# On the Geometric Ergodicity of Metropolis-Hastings Algorithms for Lattice Gaussian Sampling

Zheng Wang, Cong Ling, *Member, IEEE*

***Abstract*—Sampling from the lattice Gaussian distribution is emerging as an important problem in coding and cryptography. In this paper, the classic Metropolis-Hastings (MH) algorithm from Markov chain Monte Carlo (MCMC) methods is introduced into lattice Gaussian sampling. Two MH-based algorithms are proposed, which overcome the restriction suffered by the default Klein algorithm. The first one, referred to as the independent Metropolis-Hastings-Klein (MHK) algorithm, tries to establish a Markov chain through an independent proposal distribution. We show that the Markov chain arising from the independent MHK algorithm is uniformly ergodic, namely, it converges to the stationary distribution exponentially fast regardless of the initial Markov state. Moreover, the rate of convergence can be explicitly calculated in terms of the theta series, leading to a predictable mixing time for the underlying Markov chains. In order to further exploit the convergence potential, the symmetric Metropolis-Klein (SMK) algorithm is proposed. It is proven that the Markov chain induced by the SMK algorithm is geometrically ergodic, where a reasonable selection of the initial state is capable to enhance the convergence performance.**

***Index Terms*—Lattice Gaussian sampling, Metropolis-Hastings sampling, MCMC methods, lattice coding and decoding.**

## I. Introduction

Recently, the lattice Gaussian distribution is emerging as a common theme 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], [3], and to achieve the capacity of the Gaussian channel and the secrecy capacity of the Gaussian wiretap channel, respectively [4], [5]. 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 [6]. Meanwhile, it also has underpinned the fully-homomorphic encryption for cloud computing [7]. Algorithmically, lattice Gaussian sampling with a suitable variance allows to solve the shortest vector problem (SVP) and the closest vector problem (CVP); for example, it has led to efficient lattice decoding for multi-input multi-output (MIMO) systems [8], [9]. In theory, it has been demonstrated that lattice Gaussian sampling is equivalent to CVP via a polynomial-time dimension-preserving reduction [10], and SVP is essentially a special case of CVP.

Due to the central role of the lattice Gaussian distribution playing in these fields, its sampling algorithms become an important computational problem. However, compared to sampling from continuous Gaussian distributions, it is by no means trivial to perform the sampling even from a low-dimensional discrete Gaussian distribution. As the default sampling algorithm for lattices, Klein's algorithm [11] samples within a negligible statistical distance from the lattice Gaussian distribution if and only if the standard deviation $\sigma \geq \sqrt{\omega(\log n)} \cdot \max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$ [12], where $\omega(\log n)$ is a superlogarithmic function, $n$ denotes the lattice dimension and $\widehat{\mathbf{b}}_i$'s are the Gram-Schmidt vectors of the lattice basis $\mathbf{B}$. However, such requirement of $\sigma$ normally can be excessively large, rendering Klein's algorithm inapplicable to many scenarios of interest.

Markov chain Monte Carlo (MCMC) methods attempt to sample from the target distribution by building a Markov chain, which randomly generates the next sample conditioned on the previous samples. After a burn-in period, which is normally measured by the *mixing time*, the Markov chain will reach a stationary distribution, and successful sampling from the complex target distribution can be carried out. The complexity of each Markov move is often insignificant. Therefore, we are mostly concerned with the mixing time that the underlying Markov chain needs to get steady.

To this end, the Gibbs algorithm was introduced into lattice Gaussian sampling, which employs univariate conditional sampling to build the Markov chains [13]. It is able to sample beyond the range that Klein's algorithm. In [13], a flexible block-based Gibbs algorithm was introduced, which performs the sampling over multiple elements within a block. In this way, the correlation within the block could be exploited, leading to faster convergence especially in the case of high correlated components. Unfortunately, the related analysis of the convergence rate for the associated Markov chain in these two algorithms were lacking, resulting in an unpredictable mixing time of the Markov chains. Actually, according to *coupling technique* [14], it is easy to know that the Gibbs-based MCMC sampler converges exponentially fast for any finite state space Markov chain. However, as the Markov chain designed for lattice Gaussian sampling naturally has a countably infinite state space rather than a finite one, its convergence is still unknown.

On the other hand, Gibbs sampling has already been introduced into multiple-input multiple-output (MIMO) communications for signal detection [15]–[20]. In particular, the

This work was presented in part at the IEEE International Symposium on Information Theory (ISIT) 2015, Hong Kong, China, July 2015.

Z. Wang and C. Ling are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, United Kingdom (e-mail: z.wang10@imperial.ac.uk, cling@ieee.org).

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

selection of $\sigma$ (also referred to as "temperature") is studied in [15] and it shows that $\sigma$ should grow as fast as the signal-to-noise ratio (SNR) in general for a fast mixing time. In [16], a mixed-Gibbs sampler is proposed to achieve the near-optimal detection performance, which takes the benefits of an efficient stopping criterion and a multiple restart strategy. Moveover, Gibbs sampling is also introduced into the soft-output decoding in MIMO systems, where the extrinsic information calculated by a priori probability (APP) detector is used to produce soft outputs [17]. In [18], the investigation of the Gibbs-based MCMC receivers in different communication channels are given. Due to a finite state space formed by the modulation constellation, those Gibbs sampler converges geometrically fast to the stationary distribution. Nevertheless, the rate of convergence is still difficult to determine.

In this paper, another famous MCMC scheme, known as the Metropolis-Hastings (MH) algorithm [21], is studied in details for lattice Gaussian sampling. In particular, it makes use of a *proposal distribution* which suggests a possible state candidate and then employs a acceptance-rejection rule to decide whether to accept the suggested candidate in the next Markov move. Obviously, the art of designing an efficient MH algorithm chiefly lies in choosing an appropriate proposal distribution, and this motivates us to design the target proposal distributions based on Klein's algorithm.

In the proposed independent Metropolis-Hastings-Klein (MHK) algorithm, through Klein's algorithm, a candidate at each Markov move is generated from a Gaussian-like proposal distribution. In this case, we show that the Markov chain induced by the proposed algorithm is uniformly ergodic, namely, it converges to the stationary distribution exponentially fast irrespective of the starting state. Further analysis of its convergence rate is then performed, where the convergence rate can be explicitly estimated given the lattice basis $\mathbf{B}$, query point $\mathbf{c}$ and standard derivation $\sigma$. Thus, the mixing time of the proposed algorithm becomes tractable. Note that the proposed independent MHK algorithm is well applicable to MIMO detections, thus making the convergence rate as well as the mixing time accessible. To the best of our knowledge, this is the first time that the convergence rate of MCMC in communications and signal processing can be determined since MCMC was introduced into these fields in 1990s [22].

Different from the algorithms in [23], [24] which have an exponential lower bound $2^{n/2}$ on the space and time complexity, the proposed independent MHK algorithm has polynomial space complexity, and its time complexity varies with $\sigma$, where a larger value of $\sigma$ corresponds to smaller mixing time. This is in accordance with the fact we knew before; if $\sigma$ is large enough, then there is no need of MCMC in lattice Gaussian sampling since Klein's algorithm can be applied directly with polynomial time complexity.

The second proposed algorithm, named as symmetric Metropolis-Klein (SMH) algorithm, establishes a symmetric proposal distribution between two consecutive Markov states. We show it converges to the stationary distribution exponentially fast while the selection of the initial state also plays a role, such case is referred to as geometric ergodicity in MCMC literature [25]. Besides the geometric ergodicity, another advantage of the proposed SMH algorithm lies in its remarkable elegance and simplicity, which comes from the usage of a symmetrical proposal distribution.

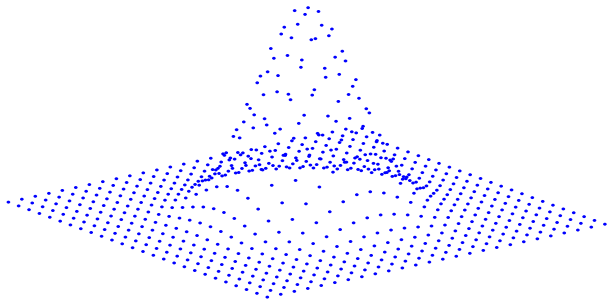

Fig. 1. Illustration of a two-dimensional lattice Gaussian distribution.

To summarize, the main contributions of this paper are the following:

1) The independent MHK algorithm is proposed for lattice Gaussian sampling, where the Markov chain arisen from it converges exponentially fast to the stationary distribution.
2) The convergence rate in the independent MHK algorithm is explicitly estimated by theta series, thereby leading to a tractable mixing time of the underlying Markov chain.
3) The SMH algorithm is further proposed for lattice Gaussian sampling, which not only achieves an exponential convergence performance, but also is greatly simplified due to its symmetry.

The rest of this paper is organized as follows. Section II introduces the lattice Gaussian distribution and briefly reviews the basics of MCMC methods. In Section III, we propose the independent MHK algorithm for lattice Gaussians, where uniform ergodicity is demonstrated. In Section IV, the convergence rate of the independent MHK algorithm is analyzed and explicitly calculated in terms of the theta series. In Section V, the proposed symmetric random walk MH algorithm for lattice Gaussian sampling is given, followed by the demonstration of geometric ergodicity. Finally, Section VI concludes the paper.

*Notation:* Matrices and column vectors are denoted by upper and lowercase boldface letters, and the transpose, inverse, pseudoinverse of a matrix $\mathbf{B}$ by $\mathbf{B}^T, \mathbf{B}^{-1}$, and $\mathbf{B}^\dagger$, respectively. We use $\mathbf{b}_i$ for the $i$th column of the matrix $\mathbf{B}$, $\widehat{\mathbf{b}}_i$ for the $i$th Gram-Schmidt vector of the matrix $\mathbf{B}$, $b_{i,j}$ for the entry in the $i$th row and $j$th column of the matrix $\mathbf{B}$. $\lceil x \rfloor$ denotes rounding to the integer closest to $x$. If $x$ is a complex number, $\lceil x \rfloor$ rounds the real and imaginary parts separately. Finally, in this paper, the computational complexity is measured by the number of arithmetic operations (additions, multiplications, comparisons, etc.).

## II. Preliminaries

In this section, we introduce the background and mathematical tools needed to describe and analyze the proposed lattice

Gaussian sampling algorithms.

### A. Lattice Gaussian Distribution

Let $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n] \subset \mathbb{R}^n$ consist of $n$ linearly independent vectors. The $n$-dimensional lattice $\Lambda$ 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 $\sigma$ are not specified, we assume that they are $\mathbf{0}$ and 1 respectively. Then, the *discrete Gaussian distribution* over $\Lambda$ 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 $\mathbf{Bx} \in \Lambda$, where $\rho_{\sigma,\mathbf{c}}(\Lambda) \triangleq \sum_{\mathbf{Bx}\in\Lambda} \rho_{\sigma,\mathbf{c}}(\mathbf{Bx})$ is just a scaling quantity needed to make the function into a probability distribution.

We remark that this definition differs slightly from the one in [6], where $\sigma$ is scaled by a constant factor $\sqrt{2\pi}$ (i.e., $s = \sqrt{2\pi}\sigma$). Fig. 1 illustrates the discrete Gaussian distribution over $\mathbb{Z}^2$. As it can be seen clearly, it resembles a continuous Gaussian distribution, but is only defined over a lattice. In fact, discrete and continuous Gaussian distributions share similar properties, if the *flatness factor* is small [5].

