Markov Chain Monte Carlo Algorithms for Lattice Gaussian Sampling

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

Guillaume Hanrot
École Normale Supérieure de Lyon
LIP (CNRS, ENS Lyon, UCBL, INRIA)
46 Allée d'Italie, 69364 Lyon Cedex 07, France
Email:Guillaume.Hanrot@ens-lyon.fr

***Abstract*—To be considered for an IEEE Jack Keil Wolf ISIT Student Paper Award.**

Sampling from a lattice Gaussian distribution is emerging as an important problem in various areas such as coding and cryptography. The default sampling algorithm — Klein's algorithm yields a distribution close to the lattice Gaussian only if the standard deviation is sufficiently large. In this paper, we propose the Markov chain Monte Carlo (MCMC) method for lattice Gaussian sampling when this condition is not satisfied. In particular, we present a sampling algorithm based on Gibbs sampling, which converges to the target lattice Gaussian distribution for any value of the standard deviation. To improve the convergence rate, a more efficient algorithm referred to as *Gibbs-Klein sampling* is proposed, which samples block by block using Klein's algorithm. We show that Gibbs-Klein sampling yields a distribution close to the target lattice Gaussian, under a less stringent condition than that of the original Klein algorithm.

I. Introduction

The lattice Gaussian distribution is emerging as a common theme in various areas. In mathematics, Banaszczyk [1] firstly used it to prove the transference theorems of lattices. In coding, it mimics Shannon's Gaussian random coding technique, yet permits lattice decoding. Forney applied the lattice Gaussian distribution to obtain the full shaping gain in lattice coding [2] (see also [3]). Recently, it has been used to achieve the capacity of the Gaussian channel [4] and to approach the secrecy capacity of the Gaussian wiretap channel [5], respectively. Sampling from the lattice Gaussian has also been used in lattice decoding for the multi-input multi-output system [6], [7]. In cryptography, lattice Gaussians have become a central tool in the construction of many primitives. Micciancio and Regev used it to propose lattice-based cryptosystems based on the worst-case hardness assumptions [8], and recently, it has underpinned the fully-homomorphic encryption for cloud computing [9]. The key fact is again that a vector distributed as a lattice Gaussian centered at $\mathbf{c}$ with a small standard deviation is typically very close to $\mathbf{c}$. To illustrate why this might be useful in cryptography, note that if one knows a short basis of the lattice, one can efficiently produce such a vector [10], while disclosing no information on the short basis—since the lattice Gaussian distribution does not depend on the particular basis.

Thus, in both coding and cryptography, efficient sampling algorithms for the lattice Gaussian as well as a good understanding on how the complexity depends on the standard deviation is an important issue. However, in contrast to sampling from the continuous Gaussian distribution, it is not at all straightforward to sample from a discrete Gaussian distribution over a lattice. At present, the default sampling algorithm for lattices is due to Klein, originally proposed for bounded-distance decoding [11] (see also [12], [13] for variations and [4] for an algorithm for lattices of Construction A). It was shown in [10] that Klein's algorithm samples within a negligible statistical distance from the lattice Gaussian distribution only if the standard deviation $\sigma \geq \omega(\sqrt{\log n}) \cdot \max_{1 \leq i \leq n} \|\widehat{\mathbf{b}}_i\|$, where n is the lattice dimension and $\widehat{\mathbf{b}}_i$'s are the Gram-Schmidt vectors of the lattice basis. Unfortunately, such a requirement of σ can be excessive, rendering Klein's algorithm inapplicable to many cases of interest.

Markov chain Monte Carlo (MCMC) methods attempt to sample from the target distribution of interest by building a Markov chain, which randomly generate the next sample conditioned on the previous samples. As a major algorithm of MCMC, Gibbs sampling [14] constructs a Markov chain which gradually converges to the target distribution by only considering univariate sampling at each step. In this paper, we introduce the Gibbs algorithm into lattice Gaussian sampling and propose a more efficient block-based algorithm named as *Gibbs-Klein sampling*. In contrast to conventional blocked sampling which is computationally more demanding, the proposed algorithm takes advantages of Klein's algorithm as a building block. The proposed algorithms are applicable in the scenario $\sigma < \omega(\sqrt{\log n}) \cdot \max_{1 \leq i \leq n} \|\widehat{\mathbf{b}}_i\|$.

