Estimation of the Success Probability of Random Sampling by the Gram-Charlier Approximation

Yoshitatsu Matsuda[1], Tadanori Teruya[2], and Kenji Kashiwabara[1]

[1] Department of General Systems Studies,
Graduate School of Arts and Sciences, The University of Tokyo,
3-8-1, Komaba, Meguro-ku, Tokyo, 153-8902, *Japan*.
matsuda@graco.c.u-tokyo.ac.jp,kashiwa@idea.c.u-tokyo.ac.jp
[2] Information Technology Research Institute,
National Institute of Advanced Industrial Science and Technology,
AIST Tokyo Waterfront Bio-IT Research Building,
2-4-7 Aomi, Koto-ku, Tokyo, 135-0064, *Japan*.
tadanori.teruya@aist.go.jp

Abstract. The lattice basis reduction algorithm is a method for solving the Shortest Vector Problem (SVP) on lattices. There are many variants of the lattice basis reduction algorithm such as LLL, BKZ, and RSR. Though BKZ has been used most widely, it is shown recently that some variants of RSR are quite efficient for solving a high-dimensional SVP (they achieved many best scores in TU Darmstadt SVP challenge). RSR repeats alternately the generation of new very short lattice vectors from the current basis (we call this procedure "random sampling") and the improvement of the current basis by utilizing the generated very short lattice vectors. Therefore, it is important for investigating and ameliorating RSR to estimate the success probability of finding very short lattice vectors by combining the current basis. In this paper, we propose a new method for estimating the success probability by the Gram-Charlier approximation, which is a basic asymptotic expansion of any probability distribution by utilizing the higher order cumulants such as the skewness and the kurtosis. The proposed method uses a "parametric" model for estimating the probability, which gives a closed-form expression with a few parameters. Therefore, the proposed method is much more efficient than the previous methods using the non-parametric estimation. This enables us to investigate the lattice basis reduction algorithm intensively in various situations and clarify its properties. Numerical experiments verified that the Gram-Charlier approximation can estimate the actual distribution quite accurately. In addition, we investigated RSR and its variants by the proposed method. Consequently, the results showed that the weighted random sampling is useful for generating shorter lattice vectors. They also showed that it is crucial for solving the SVP to improve the current basis periodically.

1 Introduction

The shortest vector problem (SVP) on a lattice is to find the shortest non-zero lattice vector. In other words, the Euclidean norm (namely, the length)

$\ell = \|\sum_{i=1}^{n} a_i \boldsymbol{b}_i\|$ is minimized with respect to non-zero $\boldsymbol{a} = (a_i) \in \mathbb{Z}^n$, where $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ ($\boldsymbol{b}_i \in \mathbb{R}^m$) is a basis vector of the lattice and there is at least one non-zero element in $\boldsymbol{a}$. Here, the full-rank integral lattice ($n = m$ and $\boldsymbol{b}_i \in \mathbb{Z}^m$) is usually assumed. The SVP is a well-known problem in the field of combinatorial theory and is useful in many applications such as cryptography [12]. Many algorithms have been proposed for finding the shortest vector, for example, enumeration [13] and sieving [2]. However, it is so hard for a high-dimensional lattice to find the true shortest vector directly. Therefore, an approximate version of the SVP is widely used in practice, where we search an extremely short $\boldsymbol{a}$ "near" to the true shortest vector. In other words, we search $\boldsymbol{a}$ satisfying $\|\sum_{i=1}^{n} a_i \boldsymbol{b}_i\| < \hat{\ell}$, where $\hat{\ell}$ is an extremely short threshold. The lattice basis reduction algorithm is a method for solving the approximate SVP for a high dimensional lattice. There are many variants such as the Lenstra-Lenstra-Lovász algorithm (LLL) [14], the block Korkine-Zolotarev algorithm (BKZ) [17], random sampling reduction (RSR) [18], and so on. Recently, a novel lattice basis reduction algorithm was proposed by Fukase and Kashiwabara (called the Fukase-Kashiwabara algorithm (FK) in this paper) [10], which is a variant of RSR. FK and its improved variants by Teruya, Kashiwabara, and Hanaoka [19] can solve the SVP efficiently and has achieved many best scores in TU Darmstadt SVP challenge [16]. Though the reason of the efficiency of FK has been investigated [10, 3], it has not been clarified sufficiently.

RSR is a lattice basis reduction algorithm, which repeats alternately the generation of very short lattice vectors by combining the current basis vectors according to a random sampling distribution and the improvement of the current basis by utilizing the generated very short lattice vectors. Therefore, it is quite important for investigating and ameliorating RSR to estimate the "success" probability of succeeding in generating very short lattice vectors from the current basis [9, 11, 10, 3]. The success probability is defined as the probability that the length of a generated lattice vector is lower than a given very short length. The geometric description inspired by the Gaussian heuristic is a widely-used principle for estimating the success probability by using the volume of the intersection between the lattice and a ball [9]. However, it is intractable in practice if the dimension of lattice n is large. Though some more efficient algorithms estimating the intersection have been proposed recently [11, 3], such methods are nevertheless time-consuming. It is because they need to numerically calculate the volume of the intersection through many small partitions by some methods such as constrained optimization and Monte Carlo simulation. We refer to these method as the "non-parametric" approach in this paper. On the other hand, the "parametric" approach is proposed in [10], where the probability distribution is approximated as a normal distribution with only two parameters (the mean and the variance) under the randomness assumption and the central limit theorem. Though the normal approximation is simple and quite efficient, it cannot estimate the actual distribution accurately as is pointed out in [3].

