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

Zheng Wang, *Member, IEEE*, and Cong Ling, *Member, IEEE*

***Abstract*— Sampling from the lattice Gaussian distribution has emerged as an important problem in coding, decoding, and cryptography. In this paper, the classic Metropolis-Hastings (MH) algorithm in Markov chain Monte Carlo methods is adopted for lattice Gaussian sampling. Two MH-based algorithms are proposed, which overcome the limitation of Klein's algorithm. The first one, referred to as the independent Metropolis-Hastings-Klein (MHK) algorithm, establishes a Markov chain via an independent proposal distribution. We show that the Markov chain arising from this independent MHK algorithm is uniformly ergodic, namely, it converges to the stationary distribution exponentially fast regardless of the initial state. Moreover, the rate of convergence is analyzed in terms of the theta series, leading to predictable mixing time. A symmetric Metropolis-Klein algorithm is also proposed, which is proven to be geometrically ergodic.**

***Index Terms*— Lattice Gaussian distribution, lattice coding, lattice decoding, MCMC methods, integer least-squares problems.**

I. Introduction

RECENTLY, the lattice Gaussian distribution has emerged as a common theme in various research domains. 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 [4]. It was also used to achieve information-theoretic security in the Gaussian wiretap channel [5], [6] and in the bidirectional relay channel [7], respectively. 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 [8]. Meanwhile, it also has underpinned the fully-homomorphic encryption for cloud computing [9]. Algorithmically, lattice Gaussian sampling with a suitable variance allows to solve the shortest vector problem (SVP) and the closest vector problem (CVP) [10], [11]; for example, it has led to efficient lattice decoding for multi-input multi-output (MIMO) systems [12], [13]. In theory, it has been demonstrated that lattice Gaussian sampling is equivalent to CVP via a polynomial-time dimension-preserving reduction [14], and SVP is essentially a special case of the CVP.

Due to the central role of the lattice Gaussian distribution playing in these fields, its sampling algorithms become an important computational problem. In contrast to sampling from a continuous Gaussian distribution, 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 [15] is capable to sample from the lattice Gaussian distribution within a negligible statistical distance only if the standard deviation is large enough [16]. However, such a requirement renders 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 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. To this end, the Gibbs algorithm was introduced into lattice Gaussian sampling, which employs univariate conditional sampling to build a Markov chain [17]. It is able to sample beyond the range of Klein's algorithm. In [17], a flexible block-based Gibbs algorithm was also presented, which performs sampling over multiple elements within a block. In this way, the correlation within the block could be exploited, leading to a faster convergence especially in the case of highly correlated components. Unfortunately, related analysis of the convergence rate for the associated Markov chains in these two algorithms was lacking, resulting in an unpredictable mixing time.

On the other hand, Gibbs sampling has already been adapted to signal detection for multi-input multi-output (MIMO) communications [18]–[23]. In particular, the selection of σ (also referred to as "temperature") is studied in [18] and it is argued that σ should grow as fast as the signal-to-noise ratio (SNR) in general. In [19], a mixed-Gibbs sampler is proposed to achieve near-optimal performance, which takes the advantages of an efficient stopping criterion and a multiple restart strategy. Moreover, Gibbs sampling is also introduced into soft-output decoding in MIMO systems, where the extrinsic information calculated by a priori probability (APP) detector is used to produce soft outputs [20]. In [21], an investigation

Manuscript received December 7, 2015; revised April 18, 2017; accepted August 1, 2017. Date of publication August 21, 2017; date of current version January 18, 2018. This work was supported by FP7 Project PHYLAWS under Grant EU FP7-ICT 31756. This paper was presented at the 2015 IEEE International Symposium on Information Theory.

Z. Wang is with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210000, China.

C. Ling is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: z.wang@ieee.org; cling@ieee.org).

Communicated by A. G. Dimakis, Associate Editor for Coding Techniques.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2017.2742509

of Gibbs-based MCMC receivers in different communication channels are given. Due to the finite state space formed by a finite modulation constellation, those Gibbs samplers converge exponentially fast to the stationary distribution. However, the rate of convergence has not yet been determined.

In this paper, another famous MCMC scheme, known as the Metropolis-Hastings (MH) algorithm [24], is studied in detail for lattice Gaussian sampling. In particular, it makes use of a *proposal distribution* which suggests a possible state candidate and then employs an 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 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, a candidate at each Markov move is generated from a Gaussian-like proposal distribution via Klein's algorithm. In this case, we show that the Markov chain induced by the independent MHK algorithm is uniformly ergodic, which implies it converges exponentially fast to the stationary distribution irrespective of the starting state. Its convergence rate is then estimated given the lattice basis $\mathbf{B}$, the query point $\mathbf{c}$ and the standard derivation σ. Thus, the mixing time of the induced Markov chain becomes predictable. To the best of our knowledge, this is the first time that the convergence rate of MCMC in communications and signal processing is determined analytically since MCMC was introduced into this field in 1990's [25].

Different from the algorithms in [10] and [11] which have exponential space and time complexity, the proposed independent MHK algorithm has polynomial space complexity, and its time complexity[1] varies with σ, where a larger value of σ corresponds to smaller mixing time. This is in agreement with the fact we knew before: if σ 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. Likewise, the proposed sampling algorithm can also be extended to lattice decoding, and more details can be found in our following work [26].

The second proposed algorithm, namely the symmetric Metropolis-Klein (SMH) algorithm, establishes a symmetric proposal distribution between two consecutive Markov states. We show it also converges to the stationary distribution exponentially fast but the selection of the initial state also plays a role. Such a case is referred to as geometric ergodicity in MCMC literature [27]. 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.

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 arising from it converges exponentially fast to the stationary distribution.
2) The convergence rate of the independent MHK algorithm is derived explicitly in terms of the theta series, thereby making the estimation of mixing time possible.
3) The SMH algorithm is further proposed for lattice Gaussian sampling, which not only achieves exponential convergence, but also is simpler 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 SMH 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 denote by $\mathbf{b}_i$ the ith column of the matrix $\mathbf{B}$, by $\widehat{\mathbf{b}}_i$ the ith Gram-Schmidt vector of $\mathbf{B}$, and by $b_{i,j}$ the entry in the ith row and jth column of $\mathbf{B}$. $\lceil x \rfloor$ denotes rounding to the integer closest to x. If x is a complex number, $\lceil x \rfloor$ rounds the real and imaginary parts separately. In addition, we use the standard *small omega* notation $\omega(\cdot)$, i.e., $f(n) = \omega(g(n))$ if for any $k > 0$, the inequality $|f(n)| > k \cdot |g(n)|$ holds for all sufficiently large n.

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

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

where $\mathbf{B}$ is called 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

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

for all $\mathbf{x} \in \mathbb{Z}^n$, where $\rho_{\sigma,\mathbf{c}}(\Lambda) \triangleq \sum_{\mathbf{Bx}\in\Lambda} \rho_{\sigma,\mathbf{c}}(\mathbf{Bx})$ is just a scaling to obtain a probability distribution.

Note that this definition differs slightly from the one in [8], where σ is scaled by a constant factor $\sqrt{2\pi}$ (i.e., $s = \sqrt{2\pi}\sigma$).

[1] In this paper, the computational complexity is measured by the number of arithmetic operations (additions, multiplications, comparisons, etc.). The time complexity of an MCMC sampler can be estimated by the mixing time times the complexity of each Markov move.

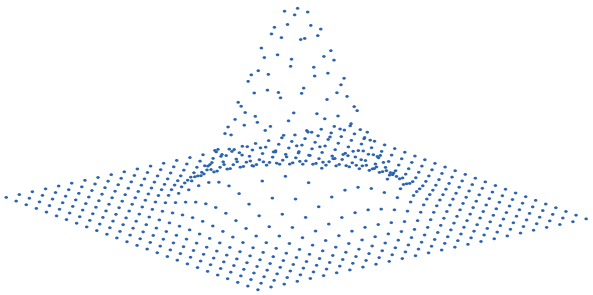

Fig. 1. Illustration of a two-dimensional lattice Gaussian distribution.

Fig. 1 illustrates the discrete Gaussian distribution over $\mathbb{Z}^2$. As 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

Intuitively, the shape 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 to solve 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 proposed for lattice decoding [15].

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 $\mathbf{Bx}$

As shown in Algorithm 1, the operation of the Klein's algorithm has polynomial complexity $O(n^2)$ excluding QR decomposition (which may be done only once in the beginning). 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

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

where $\widetilde{x}_i = \frac{c'_i - \sum_{j=i+1}^{n} r_{i,j} x_j}{r_{i,i}}$, $\sigma_i = \frac{\sigma}{|r_{i,i}|} = \frac{\sigma}{\|\widehat{\mathbf{b}}_i\|}$, $\mathbf{c}' = \mathbf{Q}^{\dagger}\mathbf{c}$, $r_{i,j}$ denotes the element of the upper triangular matrix $\mathbf{R}$ from the QR decomposition $\mathbf{B} = \mathbf{QR}$ and $\widehat{\mathbf{b}}_i$'s are the Gram-Schmidt vectors of $\mathbf{B}$ with $\|\widehat{\mathbf{b}}_i\| = |r_{i,i}|$.

Furthermore, it has been demonstrated in [16] that $P_{\text{Klein}}(\mathbf{x})$ is close to $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$ within a negligible statistical distance if

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

However, even with the help of lattice reduction[2] (e.g., LLL reduction), the standard deviation $\omega(\sqrt{\log n}) \cdot \max_{1 \le i \le n} \|\widehat{\mathbf{b}}_i\|$ can be too large to be useful.