### B. Klein's Algorithm

Obviously, an intuition of $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$ suggests that a lattice point $\mathbf{Bx}$ closer to $\mathbf{c}$ will be sampled with a higher probability. Therefore, sampling from the lattice Gaussian distribution can be naturally used in solving the CVP (where $\mathbf{c}$ is the query point) and SVP (where $\mathbf{c} = \mathbf{0}$) in lattices. Because of this, Klein's algorithm that samples from a Gaussian-like distribution was originally designed for lattice decoding [11], and small size of $\sigma$ is preferred for the consideration of the decoding efficiency. As shown in Algorithm 1, the operation of Klein's algorithm has polynomial complexity $O(n^2)$ excluding QR decomposition. More precisely, by sequentially sampling from the 1-dimensional conditional Gaussian distribution $D_{\mathbb{Z},\sigma_i,\widetilde{x}_i}$ in a backward order from $x_n$ to $x_1$, the Gaussian-like distribution arising from Klein's algorithm is given by

$$\begin{aligned} P_{\text{Klein}}(\mathbf{x}) &= \prod_{i=1}^{n} D_{\mathbb{Z},\sigma_i,\widetilde{x}_i}(x_i) = \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{Bx})}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z})} \\ &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\prod_{i=1}^{n}\sum_{\widetilde{x}_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|x_i-\widetilde{x}_i\|^2}}, \end{aligned} \tag{4}$$

where $\widetilde{x}_i = \frac{c_i' - \sum_{j=i+1}^{n} r_{i,j}x_j}{r_{i,i}}$, $\sigma_i = \frac{\sigma}{|r_{i,i}|}$, $\mathbf{c}' = \mathbf{Q}^\dagger\mathbf{c}$ and $\mathbf{B} = \mathbf{QR}$. In [12], it has been demonstrated that $P_{\text{Klein}}(\mathbf{x})$ is close to $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$ within a negligible statistical distance if

$$\sigma \geq \sqrt{\omega(\log\, n)} \cdot \max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|. \tag{5}$$

**Algorithm 1** Klein's Algorithm

**Input:** $\mathbf{B}, \sigma, \mathbf{c}$
**Output:** $\mathbf{Bx} \in \Lambda$
1: let $\mathbf{B} = \mathbf{QR}$ and $\mathbf{c}' = \mathbf{Q}^\dagger\mathbf{c}$
2: **for** $i = n, \ldots, 1$ **do**
3: let $\sigma_i = \frac{\sigma}{|r_{i,i}|}$ and $\widetilde{x}_i = \frac{c_i' - \sum_{j=i+1}^{n} r_{i,j}x_j}{r_{i,i}}$
4: sample $x_i$ from $D_{\mathbb{Z},\sigma_i,\widetilde{x}_i}$
5: **end for**
6: return $\mathbf{Bx}$

However, even with the help of lattice reduction (e.g., LLL reduction), the size of $\max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$ still can be excessively large.

### C. MCMC Methods

As for the lattice Gaussian sampling in the range $\sigma < \sqrt{\omega(\log\, n)}\cdot\max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$, MCMC methods have become an alternative solution, where the discrete Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$ is viewed as a complex target distribution lacking direct sampling methods. By establishing a Markov chain that randomly generates the next state based on the previous state, MCMC is capable of sampling from the target distribution of interest, thereby removing the restriction on $\sigma$ [13].

As an important parameter which measures the time required by a Markov chain to get close to its stationary distribution, the *mixing time* is defined as [26]

$$t_{\text{mix}}(\epsilon) = \min\{t : \max\|P^t(\mathbf{x},\cdot) - \pi(\cdot)\|_{TV} \leq \epsilon\}, \tag{6}$$

where $\|\cdot\|_{TV}$ represents the total variation distance and other measures of distance also exist [14]. It is well known that the *spectral gap* $\gamma$ of the transition matrix offers a good upper bound on the mixing time, that is

$$t_{\text{mix}}(\epsilon) \leq \frac{1}{\gamma}\log\left(\frac{1}{\pi_{\min}\epsilon}\right), \tag{7}$$

where $\pi_{\min} = \min_{\mathbf{x}\in\mathbf{\Omega}}\pi(\mathbf{x})$, $\Omega$ stands for the state space, $\gamma = 1 - |\lambda_1| > 0$ and $\lambda_1$ represents the second largest eigenvalue of the transition matrix $\mathbf{P}$ in a Markov chain. Therefore, a large value of spectral gap leads to the rapid convergence to stationarity [27].

However, the spectrum of the Markov chain is very hard to analyze directly, especially the size of the state space $\Omega$ tends to be exponentially large, making it difficult to have a compact, mathematical representation for the adjacency matrix. Thanks to the celebrated *coupling technique*, for any Markov chain with finite state space $\Omega$, exponentially fast convergence can be demonstrated if the underlying Markov chain is irreducible and aperiodic with an invariant distribution $\pi$ [26]. Nevertheless, in the case of lattice Gaussian sampling by MCMC, the countably infinite state space $\mathbf{x} \in \mathbb{Z}^n$ naturally becomes a challenge. To this end, we perform the convergence analysis from the beginning — ergodicity.

**Definition 1.** *Let* $\mathbf{P}$ *be an irreducible and aperiodic transition matrix for a Markov chain. If the chain is positive recurrent, then it is ergodic, namely, there is a unique probability distribution* $\pi$ *on* $\Omega$ *and for all* $\mathbf{x} \in \Omega$,

$$\lim_{t\to\infty}\|P^t(\mathbf{x},\cdot)-\pi\|_{TV}=0, \tag{8}$$

*where $P^t(\mathbf{x};\cdot)$ denotes a row of the transition matrix $\mathbf{P}$ for $t$ Markov moves.*

Although *ergodicity* implies asymptotic convergence to stationarity, it does not say anything about the convergence rate. To this end, the following definitions are given.

**Definition 2.** *A Markov chain having stationary distribution $\pi(\cdot)$ is uniformly ergodic if there exists $0<\delta<1$ and $M<\infty$ such that for all $\mathbf{x}$*

$$\|P^t(\mathbf{x},\cdot)-\pi(\cdot)\|_{TV}\le M(1-\delta)^t. \tag{9}$$

Obviously, the exponential decay coefficient $\delta$ is key to determine the convergence rate. As $M$ is a constant, the convergence rate does not depend on the initial state $\mathbf{x}$. As a weaker version of uniform ergodicity, *geometric ergodicity* also converges exponentially as well, but $M$ is parameterized by the initial state $\mathbf{x}$.

**Definition 3.** *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 $\mathbf{x}$*

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

Besides exponential convergence, polynomial convergence also exists [28], which goes beyond the scope of this paper due to the inefficient convergence performance. Unless stated otherwise, the state space of the Markov chain we are concerned with throughout the context is countably infinite, i.e., $\Omega=\mathbb{Z}^n$.

### D. Classical MH Algorithms

The origins of the Metropolis algorithm can be traced back to the celebrated work of [29] in 1950's. In [21], the original Metropolis algorithm was successfully extended to a more general scheme known as the Metropolis-Hastings (MH) algorithm. In particular, let us consider a target invariant distribution $\pi$ together with a proposal distribution $q(\mathbf{x},\mathbf{y})$. 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 the proposal distribution $q(\mathbf{x},\mathbf{y})$. Then the acceptance ratio $\alpha$ is computed by

$$\alpha(\mathbf{x},\mathbf{y})=\min\left\{1,\frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}\right\}, \tag{11}$$

and $\mathbf{y}$ will be accepted as the new state by $\mathbf{X}_{t+1}$ with probability $\alpha$. Otherwise, $\mathbf{x}$ will be retained by $\mathbf{X}_{t+1}$ with probability $1-\alpha$. 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:

$$P(\mathbf{x},\mathbf{y})=\begin{cases}q(\mathbf{x},\mathbf{y})\alpha(\mathbf{x},\mathbf{y}) & \text{if } \mathbf{y}\neq\mathbf{x},\\ 1-\sum_{\mathbf{z}\neq\mathbf{x}}q(\mathbf{x},\mathbf{z})\alpha(\mathbf{x},\mathbf{z}) & \text{if } \mathbf{y}=\mathbf{x}.\end{cases} \tag{12}$$

It is interesting that in MH algorithms, the proposal distribution $q(\mathbf{x},\mathbf{y})$ can be any fixed distribution from which we can conveniently draw samples. Undoubtedly, the fastest converging proposal density would be $q(\mathbf{x},\mathbf{y})=\pi(\mathbf{y})$, but in the context it is assumed that $\pi$ cannot be sampled directly. To this end, many variations of MH algorithms with different configurations of $q(\mathbf{x},\mathbf{y})$ were proposed. Nevertheless, how to generate the candidate state $\mathbf{y}$ from the proposal distribution $q$ still tends to be difficult for MH algorithms especially in high dimensions.

## III. Independent MHK Algorithm

In this section, the independent Metropolis-Hastings-Klein (MHK) algorithm for lattice Gaussian sampling is firstly presented. Then, we show that the Markov chain induced by the proposed algorithm is uniformly ergodic.

### A. Independent MHK Algorithm

In the proposed independent MHK algorithm, Klein's sampling is used to generate the state candidate $\mathbf{y}$ for the each Markov move $\mathbf{X}_{t+1}$. As shown in Algorithm 2, it consists of three basic steps:

1) *Sample from the independent proposal distribution through Klein's algorithm to obtain the candidate state $\mathbf{y}$ for $\mathbf{X}_{t+1}$,*

$$\begin{aligned}q(\mathbf{x},\mathbf{y}) &= P_{\text{Klein}}(\mathbf{y}) = \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{By})}{\prod_{i=1}^n\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})}\\ &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{By}-\mathbf{c}\|^2}}{\prod_{i=1}^n\sum_{\widetilde{y}_i\in\mathbb{Z}}e^{-\frac{1}{2\sigma_i^2}\|y_i-\widetilde{y}_i\|^2}}\\ &= q(\mathbf{y}),\end{aligned} \tag{13}$$

where $\mathbf{y}\in\mathbb{Z}^n$, $\widetilde{y}_i=\frac{c_i'-\sum_{j=i+1}^n r_{i,j}y_j}{r_{i,i}}$, $\sigma_i=\frac{\sigma}{|r_{i,i}|}$, $\mathbf{c}'=\mathbf{Q}^\dagger\mathbf{c}$ and $\mathbf{B}=\mathbf{QR}$.

2) *Calculate the acceptance ratio $\alpha(\mathbf{x},\mathbf{y})$*

$$\begin{aligned}\alpha(\mathbf{x},\mathbf{y}) &= \min\left\{1,\frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}\right\}=\min\left\{1,\frac{\pi(\mathbf{y})q(\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{y})}\right\}\\ &= \min\left\{1,\frac{\prod_{i=1}^n\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})}{\prod_{i=1}^n\rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z})}\right\},\end{aligned} \tag{14}$$

where $\pi=D_{\Lambda,\sigma,\mathbf{c}}$.

3) *Make a decision for $\mathbf{X}_{t+1}$ based on $\alpha(\mathbf{x},\mathbf{y})$ to accept $\mathbf{X}_{t+1}=\mathbf{y}$ or not.*

A salient feature of the independent MHK algorithm is that the generation of the state candidate $\mathbf{y}$ is independent of the previous one, which is completely accomplished by Klein's algorithm. Therefore, the connection between two consecutive Markov states only lies in the decision part, resulting in an independent MCMC sampler.