Our Contribution. In this paper, we propose a new parametric approach by extending the normal approximation under the randomness assumption in order

to estimate the success probability efficiently. The key feature of the proposed method is that the probability is approximated by the Gram-Charlier A series, which is a basic asymptotic expansion of any probability distribution by utilizing the higher order cumulants (such as the skewness and the kurtosis). The proposed method gives a parametric model (in other words, a closed-form expression with a few parameters) for estimating the success probability. It is much more efficient than the previous non-parametric approach and can estimate the probability much more accurately than the simple normal approximation. The accuracy of the proposed method was verified by numerical experiments. Moreover, the intensive investigations with the proposed method discovered why the variants of RSR are more efficient than other algorithms.

Road Map. This paper is organized as follows. Section 2 gives the background of this work: a brief introduction of RSR and its variant FK in Section 2.1, the explanation about both the previous non-parametric and parametric approaches for estimating the success probability in Section 2.2, and the general explanation of the Gram-Charlier approximation in Section 2.3. Section 3 explains our proposed approach which utilizes the Gram-Charlier approximation for estimating the success probability in the SVP. In Section 4, the numerical experiments show that the proposed approach can accurately estimate the success probability. In Section 5, the experimental investigations with the proposed method clarify the two reasons (the weighted random sampling and the periodical improvement of the basis) why FK is quite efficient for solving the SVP. In Section 6, the validity of the randomness assumption and the accuracy of our proposed approach are discussed. Moreover, the dependence among the indices of the natural number representation in our approach is discussed. In addition, the convergence property of the Gram-Charlier approximation is discussed experimentally. Lastly, this paper is concluded in Section 7.

2 Background

2.1 Random Sampling Reduction and Fukase-Kashiwabara Algorithm

Preliminaries. Here, the preliminary notations and definitions are introduced. A full-rank integral lattice basis is given as an $n \times n$ matrix $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, where each $\boldsymbol{b}_i = (b_{ij}) \in \mathbb{Z}^n$ is a basis vector. The lattice $L(\boldsymbol{B})$ is defined as an additive group consisting of $\sum_{i=1}^{n} a_i \boldsymbol{b}_i$ for $a_i \in \mathbb{Z}$. The Euclidean inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ is denoted by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{y}$. The Euclidean norm (length) of $\boldsymbol{x}$ is defined as $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. $\boldsymbol{b}_i$ can be orthogonalized to $\boldsymbol{b}_i^* = \left(b_{ij}^*\right)$ by the following Gram-Schmidt process:

$$\boldsymbol{b}_i^* = \boldsymbol{b}_i - \sum_{j=1}^{i-1} \eta_{ji} \boldsymbol{b}_j^* \text{ and } \eta_{ji} = \frac{\langle \boldsymbol{b}_j^*, \boldsymbol{b}_i \rangle}{\left\|\boldsymbol{b}_j^*\right\|^2}. \tag{1}$$

Then, $\langle \boldsymbol{b}_i^*, \boldsymbol{b}_j^* \rangle = 0$ holds for $i \neq j$. Note that $\|\boldsymbol{b}_i^*\|$ is not constrained to be 1. $\boldsymbol{B}$ provides the fundamental parallelpiped $\{\sum_{i=1}^{n} t_i \boldsymbol{b}_i : t_i \in [0, 1)\}$, the volume of

which is independent on the choice of the basis for the same lattice. This volume is called the determinant (or co-volume) of $L(\boldsymbol{B})$. Note that the determinant is equal to $\prod_{i=1}^{n} \|\boldsymbol{b}_i^*\|$.

A lattice vector $\sum_{i=1}^{n} a_i \boldsymbol{b}_i$ ($a_i \in \mathbb{Z}$) is given as $\sum_{i=1}^{n} \zeta_i \boldsymbol{b}_i^*$. The squared length of the lattice vector (denoted by ℓ^2) is given as

$$\ell^2\left(\sum_{i=1}^{n} a_i \boldsymbol{b}_i\right) = \left\|\sum_{i=1}^{n} \zeta_i \boldsymbol{b}_i^*\right\|^2 = \sum_{i=1}^{n} \zeta_i^2 \|\boldsymbol{b}_i^*\|^2. \tag{2}$$

Because each $\zeta_i \in \mathbb{R}$ is given as the sum of $\bar{\zeta}_i$ ($-\frac{1}{2} \leq \bar{\zeta}_i < \frac{1}{2}$) and an integer, ζ_i is uniquely determined by a natural number d_i satisfying

$$-\frac{d_i+1}{2} \leq \zeta_i < -\frac{d_i}{2} \text{ or } \frac{d_i}{2} \leq \zeta_i < \frac{d_i+1}{2}, \tag{3}$$

where the natural numbers begin with 0 (namely, $d_i = 0$ is allowed). The sequence $\boldsymbol{d} = (d_1, \ldots, d_n)$ is called the natural number representation. It was shown in [10] that any vector in the lattice is uniquely determined by $\boldsymbol{d}$ and $\boldsymbol{B}^* = (\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*)$. In other words, there is a one-to-one correspondence between a natural number representation and a lattice vector.