C. MCMC Methods

As for the lattice Gaussian sampling in the range $\sigma \neq \omega(\sqrt{\log n}) \cdot \max_{1 \le i \le n} \|\widehat{\mathbf{b}}_i\|$, 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, MCMC is capable of sampling from the target distribution of interest, thereby removing the restriction on σ [17].

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 [29]

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

where $\|\cdot\|_{TV}$ represents the total variation distance (other measures of distance also exist, see [30] for more details). It is well known that the *spectral gap* $\gamma = 1 - |\lambda_1| > 0$ of the transition matrix offers an upper bound on the mixing time, where λ_1 represents the second largest eigenvalue (in magnitude) of the transition matrix $\mathbf{P}$. A large value of the spectral gap leads to rapid convergence to stationarity [31].

However, the spectrum of a Markov chain can be hard to analyze, especially when the state space Ω becomes exponentially large, making it difficult to have a compact mathematical expression of the adjacency matrix. Thanks to the celebrated *coupling technique*, for any Markov chain with finite state space Ω, exponentially fast convergence can be demonstrated if the underlying Markov chain is irreducible and aperiodic with an invariant distribution π [29]. Nevertheless, in the case of lattice Gaussian sampling, the countably infinite state space $\mathbf{x} \in \mathbb{Z}^n$ naturally becomes a challenge. For this reason, we perform the convergence analysis from the beginning — ergodicity [32].

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* π *on* Ω *and for all* $\mathbf{x} \in \Omega$*,*

$$\lim_{t \to \infty} \|P^t(\mathbf{x}, \cdot) - \pi\|_{TV} = 0, \tag{7}$$

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 definition is given [32].

[2]It is well known that lattice reduction such as the LLL algorithm is able to significantly improve $\min_i \|\widehat{\mathbf{b}}_i\|$ while reducing $\max_i \|\widehat{\mathbf{b}}_i\|$ at the same time [28].

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

$$\|P^t(\mathbf{x}, \cdot) - \pi(\cdot)\|_{TV} \leq M(1-\delta)^t. \tag{8}$$

Obviously, the exponential decay coefficient δ 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 ergodicity, *geometric ergodicity* also converges exponentially, but M is parameterized by the initial state $\mathbf{x}$.

Definition 3: A Markov chain with stationary distribution $\pi(\cdot)$ is geometrically ergodic if there exists $0 < \delta < 1$ and $M(\mathbf{x}) < \infty$ such that for all $\mathbf{x}$

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

Besides exponential convergence, polynomial convergence also exists [33], which goes beyond the scope of this paper due to its slow convergence. Unless stated otherwise, the state space of the Markov chain we are concerned with throughout the context is the countably infinite $\mathbf{\Omega} = \mathbb{Z}^n$.

D. Classical MH Algorithms

The origin of the Metropolis algorithm can be traced back to the celebrated work of [34] in 1950's. In [24], 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 π 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 α is computed by

$$\alpha(\mathbf{x}, \mathbf{y}) = \min\left\{1, \frac{\pi(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x}, \mathbf{y})}\right\}, \tag{10}$$

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:

$$P(\mathbf{x}, \mathbf{y}) = \begin{cases} q(\mathbf{x}, \mathbf{y})\alpha(\mathbf{x}, \mathbf{y}) & \text{if } \mathbf{y} \neq \mathbf{x}, \\ 1 - \sum_{\mathbf{z} \neq \mathbf{x}} q(\mathbf{x}, \mathbf{z})\alpha(\mathbf{x}, \mathbf{z}) & \text{if } \mathbf{y} = \mathbf{x}. \end{cases} \tag{11}$$

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 distribution would be $q(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y})$ itself, but in most cases of interest π cannot be sampled directly. To this end, many variations of MH algorithms with different configurations of $q(\mathbf{x}, \mathbf{y})$ were proposed.

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:

Algorithm 2 Independent Metropolis-Hastings-Klein Algorithm for Lattice Gaussian Sampling

Input: $\mathbf{B}, \sigma, \mathbf{c}, \mathbf{X}_0, t_{\text{mix}}(\epsilon)$
Output: sample from a distribution statistically close to $\pi = D_{\Lambda,\sigma,\mathbf{c}}$

```
for t = 1,2, ..., do
    let x denote the state of X_{t-1}
    generate y by the proposal distribution q(x, y) in (12)
    calculate the acceptance ratio α(x, y) in (13)
    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 t ≥ t_mix(ε) then
        output the state of X_t
    end if
end for
```

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

$$\begin{aligned} q(\mathbf{x}, \mathbf{y}) = q(\mathbf{y}) &= P_{\text{Klein}}(\mathbf{y}) \\ &= \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{By})}{\prod_{i=1}^n \rho_{\sigma_i, \widetilde{y}_i}(\mathbb{Z})} \\ &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{By}-\mathbf{c}\|^2}}{\prod_{i=1}^n \sum_{\widetilde{y}_i \in \mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|y_i - \widetilde{y}_i\|^2}} \end{aligned} \tag{12}$$

where $\mathbf{y} \in \mathbb{Z}^n$, $\widetilde{y}_i = \frac{c_i' - \sum_{j=i+1}^n r_{i,j} y_j}{r_{i,i}}$, $\sigma_i = \frac{\sigma}{|r_{i,i}|} = \frac{\sigma}{\|\widehat{\mathbf{b}}_i\|}$, $\mathbf{c}' = \mathbf{Q}^\dagger \mathbf{c}$, $\mathbf{B} = \mathbf{QR}$ by QR decomposition and $\widehat{\mathbf{b}}_i$'s are the Gram-Schmidt vectors of $\mathbf{B}$.

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

$$\begin{aligned} \alpha(\mathbf{x}, \mathbf{y}) &= \min\left\{1, \frac{\pi(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x}, \mathbf{y})}\right\} = \min\left\{1, \frac{\pi(\mathbf{y})q(\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{y})}\right\} \\ &= \min\left\{1, \frac{\prod_{i=1}^n \rho_{\sigma_i, \widetilde{y}_i}(\mathbb{Z})}{\prod_{i=1}^n \rho_{\sigma_i, \widetilde{x}_i}(\mathbb{Z})}\right\}, \end{aligned} \tag{13}$$

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

3) *With probability* $\alpha(\mathbf{x}, \mathbf{y})$ *accept* $\mathbf{X}_{t+1} = \mathbf{y}$*; otherwise, reject* $\mathbf{y}$ *and let* $\mathbf{X}_{t+1} = \mathbf{x}$.

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. The complexity of the MCMC sampler is given by the number of Markov moves times the complexity of each move, i.e., $O(t_{\text{mix}} \cdot n^2)$.

It is easy to check that the Markov chain associated with the independent proposal distribution q shown in (12) is reversible

(by the Metropolis-Hastings construction), from which we obtain that π is a strictly positive stationary distribution, and hence, the Markov chain is positive recurrent, namely ergodic. Then, we have the following well-known result, whose proof can be found in [29] and [32].

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

$$\lim_{t\to\infty} \|P^t(\mathbf{x};\cdot) - D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV} = 0 \tag{14}$$

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

B. Uniform Ergodicity

The independent proposal distribution defined in (12) enjoys the following property.

Lemma 1: In the independent MHK algorithm for lattice Gaussian sampling from $D_{\Lambda,\sigma,\mathbf{c}}$, there exists $\delta > 0$ such that

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})} \geq \delta \tag{15}$$

for all $\mathbf{x} \in \mathbb{Z}^n$, where $q(\mathbf{x}) = P_{\text{Klein}}(\mathbf{x})$.

Proof: Using (3) and (4), we have

$$\begin{aligned}\frac{q(\mathbf{x})}{\pi(\mathbf{x})} &= \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{Bx})}{\prod_{i=1}^n \rho_{\sigma_i,\tilde{x}_i}(\mathbb{Z})} \cdot \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\rho_{\sigma,\mathbf{c}}(\mathbf{Bx})} \\ &= \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^n \rho_{\sigma_i,\tilde{x}_i}(\mathbb{Z})} \\ &\overset{(a)}{\geq} \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^n \rho_{\sigma_i}(\mathbb{Z})} = \delta \end{aligned} \tag{16}$$

where (a) holds due to the fact that [8]

$$\rho_{\sigma_i,\tilde{x}}(\mathbb{Z}) \leq \rho_{\sigma_i}(\mathbb{Z}) \triangleq \sum_{j\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2} j^2}. \tag{17}$$

As can be seen clearly, the right-hand side (RHS) of (16) is completely independent of $\mathbf{x}$, meaning it can be expressed as a constant δ determined by the given $\mathbf{B}$, $\mathbf{c}$ and σ. Therefore, the proof is completed. ∎

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

Theorem 1: Given the invariant lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain established by the independent MHK algorithm is uniformly ergodic:

$$\|P^t(\mathbf{x},\cdot) - D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV} \leq (1-\delta)^t \tag{18}$$

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

Proof: By (12) and (13), the transition probability $P(\mathbf{x},\mathbf{y})$ of the independent MHK algorithm is given by

$$P(\mathbf{x},\mathbf{y}) = \begin{cases} \min\left\{q(\mathbf{y}), \dfrac{\pi(\mathbf{y})q(\mathbf{x})}{\pi(\mathbf{x})}\right\} & \text{if } \mathbf{y} \neq \mathbf{x}, \\ q(\mathbf{x}) + \displaystyle\sum_{\mathbf{z}\neq\mathbf{x}} \max\left\{0, q(\mathbf{z}) - \dfrac{\pi(\mathbf{z})q(\mathbf{x})}{\pi(\mathbf{x})}\right\} & \text{if } \mathbf{y} = \mathbf{x}. \end{cases} \tag{19}$$

