Further Results on Independent Metropolis-Hastings-Klein Sampling

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

***Abstract*—Sampling from a lattice Gaussian distribution is emerging as an important problem in coding and cryptography. This paper gives a further analysis of the independent Metropolis-Hastings-Klein (MHK) algorithm we presented at ISIT 2015. We derive the exact spectral gap of the induced Markov chain, which dictates the convergence rate of the independent MHK algorithm. Then, we apply the independent MHK algorithm to lattice decoding and obtained the decoding complexity for solving the CVP as $\tilde{O}(e^{\|\mathbf{Bx}-\mathbf{c}\|^2/\min_i \|\widehat{\mathbf{b}}_i\|^2})$. Finally, the trade-off between decoding radius and complexity is also established.**

Keywords: Lattice Gaussian sampling, MCMC methods, Metropolis-Hastings algorithm, closest vector problem.

I. Introduction

As a core problem of lattice theory, the closest vector problem (CVP) has drawn a lot of attention. The CVP is normally solved by sphere decoding (SD) or its variants. However, SD is inapplicable in high dimensions due to the exponentially increased complexity. To this end, Klein introduced a sampling algorithm to solve the CVP [1], which performs the decoding by sampling over a Gaussian-like distribution. Klein showed its decoding complexity is $O(n^{\|\mathbf{Bx}-\mathbf{c}\|^2/\min_i \|\widehat{\mathbf{b}}_i\|^2})$, where $\|\mathbf{Bx}-\mathbf{c}\|$ denotes the Euclidean distance from the query point $\mathbf{c}$ to the lattice and $\widehat{\mathbf{b}}_i$'s are the Gram-Schmidt vectors of the lattice basis $\mathbf{B}$. In [2], [3], improved sampling algorithms were proposed for near-optimal decoding performance. Nevertheless, these algorithms only yield an approximately Gaussian distribution.

As a basic scheme in Markov chain Monte Carlo (MCMC), Gibbs sampling has been introduced to lattice decoding, which employs univariate conditional sampling to build a Markov chain [4]–[9]. As the Markov chain converges to the stationary distribution, the optimal solution of the CVP can be encountered by sampling. However, to the best of our knowledge, the analysis of the convergence rate for Gibbs sampling in these applications is still lacking.

In [10], we proposed independent Metropolis-Hastings-Klein (MHK) sampling and derived its rate of convergence using the conventional coupling technique. In this paper, we further study the independent MHK algorithm and examine its complexity in solving the CVP. We derive the exact spectral gap of the transition matrix, thus precisely determine the convergence rate of the underlying Markov chain. Moreover, we show its decoding complexity for solving the CVP is $\tilde{O}(e^{\|\mathbf{Bx}-\mathbf{c}\|^2/\min_i \|\widehat{\mathbf{b}}_i\|^2})$. We demonstrate that the independent MHK sampling decoder enjoys a flexible trade-off between decoding radius and complexly.

The rest of this paper is organized as follows. Section II gives a briefly review of the independent MHK sampling algorithm. In Section III, we derive the spectral gap of the Markov chain associated with the independent MHK algorithm. In Section IV, we examine the decoding complexity and establish the flexible decoding trade-off. Finally, Section V concludes the paper.

II. Independent MHK Algorithm

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

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

We define the Gaussian function centered at $\mathbf{c}$ 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 $\mathbf{Bx} \in \Lambda$, where $\rho_{\sigma,\mathbf{c}}(\Lambda) \triangleq \sum_{\mathbf{Bx}\in\Lambda} \rho_{\sigma,\mathbf{c}}(\mathbf{Bx})$ is just a scaling to yield a probability distribution. Intuitively, the lattice Gaussian distribution can be used to sample the closest lattice point. The closer $\mathbf{Bx}$ to $\mathbf{c}$, the larger probability it will be sampled, indicating that the optimal solution of $\mathbf{x}$ is the most likely to be sampled.