**Theorem 1.** *Given the target lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by the independent MHK algorithm is ergodic:*

$$\lim_{t\to\infty}\|P^t(\mathbf{x};\cdot)-D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV}=0 \tag{15}$$

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

The proof of Theorem 1 is provided in Appendix A.

### B. Uniform Ergodicity

**Lemma 1.** *In the independent MHK algorithm for lattice Gaussian sampling, there exists $\delta > 0$ such that*

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})} \geq \delta \tag{16}$$

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

*Proof.* Using (3) and (4), we have

$$\begin{aligned} \frac{q(\mathbf{x})}{\pi(\mathbf{x})} &= \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{Bx})}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z})} \cdot \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\rho_{\sigma,\mathbf{c}}(\mathbf{Bx})} \\ &= \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z})} \\ &\overset{(a)}{\geq} \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})} = \delta \end{aligned} \tag{17}$$

where $(a)$ holds due to the fact that [6],

$$\rho_{\sigma_i,\tilde{x}}(\mathbb{Z}) \leq \rho_{\sigma_i}(\mathbb{Z}) \triangleq \sum_{j\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}j^2}. \tag{18}$$

As it can be seen clearly, the right-hand side (RHS) of (17) is completely independent of $\mathbf{x}$, meaning it can be expressed by a constant $\delta$ determined by basis $\mathbf{B}$, query point $\mathbf{c}$ and standard deviation $\sigma$. Therefore, the proof is completed. □

We then arrive at a main Theorem to show the uniform ergodicity of the proposed algorithm.

**Theorem 2.** *Given the invariant lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain established by the independent MHK algorithm is uniformly ergodic:*

$$\|P^t(\mathbf{x},\cdot) - D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV} \leq (1-\delta)^t \tag{19}$$

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

*Proof.* Based on (13) and (14), the transition probability $P(\mathbf{x},\mathbf{y})$ of the independent MHK algorithm are given by

$$P(\mathbf{x},\mathbf{y}) = \begin{cases} \min\left\{q(\mathbf{y}), \frac{\pi(\mathbf{y})q(\mathbf{x})}{\pi(\mathbf{x})}\right\} & \text{if } \mathbf{y} \neq \mathbf{x}, \\ q(\mathbf{x}) + \sum\limits_{\mathbf{z}\neq\mathbf{x}} \max\left\{0, q(\mathbf{z}) - \frac{\pi(\mathbf{z})q(\mathbf{x})}{\pi(\mathbf{x})}\right\} & \text{if } \mathbf{y} = \mathbf{x}. \end{cases} \tag{20}$$

Using (16) in Lemma 1, it is straightforward to check that the following relationship holds

$$P(\mathbf{x},\mathbf{y}) \geq \delta\pi(\mathbf{y}) \tag{21}$$

for all cases of $\mathbf{x},\mathbf{y} \in \mathbb{Z}^n$, which indicates all the Markov transitions have a component of size $\delta$ in common. Based on (21), the coupling technique is applied to proceed the proof.

Specifically, assume two Markov chains $\mathbf{X}_n, \mathbf{Y}_n$ run independently, and both of them marginally update according to the transition probability $P(\mathbf{x},\mathbf{y})$. $\mathbf{Y}_n$ is supposed to start from the stationary distribution $\pi$, and $\mathbf{X}_n$ starts from some initial distribution $\pi_0$, which is not yet steady. According to (21), at each Markov move, $\mathbf{X}_n$ and $\mathbf{Y}_n$ have the probability at least $\delta$ to get coupled by choosing the same Markov state from the probability measure $\pi(\cdot)$. Because Markov state is only decided by its previous state, once they get coupled, these two Markov chains are naturally viewed to remain equal for all future moves.