Using Lemma 1, it is straightforward to check that the following relationship holds

$$P(\mathbf{x},\mathbf{y}) \geq \delta\pi(\mathbf{y}) \tag{20}$$

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

Now, consider the following construction of coupling of two Markov chains $\mathbf{X}_t$ and $\mathbf{X}'_t$ [27]. $\mathbf{X}'_t$ is supposed to start from the stationary distribution π, and $\mathbf{X}_t$ from a fixed (but arbitrarily) initial state $\mathbf{x}_0$. At each step $t > 0$, repeat the following procedure:

- If $\mathbf{X}_t = \mathbf{X}'_t$, choose $\mathbf{X}_{t+1} = \mathbf{X}'_{t+1}$ from distribution $P(\mathbf{X}_t,\cdot)$.
- Else,
  - With probability δ, choose $\mathbf{X}_{t+1} = \mathbf{X}'_{t+1}$ from distribution $\pi(\cdot)$;
  - With probability $1-\delta$, conditionally independently sample

$$\mathbf{X}_{t+1} \text{ from distribution } \frac{1}{1-\delta}[P(\mathbf{X}_t,\cdot) - \delta\pi(\cdot)];$$
$$\mathbf{X}'_{t+1} \text{ from distribution } \frac{1}{1-\delta}[P(\mathbf{X}'_t,\cdot) - \delta\pi(\cdot)].$$

It is easy to check that $\mathbf{X}_t$ and $\mathbf{X}'_t$ marginally update according to the same transition probability (19).

According to the *coupling inequality* [29], the total variation distance between the distributions of $\mathbf{X}_t$ and $\mathbf{X}'_t$ is upper bounded by

$$\|P^t(\mathbf{x}_0,\cdot) - \pi(\cdot)\|_{TV} \leq P(\mathbf{X}_t \neq \mathbf{X}'_t). \tag{21}$$

Note that, by construction, the two chains stay together at all times once they meet at a same state, namely,

$$\text{if } \mathbf{X}_n = \mathbf{X}'_n, \quad \text{then } \mathbf{X}_t = \mathbf{X}'_t \text{ for } t \geq n. \tag{22}$$

Therefore, given the event $\mathbf{X}_t \neq \mathbf{X}'_t$, there is no coupling in any of the t consecutive moves, and we have

$$\begin{aligned} P(\mathbf{X}_t \neq \mathbf{X}'_t) &= P(\mathbf{X}_t \neq \mathbf{X}'_t, \ldots, \mathbf{X}_0 \neq \mathbf{X}'_0) \\ &= \prod_{i=1}^t P(\mathbf{X}_i \neq \mathbf{X}'_i | \mathbf{X}_{i-1} \neq \mathbf{X}'_{i-1}) \cdot P(\mathbf{X}_0 \neq \mathbf{X}'_0) \\ &\leq \prod_{i=1}^t P(\mathbf{X}_i \neq \mathbf{X}'_i | \mathbf{X}_{i-1} \neq \mathbf{X}'_{i-1}) \\ &= \prod_{i=1}^t \left[1 - P(\mathbf{X}_i = \mathbf{X}'_i | \mathbf{X}_{i-1} \neq \mathbf{X}'_{i-1})\right] \\ &= \left[1 - \sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{X}_i = \mathbf{X}'_i = \mathbf{y} | \mathbf{X}_{i-1} \neq \mathbf{X}'_{i-1})\right]^t \\ &\overset{(b)}{\leq} \left[1 - \sum_{\mathbf{y}\in\mathbb{Z}^n} \delta\pi(\mathbf{y})\right]^t \\ &= (1-\delta)^t, \end{aligned} \tag{23}$$

where (b) is because, by construction again, for each move we have probability at least δ of making $\mathbf{X}_i$ and $\mathbf{X}'_i$ $(i = 1, 2, \ldots, t)$ equal.

Then, substituting (23) into (21), we obtain

$$\|P^t(\mathbf{x},\cdot) - \pi(\cdot)\|_{TV} \leq (1-\delta)^t, \tag{24}$$

completing the proof. ∎

Obviously, given the value of $\delta < 1$, the mixing time of the Markov chain can be calculated by (6) and (24), that is,

$$t_{\text{mix}}(\epsilon) = \frac{\ln\epsilon}{\ln(1-\delta)} \leq (-\ln\epsilon)\cdot\left(\frac{1}{\delta}\right), \quad \epsilon < 1 \tag{25}$$

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

Here, we point out that the aforementioned spectral gap γ of the transition matrix can also be used to bound the mixing time. Resorting to the *conductance* of the Markov chain [29], one obtains a lower bound on the spectral gap γ of the transition matrix (see Appendix VI for its derivation)

$$\gamma \geq \frac{\delta^2}{8}. \tag{26}$$

This yields another upper bound on the mixing time on the order of $\frac{1}{\delta^2}$, which is however looser than (25).

C. Convergence 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 σ, 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 σ is investigated.

Let the standard deviations of $q(\mathbf{x})$ and $\pi(\mathbf{x})$ be $\overline{\sigma}$ and σ respectively, then the corresponding ratio of $q(\mathbf{x})/\pi(\mathbf{x})$ in (16) can be rewritten as

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})} \geq \frac{\rho_{\sigma,\mathbf{c}}(\Lambda)}{\prod_{i=1}^{n}\rho_{\overline{\sigma}_i}(\mathbb{Z})}\cdot e^{-(\frac{1}{2\overline{\sigma}^2}-\frac{1}{2\sigma^2})\|\mathbf{Bx}-\mathbf{c}\|^2}. \tag{27}$$

Unfortunately, in the case of $\overline{\sigma} < \sigma$, as $\|\mathbf{Bx}-\mathbf{c}\|$ can be arbitrary, 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 [35].[3] Therefore, $\overline{\sigma} < \sigma$ should be avoided in practice and the corresponding convergence analysis is ignored here.

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

$$d(\Lambda,\mathbf{c}) = \min_{\mathbf{x}\in\mathbb{Z}^n}\|\mathbf{Bx}-\mathbf{c}\|, \tag{28}$$

then it follows that

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})} \geq \frac{\rho_{\sigma,\mathbf{c}}(\Lambda)}{\prod_{i=1}^{n}\rho_{\overline{\sigma}_i}(\mathbb{Z})}\cdot e^{-(\frac{1}{2\overline{\sigma}^2}-\frac{1}{2\sigma^2})d^2(\Lambda,\mathbf{c})} \tag{29}$$

for all $\mathbf{x} \in \mathbb{Z}^n$, which means the underlying Markov chain is uniformly ergodic by satisfying (15) in Lemma 1. More precisely, $q(\mathbf{x})/\pi(\mathbf{x})$ could be expressed as

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})} \geq \frac{\rho_{\sigma,\mathbf{c}}(\Lambda)}{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})}\cdot\beta \tag{30}$$

where

$$\beta = \frac{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})}{\prod_{i=1}^{n}\rho_{\overline{\sigma}_i}(\mathbb{Z})}\cdot e^{-(\frac{1}{2\overline{\sigma}^2}-\frac{1}{2\sigma^2})d(\Lambda,\mathbf{c})^2}. \tag{31}$$

[3]In theory, that $q(\mathbf{x})/\pi(\mathbf{x})$ is lower bounded by a constant for all $\mathbf{x} \in \mathbb{Z}^n$ is both sufficient and necessary to the uniform ergodicity [35].

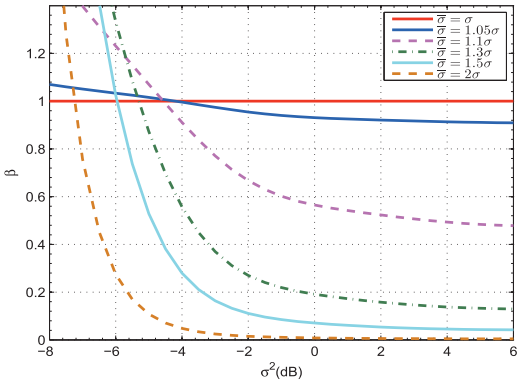

Fig. 2. Coefficient β of E_8 lattice in the case of $\overline{\sigma} > \sigma$ when $\mathbf{c} = \mathbf{0}$.

Clearly, parameter β becomes the key to govern the convergence performance. Compared to (16), if $\beta > 1$, the convergence of the Markov chain will be boosted by a larger value of δ, otherwise the convergence will be slowed down. However, in the case of $\overline{\sigma} > \sigma$, it easy to check that the value of β is monotonically decreasing with the given σ, rendering $\beta > 1$ inapplicable to the most cases of interest.

As can be seen clearly from Fig. 2, the convergence rate can be enhanced by $\beta > 1$ only for a small enough σ (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.

IV. Convergence Rate Analysis

In this section, convergence analysis about the exponential decay coefficient δ 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 cases $\mathbf{c} = \mathbf{0}$ and $\mathbf{c} \neq \mathbf{0}$ separately.