To the best of our knowledge, this is the first time that MCMC methods are used in lattice Gaussian distributions. Different from previous works on Gibbs sampling for signal detection of finite constellations [15]–[17], here we are concerned with countably infinite state spaces and with simulating Gaussian distributions over a lattice. It is worth pointing out that although the underlying Markov chain converges to the stationary distribution for all values of σ, the convergence is expected to become very slow when σ becomes small, since for very small σ we would solve the closest vector problem (CVP) and shortest vector problem (SVP) with high

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

Algorithm 1 Klein's Algorithm

Input: $\mathbf{B}, \sigma, \mathbf{c}$
Output: $\mathbf{Bx} \in \Lambda$

1: let $\mathbf{B} = \mathbf{QR}$ and $\mathbf{c}' = \mathbf{Q}^T\mathbf{c}$
2: **for** $i = n, \ldots, 1$ **do**
3: let $\alpha_i = \frac{\sigma}{|r_{i,i}|}$ 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 $D_{\mathbb{Z},\alpha_i,\widetilde{x}_i}$
5: **end for**
6: return $\mathbf{Bx}$

probability.

The rest of this paper is organized as follows. Section II introduces lattice Gaussian distributions and briefly reviews Klein's algorithm. In Section III, the conventional Gibbs and the new Gibbs-Klein sampling algorithms are proposed for lattice Gaussians, followed by a theoretical analysis in Section IV. Section V presents the simulation results.

II. Lattice Gaussian Distributions

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

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

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

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

for all $\mathbf{z} \in \mathbb{R}^n$. 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 $\rho_{\sigma,\mathbf{c}}(\Lambda) \triangleq \sum_{\mathbf{Bx}\in\Lambda} \rho_{\sigma,\mathbf{c}}(\mathbf{Bx})$.

An intuition of $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$ suggests that the closer lattice point $\mathbf{Bx}$ is to $\mathbf{c}$, the higher probability it will be sampled. Thus, lattice Gaussian sampling can be applied to solve the CVP, and Klein's algorithm was originally proposed for decoding [11]. As a randomized version of Babai's nearest-plane algorithm (i.e., successive interference cancellation), Klein's algorithm obtains a vector by sequentially sampling from a 1-dimensional conditional Gaussian distribution. As shown in Algorithm 1, its operation has polynomial complexity $O(n^2)$ excluding QR decomposition.

The parameter σ is key to the distribution produced by Klein's algorithm. Klein suggested $\sigma = \min_i\|\widehat{\mathbf{b}}_i\|/\sqrt{\log n}$ and this was followed/adapted in [6], [7]. In this case, Klein's algorithm only yields a distribution that is lower-bounded by the Gaussian distribution. On the other hand, it was demonstrated in [10] that Klein's algorithm actually samples from $D_{\Lambda,\sigma,\mathbf{c}}$ within a negligible statistical distance if

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

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

III. MCMC for Lattice Gaussian

In this section, we introduce the concept of MCMC into lattice Gaussian sampling for the range of σ where Klein's algorithm cannot reach. We further propose a more efficient sampling algorithm named as *Gibbs-Klein sampling* to improve the convergence rate.

A. Gibbs Sampling for Lattice Gaussian

Lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$ with $\sigma < \omega(\sqrt{\log n}) \cdot \max_i\|\widehat{\mathbf{b}}_i\|$ can be seen as a complex target distribution lacking direct sampling methods. MCMC makes use of the conditional distribution as a tractable alternative to work with. Here we apply the Gibbs algorithm to sample from the original joint distribution $D_{\Lambda,\sigma,\mathbf{c}}$.

Gibbs sampling employs 1-dimensional conditional distributions to construct the Markov chain [14], where all other variables in the distribution are unchanged in each step. In this way, we sample n random variables from the corresponding n univariate conditionals in a certain order instead of directly generating an n-dimensional vector. Samples drawn from the target joint distribution will be generated when the Markov chain reaches the stationary distribution.