From the MCMC perspective, the discrete Gaussian distribution can be viewed as a complex target distribution lacking of direct sampling methods. In order to obtain samples from $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$, the independent MHK sampling was proposed in [10]. Specifically, in the independent MHK sampling, a state candidate $\mathbf{y}$ for the next Markov move $\mathbf{X}_{t+1}$ is generated

978-1-5090-1806-2/16/$31.00 ©2016 IEEE

by Klein's algorithm [1], via the following backward one-dimensional conditional sampling:

$$\begin{aligned} P(y_i|\overline{\mathbf{y}}_{[-i]}) &= P(y_i|y_{i+1},\ldots,y_n) \\ &= \frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{c}}'-\overline{\mathbf{R}}\overline{\mathbf{y}}\|^2}}{\sum_{y_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{c}}'-\overline{\mathbf{R}}\overline{\mathbf{y}}\|^2}} \\ &= \frac{e^{-\frac{1}{2\sigma^2}\|c_i'-\sum_{j=i}^n r_{i,j}y_j\|^2}}{\sum_{y_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}\|c_i'-\sum_{j=i}^n r_{i,j}y_j\|^2}} \\ &= \frac{e^{-\frac{1}{2\sigma_i^2}\|y_i-\widetilde{y}_i\|^2}}{\sum_{y_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|y_i-\widetilde{y}_i\|^2}} \\ &= D_{\mathbb{Z},\sigma_i,\widetilde{y}_i}(y_i) \end{aligned} \tag{4}$$

where $\widetilde{y}_i = \frac{c_i'-\sum_{j=i+1}^n r_{i,j}y_j}{r_{i,i}}$, $\sigma_i = \frac{\sigma}{|r_{i,i}|}$, $\mathbf{c}' = \mathbf{Q}^\dagger\mathbf{c}$ and $\mathbf{B} = \mathbf{QR}$. Note that $\overline{\mathbf{y}}_{[-i]} = [y_{i+1},\ldots,y_n]$, $\overline{\mathbf{R}}$, $\overline{\mathbf{c}}'$ and $\overline{\mathbf{y}}$ are the $(n-i+1)$ segments of $\mathbf{R}$, $\mathbf{c}'$ and $\mathbf{y}$ respectively (i.e., $\overline{\mathbf{R}}$ is a $(n-i+1)\times(n-i+1)$ submatrix of $\mathbf{R}$ with $r_{i,i}$ to $r_{n,n}$ in the diagonal).

Given the current state $\mathbf{x}$ of the Markov chain $\mathbf{X}_t$, the proposal distribution $q(\mathbf{x},\mathbf{y})$ in the independent MHK sampling is given by

$$\begin{aligned} q(\mathbf{x},\mathbf{y}) &= \prod_{i=1}^n P(y_{n+1-i}|\overline{\mathbf{y}}_{[-(n+1-i)]}) \\ &= \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{By})}{\prod_{i=1}^n \rho_{\sigma_{n+1-i},\widetilde{y}_{n+1-i}}(\mathbb{Z})}. \end{aligned} \tag{5}$$

Actually, the proposal distribution is independent of $\mathbf{x}$ from $\mathbf{X}_t$, that is

$$q(\mathbf{x},\mathbf{y}) = q(\mathbf{y}), \tag{6}$$

implying the connection between two consecutive Markov moves is only due to following the decision stage.

More precisely, with the state candidate $\mathbf{y}$, the acceptance ratio α is computed by

$$\begin{aligned} \alpha(\mathbf{x},\mathbf{y}) &= \min\left\{1, \frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}\right\} \\ &= \min\left\{1, \frac{\prod_{i=1}^n \rho_{\sigma_{n+1-i},\widetilde{y}_{n+1-i}}(\mathbb{Z})}{\prod_{i=1}^n \rho_{\sigma_{n+1-i},\widetilde{x}_{n+1-i}}(\mathbb{Z})}\right\}, \end{aligned} \tag{7}$$

where $\widetilde{x}_i = \frac{c_i'-\sum_{j=i+1}^n r_{i,j}x_j}{r_{i,i}}$ and $\pi = D_{\Lambda,\sigma,\mathbf{c}}$ (these notations will be followed throughout the context). Then, $\mathbf{y}$ 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, which is summarized in Algorithm 1. Note that in MH algorithms, the proposal distribution $q(\mathbf{x},\mathbf{y})$ can be any fixed distribution from which we can conveniently draw samples. Meanwhile, the initial state $\mathbf{x}_0$ for $\mathbf{X}_0$ can be chosen from $\mathbb{Z}^n$ arbitrarily or from the output of a suboptimal algorithm.

