. Apparently, this new proposal distribution is still symmetric while the transition probability $P_{\text{opt}}(\mathbf{x}, \mathbf{y})$ with $\mathbf{y} \neq \mathbf{x}$ becomes

$$P_{\text{opt}}(\mathbf{x}, \mathbf{y}) = \min \left\{ q_{\text{opt}}(\mathbf{x}, \mathbf{y}), \frac{\pi(\mathbf{y}) q_{\text{opt}}(\mathbf{x}, \mathbf{y})}{\pi(\mathbf{x})} \right\}. \tag{36}$$

Theorem 4. *The optimized symmetrical Metropolis-within-Gibbs algorithm is statistically more efficient than the original one.*

Proof. According to (12) and (36), the following relationship holds

$$q_{\text{opt}}(\mathbf{x}, \mathbf{y}) > q(\mathbf{x}, \mathbf{y}) \tag{37}$$

for all the cases of $\mathbf{y} \neq \mathbf{x}$. Therefore, we have

$$P_{\text{opt}}(\mathbf{x}, \mathbf{y}) > P(\mathbf{x}, \mathbf{y}), \tag{38}$$

which means each of the off-diagonal elements of the transition matrix $\mathbf{P}_{\text{opt}}$ is always larger than that of $\mathbf{P}$, namely, $\mathbf{P}_{\text{opt}} > \mathbf{P}$. Furthermore, since both of the Markov chains are reversible by satisfying

$$\pi(\mathbf{x}) \cdot P(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y}) \cdot P(\mathbf{y}, \mathbf{x}), \tag{39}$$

the convergence improvement of $\mathbf{P}_{\text{opt}}$ over $\mathbf{P}$ can be verified by invoking the *Peskun's Theorem* shown below, which takes advantages of a sensible criterion in MCMC known as *asymptotic efficiency* [9].

Lemma 1 ([16])**.** *Suppose $\mathbf{P}_1$ and $\mathbf{P}_2$ are reversible transition matrices with the same invariant distribution and $\mathbf{P}_2 \geq \mathbf{P}_1$. Then, for all any function $f \in L_0^2(\pi) = \{f \in L^2(\pi) : E\{f\} = 0\}$, we have*

$$v(f, \mathbf{P}_1) \geq v(f, \mathbf{P}_2). \tag{40}$$

Here, $L^2(\pi)$ denote the set of all function $f(\cdot)$ that are square integrable with respect to π and $v(f, \mathbf{P})$ is defined as sampler's asymptotic efficiency by

$$v(f, \mathbf{P}) = \lim_{n\to\infty} \frac{1}{n} var \left\{ \sum_{t=1}^{n} f(\mathbf{X}_t) \right\}, \tag{41}$$

where $\mathbf{X}_0, \ldots, \mathbf{X}_t$ establish the corresponding Markov chain.

□

From Lemma 1, $\mathbf{P}_{\text{opt}}$ will lead to a smaller asymptotic variance for all observable than $\mathbf{P}$, leading to further convergence efficiency. In other words, the transition matrix in MCMC is encouraged to entail smaller diagonal elements and larger off-diagonal elements [17].

Acknowledgment

This work was supported in part by FP7 project PHYLAWS (EU FP7-ICT 317562).

References

[1] 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.

[2] 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.

[3] 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.

[4] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, USA, 2009.

[5] N. Stephens-Davidowitz, "Discrete Gaussian sampling reduces to CVP and SVP," submitted for publication. [Online]. Available: http://arxiv.org/abs/1506.07490.

[6] 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.

[7] 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.

[8] Z. Wang and C. Ling, "On the geometric ergodicity of Metropolis-Hastings algorithms for lattice Gaussian sampling," *Submitted to IEEE Transactions on Information Theory, 2015.*, [Online] Available:http://arxiv.org/pdf/1501.05757v2.pdf.

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

[10] K. Latuszyński, G. O. Roberts, and J. S. Rosenthal, "Adaptive Gibbs samplers and related MCMC methods," *The Annals of Applied Probability*, vol. 23, no. 1, pp. 66–98, 2013.

[11] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[12] J. S. Liu, "Peskun's theorem and a modified discrete-state Gibbs sampler," *Biometrika*, vol. 83, no. 3, pp. 681–682, 1996.

[13] G. O. Roberts, "General state space Markov chains and MCMC algorithms," *Probability Surveys*, vol. 1, pp. 20–71, 2004.

[14] G. O. Roberts and R. L. Tweedie, "Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms," *Biometrika*, vol. 83, pp. 95–110, 1996.

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

[16] P. H. Peskun, "Optimal Monte Carlo sampling using Markov chains," *Biometrika*, vol. 60, pp. 607–612, 1973.

[17] A. Frigessi, C.-R. Hwang, and L. Younes, "Optimal spectral structure of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields," *The Annals of Applied Probability*, vol. 2, no. 3, pp. 610–628, 1992.