**Algorithm 2** Independent Metropolis-Hastings-Klein 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
    let x denote the state of X_{t-1}
    generate y from the proposal distribution q(x, y) in (13)
    calculate the acceptance ratio α(x, y) in (14)
    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
```

Therefore, based on coupling technique, coupling inequality is used to bound the variation distance between two distributions formed by Markov chains [25], that is,

$$\|\mathcal{L}(\mathbf{X}_n) - \mathcal{L}(\mathbf{Y}_n)\|_{TV} \leq P(\mathbf{X}_n \neq \mathbf{Y}_n). \tag{22}$$

According to (22), the variation distance $\|\cdot\|_{TV}$ between $\mathcal{L}(\mathbf{X}_n)$ and $\mathcal{L}(\mathbf{Y}_n)$ is upper bounded by the probability that the Markov chains $\mathbf{X}_n$ and $\mathbf{Y}_n$ are not coupled. Since $\mathbf{X}_n$ and $\mathbf{Y}_n$ marginally follow the updating rules by $P(\mathbf{x},\mathbf{y})$ and $\mathbf{Y}_n$ reaches the stationary distribution $\pi(\cdot)$ from the beginning, (22) can be rewritten as

$$\|P^t(\mathbf{x},\cdot) - \pi(\cdot)\|_{TV} \leq P(\mathbf{X}_n \neq \mathbf{Y}_n). \tag{23}$$

Moreover, motivated by (21), in every single Markov move, we know that the probability of making $\mathbf{X_n}$ and $\mathbf{Y_n}$ get coupled is lower bounded by

$$P(\mathbf{X}_{n-1\to n} = \mathbf{Y}_{n-1\to n}) \geq \delta. \tag{24}$$

Therefore, during $n$ Markov moves, the probability of two independent Markov chains $\mathbf{X}$ and $\mathbf{Y}$ not coupled can be derived as

$$P(\mathbf{X}_n \neq \mathbf{Y}_n) = (1 - P(\mathbf{X} = \mathbf{Y}))^t \leq (1-\delta)^n. \tag{25}$$

Then, by substituting (25) into (23), we obtain

$$\|P^t(\mathbf{x},\cdot) - \pi(\cdot)\|_{TV} \leq (1-\delta)^t. \tag{26}$$

completing the proof. □

Obviously, given the value of $\delta < 1$, the mixing time of the Markov chain can be calculated by (6) and (26), that is

$$t_{\text{mix}}(\epsilon) = \frac{\ln\epsilon}{\ln(1-\delta)} < (-\ln\epsilon)\cdot\left(\frac{1}{\delta}\right), \quad \epsilon < 1 \tag{27}$$

where we use the bound $\ln(1-\delta) < -\delta$ for $0 < \delta < 1$. Therefore, the mixing time is proportional to $1/\delta$, and becomes

$O(1)$ if $\delta \to 1$.

### C. Convergence Rate in General Cases ($\overline{\sigma} \neq \sigma$)

In the proposed independent MHK algorithm, by default, the standard deviation of the proposal distribution $q$ is set the same as $\sigma$, namely, $\overline{\sigma} = \sigma$. Therefore, a natural question is whether a flexible standard deviation $\overline{\sigma} \neq \sigma$ still works. For this reason, in what follows, the relationship between $\overline{\sigma}$ and $\sigma$ is investigated.

Without loss of generality, by substitution, the corresponding ratio of $q(\mathbf{x})/\pi(\mathbf{x})$ in (17) can be rewritten as

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})} \geq \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^{n}\rho_{\overline{\sigma}_i}(\mathbb{Z})} \cdot e^{-(\frac{1}{2\overline{\sigma}^2}-\frac{1}{2\sigma^2})\|\mathbf{B}\mathbf{x}-\mathbf{c}\|^2}. \quad (28)$$

Unfortunately, in the case of $\overline{\sigma} < \sigma$, as $\|\mathbf{B}\mathbf{x}-\mathbf{c}\|$ could go to infinity, it is impossible to determine a constant lower bound upon $q(\mathbf{x})/\pi(\mathbf{x})$ for $\mathbf{x} \in \mathbb{Z}^n$, implying the uniform ergodicity can not be achieved. Furthermore, the Markov chains in this case are not even geometrically ergodic by not satisfying the drift condition shown in (74) [30], which is clearly explained in Section V. Therefore, for the sake of the relatively slow convergence performance, $\overline{\sigma} < \sigma$ is encouraged to be avoided in practice and the corresponding convergence analysis is omitted here.

On the other hand, in the case of $\overline{\sigma} > \sigma$, let $d(\Lambda, \mathbf{c})$ denote the Euclidean distance between the lattice $\Lambda$ and $\mathbf{c}$, that is

$$d(\Lambda, \mathbf{c}) = \min_{\mathbf{x}\in\mathbb{Z}^n} \|\mathbf{B}\mathbf{x}-\mathbf{c}\|, \quad (29)$$

then it follows that

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})} \geq \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^{n}\rho_{\overline{\sigma}_i}(\mathbb{Z})} \cdot e^{-(\frac{1}{2\overline{\sigma}^2}-\frac{1}{2\sigma^2})d(\Lambda,\mathbf{c})^2} \quad (30)$$

for all $\mathbf{x} \in \mathbb{Z}^n$, which means the underlying Markov chain is uniformly ergodic by satisfying (16) in Lemma 1. More precisely, $q(\mathbf{x})/\pi(\mathbf{x})$ could be expressed as

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})} \geq \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})} \cdot \beta \quad (31)$$

where

$$\beta = \frac{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})}{\prod_{i=1}^{n}\rho_{\overline{\sigma}_i}(\mathbb{Z})} \cdot e^{-(\frac{1}{2\overline{\sigma}^2}-\frac{1}{2\sigma^2})d(\Lambda,\mathbf{c})^2}. \quad (32)$$

Clearly, parameter $\beta$ becomes the key to govern the convergence performance. Compared to (17), if $\beta > 1$, the convergence of the Markov chain will be boosted by a larger size of $\delta$, otherwise the convergence will be slowed down. However, in the case of $\overline{\sigma} > \sigma$, it easy to check that the value of $\beta$ is monotonically decreasing with the given $\sigma$, rendering $\beta > 1$ inapplicable to the most cases of interest.

As it can be seen clearly from Fig. 2, the convergence rate can be enhanced by $\beta > 1$ only for a small enough $\sigma$ (e.g., $\sigma^2 < 0.398$, e.g., $-4$ dB), thus making the choice of $\overline{\sigma} = \sigma$ (i.e., $\beta = 1$) reasonable to maintain the convergence performance. This essentially explains the reason why the independent MHK algorithm is proposed with $\overline{\sigma} = \sigma$ as a default configuration in general.

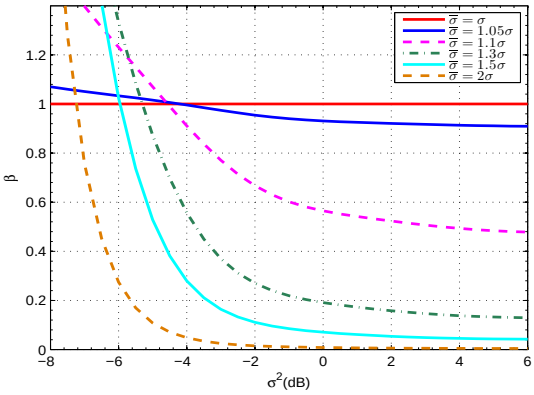


Fig. 2. Coefficient $\beta$ of $E_8$ lattice in the case of $\overline{\sigma} > \sigma$ when $\mathbf{c} = \mathbf{0}$.

## IV. Convergence Analysis

In this section, convergence analysis about the exponential decay coefficient $\delta$ in the independent MHK algorithm is performed, which leads to a quantitative estimate of the mixing time. For a better understanding, the analysis is carried out in two cases respectively, namely, $\mathbf{c} = \mathbf{0}$ and $\mathbf{c} \neq \mathbf{0}$.

### A. Convergence Rate ($\mathbf{c} = \mathbf{0}$)

Lemma 1 shows that the ratio $q(\mathbf{x})/\pi(\mathbf{x})$ in the independent MHK sampling algorithm is lower bounded by a constant $\delta$. We further derive an explicit expression of coefficient $\delta$ due to its significant impact on the convergence rate, for the case $\mathbf{c} = \mathbf{0}$.

Specifically, we have

$$\begin{aligned}\frac{q(\mathbf{x})}{\pi(\mathbf{x})} &= \frac{\rho_{\sigma,\mathbf{0}}(\mathbf{\Lambda})}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z})} \\ &\overset{(b)}{\geq} \frac{\sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{B}\mathbf{x}\|^2}}{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})} \\ &\overset{(c)}{=} \frac{\Theta_\Lambda(\frac{1}{2\pi\sigma^2})}{\prod_{i=1}^{n}\Theta_{\mathbb{Z}}(\frac{1}{2\pi\sigma_i^2})} \\ &\overset{(d)}{=} \frac{\Theta_\Lambda(\frac{1}{s^2})}{\prod_{i=1}^{n}\vartheta_3(\frac{1}{s_i^2})} = \delta. \end{aligned} \quad (33)$$

Here, for notational simplicity, $s = \sqrt{2\pi}\sigma$ and $s_i = \sqrt{2\pi}\sigma_i = s/\|\widehat{\mathbf{b}}_i\|$ are applied in the equations. In $(b)$, the inequality $\rho_{\sigma_i,\tilde{x}}(\mathbb{Z}) \leq \rho_{\sigma_i}(\mathbb{Z})$ shown in (18) is used again. Theta series $\Theta_\Lambda$ and Jacobi theta function $\vartheta_3$ are applied in $(c)$ and $(d)$ respectively, where

$$\Theta_\Lambda(\tau) = \sum_{\lambda\in\Lambda} e^{-\pi\tau\|\lambda\|^2}, \quad (34)$$

$$\vartheta_3(\tau) = \sum_{n=-\infty}^{+\infty} e^{-\pi\tau n^2} \quad (35)$$

with $\Theta_{\mathbb{Z}} = \vartheta_3$ [31].

**Proposition 1.** *If* $s \geq \sqrt{2\pi\cdot\omega(\log n)}\cdot\max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$ *or* $s \leq [\sqrt{2\pi\cdot\omega(\log n)}]^{-1}\cdot\min_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$, *then the coefficient* $\delta \approx 1$.

*Proof.* To start with, let us recall the *flatness factor* [5], which is defined as

$$\epsilon_\Lambda(\sigma) = \frac{\det(\mathbf{B})}{(\sqrt{2\pi}\sigma)^n}\Theta_\Lambda(\frac{1}{2\pi\sigma^2}) - 1 \tag{36}$$

and

$$\epsilon_\Lambda(\sigma) = \varepsilon, \quad \text{if} \quad \sigma = \eta_\varepsilon(\Lambda). \tag{37}$$

Here, $\eta_\varepsilon(\Lambda)$ is known as the *smoothing parameter* in lattices, and if $\eta_\varepsilon(\Lambda) \leq \sqrt{\omega(\log n)} \cdot \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$, then $\varepsilon$ will become negligible (Lemma 3.3, [6]).

Therefore, the exponential decay coefficient $\delta$ given in (33) can be expressed as

$$\begin{aligned}\delta &= \frac{\Theta_\Lambda(\frac{1}{2\pi\sigma^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{2\pi\sigma_i^2})} \\ &= \frac{\det(\mathbf{B})^{-1}\cdot(\sqrt{2\pi}\sigma)^n \cdot [\epsilon_\Lambda(\sigma)+1]}{\prod_{i=1}^n \sqrt{2\pi}\sigma_i \cdot [\epsilon_\mathbb{Z}(\sigma_i)+1]} \\ &= \frac{\epsilon_\Lambda(\sigma)+1}{\prod_{i=1}^n[\epsilon_\mathbb{Z}(\sigma_i)+1]},\end{aligned} \tag{38}$$

where $\det(\cdot)$ denotes the determinant of a matrix.

From (36), it is easy to verify that the flatness factor $\epsilon_\Lambda$ is a monotonically decreasing function of $\sigma$, i.e., for $\sigma_1 \geq \sigma_2$, we have $\epsilon_\Lambda(\sigma_1) \leq \epsilon_\Lambda(\sigma_2)$. Therefore, from (37), let $\sigma > \eta_\varepsilon(\Lambda)$, then the flatness factor $\epsilon_\Lambda(\sigma)$ will be upper bounded as

$$\epsilon_\Lambda(\sigma) < \varepsilon. \tag{39}$$

Furthermore, assume $\sigma > \sqrt{\omega(\log n)} \cdot \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$, then $\epsilon_\Lambda(\sigma)$ will approach 0 since its upper bound $\varepsilon$ becomes negligible. Meanwhile, it is easy to check that the same thing also happens to $\epsilon_\mathbb{Z}(\sigma_i)$ for $\sigma > \sqrt{\omega(\log n)} \cdot \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$. Hence, we have

$$\delta = \frac{\epsilon_\Lambda(\sigma)+1}{\prod_{i=1}^n[\epsilon_\mathbb{Z}(\sigma_i)+1]} \approx 1 \tag{40}$$

if $\sigma > \sqrt{\omega(\log n)} \cdot \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$.

On the other hand, according to the *Jacobi's formula* [32]

$$\Theta_\Lambda(\tau) = |\det(\mathbf{B})|^{-1}\left(\frac{1}{\tau}\right)^{\frac{n}{2}}\Theta_{\Lambda^*}\left(\frac{1}{\tau}\right), \tag{41}$$

the expression of the flatness factor shown in (36) can be rewritten as

$$\epsilon_\Lambda(\sigma) = \Theta_{\Lambda^*}(2\pi\sigma^2) - 1, \tag{42}$$

where $\Lambda^*$ is the dual lattice of $\Lambda$. Then, we have

$$\begin{aligned}\delta &= \frac{\Theta_\Lambda(\frac{1}{2\pi\sigma^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{2\pi\sigma_i^2})} \\ &= \frac{\epsilon_{\Lambda^*}(\frac{1}{2\pi\sigma})+1}{\prod_{i=1}^n[\epsilon_{\mathbb{Z}^*}(\frac{1}{2\pi\sigma_i})+1]},\end{aligned} \tag{43}$$

where $\vartheta_3^* = \Theta_{\mathbb{Z}^*}$.

With respect to $\epsilon_{\Lambda^*}(\frac{1}{2\pi\sigma})$ and $\epsilon_{\Lambda^*}(\frac{1}{2\pi\sigma_i})$ in (43), similarly, it follows that if

$$\frac{1}{2\pi\sigma} \geq \sqrt{\omega(\log n)} \cdot \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i^*\|, \tag{44}$$

then both $\epsilon_{\Lambda^*}(\frac{1}{2\pi\sigma})$ and $\epsilon_{\Lambda^*}(\frac{1}{2\pi\sigma_i})$ will become negligible, and we have

$$\delta \approx 1, \tag{45}$$

where $\widehat{\mathbf{b}}_i^*$'s are the Gram-Schmidt vectors of the dual lattice basis $\mathbf{B}^*$.

According to (44), it follows that

$$\begin{aligned}\sigma &\leq \frac{1}{2\pi}\cdot\sqrt{\omega(\log n)}^{-1}\cdot(\max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i^*\|)^{-1} \\ &\overset{(e)}{=} \frac{1}{2\pi}\cdot\sqrt{\omega(\log n)}^{-1}\cdot[\max_{1\leq i\leq n}(\|\widehat{\mathbf{b}}_{n-i+1}\|^{-1})]^{-1} \\ &= \frac{1}{2\pi}\cdot\sqrt{\omega(\log n)}^{-1}\cdot[(\min_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|)^{-1}]^{-1} \\ &= [2\pi\cdot\sqrt{\omega(\log n)}]^{-1}\cdot\min_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|,\end{aligned} \tag{46}$$

where $(e)$ comes from the fact that [33]

$$\|\widehat{\mathbf{b}}_i^*\| = \|\widehat{\mathbf{b}}_{n-i+1}\|^{-1}. \tag{47}$$

Therefore, by recalling $s = \sqrt{2\pi}\sigma$, the proof is completed. □

Obviously, according to Proposition 1, as $s$ goes to 0 and $\infty$ on both sides, the coefficient $\delta$ will converges to 1.

**Proposition 2.** *If $s \leq \min_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$, then the coefficient $\delta$ is lower bounded by*

$$\delta \geq 1.086^{-n}\cdot\Theta_\Lambda(\frac{1}{s^2}). \tag{48}$$

*Meanwhile, if $s \geq \max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$, then the coefficient $\delta$ is lower bounded by*

$$\delta \geq 1.086^{-n}\cdot\Theta_{\Lambda^*}(s^2). \tag{49}$$

*Proof.* By definition, we have

$$\vartheta_3(1) = \sum_{n=-\infty}^{+\infty} e^{-\pi n^2} = \frac{\sqrt[4]{\pi}}{\Gamma(\frac{3}{4})} \simeq 1.086^{[1]}, \tag{50}$$

where $\Gamma(\cdot)$ stands for the Gamma function. As the Jacobi theta function $\vartheta_3(\tau)$ is monotonically decreasing with $\tau$, let $1/s_i^2 \geq 1$, namely, $s \leq \|\widehat{\mathbf{b}}_i\|$, then it follows that

$$\vartheta_3(\frac{1}{s_i^2}) \leq \vartheta_3(1) \leq 1.086. \tag{51}$$

Assume $s \leq \min_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$, then the following lower bound for $\delta$ can be obtained,

$$\delta = \frac{\Theta_\Lambda(\frac{1}{s^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{s_i^2})} \geq 1.086^{-n}\cdot\Theta_\Lambda(\frac{1}{s^2}). \tag{52}$$

On the other hand, as $\mathbb{Z}$ is a self-dual lattice, i.e., $\mathbb{Z} = \mathbb{Z}^*$, then if $s_i^2 \geq 1$, namely, $s \geq \|\widehat{\mathbf{b}}_i\|$, it follows that