III. Spectral Gap

The proof of the following theorem was given in [10] using the *coupling technique*. Here, we reformulate it in terms of the *spectral gap* of the Markov chain and give a more straightforward proof following [11].

Algorithm 1 Independent Metropolis-Hastings-Klein Sampling Algorithm

Input: $\mathbf{B},\sigma,\mathbf{c},\mathbf{x}_0$;
Output: $\mathbf{x}\sim D_{\Lambda,\sigma,\mathbf{c}}$;

```
let X_0 = x_0
for t = 1,2, ..., do
    let x denote the state of X_{t-1}
    sample y from the proposal distribution q(x, y) in (5)
    calculate the acceptance ratio α(x, y) in (7)
    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
    output x = X_t
end for
```

Theorem 1 (Uniform ergodicity of independent MHK)**.** *Given the invariant lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by independent MHK sampling converges exponentially fast to the stationary distribution in total variational distance:*

$$\|P^t(\mathbf{x},\cdot) - D_{\Lambda,\sigma,\mathbf{c}}(\cdot)\|_{TV} \le (1-\delta)^t, \tag{8}$$

where δ is a lower bound on the spectral gap to be given in Lemma 1.

Proof. From (5) and (7), the transition probability $P(\mathbf{x},\mathbf{y})$ of each Markov move in the independent MHK sampling is given by

$$P(\mathbf{x},\mathbf{y}) = \begin{cases} \min\left\{q(\mathbf{y}), \frac{\pi(\mathbf{y})q(\mathbf{x})}{\pi(\mathbf{x})}\right\} & \text{if } \mathbf{y}\neq\mathbf{x}, \\ q(\mathbf{x}) + \sum\limits_{\mathbf{z}\neq\mathbf{x}} \max\left\{0, q(\mathbf{z}) - \frac{\pi(\mathbf{z})q(\mathbf{x})}{\pi(\mathbf{x})}\right\} & \text{if } \mathbf{y}=\mathbf{x}. \end{cases} \tag{9}$$

For notational simplicity, here we define the *importance weight* $w(\mathbf{x})$ as

$$w(\mathbf{x}) = \frac{\pi(\mathbf{x})}{q(\mathbf{x})}. \tag{10}$$

Then the transition probability can be rewritten as

$$P(\mathbf{x},\mathbf{y}) = \begin{cases} \min\left\{q(\mathbf{y}), \frac{\pi(\mathbf{y})}{w(\mathbf{x})}\right\} & \text{if } \mathbf{y}\neq\mathbf{x}, \\ q(\mathbf{x}) + \sum\limits_{\mathbf{z}\neq\mathbf{x}} \max\left\{0, q(\mathbf{z}) - \frac{\pi(\mathbf{z})}{w(\mathbf{x})}\right\} & \text{if } \mathbf{y}=\mathbf{x}. \end{cases} \tag{11}$$

Without loss of generality, we label the countably infinite state space $\Omega = \mathbb{Z}^n$ as $\Omega = \{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_\infty\}$ (i.e., $\mathbf{y} = \mathbf{x}_i \in \Omega$) and assume that these states are sorted according to the magnitudes of their importance weights, namely,

$$w(\mathbf{x}_1) \ge w(\mathbf{x}_2) \ge \cdots \ge w(\mathbf{x}_\infty). \tag{12}$$