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

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

Specifically, we have

$$\begin{aligned}\frac{q(\mathbf{x})}{\pi(\mathbf{x})} &= \frac{\rho_{\sigma,\mathbf{0}}(\Lambda)}{\prod_{i=1}^{n}\rho_{\sigma_i,\tilde{x}_i}(\mathbb{Z})}\\ &\overset{(c)}{\geq} \frac{\sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}\|^2}}{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})}\\ &\overset{(d)}{=} \frac{\Theta_\Lambda(\frac{1}{2\pi\sigma^2})}{\prod_{i=1}^{n}\Theta_{\mathbb{Z}}(\frac{1}{2\pi\sigma_i^2})}\\ &\overset{(e)}{=} \frac{\Theta_\Lambda(\frac{1}{s^2})}{\prod_{i=1}^{n}\vartheta_3(\frac{1}{s_i^2})} = \delta.\end{aligned} \tag{32}$$

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 (c), the inequality $\rho_{\sigma_i,\tilde{x}}(\mathbb{Z}) \le \rho_{\sigma_i}(\mathbb{Z})$ shown in (17) is used again. Theta series Θ_Λ and Jacobi theta function ϑ_3 are applied in (d) and (e) respectively, where

$$\Theta_\Lambda(\tau) = \sum_{\boldsymbol{\lambda}\in\Lambda} e^{-\pi\tau\|\boldsymbol{\lambda}\|^2}, \tag{33}$$

$$\vartheta_3(\tau) = \sum_{n=-\infty}^{+\infty} e^{-\pi\tau n^2} \tag{34}$$

with $\Theta_{\mathbb{Z}} = \vartheta_3$ [36].

Proposition 2: If $s = \omega(\sqrt{\log n}) \cdot \max_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|$ *or* $1/s = \omega(\sqrt{\log n}) \cdot \left(\min_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|\right)^{-1}$, *then the coefficient* $\delta \approx 1$.

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

$$\epsilon_\Lambda(\sigma) = \frac{\det(\mathbf{B})}{(\sqrt{2\pi}\sigma)^n}\Theta_\Lambda\left(\frac{1}{2\pi\sigma^2}\right) - 1. \tag{35}$$

and

$$\epsilon_\Lambda(\sigma) = \varepsilon, \quad \text{if } \sigma = \eta_\varepsilon(\Lambda). \tag{36}$$

Here, $\eta_\varepsilon(\Lambda)$ is known as the *smoothing parameter* and for any n-dimensional lattice Λ and positive real $\varepsilon > 0$, $\eta_\varepsilon(\Lambda)$ is defined as the smallest real $\sigma > 0$ such that $\rho_{1/\sqrt{2\pi}\sigma}(\Lambda^*\backslash\{\mathbf{0}\}) \le \varepsilon$, where Λ^* denotes the dual lattice of Λ [16].

Therefore, the exponential decay coefficient δ given in (32) can be expressed as

$$\begin{aligned}\delta &= \frac{\Theta_\Lambda(\frac{1}{2\pi\sigma^2})}{\prod_{i=1}^n \vartheta_3(\frac{1}{2\pi\sigma_i^2})} \\ &= \frac{|\det(\mathbf{B})|^{-1}\cdot(\sqrt{2\pi}\sigma)^n\cdot[\epsilon_\Lambda(\sigma)+1]}{\prod_{i=1}^n\sqrt{2\pi}\sigma_i\cdot[\epsilon_{\mathbb{Z}}(\sigma_i)+1]} \\ &= \frac{\epsilon_\Lambda(\sigma)+1}{\prod_{i=1}^n[\epsilon_{\mathbb{Z}}(\sigma_i)+1]},\end{aligned} \tag{37}$$

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

Meanwhile, from [8, Lemma 3.3], for any n-dimensional lattice Λ and positive real $\varepsilon > 0$, it follows that

$$\eta_\varepsilon(\Lambda) \le \sqrt{\frac{\log(2n(1+1/\varepsilon))}{\pi}}\cdot\max_{1\le i\le n}\|\widehat{\mathbf{b}}_i\| \tag{38}$$

and for any $\omega(\log n)$, there is a negligible $\varepsilon(n)$ such that

$$\eta_\varepsilon(\Lambda) \ne \omega(\sqrt{\log n})\cdot\max_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|. \tag{39}$$

According to (35), it is easy to verify that the flatness factor $\epsilon_\Lambda(\sigma)$ is a monotonically decreasing function of σ, i.e., for $\sigma_1 \ge \sigma_2$, we have $\epsilon_\Lambda(\sigma_1) \le \epsilon_\Lambda(\sigma_2)$. Therefore, letting $\eta_\varepsilon(\Lambda) \ne \omega(\sqrt{\log n})\cdot\max_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|$ be a benchmark of comparison, we may bound the flatness factor $\epsilon_\Lambda(\sigma)$ by a negligible $\varepsilon(n)$ if $\sigma = \omega(\sqrt{\log n})\cdot\max_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|$. On the other hand, it is also easy to check that $\epsilon_{\mathbb{Z}}(\sigma_i)$ will become negligible if $\sigma_i = \omega(\sqrt{\log n})$. Hence, we have

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

for $\sigma = \omega(\sqrt{\log n})\cdot\max_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|$.

On the other hand, according to *Jacobi's formula* [37]

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

the expression of the flatness factor shown in (35) can be rewritten as

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

where Λ^* is the dual lattice of Λ. 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 $\mathbb{Z}^* = \mathbb{Z}$.

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

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

where $\widehat{\mathbf{b}}_i^*$'s are the Gram-Schmidt vectors of the dual lattice basis $\mathbf{B}^* \triangleq (\mathbf{B}^{-1})^T\mathbf{J}$ ($\mathbf{J}$ is a column-flipping matrix), then both $\epsilon_{\Lambda^*}(\frac{1}{2\pi\sigma})$ and $\epsilon_{\mathbb{Z}^*}(\frac{1}{2\pi\sigma_i})$ will be bounded by a negligible $\varepsilon(n)$. Thus, we have

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

According to (44), it follows that

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

where (f) comes from the fact that [38]

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

Therefore, the proof is completed. ■

Obviously, according to Proposition 1, as s either goes to 0 or ∞, the coefficient δ will converge to 1. This is in line with the fact that Klein's algorithm is capable of sampling from the lattice Gaussian distribution directly when $\sigma = \omega(\sqrt{\log n})\cdot\max_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|$.

Proposition 3: If $s \le \min_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|$, *then the coefficient* δ *is lower bounded by*

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

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

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

Proof: By definition, we have

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

where $\Gamma(\cdot)$ stands for the Gamma function [39]. It is worth pointing out that the explicit values of $\vartheta_3(2)$, $\vartheta_3(3)$, ... can also be calculated [40], 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. As the Jacobi theta function $\vartheta_3(\tau)$ is monotonically decreasing with τ, let $1/s_i^2 \geq 1$, i.e., $s \leq \|\widehat{\mathbf{b}}_i\|$, then it follows that

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

Assume $s \leq \min_{1\leq i\leq n} \|\widehat{\mathbf{b}}_i\|$, then the following lower bound for δ 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 Jacobi's formula shown in (41), δ 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. ∎

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 δ.

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 4: 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* δ *is lower bounded by*

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

where I *denotes the subset of indexes* i *with* $s_i > 1$ *(i.e.,* $s > \|\widehat{\mathbf{b}}_i\|$*),* $i \in \{1, 2, \ldots, n\}$, $|I| = m$.

Proof: From the definition, we have

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

TABLE I

LOWER BOUNDS ON δ WITH RESPECT TO $s = \sqrt{2\pi}\sigma$ IN THE INDEPENDENT MHK ALGORITHM

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

where (g) 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

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

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

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

completing the proof. ∎

To summarize, the value of δ with respect to the given $s = \sqrt{2\pi}\sigma$ in the independent MHK algorithm is given in Table I.

Now, let us consider some lattices whose theta series are more understood. We have the following property for an isodual lattice, which is one that is geometrically similar to its dual [37].

Proposition 5: 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* Λ *has a multiplicative symmetry point at* $s = 1$, *and asymptotically converges to 1 on both sides when* s *either goes to* 0 *or* ∞.

Proof: Here, we note that the theta series Θ_Λ of an isodual lattice Λ and that of its dual Λ^* 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,

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

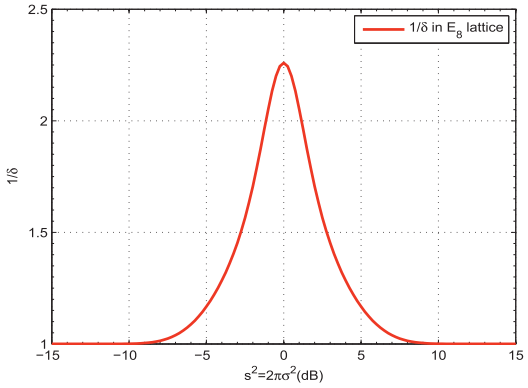

Fig. 3. Coefficient $1/\delta$ of the E_8 lattice in the case of $\mathbf{c} = \mathbf{0}$.

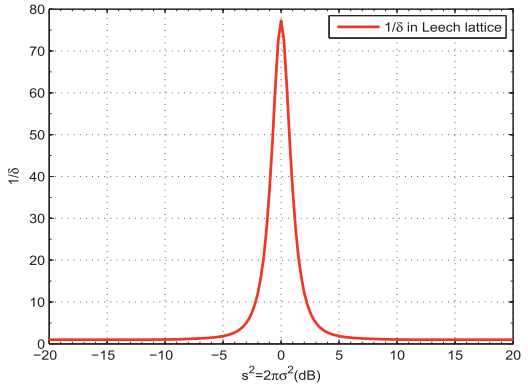

Fig. 4. Coefficient $1/\delta$ of the Leech lattice in the case of $\mathbf{c} = \mathbf{0}$.