The randomness assumption is defined as follows.

Assumption 1 (Randomness Assumption) *Each $\bar{\zeta}_i$ is uniformly distributed in $\left[-\frac{1}{2}, \frac{1}{2}\right)$ and is statistically independent of $\bar{\zeta}_j$ for $j \neq i$.*

Though there are several different definitions of the randomness assumption, we employ the above one based on Schnorr's assertion [18]. Though the randomness assumption cannot hold rigorously [3], this paper verifies that this assumption is quite useful for estimating the success probability.

By assuming that the volume of $L(\boldsymbol{B})$ is approximately equal to the volume of a ball with the diameter of the shortest lattice vector length, the length of the shortest lattice vector is estimated as

$$\ell_{\mathrm{GH}} = \frac{\left(\Gamma\left(\frac{n}{2}+1\right)\prod_{i=1}^{n}\|\boldsymbol{b}_i^*\|\right)^{\frac{1}{n}}}{\sqrt{\pi}}, \tag{4}$$

where Γ is the gamma function occurring in the calculation of the volume of an n-dimensional ball [12]. This approximation is called the Gaussian heuristic. Though the original shortest vector problem (SVP) is to find the shortest non-zero lattice vector, it is generally too difficult to solve. In the similar way as in TU Darmstadt SVP challenge [16], we define the SVP as finding an extremely short lattice vector whose length is less than $(1+\epsilon)\,\ell_{\mathrm{GH}}$ where ϵ is a small positive constant ($\epsilon = 0.05$ in TU Darmstadt SVP challenge).

Random Sampling Reduction. Here, RSR [18] is explained in brief. It is assumed that the Gram-Schmidt orthogonalized basis $\boldsymbol{B}^*$ is roughly reduced by BKZ with block size 20 (or other classical efficient algorithms such as LLL)

so that $\boldsymbol{b}_i^*$ is roughly arranged in descending order of $\|\boldsymbol{b}_i^*\|$. Then, RSR solves the SVP by alternately repeating the generation of very short lattice vectors from the current basis (we call this process "Random Sampling" (RS)) and the reduction of the basis by the generated lattice vectors. First, RSR generates the candidates for a very short lattice vector randomly by the following sampling distribution in the natural number representation:

$$d_i = \begin{cases} 0 & (i \leq n-u-1), \\ 0 \text{ or } 1 & (n-u \leq i \leq n-1), \\ 1 & (i = n), \end{cases} \tag{5}$$

where u is a constant integer ($u < n$). In other words, d_i is sampled randomly from 0 and 1 with equal probability if $n-u \leq i \leq n-1$. At most 2^u lattice vectors are sampled. Second, RSR selects a lattice vector reducing $\|\boldsymbol{b}_i^*\|$ greatly for an index i, inserts the selected lattice vector to a column of $\boldsymbol{B}$, and utilizes BKZ for generating a new lattice basis $\boldsymbol{B}$. Consequently, the basis is reduced. The brief algorithmic description of RSR and its variants is given in Algorithm 1.

Algorithm 1 The brief description of RSR and its variants.

Require: $\boldsymbol{B}$.
Roughly reduce $\boldsymbol{B}$ by BKZ (or other efficient algorithms).
while $\boldsymbol{B}$ does not converge under a given condition **do**
  Pick up some possible natural number representations randomly, and generate the corresponding lattice vectors.
  for all the generated "short" lattice vectors **do**
    Generate a new basis by inserting the lattice vector to $\boldsymbol{B}$ and applying BKZ (or other efficient algorithms) to $\boldsymbol{B}$.
    if the new basis is "better" than the current $\boldsymbol{B}$ under a given condition **then**
      Update $\boldsymbol{B}$.
    end if
  end for
end while
return the reduced $\boldsymbol{B}$.

Note that all the possible natural number representations are sampled deterministically in many practical cases. It is because a probabilistic algorithm generates a large number of duplicate samples. The upper bound of the complexity of RSR can be estimated theoretically under the randomness assumption [18], which is lower than that of other widely-used methods such as BKZ. However, it was not as efficient as BKZ in practice until its variant FK was proposed.