From (11) and (12), the transition matrix $\mathbf{P}$ of the Markov

chain can be exactly expressed as

$$\mathbf{P}=\begin{bmatrix} q(\mathbf{x}_1)+\lambda_1 & \frac{\pi(\mathbf{x}_2)}{w(\mathbf{x}_1)} & \frac{\pi(\mathbf{x}_3)}{w(\mathbf{x}_1)} & \cdots & \frac{\pi(\mathbf{x}_\infty)}{w(\mathbf{x}_1)} \\ q(\mathbf{x}_1) & q(\mathbf{x}_2)+\lambda_2 & \frac{\pi(\mathbf{x}_3)}{w(\mathbf{x}_2)} & \cdots & \frac{\pi(\mathbf{x}_\infty)}{w(\mathbf{x}_2)} \\ q(\mathbf{x}_1) & q(\mathbf{x}_2) & q(\mathbf{x}_3)+\lambda_3 & \cdots & \frac{\pi(\mathbf{x}_\infty)}{w(\mathbf{x}_3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q(\mathbf{x}_1) & q(\mathbf{x}_2) & q(\mathbf{x}_3) & \cdots & q(\mathbf{x}_\infty) \end{bmatrix}$$

where

$$\lambda_j=\sum_{i=j}^{\infty}\left(q(\mathbf{x}_i)-\frac{\pi(\mathbf{x}_i)}{w(\mathbf{x}_j)}\right) \tag{13}$$

stands for the probability of being rejected in the decision stage with the current state $\mathbf{x}_j$ for $\mathbf{X}_t$.

Let $\mathbf{q}=[q(\mathbf{x_1}),q(\mathbf{x_2}),\ldots]^T$ denote the vector of proposal probabilities. Then by decomposition, it follows that

$$\mathbf{P}=\mathbf{G}+\mathbf{e}\mathbf{q}^T, \tag{14}$$

where $\mathbf{e}=[1,1,\ldots]^T$ and $\mathbf{G}$ is an upper triangular matrix of the form

$$\mathbf{G}=\begin{bmatrix} \lambda_1 & \frac{\pi(\mathbf{x}_2)}{w(\mathbf{x}_1)}-q(\mathbf{x}_2) & \cdots & \frac{\pi(\mathbf{x}_\infty)}{w(\mathbf{x}_1)}-q(\mathbf{x}_\infty) \\ 0 & \lambda_2 & \cdots & \frac{\pi(\mathbf{x}_\infty)}{w(\mathbf{x}_2)}-q(\mathbf{x}_\infty) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

It is well-known that for a Markov chain, the largest eigenvalue of the transition matrix $\mathbf{P}$ always equals 1. Here, as $\mathbf{e}$ is a common right eigenvector for both $\mathbf{P}$ and $\mathbf{P}-\mathbf{G}$, it naturally corresponds to the largest eigenvalue 1. Meanwhile, since the rank of $\mathbf{P}-\mathbf{G}$ is 1, the other eigenvalues of $\mathbf{G}$ are exactly the same as those of $\mathbf{P}$.

On the other hand, the *mixing time*, which measures the time required by a Markov chain to get close to its stationary distribution

$$t_{\text{mix}}(\epsilon)=\min\{t:\max\|P^t(\mathbf{x},\cdot)-\pi(\cdot)\|_{TV}\le\epsilon\}, \tag{15}$$

is asymptotically dominated by the second largest eigenvalue λ^* [12], that is

$$t_{\text{mix}}(\epsilon)\le\frac{1}{1-|\lambda^*|}\log\left(\frac{1}{\epsilon}\right). \tag{16}$$

Thanks to the ascending order in (12), it is easy to verify that

$$\lambda^*=\lambda_1 \tag{17}$$

and

$$1>|\lambda_1|\ge|\lambda_2|\ge\cdots>0, \tag{18}$$

thereby raising the interest of identifying the value of λ_1.

Therefore, according to (13), we can easily get that

$$\begin{aligned}\lambda_1&=\sum_{i=1}^{\infty}\left(q(\mathbf{x}_i)-\frac{\pi(\mathbf{x}_i)}{w(\mathbf{x}_1)}\right)\\&=\sum_{i=1}^{\infty}q(\mathbf{x}_i)-\frac{1}{w(\mathbf{x}_1)}\cdot\sum_{i=1}^{\infty}\pi(\mathbf{x}_i)\\&=1-\frac{1}{w(\mathbf{x}_1)}=1-\frac{q(\mathbf{x}_1)}{\pi(\mathbf{x}_1)}.\end{aligned} \tag{19}$$

In other words, the spectral gap $1-\lambda_1$ is exactly captured by the ratio $q(\mathbf{x}_1)/\pi(\mathbf{x}_1)$. Next, we invoke the following Lemma to lower bound the ratio $q(\mathbf{x})/\pi(\mathbf{x})$.