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 $s = 1$, i.e., $\sigma^2 = \frac{1}{2\pi}$. Actually, $1/\delta$ is similar to, but not exactly the same as the *secrecy gain* defined in [37]. 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 δ, the explicit estimation of the 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 δ' as

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

then we have the following proposition.

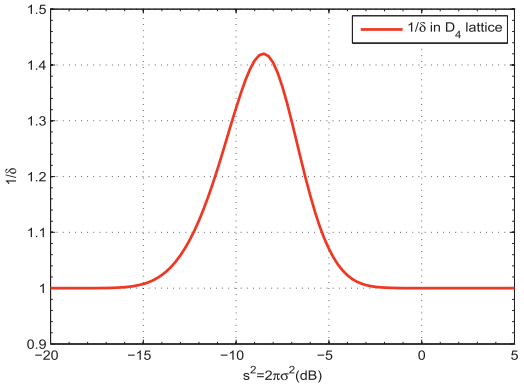

Fig. 5. Coefficient $1/\delta$ of the D_4 lattice in the case of $\mathbf{c} = \mathbf{0}$.

Proposition 6: For any $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{c} \neq \mathbf{0}$, one has

$$\delta' \geq e^{-\frac{d^2(\Lambda,\mathbf{c})}{2\sigma^2}} \cdot \delta \tag{64}$$

where δ is exponential decay coefficient for the case $\mathbf{c} = \mathbf{0}$.

Proof: Let $\mathbf{c}' = \mathbf{c} \mod \Lambda$ stand for the modular operation of $\mathbf{c}$ over lattice Λ. Then it follows that

$$\begin{aligned}
\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda}) &= \sum_{\mathbf{z}\in\mathbf{\Lambda}} e^{-\frac{1}{2\sigma^2}\|\mathbf{z}-\mathbf{c}\|^2} \\
&= \sum_{\mathbf{z}\in\mathbf{\Lambda}} e^{-\frac{1}{2\sigma^2}\|\mathbf{z}-\mathbf{c}'\|^2} \\
&= e^{-\frac{\|\mathbf{c}'\|^2}{2\sigma^2}} \cdot \sum_{\mathbf{z}\in\mathbf{\Lambda}} e^{-\frac{\|\mathbf{z}\|^2}{2\sigma^2}} \cdot \frac{1}{2} \cdot \left(e^{-\frac{1}{\sigma^2}\langle \mathbf{z},\mathbf{c}'\rangle} + e^{\frac{1}{\sigma^2}\langle \mathbf{z},\mathbf{c}'\rangle} \right) \\
&\overset{(h)}{\geq} e^{-\frac{\|\mathbf{c}'\|^2}{2\sigma^2}} \cdot \sum_{\mathbf{z}\in\mathbf{\Lambda}} e^{-\frac{\|\mathbf{z}\|}{2\sigma^2}} \\
&= e^{-\frac{d^2(\Lambda,\mathbf{c})}{2\sigma^2}} \cdot \rho_\sigma(\mathbf{\Lambda}),
\end{aligned} \tag{65}$$

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

Thus, the value of δ' is reduced by a factor of $e^{-\frac{d^2(\Lambda,\mathbf{c})}{2\sigma^2}}$ from δ. Clearly, if $\mathbf{c} = \mathbf{0}$, then $\delta' = \delta$, implying $\mathbf{c} \neq \mathbf{0}$ is a general case of $\mathbf{c} = \mathbf{0}$.[4] Hence, according to (65), as long as $\mathbf{c}$ is not too far from Λ, δ' has a similar lower bound.

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 $q(\mathbf{x}, \mathbf{y}) = q(\mathbf{y}, \mathbf{x})$ [34]. In the proposed algorithm, we again use Klein's algorithm to generate the symmetric

[4] In fact, as $\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})$ is periodic, all $\mathbf{c} \in \Lambda$ will lead to $d(\Lambda, \mathbf{c}) = \mathbf{0}$, thus corresponding to the case of $\mathbf{c} = \mathbf{0}$.

proposal distribution. Yet, the generation of the state candidate $\mathbf{y}$ depends on the current state $\mathbf{x}$, which is different from the independent MHK algorithm. Specifically, as shown in Algorithm 3, its sampling procedure at each Markov move can be summarized by the following steps:

Algorithm 3 Symmetric Metropolis-Klein Algorithm for Lattice Gaussian Sampling

Input: $\mathbf{B}, \sigma, \mathbf{c}, \mathbf{X}_0, t_{\text{mix}}(\epsilon)$
Output: sample from a distribution statistically close to $\pi = D_{\Lambda,\sigma,\mathbf{c}}$

```
for t = 1,2, ..., do
    let x denote the state of X_{t-1}
    generate y by the proposal distribution q(x, y) in (66)
    calculate the acceptance ratio α(x, y) in (67)
    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 t ≥ t_mix(ε) then
        output the state of X_t
    end if
end for
```

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 $\mathbf{X}_{t+1}$,*

$$q(\mathbf{x},\mathbf{y}) = \frac{\rho_{\sigma,\mathbf{Bx}}(\mathbf{By})}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{By}\|^2}}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} \overset{(i)}{=} q(\mathbf{y},\mathbf{x}), \quad (66)$$

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 (i) holds due to the inherent symmetry (see Lemma 2 in the following).

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

$$\begin{aligned}\alpha(\mathbf{x},\mathbf{y}) &= \min\left\{1, \frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}\right\} = \min\left\{1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right\} \\ &= \min\left\{1, e^{\frac{1}{2\sigma^2}\left(\|\mathbf{Bx}-\mathbf{c}\|^2 - \|\mathbf{By}-\mathbf{c}\|^2\right)}\right\}, \end{aligned} \quad (67)$$

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

3) *With probability $\alpha(\mathbf{x},\mathbf{y})$ accept $\mathbf{X}_{t+1} = \mathbf{y}$; otherwise, reject $\mathbf{y}$ and let $\mathbf{X}_{t+1} = \mathbf{x}$.*

Lemma 2: The proposal distribution q shown in (66) is symmetric and only depends on $\mathbf{x}-\mathbf{y}$, namely,

$$q(\mathbf{x},\mathbf{y}) = q(\mathbf{y},\mathbf{x}) = q(\mathbf{x}-\mathbf{y}) \quad (68)$$

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

The proof of Lemma 2 is provided in Appendix VI. Such a special case is called the "random-walk" Metropolis-Hastings algorithm [27].

At each Markov move, the state candidate $\mathbf{y}$ for $\mathbf{X}_{t+1}$ is sampled from a Gaussian-like distribution centered at the current state $\mathbf{x}$. Since the chain is symmetric, the calculation of the acceptance ratio α is greatly simplified. From (67), 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 $\mathbf{X}_{t+1}$ since $\alpha = 1$; otherwise it will be accepted with a probability depending on the distance from $\mathbf{By}$ to $\mathbf{c}$, thus forming a Markov chain.[5]

Again, we recall the following standard result (see, e.g., [29] for a proof).

Proposition 7: Given the target lattice Gaussian distribution $\pi = D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by the proposed symmetric Metropolis-Klein algorithm is ergodic:

$$\lim_{t\to\infty} \|P^t(\mathbf{x};\cdot) - D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV} = 0 \quad (69)$$

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

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 υ on Ω such that

$$P^k(\mathbf{x},\mathcal{B}) \geq \delta\upsilon(\mathcal{B}), \quad \forall \mathbf{x} \in C \quad (70)$$

for all measurable subsets $\mathcal{B} \subseteq \Omega$. This is also known as the *minorisation condition* in literature [32]. Actually, uniform ergodicity is a special case where the minorisation condition is satisfied with $C = \Omega$. For a bounded small set C, the *drift condition* of discrete state space Markov chains is defined as follows [27]:

Definition 4: 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

$$\sum_{\mathbf{y}\in\Omega} P(\mathbf{x},\mathbf{y})V(\mathbf{y}) \leq \lambda V(\mathbf{x}) + b\mathbf{1}_C(\mathbf{x}) \quad (71)$$

for all $\mathbf{x} \in \Omega$, where $C \subseteq \Omega$ is a small set, and the indicator function $\mathbf{1}_C(\mathbf{x}) = 1$ if $\mathbf{x} \in C$ and 0 otherwise.

It is well-known that the drift condition implies geometric ergodicity [32]. Equipped with minorisation and drift conditions, we are now in a position to prove the following theorem:

Theorem 2: Given the invariant lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain established by the symmetric Metropolis-Klein algorithm is geometrically ergodic.

Proof: The proof boils down to verifying the drift condition (71).

First of all, the distribution $\pi(\mathbf{x}) = D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$ is clearly bounded between 0 and 1 over any bounded set. Besides, for any $\|\mathbf{Bx}-\mathbf{By}\| \leq \delta_q$, where $\delta_q > 0$ is a constant, the proposal distribution $q(\mathbf{x},\mathbf{y})$ can always be lower bounded by a constant $\epsilon_q > 0$ as follows,

$$\begin{aligned} q(\mathbf{x},\mathbf{y}) &\geq \frac{e^{-\frac{\delta_q^2}{2\sigma^2}}}{\prod_{i=1}^{n}\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} \\ &\overset{(j)}{\geq} \frac{e^{-\frac{\delta_q^2}{2\sigma^2}}}{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})} = \epsilon_q, \end{aligned} \quad (72)$$

[5] 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 tedious to analyze, we omit its analysis here.

where (j) holds due to (17). Thus, by [41, Th. 2.1], every non-empty bounded set $C \subseteq \mathbb{Z}^n$ in the underlying Markov chain of the SMK algorithm is a small set. Then we may define a small set C as

$$C = \{\mathbf{x} \in \mathbb{Z}^n : \pi(\mathbf{x}) \geq \epsilon\} \tag{73}$$

for sufficiently small ϵ.