Specifically, in Gibbs sampling, each coordinate of $\mathbf{x}$ is sampled from the following 1-dimensional conditional distribution

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

where $1 \leq i \leq n$ denotes the coordinate index of $\mathbf{x}$, $\mathbf{x}_{[-i]}^t \triangleq [x_1^t, \ldots, x_{i-1}^t, x_{i+1}^t, \ldots, x_n^t]^T$, and t is the time index of the Markov chain. It is noteworthy that there are many scan schemes in Gibbs sampling and we apply the random-scan in this paper, which means the index i is randomly chosen at each step. The extension to other scan strategies is possible.

By repeating such a procedure, an underlying Markov chain $\mathbf{x}^{t+1} = [x_1^t, \ldots, x_{i-1}^t, x_i^{t+1}, x_{i+1}^t, \ldots, x_n^t]^T$ is induced, whose transition probability between two adjacent states is defined by the univariate Gibbs sampler,

$$P(\mathbf{x}^t; \mathbf{x}^{t+1}) = P(x_i^{t+1}|\mathbf{x}_{[-i]}^t). \tag{6}$$

Clearly, every two adjacent states of $\mathbf{x}$ differ from each other by only one coordinate and it is easy to see that $D_{\Lambda,\sigma,\mathbf{c}}$ stays invariant under such transitions. Algorithm 2 gives the operation of Gibbs sampling for lattice Gaussian distributions. The initial random variable $\mathbf{x}^0$ can be chosen from $\mathbb{Z}^n$ arbitrarily or from the output of a suboptimal algorithm, while the time bound T is large enough to reach the stationary distribution $D_{\Lambda,\sigma,\mathbf{c}}$.

With the transition probabilities (6), we may form the infinite transition matrix $\mathbf{P}$, whose (i,j)-th entry $P(s_i; s_j)$ represents the probability of transferring to state s_j from the previous state s_i. Denote by $\mathbf{P}^t$ the transition matrix after t steps. We group in the following theorem standard results about Gibbs sampling [18].

Algorithm 2 Gibbs sampling for lattice Gaussian

Input: $\mathbf{B}, \sigma, \mathbf{c}, \mathbf{x}^0$
Output: $\mathbf{x} \sim D_{\Lambda,\sigma,\mathbf{c}}$ as $T \to \infty$
1: **for** t =1, …, T **do**
2: randomly choose coordinate index i from $\{1, 2, \ldots, n\}$
3: sample x_i from $P(x_i^t | \mathbf{x}_{[-i]}^{t-1})$
4: update $\mathbf{x}^t = [x_1^{t-1}, \ldots, x_{i-1}^{t-1}, x_i, x_{i+1}^{t-1}, \ldots, x_n^{t-1}]^T$
5: **if** Markov chain has reached stationarity **then**
6: output $\mathbf{x}^t$
7: **end if**
8: **end for**

Proposition 1. *Given the invariant distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by the Gibbs sampler is irreducible, aperiodic and reversible (hence positive recurrent), and converges to the stationary distribution in the total variation (TV) distance as $t \to \infty$:*

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

for all states $\mathbf{x} \in \mathbb{Z}^n$, where $P^t(\mathbf{x}; \cdot)$ denotes the row of $\mathbf{P}^t$ corresponding to initial state $\mathbf{x}$.

According to Proposition 1, if time permits to reach the stationary distribution, the proposed Gibbs sampler will draw samples from $D_{\Lambda,\sigma,\mathbf{c}}$ no matter what value σ takes, which means the obstacle encountered by Klein's algorithm is overcome.

B. Gibbs-Klein Sampling for Lattice Gaussian

Although the afore-mentioned Gibbs sampler will converge to the stationary distribution eventually, the way it functions by individually sampling only one component each time leads to slow convergence. Especially, for lattice bases whose components are highly correlated with each other, the Markov chain induced by the standard Gibbs sampling can be trapped for a long time. To hasten convergence of the Markov chain, a new sampling algorithm combining Gibbs and Klein algorithms is proposed in the sequel.

The idea of blocked sampling is to sample a block of components of $\mathbf{x}$ at each step [19]. Intuitively, this will lead to a faster convergence rate, which is already shown in [14]. However, sampling a block is generally more costly than componentwise sampling. We propose to use Klein's algorithm for block sampling; this leads to the Gibbs-Klein.