$$\vartheta_3^*(s_i^2) = \vartheta_3(s_i^2) \leq \vartheta_3(1) \leq 1.086. \tag{53}$$

Therefore, let $s \geq \max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$, according to the Jacobi's

[1]It is worth pointing out that the explict values of $\vartheta_3(2)$, $\vartheta_3(3)$, ...can also be calculated, where the same derivation in the following can also be carried out. Here we choose $\vartheta_3(1)$ as the benchmark due to its simplicity.

TABLE I
VALUE OF $\delta$ WITH RESPECT TO THE GIVEN $s = \sqrt{2\pi}\sigma$ IN THE INDEPENDENT MHK ALGORITHM.

| | |
|---|---|
| $s \leq [\sqrt{2\pi\omega(\log n)}]^{-1} \cdot \min_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$ | $\delta \approx 1$ |
| $s \leq \min_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$ | $\delta \geq 1.086^{-n} \cdot \Theta_\Lambda(\frac{1}{s^2})$ |
| $\min_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\| \leq s \leq \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$ | $\delta \geq 1.086^{-(n-m)} \cdot 2^{-m} \cdot \frac{\prod_{i\in I} \|\widehat{\mathbf{b}}_i\|}{s^m} \cdot \Theta_\Lambda(\frac{1}{s^2})$ |
| $s \geq \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$ | $\delta \geq 1.086^{-n} \cdot \Theta_{\Lambda^*}(s^2)$ |
| $s \geq \sqrt{2\pi\omega(\log n)} \cdot \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$ | $\delta \approx 1$ |

formula shown in (41), $\delta$ can be lower bounded as

$$\begin{aligned} \delta &= \frac{\Theta_\Lambda(\frac{1}{s^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{s_i^2})} \\ &= \frac{|\det(\mathbf{B})|^{-1}(s^2)^{\frac{n}{2}}\Theta_{\Lambda^*}(s^2)}{\prod_{i=1}^n (s_i^2)^{\frac{n}{2}}\vartheta_3^*(s_i^2)} \\ &= \frac{\Theta_{\Lambda^*}(s^2)}{\prod_{i=1}^n \vartheta_3^*(s_i^2)} \\ &\geq 1.086^{-n} \cdot \Theta_{\Lambda^*}(s^2), \end{aligned} \tag{54}$$

completing the proof. □

Next, with respect to the range of $\min_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\| \leq s \leq \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$, we arrive at the following proposition.

**Proposition 3.** *If* $\min_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\| \leq s \leq \max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$, *then the coefficient* $\delta$ *is lower bounded by*

$$\delta \geq 1.086^{-(n-m)} \cdot 2^{-m} \cdot \frac{\prod_{i\in I} \|\widehat{\mathbf{b}}_i\|}{s^m} \cdot \Theta_\Lambda(\frac{1}{s^2}), \tag{55}$$

*where* $m$ *denotes the number of terms* $\vartheta_3(\frac{1}{s_i^2})$ *with* $s_i > 1$ *(i.e.,* $s > \|\widehat{\mathbf{b}}_i\|$*),* $i \in \{1, 2, \ldots, n\}$ *and set* $I \subseteq [1, \ldots, n]$, $|I| = m$.

*Proof.* From the definition, we have

$$\begin{aligned} \vartheta_3(\tau) &= \sum_{n=-\infty}^{+\infty} e^{-\pi\tau n^2} \\ &= 1 + 2\sum_{n\geq 1} e^{-\pi\tau n^2} \\ &\leq 1 + 2\int_0^\infty e^{-\pi\tau x^2} dx \\ &\overset{(f)}{=} 1 + \sqrt{\frac{1}{\tau}}, \end{aligned} \tag{56}$$

where $(f)$ holds due to the *Gaussian integral* $\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}$.

Hence, for terms $\vartheta_3(\frac{1}{s_i^2})$ with $1/s_i^2 \leq 1$, namely, $s \geq \|\widehat{\mathbf{b}}_i\|$, we have

$$\vartheta_3(\frac{1}{s_i^2}) \leq 1 + |s_i| \leq 2s_i = 2\frac{s}{\|\widehat{\mathbf{b}}_i\|}. \tag{57}$$

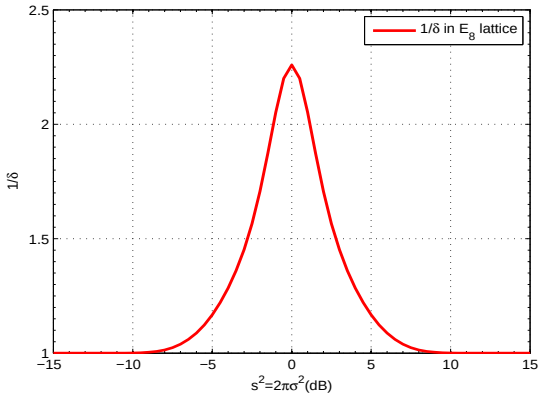


Fig. 3. Exponential decay coefficient $1/\delta$ of the $E_8$ lattice in the case of $\mathbf{c} = \mathbf{0}$.

Therefore, from (51) and (57), if follows that

$$\prod_{i=1}^n \vartheta_3(\frac{1}{s_i^2}) \leq 1.086^{(n-m)} \cdot 2^m \cdot \frac{s^m}{\prod_{i\in I} \|\widehat{\mathbf{b}}_i\|}, \tag{58}$$

completing the proof. □

*Remark:* We emphasize that the significance of lattice reduction (e.g., LLL or HKZ) can be seen here, as increasing $\min_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$ and decreasing $\max_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$ simultaneously will greatly enhance the convergence performance due to a better lower bound of $\delta$.

Now, let us consider some lattices whose theta series are more understood.

**Proposition 4.** *The coefficient* $\delta = \frac{\Theta_\Lambda(\frac{1}{s^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{s_i^2})}$ *for an isodual lattice* $\Lambda$ *has a multiplicative symmetry point at* $s = 1$*, and asymptotically converges to 1 on both sides when* $s$ *goes to* $0$ *and* $\infty$.

*Proof.* An isodual lattice is one that is geometrically similar to its dual. Here, we note that the theta series $\Theta_\Lambda$ of an isodual lattice $\Lambda$ and that of its dual $\Lambda^*$ are the same, i.e., $\Theta_\Lambda(\tau) = \Theta_{\Lambda^*}(\tau)$, and the volume of an isodual lattice $|\det(\mathbf{B})|$ naturally equals 1. Therefore, we have

$$\Theta_\Lambda\left(\frac{1}{s^2}\right) = s^n\Theta_\Lambda(s^2), \tag{59}$$

$$\vartheta_3\left(\frac{1}{s_i^2}\right) = s_i\vartheta_3(s_i^2), \tag{60}$$

then from (59) and (60), the symmetry with respect to $s = 1$ can be obtained as follows,

$$\begin{aligned} \frac{\Theta_\Lambda(\frac{1}{s^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{s_i^2})} &= \frac{s^n\Theta_\Lambda(s^2)}{\prod_{i=1}^n s_i\vartheta_3(s_i^2)} \\ &= \frac{\Theta_\Lambda(s^2)}{\prod_{i=1}^n \frac{1}{\|\widehat{\mathbf{b}}_i\|}\vartheta_3(s_i^2)} \\ &= \frac{\Theta_\Lambda(s^2)}{\frac{1}{|\det(\mathbf{B})|} \cdot \prod_{i=1}^n \vartheta_3(s_i^2)} \\ &= \frac{\Theta_\Lambda(s^2)}{\prod_{i=1}^n \vartheta_3(s_i^2)}. \end{aligned} \tag{61}$$

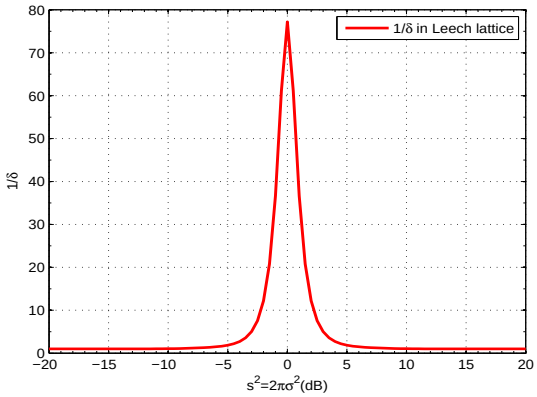


Fig. 4. Exponential decay coefficient $1/\delta$ of the Leech lattice in the case of $\mathbf{c}=\mathbf{0}$.

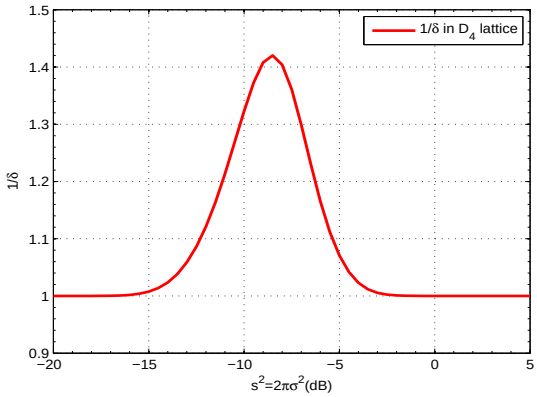


Fig. 5. Exponential decay coefficient $1/\delta$ of the $D_4$ lattice in the case of $\mathbf{c}=\mathbf{0}$.

By definition, it is straightforward to verify that

$$\frac{\Theta_\Lambda(\frac{1}{s^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{s_i^2})} \to 1, \text{ when } s \to 0. \tag{62}$$

Then because of the symmetry, $\frac{\Theta_\Lambda(\frac{1}{s^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{s_i^2})}$ will also asymptotically approach 1 when $s \to \infty$, completing the proof. □

Examples of the coefficient $1/\delta$ for the isodual $E_8$ and Leech lattice are shown in Fig. 3 and Fig. 4, respectively. It is worth pointing out that $1/\delta$ has a maximum at the symmetry point $s=1$, namely $\sigma^2=\frac{1}{2\pi}$. Actually, $1/\delta$ is similar to, but not exactly the same as the *secrecy gain* defined in [32]. In our context, $1/\delta$ roughly estimates the number of the Markov moves required to reach the stationary distribution. On the other hand, as for non-isodual lattices, $D_4$ lattice is applied to give the illustration in Fig. 5, where the symmetry still holds but centers at $s=0.376$. Therefore, with the exact value of $\delta$, the explicit estimation of mixing time for the underlying Markov chain can be obtained.

### *B. Convergence Rate ($\mathbf{c}\neq\mathbf{0}$)*

As for the convergence analysis in the case of $\mathbf{c}\neq\mathbf{0}$, we firstly define the exponential decay coefficient $\delta'$ as

$$\delta' = \frac{q(\mathbf{x})}{\pi(\mathbf{x})} = \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^n \rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z})}, \tag{63}$$

then we have the following proposition.

**Proposition 5.** *For any $\mathbf{c}\in\mathbb{R}^n$ and $\mathbf{c}\neq\mathbf{0}$, one has*

$$\delta' \geq e^{-\frac{1}{2\sigma^2}\|\mathbf{c}\|^2}\cdot\delta. \tag{64}$$

*Proof.* Specifically, we have

$$\begin{aligned}
\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda}) &= \sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2} \\
&= \frac{1}{2}\sum_{\mathbf{x}\in\mathbb{Z}^n}\left(e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}+e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}+\mathbf{c}\|^2}\right) \\
&= e^{-\frac{1}{2\sigma^2}\|\mathbf{c}\|^2}\sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}\|^2}\cdot\frac{1}{2}\cdot\left(e^{-\frac{1}{\sigma^2}\langle\mathbf{Bx},\mathbf{y}\rangle}+e^{\frac{1}{\sigma^2}\langle\mathbf{Bx},\mathbf{y}\rangle}\right) \\
&\overset{(i)}{\geq} e^{-\frac{1}{2\sigma^2}\|\mathbf{c}\|^2}\sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}\|^2} \\
&= e^{-\frac{1}{2\sigma^2}\|\mathbf{c}\|^2}\cdot\rho_\sigma(\mathbf{\Lambda}),
\end{aligned} \tag{65}$$

where $(i)$ follows from the fact that for any positive real $a>0$, $a+1/a\geq 2$. □

Thus, the value of $\delta'$ can be reduced by a factor of $e^{-\frac{1}{2\sigma^2}\|\mathbf{c}\|^2}$ from $\delta$. Clearly, if $\mathbf{c}=\mathbf{0}$, then $\delta'=\delta$, implying $\mathbf{c}\neq\mathbf{0}$ is a general case of $\mathbf{c}=\mathbf{0}$. However, here comes a problem: if $\mathbf{c}$ is far away from the original point, then the lower bound given in (64) will be mildly small. To this end, in such cases, the Euclidean distance between $\Lambda$ and $\mathbf{c}$, namely, $d(\Lambda,\mathbf{c})$, can be applied for the convergence rate assessment.

**Proposition 6** ([34])**.** *For any $n$-dimensional lattice $\Lambda,\mathbf{c}\in\mathbb{R}^n$, if $d(\Lambda,\mathbf{c})\leq k\sqrt{\log n}$ and $k>0$, one has*

$$\frac{\sum_{\mathbf{v}\in\Lambda} e^{-\pi\|\mathbf{v}-\mathbf{c}\|^2}}{\sum_{\mathbf{v}\in\Lambda} e^{-\pi\|\mathbf{v}\|^2}} > n^{-10k^2}. \tag{66}$$

According to Proposition 6, we have

$$\begin{aligned}
\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda}) &= \sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\pi\|\frac{1}{\sqrt{2\pi}\sigma}(\mathbf{Bx}-\mathbf{c})\|^2} \\
&> n^{-10k^2}\cdot\sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\pi\|\frac{1}{\sqrt{2\pi}\sigma}\mathbf{Bx}\|^2} \\
&= n^{-10k^2}\cdot\rho_\sigma(\mathbf{\Lambda})
\end{aligned} \tag{67}$$

if $d(\mathbf{Bx},\mathbf{c})\leq k\sqrt{2\pi\log n}\sigma$. Then, by substitution, the exponential decay coefficient $\delta'$ in the case $\mathbf{c}\neq\mathbf{0}$ can be further derived as

$$\begin{aligned}
\delta' &= \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^n \rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z})} \\
&\geq n^{-10k^2}\cdot\frac{\rho_\sigma(\mathbf{\Lambda})}{\prod_{i=1}^n \rho_{\sigma_i}(\mathbb{Z})} \\
&= n^{-10k^2}\cdot\delta.
\end{aligned} \tag{68}$$

## V. Symmetric Metropolis-Klein Algorithm

In this section, we propose the symmetrical Metropolis-Klein (SMK) algorithm for lattice Gaussian sampling. The

underlying Markov chain is proved to be geometrically ergodic, which not only converges exponentially fast, but also depends on the selection of the initial state.

### A. Symmetric Metropolis-Klein Algorithm

The Metropolis algorithm can be viewed as a special case of the MH algorithm by utilizing a symmetric proposal distribution $q(\mathbf{x},\mathbf{y})$ [29]. In the proposed algorithm, we use Klein's algorithm to generate the symmetric proposal distribution. By doing this, the generation of the state candidate $\mathbf{y}$ highly depends on the previous state $\mathbf{x}$, thus leading to a more effective sampling compared to the independent one. Moreover, since the chain is symmetric, the calculation of the acceptance ratio $\alpha$ is also greatly simplified.

Specifically, as shown in Algorithm 3, its sampling procedure at each Markov move can be summarized by the following steps:

1) *Given the current Markov state $\mathbf{X}_t = \mathbf{x}$, sample from the symmetric proposal distribution through Klein's algorithm to obtain the candidate state $\mathbf{y}$ for $\mathbf{X}_{t+1}$,*

$$q(\mathbf{x},\mathbf{y}) = \frac{\rho_{\sigma,\mathbf{Bx}}(\mathbf{By})}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{By}\|^2}}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} \overset{(j)}{=} q(\mathbf{y},\mathbf{x}), \quad (69)$$

where $\widetilde{y}_i = \frac{c'_i - \sum_{j=i+1}^{n} r_{i,j}y_j}{r_{i,i}}$, $\mathbf{c}' = \mathbf{Q}^{\dagger}\mathbf{Bx}$ and $\mathbf{B} = \mathbf{QR}$. Note that equality $(j)$ holds due to the symmetry stated in Lemma 2.

2) *Calculate the acceptance ratio $\alpha(\mathbf{x},\mathbf{y})$*

$$\begin{aligned}\alpha &= \min\left\{1, \frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}\right\} = \min\left\{1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right\} \\ &= \min\left\{1, e^{\frac{1}{2\sigma^2}\left(\|\mathbf{Bx}-\mathbf{c}\|^2 - \|\mathbf{By}-\mathbf{c}\|^2\right)}\right\}, \quad (70)\end{aligned}$$

where $\pi = D_{\Lambda,\sigma,\mathbf{c}}$.

3) *Make a decision for $\mathbf{X}_{t+1}$ based on $\alpha(\mathbf{x},\mathbf{y})$ to accept $\mathbf{X}_{t+1} = \mathbf{y}$ or not.*

**Lemma 2.** *The proposal distribution $q$ shown in (69) is symmetric, namely,*

$$q(\mathbf{x},\mathbf{y}) = q(\mathbf{y},\mathbf{x}) \quad (71)$$

*for all $\mathbf{x},\mathbf{y} \in \mathbb{Z}^n$.*

The proof of Lemma 2 is provided in Appendix B.

Intuitively, at each Markov move, the state candidate $\mathbf{y}$ for $\mathbf{X}_{t+1}$ is sampled based on a Gaussian-like distribution centered at the point made up by the previous Markov state $\mathbf{x}$. Given the acceptance ratio shown in (70), it is quite straightforward to see that if $\mathbf{By}$ is closer to the given point $\mathbf{c}$ than $\mathbf{Bx}$, then state candidate $\mathbf{y}$ must be accepted by $\mathbf{X}^{t+1}$ due to $\alpha = 1$, otherwise it will be accepted with a probability depending on the distance from $\mathbf{By}$ to $\mathbf{c}$, thus forming a Markov chain [2].

[2]A query about the SMK algorithm is whether a flexible standard deviation $\overline{\sigma}$ in the proposal distribution $q$ works, i.e., $\overline{\sigma} \neq \sigma$. The answer is yes. However, since the explicit convergence rate is difficult to get, we omit the corresponding analysis here.

**Algorithm 3** Symmetric Metropolis-Klein 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
    let x denote the state of X_{t-1}
    generate y by the proposal distribution q(x, y) in (69)
    calculate the acceptance ratio α(x, y) in (70)
    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
```