Fukase-Kashiwabara Algorithm. FK was a lattice basis reduction algorithm for solving the SVP, which was originally proposed by Fukase and Kashiwabara

[10] and has been developed by Teruya, Kashiwabara, and Hanaoka [19]. FK and its variants are currently known to be the most efficient algorithm which achieves the best scores for $n = 132, \ldots, 150$ in TU Darmstadt SVP challenge [16]. It has not been clarified sufficiently why FK is quite efficient. One main motivation of this work is to clarify the reason. Here, FK is briefly described in comparison with RSR. FK is a variant of RSR, which generates new lattice vectors randomly from the current basis and update the basis by using a new lattice vector. On the other hand, FK and its variants differ from RSR in many aspects, for example, the sampling of new lattice vectors, the evaluation of the reduced bases, the utilization of parallel processing, the storing of the candidate lattice vectors, and so on. In this paper, we focus on only the two aspects: the sampling distribution of new lattice vectors and the evaluation of the reduced bases. They are emphasized in the original paper of FK [10].

First, FK generalizes the sampling distribution in the natural number representation. The random sampling of the original RSR (which is called the standard RS) chooses only 0 or 1 with the equal probability. An extended version of the random sampling (the extended RS) is proposed in [4, 15], which extended the possible natural numbers $d_i \in \{0, 1\}$ to $d_i \in \{0, 1, \ldots, T_i\}$ with equal probability. Here, T_i is a small natural number such as $T_i = 3$. FK generalizes the extended random sampling furthermore so that a possible number p for the index i occurs by a probability α_{ip}, where $\alpha_{ip} \geq 0$ and $\sum_{p=0}^{T_i} \alpha_{ip} = 1$. In other words, the sampling distribution is determined by an $n \times T$ matrix $\boldsymbol{\alpha} = (\alpha_{ip})$ where T is the maximum of T_i over i. It is called the weighted RS. As ζ_i is assumed to be a uniformly distributed random variable, the probability of ζ_i is given as

$$P_\zeta(\zeta_i) = \begin{cases} \alpha_{ip} & -\frac{p+1}{2} \leq \frac{\zeta_i}{2} < -\frac{p}{2} \text{ or } \frac{p}{2} \leq \frac{\zeta_i}{2} < \frac{p+1}{2} \ (p \leq T), \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Fig. 1 shows the examples of the probability distributions of ζ_i for the standard RS, the extended RS, and the weighted RS. Note that the standard RS and the extended RS can sample all the possible natural number representations deterministically. Such deterministic sampling methods are mainly used in practice. On the other hand, the weighted RS cannot be actualized deterministically. Although the original FK employs a deterministic sampling method by allowing the dependence among ζ_i's, it is assumed in this paper that every ζ_i is independent of each other in order to facilitate the efficient calculation. It is because this paper focuses on the efficient estimation of the success probability more heavily than the construction of the deterministic sampling. Moreover, we discuss the relationship between our sampling distribution without the dependence and the previous deterministic methods in Section 6.2.

Second, FK improves the basis periodically so that the following target function $\Phi(\boldsymbol{B})$ is reduced:

$$\Phi(\boldsymbol{B}) = \sum_{i=1}^{n} \|\boldsymbol{b}_i^*\|^2. \tag{7}$$

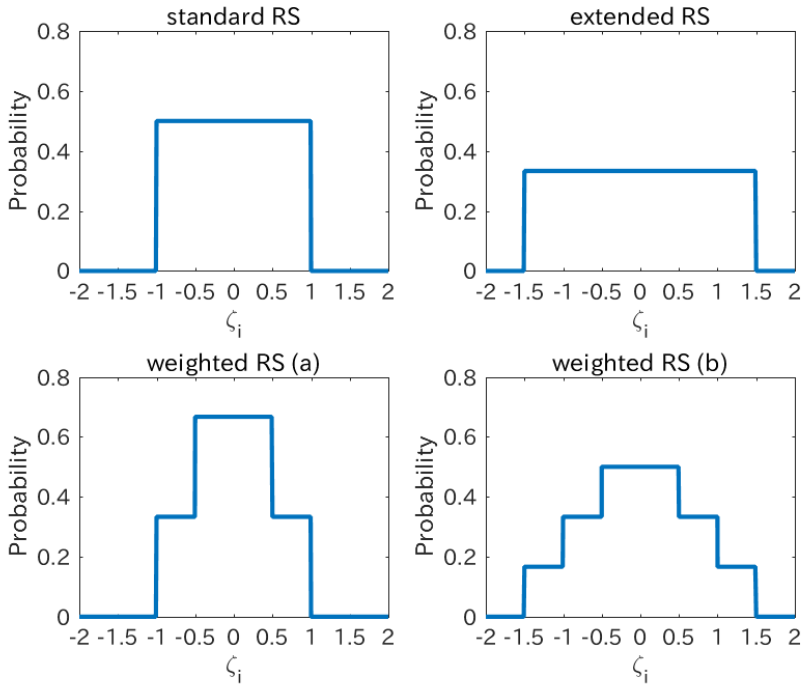


Fig. 1. Examples of probability distributions of ζ_i for the standard RS ($T = 1$, $\alpha_{ip} = \frac{1}{2}$), the extended RS ($T = 2$, $\alpha_{ip} = \frac{1}{3}$), the weighted RS (a) ($T = 1$, $\alpha_{i0} = \frac{2}{3}$, $\alpha_{i1} = \frac{1}{3}$), and the weighted RS (b) ($T = 2$, $\alpha_{i0} = \frac{1}{2}$, $\alpha_{i1} = \frac{1}{3}$, $\alpha_{i1} = \frac{1}{6}$).