At each step of the Markov chain, the proposed Gibbs-Klein sampling randomly picks up a block of m components of $\mathbf{x}$ to update. For convenience, an $n \times n$ permutation matrix $\mathbf{E}$ is applied before blocking so that the blocks are updated in a fixed order.

Specifically, if $\mathbf{E}$ is random, then Gibbs-Klein sampling on m randomly chosen components will be equivalent to sample m consecutive components of $\mathbf{z}$ in a fixed order, where $\mathbf{z} = \mathbf{E}^{-1}\mathbf{x}$ and $\widetilde{\mathbf{B}} = \mathbf{B}\mathbf{E}$. For simplicity, we always consider the block formed by the first m components of $\mathbf{z}$, namely $\mathbf{z}_{\text{block}} = [z_1, \ldots, z_m]^T$. After QR-decomposition $\widetilde{\mathbf{B}} = \mathbf{Q}\mathbf{R}$ and calculating $\mathbf{c}' = \mathbf{Q}^T\mathbf{c}$, z_i in the block is sampled from the following 1-dimensional distribution with the backward order from z_m to z_1:

$$P(z_i^{t+1} | \overline{\mathbf{z}}_{[-i]}^t) = D_{\mathbb{Z},\alpha_i,\widetilde{z}_i^t}, \tag{8}$$

where $\alpha_i = \frac{\sigma}{|r_{i,i}|}$, $\overline{\mathbf{z}}_{[-i]}^t = [z_{i+1}^{t+1}, \ldots, z_m^{t+1}, z_{m+1}^t, \ldots, z_n^t]^T$ and $\widetilde{z}_i^t = \frac{c_i' - \sum_{j=i+1}^m r_{i,j} z_j^{t+1} - \sum_{j'=m+1}^n r_{i,j'} z_{j'}^t}{r_{i,i}}$. Algorithm 3 gives the proposed Gibbs-Klein sampling, where $\mathbf{z}^{t+1} = [\mathbf{z}_{\text{block}}^{t+1}; \mathbf{z}_{[-\text{block}]}^t]$ is obtained after each step, and $\mathbf{z}_{[-\text{block}]}^t = [z_{m+1}^t, \ldots, z_n^t]^T$. The implementation given in Algorithm 3 is not so efficient due to repeated QR decompositions; Optimizing for better efficiency will be pursued in the future. Note that the extension to other scan strategies is also possible.

Algorithm 3 Gibbs-Klein sampling for lattice Gaussian

Input: $\mathbf{B}, \sigma, \mathbf{c}, m, \mathbf{x}^0$;
Output: $\mathbf{x}$ from a distribution close to $D_{\Lambda,\sigma,\mathbf{c}}$ as $T \to \infty$
1: **for** t =1, …, T **do**
2: randomly generate a permutation matrix $\mathbf{E}$
3: Let $\widetilde{\mathbf{B}} = \mathbf{B}\mathbf{E}$ and $\mathbf{z} = \mathbf{E}^{-1}\mathbf{x}$
4: Let $\widetilde{\mathbf{B}} = \mathbf{Q}\mathbf{R}$ and $\mathbf{c}' = \mathbf{Q}^T\mathbf{c}$
5: **for** $i = m$, …, 1 **do**
6: let $\alpha_i = \frac{\sigma}{|r_{i,i}|}$
7: let $\widetilde{z}_i^{t-1} = \frac{c_i' - \sum_{j=i+1}^m r_{i,j} z_j^t - \sum_{j'=m+1}^n r_{i,j'} z_{j'}^{t-1}}{r_{i,i}}$
8: sample z_i^t from $D_{\mathbb{Z},\alpha_i,\widetilde{z}_i^{t-1}}$
9: **end for**
10: update $\mathbf{z}^t = [\mathbf{z}_{\text{block}}^t; \mathbf{z}_{[-\text{block}]}^{t-1}]^T$
11: return $\mathbf{x}^t = \mathbf{E}\mathbf{z}^t$
12: **if** Markov chain has reached stationarity **then**
13: output $\mathbf{x}^t$
14: **end if**
15: **end for**

IV. Analysis of Gibbs-Klein sampling

In this section, we show that the proposed Gibbs-Klein sampling algorithm can induce a reversible Markov chain within a negligible error. From (8) and by induction, the sampling probability of $\mathbf{z}_{\text{block}}^{t+1}$ conditioned on $\mathbf{z}_{[-\text{block}]}^t$ is given by

$$P(\mathbf{z}_{\text{block}}^{t+1} \mid \mathbf{z}_{[-\text{block}]}^t) = \prod_{i=1}^m P(z_{m+1-i}^{t+1} | \overline{\mathbf{z}}_{[-(m+1-i)]}^t). \tag{9}$$

The following lemma gives a closed-form expression of this conditional probability within a negligible error and the proof follows [10].