**Theorem 3.** *Given the target lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by the proposed symmetric Metropolis-Klein algorithm is ergodic:*

$$\lim_{t\to\infty} \|P^t(\mathbf{x};\cdot) - D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV} = 0 \quad (72)$$

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

The proof of the Theorem 3 is similar to that of the Theorem 1, which is omitted here.

### B. Geometric Ergodicity

In MCMC, a set $C \subseteq \Omega$ is referred to as a *small set*, if there exist $k > 0$, $1 > \delta > 0$ and a probability measure $v$ on $\Omega$ such that

$$P^k(\mathbf{x},\cdot) \geq \delta v(\cdot) \quad (73)$$

for all $\mathbf{x} \in C$ [35]. This is also known as the *minorisation condition* in literature. Actually, the uniform ergodicity can be achieved as a special case of the minorisation condition with $C = \Omega$. For a bounded small set $C$, the drift condition of Markov chains is defined as follows [25]:

**Definition 4.** *A Markov chain satisfies the drift condition if there are constants $0 < \lambda < 1$ and $b < \infty$, and a function $V : \Omega \to [1,\infty]$, such that*

$$\int_{\Omega} P(\mathbf{x},d\mathbf{y})V(\mathbf{y}) \leq \lambda V(\mathbf{x}) + b\mathbf{1}_C(\mathbf{x}) \quad (74)$$

*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.*

In principle, the *drift condition* is the standard way to prove geometric ergodicity [36]. Here, we recall the following theorem to show the relationship between drift condition and geometric ergodicity.

**Theorem 4** ([25])**.** *Consider an irreducible, aperiodic Markov chain with stationary distribution $\pi$, if the drift condition shown in (74) is satisfied for any small set $C \subseteq \Omega$ with $\delta > 0$, then the chain is geometrically ergodic.*

Here, we highlight that the discrete version of drift condition still holds for discrete state space Markov chains [37]. According to Theorem 4, now we check whether the drift condition is satisfied in the proposed algorithm. Then, we arrive at the following Theorem.

**Theorem 5.** *Given the invariant lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain established by the symmetric Metropolis-Klein algorithm satisfies the drift condition. Therefore, the proposed symmetric Metropolis-Klein algorithm is geometrically ergodic.*

*Proof.* By definition, for any $\|\mathbf{Bx}-\mathbf{By}\| \leq K$, where $K>0$ is a constant, the proposal distribution $q(\mathbf{x},\mathbf{y})$ can always be lower bounded by a constant $\epsilon>0$ as follows,

$$\begin{aligned} q(\mathbf{x},\mathbf{y}) &\geq \frac{e^{-\frac{K^2}{2\sigma^2}}}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} \\ &\overset{(k)}{\geq} \frac{e^{-\frac{K^2}{2\sigma^2}}}{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})} = \epsilon, \end{aligned} \tag{75}$$

where $(k)$ holds due to (18).

Since $q(\mathbf{x},\mathbf{y})$ can always be lower bounded and $\pi(\mathbf{x})$ is bounded away from 0 and 1 on any bounded sets, every non-empty bounded set $C \subseteq \mathbb{Z}^n$ is a small set:

$$P(\mathbf{x},\mathbf{y}) = q(\mathbf{x},\mathbf{y})\cdot\alpha(\mathbf{x},\mathbf{y}) \geq \epsilon\cdot\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})} \geq \epsilon\pi(\mathbf{y}) \tag{76}$$

for all $\mathbf{x}\in C$. Next, in order to specify the small set $C$, we define

$$C = \{\mathbf{x}\in\mathbb{Z}^n : \pi(\mathbf{x}) \geq \frac{1}{d^2}\}, \tag{77}$$

where $d>1$ is a constant.

At each Markov move, given the acceptance ratio $\alpha$ shown in (70), the acceptance region $A_{\mathbf{x}}$ and the potential rejection region $R_{\mathbf{x}}$ for the chain started from $\mathbf{x}$ are defined as

$$A_{\mathbf{x}} = \{\mathbf{y}\in\mathbb{Z}^n | \pi(\mathbf{y}) \geq \pi(\mathbf{x})\}; \tag{78}$$

$$R_{\mathbf{x}} = \{\mathbf{y}\in\mathbb{Z}^n | \pi(\mathbf{y}) < \pi(\mathbf{x})\}. \tag{79}$$

Apparently, state candidate $\mathbf{y}\in A_{\mathbf{x}}$ will be accepted by $\mathbf{X}_{t+1}$ without uncertainty while state candidate $\mathbf{y}\in R_{\mathbf{x}}$ has a certain risk to be rejected. Then, based on $A_{\mathbf{x}}$ and $R_{\mathbf{x}}$, the discrete term $\sum_{\mathbf{y}\in\mathbb{Z}^n}P(\mathbf{x},\mathbf{y})V(\mathbf{y})$ on the LHS of (74) can be expressed as (80). Next, let $V(\mathbf{x}) = \pi(\mathbf{x})^{-\frac{1}{2}}$. We will discuss the two cases $\mathbf{x}\notin C$ and $\mathbf{x}\in C$, respectively.