Though it was partially explained by the normal approximation in [10] why this target function is effective, the explanation was insufficient (see Section 2.2).

2.2 Estimation of the Success Probability of Finding Very Short Lattice Vectors

In order to investigate the behaviors of RSR and its variants, it is important to estimate accurately the success probability that the current basis generates very short lattice vectors over a given random sampling. Here, we explain the two previous approaches for estimating the probability: the non-parametric approach using the geometric description and the parametric one using the normal approximation.

Non-parametric Approach. The geometric description inspired by the Gaussian heuristic estimates the success probability by approximately counting the number of very short generated lattice vectors. The number can be defined as the intersection between all the generated lattice vectors and a ball with a very short diameter R. For example, R is given as $(1+\epsilon)\,\ell_{\mathrm{GH}}$ at the final stage of the SVP, where ϵ is a small constant. The geometric description is a non-parametric estimation of the probability distribution, which gives a direct and accurate estimation under an arbitrary sampling distribution. However, it was generally intractable to count the actual intersection because quite many lattice vectors are generated from a given basis. Recently, some state of the art methods were proposed for estimating the intersection in practice by dividing the lattice space into small partitions such as cylinders [11] and boxes [3] and utilizing various techniques for acceleration. Nevertheless, they are still time-consuming because their complexity depends on a large number of generated lattice vectors and small partitions. Moreover, it is difficult to investigate their estimation analytically because it is non-parametric.

Parametric Approach. The normal approximation is a parametric approach under the randomness assumption and the central limit theorem [10]. It approximates the probability distribution of the squared length ℓ^2 of a generated lattice vector as a normal distribution whose parameters are only the mean and the variance. The mean of ℓ^2 for a generated lattice vector is given as the first order moment of ℓ^2:

$$E\left(\ell^2\right) = \sum_{i=1}^{n} E\left(\zeta_i^2\right) \left\|\boldsymbol{b}_i^*\right\|^2, \tag{8}$$

where $E\left(\right)$ is the expectation operator over all the possible bases. ζ_i is the sum of $\bar{\zeta}_i$ (a uniformly distributed random variable in $[-\frac{1}{2}, \frac{1}{2})$) and $\frac{\pm d_i}{2}$ (a half integer). Fukase and Kashiwabara calculated $E\left(\ell^2\right)$ analytically by assuming that every d_i is equal to 0 [10]. The distribution of $\bar{\zeta}_i$ is formally given by

$$P_{\bar{\zeta}}(x) = \begin{cases} 1 & -\frac{1}{2} \le x < \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Then, the mean $\mu = E\left(\ell^2\right)$ is given as

$$\mu = \sum_{i=1}^{n} \|\boldsymbol{b}_i^*\|^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 \mathrm{d}x = \frac{\sum_{i=1}^{n} \|\boldsymbol{b}_i^*\|^2}{12}. \tag{10}$$

Similarly, the second order moment $E\left(\ell^4\right)$ is given as

$$E\left(\ell^4\right) = \sum_{i=1}^{n} E\left(\zeta_i^4\right) \|\boldsymbol{b}_i^*\|^4 + \sum_{i=1}^{n} E\left(\zeta_i^2\right) \|\boldsymbol{b}_i^*\|^2 \sum_{j=1, j\neq i}^{n} E\left(\zeta_j^2\right) \left\|\boldsymbol{b}_j^*\right\|^2, \tag{11}$$

where we utilize the statistical independence between ζ_i and ζ_j $(i \neq j)$ under the randomness assumption. Then, the variance $\sigma^2 = E\left(\ell^4\right) - \mu^2$ is given as

$$\sigma^2 = \sum_{i=1}^{n} E\left(\zeta_i^4\right) \|\boldsymbol{b}_i^*\|^4 - \frac{\sum_{i=1}^{n} \|\boldsymbol{b}_i^*\|^4}{144} = \frac{\sum_{i=1}^{n} \|\boldsymbol{b}_i^*\|^4}{180}. \tag{12}$$

μ and σ^2 determine the simple normal distribution function which can be investigated both numerically and analytically [10]. However, the serious weakness of this approximation is that the estimation is different from the actual distribution especially when the length of a generated lattice vector is very short as is pointed out in [3]. In other words, the normal approximation is too rough to estimate accurately the success probability. In addition, it does not consider any sampling distribution because every d_i is assumed to be a constant (0 in FK).

2.3 Gram-Charlier A Series

The normal approximation cannot estimate the actual distribution sufficiently accurately. In this paper, we improve the simple normal approximation by utilizing higher order cumulants. We employ the Gram-Charlier A series [7, 21] for this purpose. Here, we explain this technique in brief.