Lemma 1 ([10])**.** *In the independent MHK algorithm, there exists $\delta>0$ such that*

$$\frac{q(\mathbf{x})}{\pi(\mathbf{x})}\ge\delta \tag{20}$$

for all $\mathbf{x}\in\Omega$, where

$$\delta\triangleq\frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\prod_{i=1}^{n}\rho_{\sigma_i}(\mathbb{Z})}. \tag{21}$$

Therefore, from Lemma 1, we can immediately get

$$\lambda_1\le1-\delta, \tag{22}$$

and by substituting (22) to (16), it follows that

$$t_{\text{mix}}(\epsilon)\le\frac{1}{\delta}\log\left(\frac{1}{\epsilon}\right). \tag{23}$$

Now, assume two independent Markov chains $\{\mathbf{X}_0,\mathbf{X}_1,\ldots\}$ and $\{\mathbf{Y}_0,\mathbf{Y}_1,\ldots\}$ proceed in the same update rule, where $\{\mathbf{X}_{i's}\}$ starts from some initial distribution π_0 and $\{\mathbf{Y}_{i's}\}$ is supposed to start from the stationary distribution π. Without loss of generality, assume $\mathbf{X}_t=\mathbf{x}$ and $\mathbf{Y}_t=\mathbf{y}$, then a state candidate $\mathbf{z}$ generated by both Markov chains at step $t+1$ will be accepted simultaneously can be expressed as

$$\begin{aligned}Pr(\mathbf{X}_{t+1}{=}\mathbf{Y}_{t+1}|\mathbf{X}_t{=}\mathbf{x},\mathbf{Y}_t{=}\mathbf{y})&=\sum_{\mathbf{z}\in\mathcal{X}^n}q(\mathbf{z})\min\left\{1,\frac{\omega(\mathbf{z})}{\omega(\mathbf{x})},\frac{\omega(\mathbf{z})}{\omega(\mathbf{y})}\right\}\\&=\sum_{\mathbf{z}\in\mathcal{X}^n}\pi(\mathbf{z})\min\left\{\frac{1}{\omega(\mathbf{z})},\frac{1}{\omega(\mathbf{x})},\frac{1}{\omega(\mathbf{y})}\right\}\\&\ge\frac{1}{\omega(\mathbf{x}_1)}\\&\ge\ \ \delta,\end{aligned} \tag{24}$$

where $\omega(\mathbf{x}_1)$ is defined as the largest importance ratio in (12).

According to the *coupling technique*, it is obvious that the first time that both chains accept a same state refers to the coupling time, where these two chains are deemed as identical thereafter. Therefore, based on (24), the number of steps for these two chains getting coupled, is upper bounded by

$$Pr(T\ge t)\le(1-\delta)^t, \tag{25}$$

and by *coupling inequality*, we can easily derive that

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

completing the proof. □

The rate of convergence derived here is the same as that given in [10]. Since the convergence rate derived from the spectral gap is precise, this means that the analysis in [10] based on the coupling technique is tight. More details on the exponential decay coefficient δ can be found in [13].

IV. Decoding Complexity

In this section, we apply the independent MHK sampling to lattice decoding and examine its complexity. We evaluate the complexity with the number of Markov moves, since the complexity of each Markov move is often insignificant in practice.