Lemma 1. *For a given invariant distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the transition probability $P(\mathbf{z}_{block}^{t+1} \mid \mathbf{z}_{[-block]}^t)$ of Gibbs-Klein algorithm is within negligible statistical distance of the following distribution*

$$D' = \frac{e^{-\frac{1}{2\sigma^2}\|\widetilde{\mathbf{B}}\mathbf{z}^{t+1} - \mathbf{c}\|^2}}{\sum_{\mathbf{z}_{\text{block}}^{t+1} \in \mathbb{Z}^m} e^{-\frac{1}{2\sigma^2}\|\widetilde{\mathbf{B}}\mathbf{z}^{t+1} - \mathbf{c}\|^2}} \tag{10}$$

if $\sigma \geq \omega(\sqrt{\log m}) \cdot \max_{1\leq i\leq m}\|r_{i,i}\|$, *where* $\mathbf{z}^{t+1}=[\mathbf{z}^{t+1}_{\text{block}};\mathbf{z}^{t}_{[-\text{block}]}]$.

Proof: According to (8) and (9), we have

$$\begin{aligned}
P(\mathbf{z}^{t+1}_{\text{block}}|\mathbf{z}^{t}_{[-\text{block}]}) &= \prod_{i=1}^{m} D_{\mathbb{Z},\alpha_{m+1-i},\widetilde{z}^{t}_{m+1-i}}(z^{t+1}_{m+1-i}) \\
&= \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(\overline{c}_{m+1-i}-\sum_{j=m+1-i}^{m} r_{m+1-i,j} z_j^{t+1}\right)^2}}{\prod_{i=1}^{m}\sum_{z^{t+1}_{m+1-i}\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}\left(\overline{c}_{m+1-i}-\sum_{j=m+1-i}^{m} r_{m+1-i,j} z_j^{t+1}\right)^2}} \\
&= \frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{c}}-\overline{\mathbf{r}}\mathbf{z}^{t+1}_{\text{block}}\|^2}}{\prod_{i=1}^{m}\sum_{z^{t+1}_{m+1-i}\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}\left(r_{m+1-i,m+1-i} z^{t+1}_{m+1-i}-\overline{c}_{m+1-i}+\sum_{j=m+2-i}^{m} r_{m+1-i,j} z_j^{t+1}\right)^2}} \\
&= \frac{\rho_{\mathcal{L}(\overline{\mathbf{r}}),\sigma,\overline{\mathbf{c}}}(\mathbf{z}^{t+1}_{\text{block}})}{\prod_{i=1}^{m}\rho_{\sigma}(r_{m+1-i,m+1-i}\mathbb{Z}+\xi)},
\end{aligned} \tag{11}$$

where $\overline{c}_i = c'_i - \sum_{j'=m+1}^{n} r_{i,j'} z^{t}_{j'}$, $\overline{\mathbf{c}} = [\overline{c}_1,\ldots,\overline{c}_m]^T$, $\xi = \sum_{j=m+2-i}^{m} r_{m+1-i,j} z_j^{t+1} - \overline{c}_{m-i+i}$ and $\overline{\mathbf{r}}$ is the $m\times m$ segment of $\mathbf{R}$ with $r_{1,1}$ to $r_{m,m}$ in the diagonal. Clearly, the effect of the subvector $\mathbf{z}^{t}_{[-\text{block}]}$ is hidden in $\overline{c}_i$. In [20], it has been demonstrated that if $\sigma > \eta_{\varepsilon}(\mathcal{L}(\overline{\mathbf{r}}))$, then

$$\frac{\prod_{i=1}^{m}\rho_{\sigma}(r_{i,i}\mathbb{Z}+\xi)}{\prod_{i=1}^{m}\rho_{\sigma}(r_{i,i}\mathbb{Z})} \in \left(\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{m}, 1\right] \tag{12}$$

which means $\prod_{i=1}^{m}\rho_{\sigma}(r_{i,i}\mathbb{Z}+\xi)$ can be substituted by $\prod_{i=1}^{m}\rho_{\sigma}(r_{i,i}\mathbb{Z})$ within negligible errors when ε is sufficiently small.