Meanwhile, at each Markov move, the acceptance ratio (67) suggests the acceptance region $A_{\mathbf{x}}$ and the potential rejection region $R_{\mathbf{x}}$ for current state $\mathbf{x}$ as follows:

$$A_{\mathbf{x}} = \{\mathbf{y} \in \mathbb{Z}^n | \pi(\mathbf{y}) \geq \pi(\mathbf{x})\}; \tag{74}$$

$$R_{\mathbf{x}} = \{\mathbf{y} \in \mathbb{Z}^n | \pi(\mathbf{y}) < \pi(\mathbf{x})\}. \tag{75}$$

Obviously, state candidate $\mathbf{y} \in A_{\mathbf{x}}$ will surely be accepted by $\mathbf{X}_{t+1}$ while state candidate $\mathbf{y} \in R_{\mathbf{x}}$ has a certain probability to be rejected. Then, the LHS of the drift condition (71) can be rewritten as (76), as shown at the bottom of this page, where the second and third terms result from whether state candidate $\mathbf{y} \in R_{\mathbf{x}}$ is accepted or rejected, respectively.

Set the potential function $V(\mathbf{x}) = \pi(\mathbf{x})^{-\frac{1}{2}}$. Dividing (76) by $V(\mathbf{x})$ on both sides, we then arrive at the results shown in (77), as shown at the bottom of this page. Furthermore, since the ratios on the RHS of (77) are at most 1, we obtain[6]

$$\frac{\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x}, \mathbf{y})V(\mathbf{y})}{V(\mathbf{x})} \leq \frac{5}{4}. \tag{78}$$

Depending on whether $\mathbf{x} \in C$ or not, the drift condition can be rewritten as

$$\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x}, \mathbf{y})V(\mathbf{y}) \leq \lambda V(\mathbf{x}) \quad \text{for } \mathbf{x} \notin C \tag{79}$$

and

$$\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x}, \mathbf{y})V(\mathbf{y}) \leq \lambda V(\mathbf{x}) + b \quad \text{for } \mathbf{x} \in C. \tag{80}$$

The two cases are illustrated in Fig. 6. We proceed case by case.

(i). In the case $\mathbf{x} \in C$,

$$V(\mathbf{x}) \leq \frac{1}{\sqrt{\epsilon}}. \tag{81}$$

By (78) we have

$$\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x}, \mathbf{y})V(\mathbf{y}) \leq \frac{1}{\sqrt{\epsilon}} \cdot \frac{5}{4} = b \quad \text{for } \mathbf{x} \in C \tag{82}$$

and thus condition (80) is satisfied.

[6]Note that $1 \leq 1 - a^2 + a \leq \frac{5}{4}$ for $0 \leq a \leq 1$.

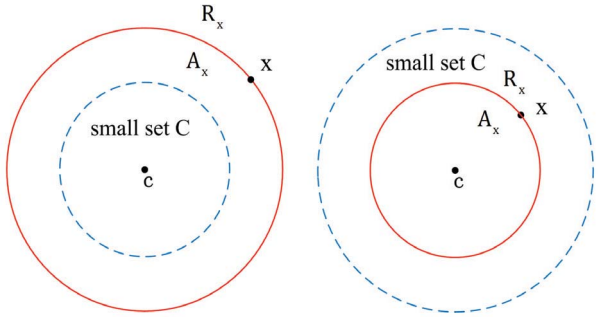

Fig. 6. Illustration of cases (a) $\mathbf{x} \notin C$ and (b) $\mathbf{x} \in C$ in the Markov move induced by SMK. The blue dash circle represents the area of the small set while the red solid circle denotes the acceptance region $A_{\mathbf{x}}$.

(ii). In the case $\mathbf{x} \notin C$, we consider

$$\lambda = \limsup_{\|\mathbf{x}\|\to\infty} \frac{\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x}, \mathbf{y})V(\mathbf{y})}{V(\mathbf{x})}. \tag{83}$$

If $\lambda < 1$, then (79) is satisfied for sufficiently small ϵ.

It is easy to verify that

$$\lim_{\|\mathbf{x}\|\to\infty} \ell(\mathbf{x}) \cdot \nabla \log \pi(\mathbf{x}) = -\infty, \tag{84}$$

where $\ell(\mathbf{x})$ denotes the unit vector $\mathbf{x}/\|\mathbf{x}\|$ and ∇ represents the gradient. This condition implies that for any $\gamma > 0$, there exists $d_\gamma > 0$ such that for $\|\mathbf{x}\| \geq d_\gamma$

$$\frac{\pi(\mathbf{x} + a \cdot \ell(\mathbf{x}))}{\pi(\mathbf{x})} \leq e^{-a\cdot\gamma}, \tag{85}$$

where $a \geq 0$ represents a constant. In other words, as $\|\mathbf{x}\|$ goes to infinity, the above ratio is at least exponentially decaying with a rate γ tending to infinity.

Let $C_\zeta = \{\mathbf{x} \in \mathbb{R}^n \mid \pi(\mathbf{x}) = \zeta\}$. We define the radial μ-zone around $C_{\pi(\mathbf{x})}$ as (See Fig. 7)

$$C_{\pi(\mathbf{x})}(\mu) = \{\mathbf{z} + s \cdot \ell(\mathbf{z}) \mid \mathbf{z} \in C_{\pi(\mathbf{x})}, -\mu \leq s \leq \mu\}.$$

Denote by $B(\mathbf{x}, K)$ a Euclidean ball of radius K, centered at $\mathbf{x}$. As in [42], for arbitrary but fixed $\epsilon_1 > 0$, choose $K > 0$ such that

$$\sum_{\substack{\mathbf{y}\in\mathbb{Z}^n \\ \mathbf{By}\notin B(\mathbf{Bx},K)}} q(\mathbf{x}, \mathbf{y}) \leq \epsilon_1. \tag{86}$$

$$\begin{aligned}\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x}, \mathbf{y})V(\mathbf{y}) &= \sum_{\mathbf{y}\in A_{\mathbf{x}}} P(\mathbf{x}, \mathbf{y})V(\mathbf{y}) + \sum_{\mathbf{y}\in R_{\mathbf{x}}} P(\mathbf{x}, \mathbf{y})V(\mathbf{y}) \\ &= \sum_{\mathbf{y}\in A_{\mathbf{x}}} q(\mathbf{x}, \mathbf{y})V(\mathbf{y}) + \sum_{\mathbf{y}\in R_{\mathbf{x}}} q(\mathbf{x}, \mathbf{y})\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}V(\mathbf{y}) + \sum_{\mathbf{y}\in R_{\mathbf{x}}} q(\mathbf{x}, \mathbf{y})\left[1 - \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right]V(\mathbf{x}) \end{aligned} \tag{76}$$

$$\frac{\sum_{\mathbf{y}\in\mathbb{Z}^n} P(\mathbf{x}, \mathbf{y})V(\mathbf{y})}{V(\mathbf{x})} = \sum_{\mathbf{y}\in A_{\mathbf{x}}} q(\mathbf{x}, \mathbf{y})\frac{\pi(\mathbf{x})^{1/2}}{\pi(\mathbf{y})^{1/2}} + \sum_{\mathbf{y}\in R_{\mathbf{x}}} q(\mathbf{x}, \mathbf{y})\left[1 - \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})} + \frac{\pi(\mathbf{y})^{1/2}}{\pi(\mathbf{x})^{1/2}}\right] \tag{77}$$

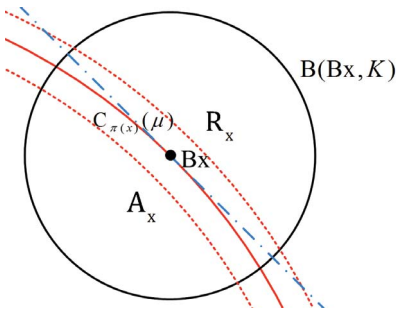

Fig. 7. Illustration of the contour $C_{\pi(\mathbf{x})}$ (solid curve), radial μ-zone $C_{\pi(\mathbf{x})}(\mu)$ (area between the two dashed curves) and ball $B(\mathbf{Bx}, K)$ in the case $\mathbf{x} \notin C$.

This can be assured by noting that

$$q(\mathbf{x},\mathbf{y}) = \frac{e^{-\frac{\|\mathbf{Bx}-\mathbf{By}\|^2}{2\sigma^2}}}{\prod_{i=1}^n \rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} \overset{(l)}{\leq} \frac{e^{-\frac{\|\mathbf{Bx}-\mathbf{By}\|^2}{2\sigma^2}}}{\prod_{i=1}^n \rho_{\sigma_i,1/2}(\mathbb{Z})}, \tag{87}$$

where (l) is because $\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})$ has a minimum at $\widetilde{y}_i = 1/2$ and then applying a tail bound of lattice Gaussian distribution [1, Lemma 1.5].