Consider the decoding of an $n \times n$ real-valued system. The extension to the complex-valued system is straightforward [2]. Let $\mathbf{x}$ denote the transmitted signal $\in \mathbb{Z}^n$. The corresponding received signal $\mathbf{c}$ is given by

$$\mathbf{c} = \mathbf{B}\mathbf{x} + \mathbf{w} \tag{27}$$

where $\mathbf{w}$ is the noise vector with zero mean and variance σ_w^2. The ML decoding is given by

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{Z}^n} \|\mathbf{c} - \mathbf{B}\mathbf{x}\|^2 \tag{28}$$

where $\|\cdot\|$ denotes the Euclidean norm. Intuitively, this CVP can be solved with the independent MHK algorithm, since the closest lattice point will be sampled with the highest probability. Note that another lattice problem — the shortest vector problem (SVP), is essentially a special case of CVP by $\mathbf{c} = \mathbf{0}$.

Because samples tend to be correlated with each other, we leave a gap t_{mix} to pick up samples from the stationary distribution. Therefore, we define the complexity of solving CVP as follows:

Definition 1. *With independent MHK sampling, the complexity of solving CVP is*

$$C_{\text{MHK}} = \frac{t_{\text{mix}}}{D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})}. \tag{29}$$

Note that the mixing time t_{mix} serves here as the gap for independent samples, which is equivalent to operate the Markov chains in parallel. In other words, it corresponds to the decoding in the worst case for MCMC. Therefore, it may be possible to reduce the gap to obtain lower complexity.

With the mixing time given in (23), C_{MHK} can be upper bounded by

$$\begin{aligned} C_{\text{MHK}} &< \log\left(\frac{1}{\epsilon}\right) \cdot \frac{\prod_{i=1}^n \rho_{\sigma_{n+1-i},\widetilde{x}_{n+1-i}}(\mathbb{Z})}{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})} \cdot \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{\Lambda})}{\rho_{\sigma,\mathbf{c}}(\mathbf{B}\mathbf{x})} \\ &= a \cdot \frac{\prod_{i=1}^n \rho_{\sigma_{n+1-i},\widetilde{x}_{n+1-i}}(\mathbb{Z})}{\rho_{\sigma,\mathbf{c}}(\mathbf{B}\mathbf{x})} \\ &= a \cdot C. \end{aligned} \tag{30}$$

As $a = \log\left(\frac{1}{\epsilon}\right)$ is a constant depending on the value of ϵ, we pay attention to the complexity

$$C = \frac{\prod_{i=1}^n \rho_{\sigma_{n+1-i},\widetilde{x}_{n+1-i}}(\mathbb{Z})}{\rho_{\sigma,\mathbf{c}}(\mathbf{B}\mathbf{x})}. \tag{31}$$

For the term $\prod_{i=1}^n \rho_{\sigma_{n+1-i},\widetilde{x}_{n+1-i}}(\mathbb{Z})$, we have

$$\begin{aligned} \prod_{i=1}^n \rho_{\sigma_{n+1-i},\widetilde{x}_{n+1-i}}(\mathbb{Z}) &\overset{(a)}{\leq} \prod_{i=1}^n \rho_{\sigma_i}(\mathbb{Z}) \\ &= \prod_{i=1}^n \sum_{x_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|x_i\|^2} \\ &= \prod_{i=1}^n \sum_{x_i\in\mathbb{Z}} e^{-\frac{\alpha\|\widehat{\mathbf{b}}_i\|^2\pi}{2\pi\alpha\sigma^2}\|x_i\|^2} \\ &\overset{(b)}{\leq} \prod_{i=1}^n (2\pi\alpha\sigma^2)^{\frac{n}{2}} \sum_{x_i\in\mathbb{Z}} e^{-\alpha\|\widehat{\mathbf{b}}_i\|^2\pi\|x_i\|^2} \\ &\overset{(c)}{=} (\sqrt{2\pi\alpha}\sigma)^{n^2} \cdot \prod_{i=1}^n \vartheta_3(\alpha\|\widehat{\mathbf{b}}_i\|^2). \end{aligned} \tag{32}$$

Here, (a) holds due to the fact that [14]

$$\rho_{\sigma,\mathbf{c}}(\Lambda) \leq \rho_\sigma(\Lambda), \tag{33}$$

(c) applies the *Jacobi theta function* ϑ_3 [15]

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

and (b) follows the inequality from [16, Lemma 1.4]

$$\sum_{\mathbf{x}\in\Lambda} e^{-\pi s^{-1}x^2} \leq s^{\frac{n}{2}} \cdot \sum_{\mathbf{x}\in\Lambda} e^{-\pi\mathbf{x}^2}, \tag{35}$$

where $s \geq 1$.