Overview. Let $P(x)$ be a probabilistic distribution function satisfying $P(x) \geq 0$ and $\int_{-\infty}^{\infty} P(x)\,\mathrm{d}x = 1$. There are three well-known series expansion of $P(x)$: the Edgeworth series, the Gram-Charlier A series, and the Gram-Charlier B series [7]. The Edgeworth series expansion is not suitable to estimate the success probability because it assumes that x is the sum of the independent and identically distributed random variables. The Gram-Charlier B series expansion is also not suitable because the principal probability distribution is the exponential one. Therefore, we employ the Gram-Charlier A series. We assume that the random variable x is normalized. In other words, its mean μ and its variance σ^2 are 0 and 1, respectively. This assumption does not lose the generality because the variable of $P(x)$ is easily transformed by

$$P(z) = \frac{P(x)}{\sigma_z}, \tag{13}$$

where $x = (z - \mu_z)/\sigma_z$ (μ_z and σ_z are the mean and the standard deviation of z, respectively). Then, the Gram-Charlier A series of $P(x)$ is given as

$$P(x) = \left(1 + \sum_{r=3}^{\infty} c_r H_r(x)\right) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \tag{14}$$

where $H_r(x)$ is the r-th degree Hermite polynomial defined as

$$H_r(x)\, e^{-\frac{x^2}{2}} = (-1)^r \frac{\mathrm{d}^r}{\mathrm{d}x^r} e^{-\frac{x^2}{2}}. \tag{15}$$

c_r is the r-th coefficient depending on the r-th and lower order cumulants (the details are described below). The series is guaranteed to converge if $P(x)$ decreases faster than $e^{-\frac{x^2}{4}}$ [6, 21]. This condition is generally satisfied in the search in the natural number representation because there exists a bound T.

Cumulants. Here, we explain the cumulants, which are the most important statistics for the Gram-Charlier approximation. The cumulants of $P(x)$ are formally defined as follows. The cumulant generating function $K(t)$ is defined as

$$K(t) = \log \int_{-\infty}^{\infty} e^{tx} P(x)\, \mathrm{d}x. \tag{16}$$

The power series expansion of $K(t)$ is given as

$$K(t) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!}, \tag{17}$$

where κ_r is the r-th order cumulant of $P(x)$. By using the r-th derivative of $K(t)$, κ_r is given as

$$\kappa_r = K^{(r)}(0). \tag{18}$$

κ_r is calculated in practice by the moments of $P(x)$. The r-th order moment of $P(x)$ (denoted by μ_r) is given as

$$\mu_r = \int_{-\infty}^{\infty} x^r P(x)\, \mathrm{d}x. \tag{19}$$

Then, the r-th order cumulant is recursively given as

$$\kappa_r = \mu_r - \sum_{m=1}^{r-1} \binom{r-1}{m-1} \kappa_m \mu_{r-m}. \tag{20}$$

$\kappa_1 = \mu$ and $\kappa_2 = \sigma^2$ are the mean and the variance. κ_3 and κ_4 are called the skewness and the kurtosis, respectively. The cumulants are the important statistics characterizing any probability distribution and have various good mathematical properties. In this paper, we utilize the homogeneity and the additivity in the

following. Let $\kappa_r(x)$ be the r-th order cumulant of a random variable x. Then, the following equation holds:

$$\kappa_r(ax) = a^r \kappa_r(x), \tag{21}$$

where a is an arbitrary constant. This property is called the homogeneity. If x and y are statistically independent random variables, the following equation holds:

$$\kappa_r(x+y) = \kappa_r(x) + \kappa_r(y). \tag{22}$$

This property is called the additivity. By utilizing these two properties under the randomness assumption, the distribution of the length of a generated lattice vector in the SVP is easily estimated. In addition, the following important property holds for any normal distribution: $\kappa_r = 0$ for $r \geq 3$. Therefore, the cumulant generating function of a normal distribution is given as

$$K_{\text{normal}}(t) = \kappa_1 t + \frac{\kappa_2 t^2}{2}. \tag{23}$$

When a random variable z is normalized to $x = (z - \kappa_1(z)) / \sqrt{\kappa_2(z)}$, the (normalized) cumulants of x (denoted by λ_r) are given as

$$\lambda_r(x) = \frac{\kappa_r(z)}{(\kappa_2(z))^{\frac{r}{2}}} \tag{24}$$

for $r \geq 3$ ($\lambda_1 = 0$ and $\lambda_2 = 1$).

Coefficients. Here, we briefly explain the derivation of the Gram-Charlier A series and give its coefficient c_r. Let $P(x)$ be any probability distribution function of a normalized random variable x (see [21] for the details). The characteristic function of $P(x)$ (denoted by $f(t)$) is defined as

$$f(t) = \int_{-\infty}^{\infty} e^{itx} P(x)\, \mathrm{d}x, \tag{25}$$

where i is the imaginary unit. Using the cumulant generating function $K(t)$, $f(t)$ is given as

$$f(t) = e^{K(it)} = \exp\left(it + \frac{-t^2}{2} + \sum_{r=3}^{\infty} \lambda_r \frac{(it)^r}{r!} \right). \tag{26}$$