As shown in [10], $\eta_{\varepsilon}(\Lambda)$ with negligible ε is upper bounded as $\eta_{\varepsilon}(\Lambda) \leq \omega(\sqrt{\log n}) \cdot \max_{1\leq i\leq n}\|\widehat{\mathbf{b}}_i\|$. Therefore, if $\sigma \geq \omega(\sqrt{\log m}) \cdot \max_{1\leq i\leq m}\|r_{i,i}\|$, $P(\mathbf{z}^{t+1}_{\text{block}} \mid \mathbf{z}^{t}_{[-\text{block}]})$ shown in (11) can be rewritten as

$$P(\mathbf{z}^{t+1}_{\text{block}}|\mathbf{z}^{t}_{[-\text{block}]}) \simeq \frac{\rho_{\mathcal{L}(\overline{\mathbf{r}}),\sigma,\overline{\mathbf{c}}}(\mathbf{z}^{t+1}_{\text{block}})}{\prod_{i=1}^{m}\rho_{\sigma}(r_{i,i}\mathbb{Z})}, \tag{13}$$

where "$\simeq$" represents equality up to a negligible error. Because the denominator is independent of $\mathbf{z}^{t+1}_{\text{block}}$, $\mathbf{z}^{t}_{[-\text{block}]}$ and $\mathbf{c}$, it can be viewed as a constant and the output has a lattice Gaussian distribution $D_{\mathcal{L}(\overline{\mathbf{r}}),\sigma,\overline{\mathbf{c}}}(\mathbf{z}^{t+1}_{\text{block}})$. ■

Then we arrive at the following proposition.

Proposition 2. *Suppose* $\sigma \geq \omega(\sqrt{\log m}) \cdot \max_{1\leq i\leq m}\|\widehat{\widetilde{\mathbf{b}}}_i\|$ *at each step so that the negligible statistical distance is absorbed by numerical errors. Then, within numerical errors, the Markov chain induced by the Gibbs-Klein sampler is irreducible, aperiodic and reversible (hence positive recurrent) and converges to the stationary distribution in the total variation distance as* $t\to\infty$*:*

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

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

Proof: Let s_i and s_j be two adjacent states in Gibbs-Klein sampling. For block size m, every two adjacent states in Gibbs-Klein sampling differ from each other by at most m components. For convenience, we express them as

$$s_i = [\mathbf{x}_{\text{block}(i)}, \mathbf{x}_{[-\text{block}]}] \quad \text{and} \quad s_j = [\mathbf{x}_{\text{block}(j)}, \mathbf{x}_{[-\text{block}]}], \tag{15}$$

where $\mathbf{x}_{\text{block}(i)}$ and $\mathbf{x}_{\text{block}(j)}$ denote the m components belonging to s_i and s_j, respectively. Then, the transition probability of Gibbs-Klein sampling is

$$\begin{aligned}
P(s_i;s_j) &= P(\mathbf{x}^{t+1}=s_j|\mathbf{x}^{t}=s_i) \\
&= P(\mathbf{x}^{t}_{\text{block}(i)} \to \mathbf{x}^{t+1}_{\text{block}(j)}|\mathbf{x}^{t}_{[-\text{block}]}) \\
&\overset{(a)}{=} P(\mathbf{x}^{t+1}_{\text{block}(j)}|\mathbf{x}^{t}_{[-\text{block}]}) \\
&\simeq \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{B}s_j-\mathbf{c}\|^2}}{\sum_{\mathbf{x}^{t+1}_{\text{block}}\in\mathbb{Z}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{B}\mathbf{x}^{t+1}-\mathbf{c}\|^2}},
\end{aligned} \tag{16}$$

where (a) is due to the fact that $\mathbf{x}^{t+1}_{\text{block}}$ is sampled only conditioned on $\mathbf{x}^{t}_{[-\text{block}]}$.