Note the condition for (b) in (32), namely,

$$(2\pi\alpha\sigma^2) \geq 1, \tag{36}$$

where the equality holds if

$$\sigma = \frac{1}{\sqrt{2\pi\alpha}}. \tag{37}$$

By substituting (32) to (31), the complexity C is upper bounded as

$$C \leq \frac{(\sqrt{2\pi\alpha}\sigma)^{n^2} \cdot \prod_{i=1}^n \vartheta_3(\alpha\|\widehat{\mathbf{b}}_i\|^2)}{e^{-\frac{1}{2\sigma^2}\|\mathbf{B}\mathbf{x}-\mathbf{c}\|^2}}. \tag{38}$$

It is natural that σ is chosen to minimize the above lower bound. However, as can be seen clearly, once $2\pi\alpha\sigma^2 > 1$, then the power of $2\pi\alpha\sigma^2$ will be the unaffordable n^2. Therefore, in order to remove the impact of such an item, we choose σ to make

$$2\pi\alpha\sigma^2 = 1 \tag{39}$$

and

$$\alpha = \frac{1}{2\pi\sigma^2}. \tag{40}$$

Then, we can rewrite the upper bound (38) as

$$C \leq \frac{\prod_{i=1}^n \vartheta_3(\|\widehat{\mathbf{b}}_i\|^2/2\pi\sigma^2)}{e^{-\frac{1}{2\sigma^2}\|\mathbf{B}\mathbf{x}-\mathbf{c}\|^2}}. \tag{41}$$

Now we recall some facts about the Jacobi theta function $\vartheta_3(\tau)$ from (34). It is obvious that $\vartheta_3(\tau)$ is monotonically decreasing with τ, and particularly the minimum of $\vartheta_3(\tau)$ is 1, namely,

$$\lim_{\tau\to\infty} \inf \vartheta_3(\tau) = 1. \tag{42}$$

By simple calculation, we can get that

$$\vartheta_3(2) = \sum_{n=-\infty}^{+\infty} e^{-2\pi n^2} = \frac{\sqrt[4]{6\pi + 4\sqrt{2}\pi}}{2\Gamma(\frac{3}{4})} = 1.0039, \quad (43)$$

where $\Gamma(\cdot)$ stands for the Gamma function. Clearly, if

$$\frac{\min_{1\le i\le n} \|\widehat{\mathbf{b}}_i\|^2}{2\pi\sigma^2} \ge 2 \quad (44)$$

then it turns out that

$$\prod_{i=1}^{n} \vartheta_3(\|\widehat{\mathbf{b}}_i\|^2/2\pi\sigma^2) \le \vartheta_3^n(2) = 1.0039^n, \quad (45)$$

is really small even for values of n up to hundreds (e.g., $1.0039^{100} = 1.4759$).

Therefore, let σ satisfy the condition given in (44), that is

$$\sigma \le \min_{1\le i\le n} \|\widehat{\mathbf{b}}_i\|/(2\sqrt{\pi}), \quad (46)$$

then we have

$$C \le 1.0039^n \cdot e^{\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}. \quad (47)$$

Here, we can also apply $\vartheta_3(3) = 1.0002$ (or $\vartheta_3(4)$, etc.) so that $1.0002^{1000} = 1.2214$. The key point is that the pre-exponential factor is rather small. Therefore, set $\sigma = \min_i \|\widehat{\mathbf{b}}_i\|/(2\sqrt{\pi})$, then the decoding complexity is given by

$$C = \tilde{O}(e^{\|\mathbf{Bx}-\mathbf{c}\|^2/\min_i \|\widehat{\mathbf{b}}_i\|^2}). \quad (48)$$

By law of large numbers, $\|\mathbf{Bx}-\mathbf{c}\|^2 \approx n\sigma_w^2$. Thus $C \approx O(e^{n\sigma_w^2/\min_i \|\widehat{\mathbf{b}}_i\|^2})$. We highlight the significance of lattice reduction (i.e., LLL reduction) here, since increasing $\min_i \|\widehat{\mathbf{b}}_i\|$ will significantly decrease the complexity.