On the other hand, the characteristic function of the unit normal distribution ($\frac{1}{2\pi} e^{\frac{-x^2}{2}}$) is given as

$$g(t) = \exp\left(it + \frac{-t^2}{2} \right). \tag{27}$$

Then, $f(t)$ is given as

$$f(t) = \exp\left(\sum_{r=3}^{\infty} \lambda_r \frac{(it)^r}{r!}\right) g(t) = \left(\sum_{p=0}^{\infty} \frac{1}{p!} \left(\sum_{r=3}^{\infty} \lambda_r \frac{(it)^r}{r!}\right)^p\right) g(t)$$
$$= \left(1 + \sum_{r=3}^{\infty} c_r (it)^r\right) g(t), \tag{28}$$

where the Maclaurin series of the exponential function is utilized and c_r is a coefficient. Because the characteristic function $f(t)$ (replacing t with $-t$) can be regarded as the inverse Fourier transform of any $P(x)$, the following equation holds:

$$(-it)^r f(t) = \int_{-\infty}^{\infty} \frac{\mathrm{d}^r P(x)}{\mathrm{d}x^r} e^{itx} \mathrm{d}x. \tag{29}$$

By applying the Fourier transformation to $f(t)$, the following equation is derived:

$$P(x) = \left(1 + \sum_{r=3}^{\infty} c_r (-1)^r \frac{\mathrm{d}^r}{\mathrm{d}x^r}\right) \frac{e^{\frac{-x^2}{2}}}{2\pi} = \left(1 + \sum_{r=3}^{\infty} c_r H_r(x)\right) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}. \tag{30}$$

This is the Gram-Charlier A series. Each c_q is the coefficient of u^q of the following polynomial $\psi(u)$:

$$\psi(u) = \sum_{p=0}^{\infty} \frac{1}{p!} \left(\sum_{r=3}^{\infty} \lambda_r \frac{u^r}{r!}\right)^p. \tag{31}$$

Though $\psi(u)$ includes the infinite summation, c_q can be calculated from a finite sets of the terms with $r \leq q$ and $p \leq \frac{q}{3}$. The complexity of calculating c_q is $O\left(q^{\frac{q}{3}}\right)$ if every λ_r $(r \leq q)$ is given.

There is another formulation of the coefficients, which is often more efficient than this usual formulation. See Section 6.4 for the details.

Cumulative Distribution Function. The cumulative distribution function of $P(x)$ is defined as

$$F(x) = \int_{-\infty}^{x} P(x)\, \mathrm{d}x. \tag{32}$$

The integration of the $H_r(x) e^{-\frac{x^2}{2}}$ is given as

$$\int_{-\infty}^{x} H_r(u) e^{-\frac{u^2}{2}} \mathrm{d}u = -\int_{-\infty}^{x} \frac{\mathrm{d}H_{r-1}(u) e^{-\frac{u^2}{2}}}{\mathrm{d}u} \mathrm{d}u = -H_{r-1}(x) e^{-\frac{x^2}{2}}. \tag{33}$$

Therefore, the Gram-Charlier A series of $F(x)$ is given as

$$F(x) = \int_{\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} \mathrm{d}u - \left(\sum_{r=3}^{\infty} c_r H_{r-1}(x)\right) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \tag{34}$$

where the first term of the right side is a sort of the Gaussian error function. The Gaussian error function is easily calculated numerically and has some approximate forms.

Examples. Lastly, we show the concrete examples of the Hermite polynomials $H_r(x)$ for $r = 2, \ldots, 9$ and the coefficients c_r for $r = 3, \ldots, 10$ in the following:

$$\begin{aligned}
H_2(x) &= x^2 - 1, \\
H_3(x) &= x^3 - 3x, \\
H_4(x) &= x^4 - 6x^2 + 3, \\
H_5(x) &= x^5 - 10x^3 + 15x, \\
H_6(x) &= x^6 - 15x^4 + 45x^2 - 15, \\
H_7(x) &= x^7 - 21x^5 + 105x^3 - 105x, \\
H_8(x) &= x^8 - 28x^6 + 210x^4 - 420x^2 + 105, \\
H_9(x) &= x^9 - 36x^7 + 378x^5 - 1260x^3 + 945x,
\end{aligned} \tag{35}$$

and

$$\begin{aligned}
c_3 &= \frac{\lambda_3}{3!}, \\
c_4 &= \frac{\lambda_4}{4!}, \\
c_5 &= \frac{\lambda_5}{5!}, \\
c_6 &= \frac{\lambda_6 + 10\lambda_3^2}{6!}, \\
c_7 &= \frac{\lambda_7 + 35\lambda_3\lambda_4}{7!}, \\
c_8 &= \frac{\lambda_8 + 56\lambda_3\lambda_5 + 35\lambda_4^2}{8!}, \\
c_9 &= \frac{\lambda_9 + 84\lambda_3\lambda_6 + 126\lambda_4\lambda_5 + 280\lambda_3^3}{9!}, \\
c_{10} &= \frac{\lambda_{10} + 120\lambda_3\lambda_7 + 210\lambda_4\lambda_6 + 126\lambda_5^2 + 2100\lambda_3^2\lambda_4}{10!}.
\end{aligned} \tag{36}$$