From the fact that the Euclidean norms $\{\|\mathbf{Bx}\|,\ \mathbf{x} \in \mathbb{Z}^n\}$ of a lattice are discrete, it follows that for any $K > 0$ there exists $\mu > 0$ such that

$$\limsup_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in\mathbb{Z}^n \cap C_{\pi(\mathbf{x})}(\mu)\mathbf{By}\in B(\mathbf{Bx},K)} q(\mathbf{x},\mathbf{y}) = \limsup_{\|\mathbf{x}\|\to\infty} \sum_{\substack{\mathbf{y}\in\mathbb{Z}^n,\|\mathbf{By}\|=\|\mathbf{Bx}\| \\ \mathbf{By}\in B(\mathbf{Bx},K)}} q(\mathbf{x},\mathbf{y}). \tag{88}$$

In words, one may choose small enough μ such that only those lattice points of the same norm $\|\mathbf{x}\|$ count.

Substitute (83) into (77) and rearrange it as

$$\lambda = \limsup_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in R_{\mathbf{x}}} q(\mathbf{x},\mathbf{y}) + \sum_{\mathbf{y}\in A_{\mathbf{x}}} q(\mathbf{x},\mathbf{y}) \frac{\pi(\mathbf{x})^{1/2}}{\pi(\mathbf{y})^{1/2}} + \sum_{\mathbf{y}\in R_{\mathbf{x}}} q(\mathbf{x},\mathbf{y}) \left[-\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})} + \frac{\pi(\mathbf{y})^{1/2}}{\pi(\mathbf{x})^{1/2}} \right]. \tag{89}$$

We will keep the first term and bound the sum of the other two. To do so, we consider three regions:

1) By $\notin$ B(Bx, K). Since all the ratios in (89) are at most 1 (in fact $0 \leq -\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})} + \frac{\pi(\mathbf{y})^{1/2}}{\pi(\mathbf{x})^{1/2}} \leq \frac{1}{4}$ for $\mathbf{y} \in R_{\mathbf{x}}$), the sum of the last two terms in (89) is upper bounded by ϵ_1 by choosing K such that (86) holds;
2) By $\in$ B(Bx, K) but $\mathbf{y} \notin C_{\pi(\mathbf{x})}(\mu)$. As $\|\mathbf{x}\| \to \infty$, all the ratios in (89) tend to 0 outside of any radial μ-zone for any K (cf. (85)). Thus the sum of the last two terms in (89) can be bounded by some ϵ_2 in this region;
3) By $\in$ B(Bx, K) and $\mathbf{y} \in C_{\pi(\mathbf{x})}(\mu)$. Again, since all the ratios in (89) are at most 1, the limit of the sum of the last two terms in (89) is given by (88).

In words, as $\|\mathbf{x}\| \to \infty$, only those lattice points of the same norm $\|\mathbf{x}\|$ in the last region count, when one evaluates the second and third sums of (89).

For notational convenience, define two regions $\underline{A}_{\mathbf{x}} = \{\mathbf{y} \in \mathbb{Z}^n | \pi(\mathbf{y}) > \pi(\mathbf{x})\}$ and $\overline{R}_{\mathbf{x}} = \{\mathbf{y} \in \mathbb{Z}^n | \pi(\mathbf{y}) \leq \pi(\mathbf{x})\}$, which are slightly different from (74), (75), i.e., $\underline{A}_{\mathbf{x}}$ does not include the boundary but $\overline{R}_{\mathbf{x}}$ does. Then we arrive at

$$\begin{aligned} \lambda &\leq \limsup_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in\overline{R}_{\mathbf{x}}} q(\mathbf{x},\mathbf{y}) \\ &= 1 - \liminf_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in\underline{A}_{\mathbf{x}}} q(\mathbf{x},\mathbf{y}) \\ &\overset{(k)}{<} 1 \end{aligned} \tag{90}$$

where inequality (k) holds because

$$\liminf_{\|\mathbf{x}\|\to\infty} \sum_{\mathbf{y}\in\underline{A}_{\mathbf{x}}} q(\mathbf{x},\mathbf{y}) > 0 \tag{91}$$

due to symmetry of $q(\mathbf{x},\mathbf{y})$. In fact, as shown in Fig. 7, it follows from the symmetry (i.e., depicted by the blue dash dot line) that

$$\sum_{\mathbf{y}\in\underline{A}_{\mathbf{x}}} q(\mathbf{x},\mathbf{y}) < \frac{1}{2} < \sum_{\mathbf{y}\in\overline{R}_{\mathbf{x}}} q(\mathbf{x},\mathbf{y}), \tag{92}$$

and the two probabilities can approach $\frac{1}{2}$ as $\|\mathbf{x}\| \to \infty$. This completes the proof in the case $\mathbf{x} \notin C$. ■

In essence, the convergence of geometric ergodicity can be divided into two stages. On one hand, if $\mathbf{x} \notin C$, the drift condition guarantees the Markov chain shrinks geometrically towards the small set C. On the other hand, if $\mathbf{x} \in C$, the minorisation condition shown in (70) implies the Markov chain will converge to the stationary distribution exponentially fast. This can be demonstrated by using the coupling technique as in the previous section and δ is just the exponential decay coefficient, which depends on C. It was shown in [43] 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 follows

$$\|P^n(\mathbf{x}_0,\cdot) - \pi(\cdot)\|_{TV} \leq (1-\delta)^{rn} + \left(\frac{U^r}{\alpha^{1-r}}\right)^n \times \left(1 + \frac{b}{1-\lambda} + V(\mathbf{x}_0)\right), \tag{93}$$

where $0 < r < 1$,

$$\alpha = \frac{1+d}{1+2b+\lambda d} \quad \text{and} \quad U = 1 + 2(d+b). \tag{94}$$

Clearly, there is a trade-off between these two convergence stages: a larger set C indicates a smaller δ in the minorisation condition for $\mathbf{x} \in C$ but a faster shrink speed λ 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 δ and λ sensitive to a slight change of d. Moreover, a closed-form expression of λ is difficult to get even for a specific C. Therefore, although geometric ergodicity can be achieved by the proposed SMK algorithm, it is difficult to obtain quantitative bounds on δ and λ.

Finally, (93) indicates that the convergence of the Markov chain arising from the SMK algorithm also highly depends on the starting state $\mathbf{x}_0$, which follows the definition of geometric

ergodicity given in (9). 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, starting the Markov chain with $\mathbf{x}_0$ as close to the center of the distribution as possible would be a judicious choice. This is actually in accordance with the result shown in (93), implying the closest point to $\mathbf{c}$ is the optimal choice. As a simple solution, Babai's nearest plane algorithm is recommended here to output $\mathbf{x}_0$ [44].

VI. Conclusions

In this paper, two MH-based algorithms were proposed to sample from lattice Gaussian distributions. As the proposal distribution in the MH algorithms can be set freely, an independent proposal distribution and a symmetric proposal distribution were exploited respectively for geometric convergence. In addition, it was proven that the Markov chain arising from the independent MHK algorithm is uniformly ergodic, leading to exponential convergence regardless of the starting state. We showed its convergence rate can be explicitly calculated via theta series, making the mixing time predictable. On the other hand, the proposed SMK algorithm was demonstrated to be geometrically ergodic, where the selection of the starting state matters. Due to its inherent symmetry, it not only converges exponentially fast, but also is simple to implement.

Appendix A
Proof of Inequality in (26)

Proof: To start with, let us recall the definition of *conductance* (also known as *bottleneck ratio*) in Markov chains [29].

Definition 5: The conductance Φ *of a Markov chain is defined as*

$$\Phi(S) = \min_{S\subseteq\Omega,\pi(S)\le 1/2} \frac{Q(S,S^c)}{\pi(S)}, \tag{95}$$

where subset S^c *stands for the complement set of* S *(i.e.,* $S\bigcup S^c = \Omega$, $S\bigcap S^c = \emptyset$*), and the edge measure* Q *is defined by*

$$Q(x,y) = \pi(x)P(x,y) \tag{96}$$

and

$$Q(S,S^c) = \sum_{x\in S, y\in S^c} Q(x,y). \tag{97}$$

It is this value $0 < \Phi \le 1$ that has been used to bound the spectral gap γ of Markov chains. More precisely, in the independent MHK algorithm, we have

$$\begin{aligned}
\Phi &= \min_{S\subseteq\Omega,\pi(S)\le 1/2} \frac{\sum_{\mathbf{x}\in S,\mathbf{y}\in S^c}\pi(\mathbf{x})P(\mathbf{x},\mathbf{y})}{\pi(S)} \\
&\overset{(m)}{\ge} \min_{S\subseteq\Omega,\pi(S)\le 1/2} \frac{\sum_{\mathbf{x}\in S,\mathbf{y}\in S^c}\pi(\mathbf{x})\cdot\delta\pi(\mathbf{y})}{\pi(S)} \\
&= \min_{S\subseteq\Omega,\pi(S)\le 1/2} \frac{\delta\cdot\sum_{\mathbf{x}\in S}\pi(\mathbf{x})\cdot\sum_{\mathbf{y}\in S^c}\pi(\mathbf{y})}{\pi(S)} \\
&= \min_{S\subseteq\Omega,\pi(S)\le 1/2} \delta\cdot\pi(S^c) \\
&\ge \frac{\delta}{2},
\end{aligned} \tag{98}$$

where inequality (m) holds due to (20).