To show the Markov chain is irreducible, we note that given a state s one can attain with positive probability in one step any state s' which shares $>= (n-m)$ components with s. Now, if s and s' have, say, $d < n-m$ components in common, there is always a positive probability that after each step they get exactly one more component in common. So we can go in $n-d$ steps from one to the other. But as soon as $m >= 2$, we can assume that at the first step we get two more components in common, and then one at each further step, so we can go with positive probability in $n-d-1$ steps.

On the other hand, it is clear to see that the number of steps required to move between any two states (can be the same state) is arbitrary without any limitation to be a multiple of some integer. Put another way, the chain is not forced into some cycle with fixed period between certain states. Therefore, the Markov chain is aperiodic.

As for reversibility, it is no hard to check that the following relationship holds

$$D_{\Lambda,\sigma,\mathbf{c}}(s_i)P(s_i;s_j) \simeq D_{\Lambda,\sigma,\mathbf{c}}(s_j)P(s_j;s_i) \tag{17}$$

with the same expression

$$\frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{B}s_i-\mathbf{c}\|^2}}{\sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{B}\mathbf{x}-\mathbf{c}\|^2}} \cdot \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{B}s_j-\mathbf{c}\|^2}}{\sum_{\mathbf{x}^{t+1}_{\text{block}}\in\mathbb{Z}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{B}\mathbf{x}^{t+1}-\mathbf{c}\|^2}}, \tag{18}$$

within negligible errors. Thus, the conclusion follows, completing the proof. ■

The advantages of Gibbs-Klein sampling are two-fold: compared with the conventional Gibbs sampling which only processes a single variate each time, it is more efficient to sample multiple variates in a block, improving the convergence rate; on the other hand, it overcomes the limitation of Klein's sampling which requires large values of σ and extends lattice Gaussian sampling to the more general case.

V. Simulation Results

In this section, the performances of various sampling schemes are exemplified in the context of MIMO decoding. Specifically, we examine the decoding error probabilities to assess the convergence rates. By sampling from $D_{\Lambda,\sigma,\mathbf{c}}$, the closest lattice point will be returned with the highest probability, which implies an effective approach to lattice decoding.

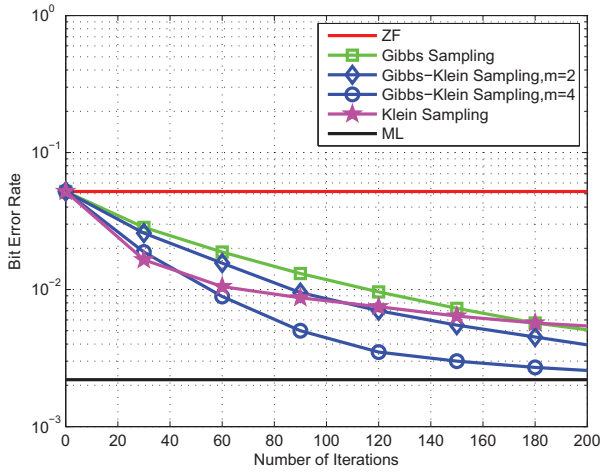

Fig. 1. Bit error rate versus the number of iterations for the uncoded 4×4 MIMO system using 16-QAM.

Fig. 1 depicts the bit error rates (BER) of different Gibbs samplers in a 4×4 uncoded MIMO system with 16-QAM. This corresponds to lattice dimension $n = 8$. The performances of zero-forcing (ZF) and maximum-likelihood (ML) decoding are also shown as benchmarks. We assume a flat fading environment with fixed SNR ($E_b/N_0 = 15$ dB). The channel matrix $\mathbf{H}$ consists of uncorrelated complex Gaussian fading gains with unit variance. $\mathbf{Hx}$ can be viewed as a lattice point in lattice $\Lambda = \mathcal{L}(\mathbf{H})$ and detecting the transmitted signal $\mathbf{x}$ corresponds to solving the CVP. Due to the finite constellation size, the implementation for discrete Gaussian sampling given in [6] is followed.