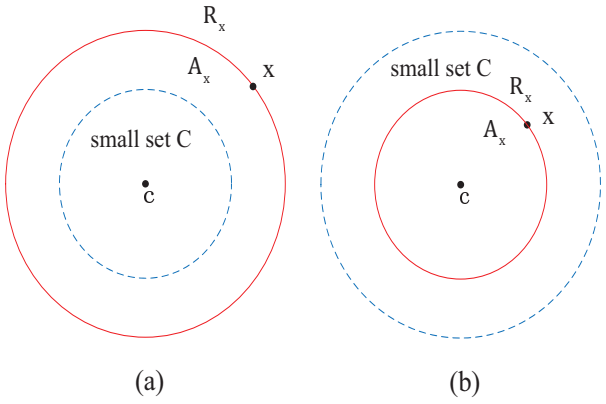


Fig. 6. Illustration of cases (a) $\mathbf{x}\notin C$ and (b) $\mathbf{x}\in C$ in the Markov move induced by SMK. The blue dash curve represents the area of small set while the red solid curve denotes the acceptance region $A_x$.

(i). In the case of $\mathbf{x}\notin C$, i.e., $V(\mathbf{x})>d$, we have the following derivation shown in (81), where $\pi'(\mathbf{x}) = e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}$. It is easy to verify that

$$\lim_{\|\mathbf{x}\|\to\infty} l(\mathbf{x})\cdot\nabla\log\pi'(\mathbf{x}) = -\infty, \tag{82}$$

where $l(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ denotes the unit vector and $\nabla$ represents the gradient. This condition implies that for any arbitrarily large $\gamma>0$, there exists $R>0$ such that

$$\frac{\pi'(\mathbf{x}+a\cdot l(\mathbf{x}))}{\pi'(\mathbf{x})} \leq e^{-a\cdot\gamma}, \tag{83}$$

where $\|\mathbf{x}\|\geq R, a\geq 0$. In other words, as $\|\mathbf{x}\|$ goes to infinity, $\pi'$ is at least exponentially decaying with a rate $\gamma$ tending to infinity. Therefore, once $\|\mathbf{x}\|$ is large enough, e.g., $\|\mathbf{x}\|\to\infty$, even a minimum discrete integer increment, namely, $\Delta\in\mathbb{Z}^n$ and $\|\Delta\|=1$, can make $\pi'(\mathbf{x}+\Delta)$ become extremely smaller than $\pi'(\mathbf{x})$.

---

$$\begin{aligned} \sum_{\mathbf{y}\in\mathbb{Z}^n}P(\mathbf{x},\mathbf{y})V(\mathbf{y}) &= \sum_{\mathbf{y}\in A_{\mathbf{x}}}P(\mathbf{x},\mathbf{y})V(\mathbf{y}) + \sum_{\mathbf{y}\in R_{\mathbf{x}}}P(\mathbf{x},\mathbf{y})V(\mathbf{y}) \\ &= \sum_{\mathbf{y}\in A_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})V(\mathbf{y}) + \sum_{\mathbf{y}\in R_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}V(\mathbf{y}) + \sum_{\mathbf{y}\in R_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\left[1-\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right]V(\mathbf{x}). \end{aligned} \tag{80}$$

$$\begin{aligned} \frac{\sum_{\mathbf{y}\in\mathbb{Z}^n}P(\mathbf{x},\mathbf{y})V(\mathbf{y})}{V(\mathbf{x})} &= \sum_{\mathbf{y}\in A_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\frac{V(\mathbf{y})}{V(\mathbf{x})} + \sum_{\mathbf{y}\in R_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\cdot\frac{V(\mathbf{y})}{V(\mathbf{x})} + \sum_{\mathbf{y}\in R_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\left[1-\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right] \\ &= \sum_{\mathbf{y}\in A_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\left[\frac{V(\mathbf{y})}{V(\mathbf{x})}-1\right] + \sum_{\mathbf{y}\in A_{\mathbf{x}}}q(\mathbf{x},\mathbf{y}) + \sum_{\mathbf{y}\in R_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\cdot\frac{V(\mathbf{y})}{V(\mathbf{x})} + \sum_{\mathbf{y}\in R_{\mathbf{x}}}q(\mathbf{x},\mathbf{y}) - \sum_{\mathbf{y}\in R_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})} \\ &= 1 - \sum_{\mathbf{y}\in A_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\left[1-\frac{\pi'(\mathbf{x})^{\frac{1}{2}}}{\pi'(\mathbf{y})^{\frac{1}{2}}}\right] + \sum_{\mathbf{y}\in R_{\mathbf{x}}}q(\mathbf{x},\mathbf{y})\left[\frac{\pi'(\mathbf{y})^{\frac{1}{2}}}{\pi'(\mathbf{x})^{\frac{1}{2}}} - \frac{\pi'(\mathbf{y})}{\pi'(\mathbf{x})}\right]. \end{aligned} \tag{81}$$

Now, let $\mathbf{y}_1 = \mathbf{x} + \Delta \in R_x$, with large enough $\|\mathbf{x}\|$, the ratio of $\pi'(\mathbf{y_1})/\pi'(\mathbf{x})$ could be arbitrarily small, that is

$$\frac{\pi'(\mathbf{y_1})}{\pi'(\mathbf{x})} \to 0 \text{ for } \|\mathbf{x}\| \to \infty. \tag{84}$$

As $\mathbf{y}_1$ is the closest candidate to $\mathbf{x}$ in set $R_x$ by the minimum integer increment $\Delta$, then the following relationship holds due to exponential decay of $\pi'$

$$\pi(\mathbf{y}_2) \ll \pi(\mathbf{y}_1) \text{ for } \|\mathbf{x}\| \to \infty, \tag{85}$$

where $\mathbf{y}_2 \in R_x$, $\mathbf{y}_2 \neq \mathbf{y}_1$ denotes another candidate in set $R_x$. Consequently, we have

$$\frac{\pi'(\mathbf{y_2})}{\pi'(\mathbf{x})} \ll \frac{\pi'(\mathbf{y_1})}{\pi'(\mathbf{x})} \to 0 \text{ for } \|\mathbf{x}\| \to \infty, \tag{86}$$

which means the third term of RHS of (81) can be made arbitrarily small by large enough $\|\mathbf{x}\|$.

Similarly, the same derivation can be made for the case of $\mathbf{y} \in A_x$. Since $\mathbf{x}$ is also contained in $A_x$, care must be taken to perform the analysis and we have

$$\frac{\pi'(\mathbf{x})}{\pi'(\mathbf{y_4})} \ll \frac{\pi'(\mathbf{x})}{\pi'(\mathbf{y_3})} \to 0 \text{ for } \|\mathbf{x}\| \to \infty, \tag{87}$$

where $\mathbf{y}_3 = \mathbf{x} - \Delta \in A_x$ stands for the closest non-$\mathbf{x}$ candidate in $A_x$ to $\mathbf{x}$ and $\mathbf{y}_4$ represents any other candidate in $A_x$ except $\mathbf{y}_3$ and $\mathbf{x}$.

In the case of $\mathbf{x} \notin C$, since the indicator function $\mathbf{1}_C(\mathbf{x})$ is 0, $\lambda$ can be expressed directly as the ratio between $\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x},\mathbf{y})V(\mathbf{y})$ and $V(\mathbf{x})$. Therefore, based on (86) and (87), as $\|\mathbf{x}\|$ goes to infinity, we have

$$\begin{aligned}
\lambda &= \limsup_{\|\mathbf{x}\|\to\infty} \frac{\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x},\mathbf{y})V(\mathbf{y})}{V(\mathbf{x})} \\
&= 1 - \liminf_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in A_\mathbf{x}} q(\mathbf{x},\mathbf{y}) \left[1 - \frac{\pi'(\mathbf{x})^{\frac{1}{2}}}{\pi'(\mathbf{y})^{\frac{1}{2}}}\right] \\
&= 1 - \liminf_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in A_\mathbf{x}, \mathbf{y}\neq\mathbf{x}} q(\mathbf{x},\mathbf{y}) \\
&\overset{(l)}{<} 1,
\end{aligned} \tag{88}$$

where inequality $(l)$ holds due to the following derivation

$$\liminf_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in A_\mathbf{x}, \mathbf{y}\neq\mathbf{x}} q(\mathbf{x},\mathbf{y}) = \liminf_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in A_\mathbf{x}, \mathbf{y}\neq\mathbf{x}} \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{By}\|^2}}{\prod_{i=1}^n \rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})}$$

$$> 0. \tag{89}$$

Therefore, according to (88), the drift condition shown in (74) is satisfied for $\mathbf{x} \notin C$.

(ii). As for the case of $\mathbf{x} \in C$, function $V$ for $\mathbf{y} \in A_\mathbf{x}$ is upper bounded as

$$V(\mathbf{y}) \leq V(\mathbf{x}) \leq d \tag{90}$$

due to $A_\mathbf{x} \subseteq C$. Therefore, the first term in the RHS of (80) always can be upper bounded by a constant $b > 0$, namely,

$$\sum_{\mathbf{y}\in A_\mathbf{x}} q(\mathbf{x},\mathbf{y})V(\mathbf{y}) < b < \infty. \tag{91}$$

Then as $\|\mathbf{x}\|$ goes to infinity, based on (86) and (87), it follows that

$$\begin{aligned}
\limsup_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x},\mathbf{y})V(\mathbf{y}) &\leq b + \limsup_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in R_\mathbf{x}} q(\mathbf{x},\mathbf{y})V(\mathbf{x}) \\
&= b + \lambda V(\mathbf{x}),
\end{aligned} \tag{92}$$

and it is easy to verify that

$$\lambda = \limsup_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in R_\mathbf{x}} q(\mathbf{x},\mathbf{y}) < 1, \tag{93}$$

completing the proof. □

In essence, the convergence of the geometric ergodicity can be classified into two stages. On one hand, for $\mathbf{x} \notin C$, due to $\lambda < 1$, the drift condition guarantees the Markov chain shrinks geometrically towards the small set $C$. In fact, as shown in Fig. 6, given $\mathbf{x} \notin C$, by symmetry, the probability of $\mathbf{y}$ locating in $R_\mathbf{x}$ is always larger than that in $A_\mathbf{x}$. Put it another way, the size of $\sum_{\mathbf{y}\in A_\mathbf{x}, \mathbf{y}\neq\mathbf{x}} q(\mathbf{x},\mathbf{y})$ would be always smaller than $1/2$, which means the coefficient $\lambda < 1$ shown in (88) is always lower bounded by

$$\lambda > \frac{1}{2}. \tag{94}$$

Apparently, $\lambda$ increases gradually with the decrease of $\|\mathbf{x}\|$ and vice versa.