Remark 1. *In fact, such an analysis also holds for Klein's algorithm, where the probability of sampling* $\mathbf{x}$ *is [1]*

$$P(\mathbf{x}) \ge \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\prod_{i=1}^{n} \vartheta_3(\|\widehat{\mathbf{b}}_i\|^2/2\pi\sigma^2)}, \quad (49)$$

which is exactly the inverse of (41). Klein chose $\sigma = \min_i \|\widehat{\mathbf{b}}_i\|/\sqrt{\log n}$ *and showed the decoding complexity* $O(n^{\|\mathbf{Bx}-\mathbf{c}\|^2/\min_i \|\widehat{\mathbf{b}}_i\|^2})$.

Here, we have shown that the decoding complexity can be reduced to $\tilde{O}(e^{\|\mathbf{Bx}-\mathbf{c}\|^2/\min_i \|\widehat{\mathbf{b}}_i\|^2})$, *by setting* $\sigma = \min_i \|\widehat{\mathbf{b}}_i\|/(2\sqrt{\pi})$.

In general, the complexity of solving the CVP is exponential or higher. On the other hand, one may be interested in the performance with a fixed number of Markov moves $C \gg 1$. It follows from (47) that

$$\|\mathbf{Bx}-\mathbf{c}\| \ge \sigma\sqrt{2\ln\frac{C}{1.0039^n}}. \quad (50)$$

Typically, the sampling decoder will find the closest vector point $\mathbf{Bx}$ if the distance from $\mathbf{c}$ to the lattice is less than the right-hand-side (RHS) of (50). In this regard, the RHS of (50) can be interpreted as the *decoding radius* of bounded distance decoding (BDD), that is

$$d = \|\mathbf{Bx}-\mathbf{c}\| \triangleq \sigma\sqrt{2\ln\frac{C}{1.0039^n}}. \quad (51)$$

When $\sigma = \min_i \|\widehat{\mathbf{b}}_i\|/(2\sqrt{\pi})$, we have

$$d = \sqrt{\frac{1}{2\pi}\cdot\ln\frac{C}{1.0039^n}}\cdot\min_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|. \quad (52)$$

Clearly, the decoding radius d monotonically increases with C, implying a flexible trade-off between the decoding radius and complexity.

ACKNOWLEDGMENT

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

REFERENCES

[1] P. Klein, "Finding the closest lattice vector when it is unusually close," in *ACM-SIAM Symp. Discr. Algorithms*, 2000, pp. 937–941.
[2] 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.
[3] 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.
[4] 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.
[5] 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.
[6] 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.
[7] R. Chen, J. Liu, and X. Wang, "Convergence analysis and comparisons of Markov chain Monte Carlo algorithms in digital communications," *IEEE Trans. on Signal Process.*, vol. 50, no. 2, pp. 255–270, 2002.
[8] 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.
[9] 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.
[10] Z. Wang and C. Ling, "Independent Metropolis-Hastings-Klein algorithm for lattice Gaussian sampling," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, China, June 2015, pp. 2470–2474.
[11] J. S. Liu, "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Statistics and Computing*, vol. 6, pp. 113–119, 1996.
[12] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Time*, American Mathematical Society, 2008.
[13] Z. Wang and C. Ling, "On the geometric ergodicity of Metropolis-Hastings algorithms for lattice Gaussian sampling," *Submitted to IEEE Transactions on Information Theory, 2015.*, [Online] Available:http://arxiv.org/pdf/1501.05757v2.pdf.
[14] 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.
[15] J. H. Conway and N. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1998.
[16] W. Banaszczyk, "New bounds in some transference theorems in the geometry of numbers," *Math. Ann.*, vol. 296, pp. 625–635, 1993.