Klein chose $\sigma = \min_{1\le i\le n}\|\widehat{\mathbf{b}}_i\|/\sqrt{\log\, n}$ and derived polynomial complexity $O(n^{\|\mathbf{Bx}-\mathbf{c}\|^2/\min_i\|\widehat{\mathbf{b}}_i\|^2})$ for his algorithm to find the closest lattice point when it is not far from $\mathbf{c}$ [11]. His derivation is essentially based on the assumption of a Gaussian distribution. However, we now know this choice of σ does not satisfy the smoothing condition and thus his sampler does not really produce Gaussian samples [10].

Here, we follow Klein's choice of σ and apply the proposed Gibbs and Gibbs-Klein samplers to produce Gaussian samples from the lattice. For a fair comparison, when the block size is m, we run block sampling for n/m times, and count this as a full iteration. This corresponds to one run of Klein's original algorithm which samples n components. As shown in Fig. 1, the decoding performance of all the sampling schemes improve with the number of iterations. With the same number of iterations (hence the same complexity), the decoding performance improves with the block size, which implies a faster convergence rate.

ACKNOWLEDGMENT

The authors would like to thank Damien Stehlé for helpful discussions.

REFERENCES

[1] W. Banaszczyk, "New bounds in some transference theorems in the geometry of numbers," *Math. Ann.*, vol. 296, pp. 625–635, 1993.
[2] G. Forney and L.-F. Wei, "Multidimensional constellations–Part II: Voronoi constellations," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 941–958, Aug 1989.
[3] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 913–929, May 1993.
[4] C. Ling and J.-C. Belfiore, "Achieiving the AWGN channel capacity with lattice Gaussian coding," submitted to IEEE Trans. Inform. Theory, Mar. 2012, revised, Nov. 2013. [Online]. Available: http://arxiv.org/abs/1302.5906
[5] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, "Semantically secure lattice codes for the Gaussian wiretap channel," submitted to IEEE Trans. Inform. Theory, Oct. 2012, revised, Oct. 2013. [Online]. Available: http://arxiv.org/abs/1210.6673
[6] S. Liu, C. Ling, and D. Stehle, "Decoding by sampling: a randomized lattice algorithm for bounded distance decoding," *IEEE Trans. Inform. Theory*, vol. 57, pp. 5933–5945, Sep. 2011.
[7] Z. Wang 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.
[8] D. Micciancio and O. Regev, "Worst-case to average-case reuctions based on Gaussian measures," in *Proc. Ann. Symp. Found. Computer Science*, Rome, Italy, Oct. 2004, pp. 372–381.
[9] C. Gentry, A. Sahai, and B. Waters, "Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based," in *CRYPTO*, 2013.
[10] C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapfoors for hard lattices and new cryptographic constructions," in *Proc. 40th Ann. ACM Symp. Theory of Comput.*, Victoria, Canada, 2008, pp. 197–206.
[11] P. Klein, "Finding the closest lattice vector when it is unusually close," in *ACM-SIAM Symp. Discr. Algorithms*, 2000, pp. 937–941.
[12] C. Peikert, "An efficient and parallel Gaussian sampler for lattices," in *CRYPTO*, 2010, pp. 80–97.
[13] Z. Brakerski, A. Langlois, C. Peikert, O. Regev, and D. Stehlé, "Classical hardness of learning with errors," in *STOC*, 2013, pp. 575–584.
[14] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, New York: Springer-Verlag, 2001.
[15] 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.
[16] 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 stationary distribution and mixing time," 2013. [Online]. Available: http://arxiv.org/pdf/1310.7305.pdf.
[17] R. Chen, J. Liu, and X. Wang, "Convergence analyses and comparisons of Markov chain Monte Carlo algorithms in digital communications," *IEEE Trans. on Signal Process.*, vol. 50, no. 2, pp. 255–270, 2002.
[18] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Time*, American Mathematical Society, 2008.
[19] G. O. Roberts and S. K. Sahu, "Updating schemes, correlation structure, blocking and parameterization for Gibbs sampler," *J. Roy. Statist. Soc. Series B*, **59**(2): 291-317, 1997.
[20] O. Regev, "On lattice, learning with errors, random linear codes, and cryptography," *J. ACM*, vol. 56, no. 6, pp. 34:1–34:40, 2009.