Next, by invoking the *cheeger inequality* [45] of Markov chains

$$\frac{\Phi^2}{2} \le \gamma \le 2\Phi, \tag{99}$$

we have

$$\gamma \ge \frac{\delta^2}{8}, \tag{100}$$

completing the proof. ∎

Appendix B
Proof of Lemma 2

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

$$q(\mathbf{x},\mathbf{y}) = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{By}\|^2}}{\prod_{i=1}^n \rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Rx}-\mathbf{Ry}\|^2}}{\prod_{i=1}^n \rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z})} \tag{101}$$

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 (101) can be expressed as

$$\begin{aligned}
\rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z}) &= \sum_{z_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(z_i - \frac{c_i' - \sum_{j=i+1}^n r_{i,j}y_j}{r_{i,i}})^2} \\
&= \sum_{z_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(z_i - \frac{\sum_{j=i}^n r_{i,j}x_j - \sum_{j=i+1}^n r_{i,j}y_j}{r_{i,i}})^2} \\
&= \sum_{z_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(x_i - z_i + \sum_{j=i+1}^n \frac{r_{i,j}}{r_{i,i}}(x_j - y_j))^2} \\
&= \sum_{z_i'\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(z_i' - \phi)^2} \\
&= \rho_{\sigma_i,\phi}(\mathbb{Z}),
\end{aligned} \tag{102}$$

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

Similarly, we can easily get that

$$\begin{aligned}
\rho_{\sigma_i,\widetilde{x}_i}(\mathbb{Z}) &= \sum_{z_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(y_i - z_i + \sum_{j=i+1}^n \frac{r_{i,j}}{r_{i,i}}(y_j - x_j))^2} \\
&= \sum_{z_i'\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}(z_i' - \phi)^2} \\
&= \rho_{\sigma_i,\phi}(\mathbb{Z}) = \rho_{\sigma_i,\widetilde{y}_i}(\mathbb{Z}),
\end{aligned} \tag{103}$$

where $\widetilde{x}_i = \frac{c_i'' - \sum_{j=i+1}^n r_{i,j}x_j}{r_{i,i}}$, $\mathbf{c}'' = \mathbf{Ry}$. Therefore, we have $q(\mathbf{x},\mathbf{y}) = q(\mathbf{y},\mathbf{x})$.

In fact, (102) shows that $q(\mathbf{x},\mathbf{y})$ is a function of $\mathbf{x}-\mathbf{y}$ only; moreover, since $\rho_{\sigma_i,\phi}(\mathbb{Z})$ is even in ϕ, $q(\mathbf{x},\mathbf{y}) = q(\mathbf{x}-\mathbf{y}) = q(\mathbf{y}-\mathbf{x})$, completing the proof. ∎

Acknowledgment

The authors would like to thank Damien Stehlé, Guillaume Hanrot and Antonio Campello for fruitful discussions.

References

[1] W. Banaszczyk, "New bounds in some transference theorems in the geometry of numbers," *Math. Ann.*, vol. 296, no. 1, pp. 625–635, 1993.

[2] G. D. Forney, "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. Inf. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.
[4] C. Ling and J.-C. Belfiore, "Achieving AWGN Channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. 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. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.
[6] H. Mirghasemi and J.-C. Belfiore, "Lattice code design criterion for MIMO wiretap channels," in *Proc. IEEE Inf. Theory Workshop—Fall (ITW)*, Oct. 2015, pp. 277–281.
[7] S. Vatedka, N. Kashyap, and A. Thangaraj, "Secure compute-and-forward in a bidirectional relay," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2531–2556, May 2015.
[8] D. Micciancio and O. Regev, "Worst-case to average-case reductions based on Gaussian measures," in *Proc. Ann. Symp. Found. Comput. Sci.*, Rome, Italy, Oct. 2004, pp. 372–381.
[9] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2009.
[10] D. Aggarwal, D. Dadush, O. Regev, and N. Stephens-Davidowitz, "Solving the shortest vector problem in 2^n time using discrete Gaussian sampling," in *Proc. STOC*, 2015, pp. 733–742.
[11] D. Aggarwal, D. Dadush, and N. Stephens-Davidowitz, "Solving the closest vector problem in 2^n time—The discrete Gaussian strikes again!" in *Proc. FOCS*, Oct. 2015, pp. 563–582.
[12] S. Liu, C. Ling, and D. Stehlé, "Decoding by sampling: A randomized lattice algorithm for bounded distance decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5933–5945, Sep. 2011.
[13] 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.
[14] N. Stephens-Davidowitz. (Apr. 2016). "Discrete Gaussian sampling reduces to CVP and SVP." [Online]. Available: https://arxiv.org/abs/1506.07490
[15] P. Klein, "Finding the closest lattice vector when it's unusually close," in *Proc. ACM-SIAM Symp. Discrete Algorithms*, 2000, pp. 937–941.
[16] C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapdoors for hard lattices and new cryptographic constructions," in *Proc. 40th Annu. ACM Symp. Theory Comput.*, Victoria, BC, Canada, 2008, pp. 197–206.
[17] Z. Wang, C. Ling, and G. Hanrot, "Markov chain Monte Carlo algorithms for lattice Gaussian sampling," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun./Jul. 2014, pp. 1489–1493.
[18] B. Hassibi, M. Hansen, A. G. Dimakis, H. A. J. 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 Trans. Signal Process.*, vol. 62, no. 17, pp. 4436–4450, Sep. 2014.
[19] T. Datta, N. A. Kumar, A. Chockalingam, and B. S. Rajan, "A novel Monte-Carlo-sampling-based receiver for large-scale uplink multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3019–3038, Sep. 2013.
[20] 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, May 2006.
[21] R. Chen, J. S. Liu, and X. Wang, "Convergence analyses and comparisons of Markov chain Monte Carlo algorithms in digital communications," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 255–270, Feb. 2002.
[22] P. Aggarwal and X. Wang, "Multilevel sequential Monte Carlo algorithms for MIMO demodulation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 750–758, Feb. 2007.
[23] H. Zhu, B. Farhang-Boroujeny, and R.-R. Chen, "On performance of sphere decoding and Markov chain Monte Carlo detection methods," *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 669–672, Oct. 2005.
[24] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970.
[25] A. Doucet and X. Wang, "Monte Carlo methods for signal processing: A review in the statistical signal processing context," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 152–170, Nov. 2005.
[26] Z. Wang and C. Ling. (Apr. 2017). "Lattice Gaussian sampling by Markov chain Monte Carlo: Convergence rate and decoding complexity." [Online] Available: https://arxiv.org/abs/1704.02673
[27] G. O. Roberts and J. S. Rosenthal, "General state space Markov chains and MCMC algorithms," *Probab. Surv.*, vol. 1, pp. 20–71, 2004.
[28] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, no. 4, pp. 515–534, Dec. 1982.
[29] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*. Providence, RI, USA: AMS, 2008.
[30] J. S. Rosenthal, "Optimal proposal distributions and adaptive MCMC," *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. Jones, X.-L. Meng, Eds. Boca Raton, FL, USA: CRC Press, 2010.
[31] D. Randall, "Rapidly mixing Markov chains with applications in computer science and physics," *Comput. Sci. Eng.*, vol. 8, no. 2, pp. 30–41, Mar. 2006.
[32] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[33] S. F. Jarner and G. O. Roberts, "Polynomial convergence rates of Markov chains," *Ann. Appl. Probab.*, vol. 12, no. 1, pp. 224–247, 2002.
[34] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1091, 1953.
[35] K. L. Mengersen and R. L. Tweedie, "Rates of convergence of the hastings and metropolis algorithms," *Ann. Statist.*, vol. 24, no. 1, pp. 101–121, 1996.
[36] J. H. Conway and N. A. Sloane, *Sphere Packings, Lattices and Groups*. New York, NY, USA: Springer-Verlag, 1998.
[37] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5690–5708, Oct. 2016.
[38] C. Ling, "On the proximity factors of lattice reduction-aided decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2795–2808, Jun. 2011.
[39] J. Yi, "Theta-function identities and the explicit formulas for theta-function and their applications," *Proc. J. Math. Anal. Appl.*, vol. 292, no. 2, pp. 381–400, 2004.
[40] B. C. Berndt, *Ramanujan's Notebooks: Part V*. New York, NY, USA: Springer-Verlag, 1998.
[41] 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, Mar. 1996.
[42] S. F. Jarner and E. Hansen, "Geometric ergodicity of Metropolis algorithms," *Stochastic Process. Appl.*, vol. 85, no. 2, pp. 341–361, 2000.
[43] J. S. Rosenthal, "Minorization conditions and convergence rates for Markov chain Monte Carlo," *J. Amer. Statist. Assoc.*, vol. 90, no. 430, pp. 558–566, 1995.
[44] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.
[45] A. Sinclair, "Improved bounds for mixing rates of Markov chains and multicommodity flow," *Combinat. Probab. Comput.*, vol. 1, no. 4, pp. 351–370, 1992.

Zheng Wang received the B.S. degree in electronic and information engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009, and the M.S. degree in communications from the Department of Electrical and Electronic Engineering, University of Manchester, Manchester, U.K., in 2010. He received the Ph.D degree in communication engineering from Imperial College London, UK, in 2015.

From 2015 to 2016 he served as a Research Associate at Imperial College London, UK and from 2016 to 2017 he was an senior engineer with Radio Access Network R&D division, Huawei Technologies Co. He is currently an Assistant Professor at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. His current research interests include lattice methods for wireless communications, cognitive radio and physical layer security.

Cong Ling received the B.S. and M.S. degrees in electrical engineering from the Nanjing Institute of Communications Engineering, Nanjing, China, in 1995 and 1997, respectively, and the Ph.D. degree in electrical engineering from the Nanyang Technological University, Singapore, in 2005.

He is currently a Senior Lecturer in the Electrical and Electronic Engineering Department at Imperial College London. His research interests are coding, signal processing, and security. Before joining Imperial College, he had been on the faculties of Nanjing Institute of Communications Engineering and King's College.

Dr. Ling is an Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS. He has also served as an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.