On the other hand, if $\mathbf{x} \in C$, the *minorisation condition* shown in (73) implies the Markov chain will converge to the stationary distribution exponentially fast. Specifically, with respect to $C$, the exponential convergence can be demonstrated via coupling technique and $\delta$ is just the exponential decay coefficient. However, as $\delta$ is determined by $C$, rendering it flexible to the different settings of $C$. In [38], Rosenthal pointed out that for $C = \{\mathbf{x} : V(\mathbf{x}) \leq d\}$ and $d > 2b/(1-\lambda)$, Markov chains satisfying the drift condition will converge exponentially to the stationary distribution as

$$\|P^n(\mathbf{x}_0,\cdot) - \pi(\cdot)\|_{TV} \leq (1-\delta)^{rn} + \left(\frac{U^r}{\alpha^{1-r}}\right)^n \left(1 + \frac{b}{1-\lambda} + V(\mathbf{x}_0)\right), \tag{95}$$

where $0 < r < 1$,

$$\alpha = \frac{1+d}{1+2b+\lambda d} \quad \text{and} \quad U = 1 + 2(d+b). \tag{96}$$

Clearly, there is a trade-off between these two convergence stages: a larger set $C$ indicates a smaller $\delta$ in the minorisation condition for $\mathbf{x} \in C$ but a faster shrink speed $\lambda$ towards $C$ for $\mathbf{x} \notin C$ (close to $1/2$ when $\|\mathbf{x}\| \to \infty$). However, the size of $C$, measured by $d$ here, is determined artificially, making both $\delta$ and $\lambda$ not only flexible but also sensitive to the slight change of $d$. Moreover, even with a specific $C$, $\lambda$ is still difficult to get. Therefore, although geometric ergodicity can be achieved by the proposed SMK algorithm, it is difficult to obtain the quantitative bounds on $\delta$ and $\lambda$. Even though simulation techniques can be applied to make $\delta$ and $\lambda$ accessible, the complexity is too computationally expensive to afford in practice [39].

According to (95), it is worthy to point out that the convergence of the Markov chain arisen from the SMK algorithm

also highly depends on the starting state $\mathbf{x}_0$, which follows the definition of geometric ergodicity given in (10). In theory, $\mathbf{x}_0$ could be any candidate from the state space but a poor choice may intensively increase the required mixing time. To this end, it is widely suggested to start the Markov chain with $\mathbf{x}_0$ as close to the center of the distribution as possible. This is actually in accordance with the result shown in (95), implying the solution of the CVP between $\mathbf{Bx}$ to $\mathbf{c}$ is the optimal choice for lattice Gaussian sampling. For the consideration of the complexity, Babai's nearest plane algorithm is recommended here as a preprocessing stage to output the Babai's point as $\mathbf{x}_0$ [40].

## VI. Conclusions

In this paper, two MH-based algorithms were proposed to sample from lattice Gaussian distributions. As the proposal distribution in MH algorithms can be set freely, an independent proposal distribution and a symmetric proposal distribution are exploited respectively for the geometric convergence performance. Because the state space of the Markov chain in lattices is countably infinite, convergence analysis starting from the ergodicity turns out to be necessary. It is proven that the Markov chain arising from the independent MHK algorithm is uniformly ergodic, thus leading to an exponential convergence performance regardless of the starting state. Furthermore, we show its convergence rate can be explicitly calculated by theta series, implying a tractable mixing time to the stationary distribution. On the other hand, the proposed SMK algorithm is proven to be geometrically ergodic, where the selection of the initial state matters. Due to the inherent symmetry, it not only converges exponentially fast, but also is greatly simplified for the implementation.

## Acknowledgment

The authors would like to thank Antonio Campello for the fruitful discussions.

## Appendix A
## Proof of Theorem 1

*Proof.* It is easy to verify that the Markov chain with the independent proposal distribution $q$ shown in (13) is irreducible and aperiodic. Therefore, according to the definition of ergodicity, the key in proof of ergodicity turns out to show the underlying Markov chain is *positive recurrent*.

**Definition 5.** *An irreducible Markov chain with transition matrix $\mathbf{P}$ is positive recurrent if and only if there exists a probability distribution $\pi$ on $\Omega$ such that $\pi = \pi\mathbf{P}$.*

In principle, the Markov chain produced by the independent MHK algorithm is inherently reversible with respect to $\pi$, since

$$\begin{aligned}\pi(\mathbf{x})P(\mathbf{x},\mathbf{y}) &= \pi(\mathbf{x})q(\mathbf{x},\mathbf{y})\alpha(\mathbf{x},\mathbf{y}) \\ &= \min\{\pi(\mathbf{x})q(\mathbf{y}), \pi(\mathbf{y})q(\mathbf{x})\} \\ &= \pi(\mathbf{y})P(\mathbf{y},\mathbf{x}), \end{aligned} \tag{97}$$

where the assumption $\mathbf{y} \neq \mathbf{x}$ is sufficient because the above equation holds trivially in the case of $\mathbf{y} = \mathbf{x}$.

Next, based on the reversibility, by fixing a state $\mathbf{y}$ and summing over all states, we have

$$\begin{aligned}\sum_{\mathbf{x}\in\mathbb{Z}^n} \pi(\mathbf{x})P(\mathbf{x},\mathbf{y}) &= \sum_{\mathbf{x}\in\mathbb{Z}^n} \pi(\mathbf{y})P(\mathbf{y},\mathbf{x}) \\ &= \pi(\mathbf{y}) \sum_{\mathbf{x}\in\mathbb{Z}^n} P(\mathbf{y},\mathbf{x}) \\ &= \pi(\mathbf{y}), \end{aligned} \tag{98}$$

where $\pi$=$\pi\mathbf{P}$ therefore can be easily obtained. Therefore, complete the proof.

□

## Appendix B
## Proof of Lemma 2

*Proof.* According to the QR-decomposition $\mathbf{B} = \mathbf{QR}$, we have

$$q(\mathbf{x},\mathbf{y}) = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{By}\|^2}}{\prod_{i=1}^n \rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Rx}-\mathbf{Ry}\|^2}}{\prod_{i=1}^n \rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} \tag{99}$$

by removing the orthogonal matrix $\mathbf{Q}$, where $\widetilde{y}_i = \frac{c'_i - \sum_{j=i+1}^n r_{i,j}y_j}{r_{i,i}}$, $\mathbf{c}' = \mathbf{Rx}$.

Specifically, the term $\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})$ in the denominator of (99) can be expressed as

$$\begin{aligned}\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z}) &= \sum_{y_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(y_i - \frac{c'_i - \sum_{j=i+1}^n r_{i,j}y_j}{r_{i,i}})^2} \\ &= \sum_{y_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(y_i - \frac{\sum_{j=i}^n r_{i,j}x_j - \sum_{j=i+1}^n r_{i,j}y_j}{r_{i,i}})^2} \\ &= \sum_{y_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(x_i - y_i + \sum_{j=i+1}^n \frac{r_{i,j}}{r_{i,i}}(x_j - y_j))^2} \\ &= \sum_{z_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(z_i - \phi)^2} \\ &= \rho_{\sigma_i,\phi}(\mathbb{Z}), \end{aligned} \tag{100}$$

where $z_i = y_i - x_i$ and $\phi = \sum_{j=i+1}^n \frac{r_{i,j}}{r_{i,i}}(x_j - y_j)$.

Clearly, we can easily get that

$$\begin{aligned}\rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z}) &= \sum_{x_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(y_i - x_i + \sum_{j=i+1}^n \frac{r_{i,j}}{r_{i,i}}(y_j - x_j))^2} \\ &= \sum_{z_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(z_i - \phi)^2} \\ &= \rho_{\sigma_i,\phi}(\mathbb{Z}) = \rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z}), \end{aligned} \tag{101}$$

where $\widetilde{x}_i = \frac{c''_i - \sum_{j=i+1}^n r_{i,j}x_j}{r_{i,i}}$, $\mathbf{c}'' = \mathbf{Ry}$. □

## 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] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 913–929, May. 1993.

[4] 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.

[5] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, "Semantically secure lattice codes for the Gaussian wiretap channel," *IEEE Trans. Inform. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.

[6] 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.

[7] 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.

[8] S. Liu, C. Ling, and D. Stehlé, "Decoding by sampling: A randomized lattice algorithm for bounded distance decoding," *IEEE Trans. Inform. Theory*, vol. 57, pp. 5933–5945, Sep. 2011.

[9] 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.

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

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

[12] 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.

[13] 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.

[14] J. S. Rosenthal, "Optimal proposal distributions and adaptive MCMC," *Handbook of Markov chain Monte Carlo: Methods and Applications.*, Brooks, S.P., Gelman, A., Jones, G., Meng, X.-L. (eds.) Chapman and Hall/CRC Press, Florida, USA. 2010.

[15] B. Hassibi, M. Hansen, A. Dimakis, H. Alshamary, and W. Xu, "Optimized Markov Chain Monte Carlo for signal detection in MIMO systems: An analysis of the stationary distribution and mixing time," *IEEE Transactions on Signal Processing,*, vol. 62, no. 17, pp. 4436–4450, Sep. 2014.

[16] T. Datta, N. Kumar, A. Chockalingam, and B. Rajan, "A novel Monte Carlo sampling based receiver for large-scale uplink multiuser MIMO systems," *IEEE Transactions on Vehicular Technology,*, vol. 62, no. 7, pp. 3019–3038, Sep. 2013.

[17] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, "Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1896–1909, 2006.

[18] R. Chen, J. Liu, and X. Wang, "Convergence analyses and comparisons of Markov chain Monte Carlo algorithms in digital communications," *IEEE Trans. on Signal Process.*, vol. 50, no. 2, pp. 255–270, 2002.

[19] P. Aggarwal and X. Wang, "Multilevel sequential Monte Carlo algorithms for MIMO demodulation," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 750–758, Feb. 2007.

[20] H. Zhu, B. Farhang-Boroujeny, and R.-R. Chen, "On performance of sphere decoding and Markov chain Monte Carlo detection methods," *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 669–672, 2005.

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

[22] A. Doucet and X. Wang, "Monte Carlo methods for signal processing," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 152 – 170, Nov. 2005.

[23] 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.

[24] D. Aggarwal, D. Dadush, and N. Stephens-Davidowitz, "Solving the closest vector problem in $2^n$ time — the discrete Gaussian strike again!" *FOCS*, 2015.

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

[26] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Time*, American Mathematical Society, 2008.

[27] D. Randall, "Rapidly mixing Markov chains with applications in computer science and physics," *Computing in Science and Engineering*, vol. 8, no. 2, pp. 30–41, 2006.

[28] S. F. Jarner and G. O. Roberts, "Polynomial convergence rates of Markov chains," *Ann. Appl. Probab.*, vol. 12, pp. 224–247, 2002.

[29] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1091, 1953.

[30] K. L. Mengersen and R. L. Tweedie, "Rates of convergence of the Hastings and Metropolis algorithms," *Ann. Statist.*, vol. 24, pp. 101–121, 1996.

[31] J. H. Conway and N. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1998.

[32] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," *IEEE Trans. Inform. Theory.*, no. 99, pp. 1–22, 2015.

[33] C. Ling, "On the proximity factors of lattice reduction-aided decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2795–2808, Jun. 2011.

[34] D. Aharonov and O. Regev, "Lattice problems in NP $\cap$ coNP." *J. ACM*, vol. 52, no. 5, pp. 749–765, 2005.

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

[36] 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.

[37] S. F. Jarner and E. Hansen, "Geometric ergodicity of Metropolis algorithm," *Stochastic Process*, vol. 85, pp. 341–361, 2000.

[38] J. S. Rosenthal, "Minorization conditions and convergence rates for Markov chain Monte Carlo," *J. Amer. Statist. Assoc.*, vol. 90, pp. 558–566, 1995.

[39] M. K. Cowles and J. S. Rosenthal, "A simulation approach to convergence rates for Markov chain Monte Carlo," *Statistics and Computing*, vol. 8, pp. 115–124, 1998.

[40] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.