3 Method

Here, we propose a method for estimating the success probability of succeeding in generating very short lattice vectors. This method is much more efficient than the other estimation methods because it uses a few parameters depending on the sampling distribution ($\boldsymbol{\alpha}$) and the norm of each column of the orthogonalized lattice basis ($\|\boldsymbol{b}_i^*\|$). The key idea of the proposed method is to approximate the probability distribution of the squared length $\ell^2 = \sum_i \zeta_i^2 \|\boldsymbol{b}_i^*\|^2$ of a generated lattice vector from a given lattice basis by using the Gram-Charlier A series and

assuming the randomness assumption. For this purpose, we need to calculate the r-th order moment of ζ_i^2. As the probability of ζ_i is generally given by Eq. (6) under the randomness assumption, the r-th order moment $\mu_r\left(\zeta_i^2\right)$ is given as

$$\mu_r\left(\zeta_i^2\right) = \sum_{p=0}^{T} 2\int_{\frac{p}{2}}^{\frac{p+1}{2}} \alpha_{ip}\zeta_i^{2r}\mathrm{d}\zeta_i = \sum_{p=0}^{T} \alpha_{ip}\beta_{pr}, \tag{37}$$

where β_{pr} can be analytically calculated as

$$\beta_{pr} = \frac{\left((p+1)^{2r+1} - p^{2r+1}\right)}{(2r+1)\,2^{2r}}. \tag{38}$$

Then, the r-th order cumulant $\kappa_r\left(\zeta_i^2\right)$ can be calculated by the following recursion:

$$\kappa_r = \mu_r - \sum_{m=1}^{r-1} \binom{r-1}{m-1} \kappa_m \mu_{r-m}. \tag{39}$$

Note that ζ_i^2 and ζ_j^2 are independent for $i \neq j$ under the randomness assumption. Therefore, the r-th order cumulant of ℓ^2 is given as

$$\kappa_r\left(\ell^2\right) = \sum_{i=1}^{n} \kappa_r\left(\zeta_i^2\right) \left\|\boldsymbol{b}_i^*\right\|^{2r}, \tag{40}$$

where we utilize the homogeneity and the additivity of the cumulants. The r-th order normalized cumulant of ℓ^2 is given by $\lambda_r\left(\ell^2\right) = \frac{\kappa_r}{\sqrt{\kappa_2^r}}$. Then, the coefficients of the Gram-Charlier A series (c_r) are calculated by $\lambda_3, \ldots, \lambda_r$. Let $Q \geq 3$ be a positive integer determining the degree of approximation. When ℓ^2 is normalized to a random variable x, the Gram-Charlier approximation of the cumulative distribution function $F(x)$ is given as

$$\bar{F}_Q(x) = \int_{\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\mathrm{d}u - \left(\sum_{r=3}^{Q} c_r H_{r-1}(x)\right) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}. \tag{41}$$

As the shape of a cumulative distribution function is invariant under an affine transformation of the variable, the Gram-Charlier approximation of $F(z)$ for $z = \ell^2$ is given as

$$F_Q(z) = \bar{F}_Q\left(\frac{z-\kappa_1}{\sqrt{\kappa_2}}\right). \tag{42}$$

$F_Q(z)$ is uniquely determined by the sampling distribution $\boldsymbol{\alpha} = (\alpha_{ip})$, the orthogonalized lattice basis $\boldsymbol{B}^* = \left(b_{ij}^*\right)$, and the degree of approximation Q. Finally, $F_Q\left(\hat{\ell}^2\right)$ gives the approximation of the success probability of generating a lattice vector whose squared length is shorter than a threshold $\hat{\ell}^2$. Note that $\hat{\ell}^2$ can be changed easily because $F_Q(z)$ can be calculated for any z. For example, the approximation of the success probability in TU Darmstadt SVP challenge

is given as $F_Q\left(\left(1.05\right)^2 \ell_{\mathrm{GH}}^2\right)$. The probability distribution $P_Q\left(z\right) = \mathrm{d}F_Q\left(z\right)/\mathrm{d}z$ is also easily calculated by Eq. (14) if necessary. The algorithmic description of the above process is given in Algorithm 2.

Algorithm 2 The Gram-Charlier approximation of the cumulative distribution of the squared lengths of generated lattice vectors.

Require: $\boldsymbol{\alpha}$, $\boldsymbol{B}^*$, and Q.
Calculate the moments $\mu_r\left(\zeta_i^2\right)$ by $\boldsymbol{\alpha}$ ($r \leq Q$, the same hereinafter).
Calculate the cumulants $\kappa_r\left(\zeta_i^2\right)$ by $\mu_r\left(\zeta_i^2\right)$.
Calculate the cumulants $\kappa_r\left(\ell^2\right)$ and the normalized ones $\lambda_r\left(\ell^2\right)$ by $\kappa_r\left(\zeta_i^2\right)$ and $\boldsymbol{B}^*$.
Calculate the coefficients c_r by $\lambda_r\left(\ell^2\right)$.
return the approximate cumulative distribution $F_Q\left(z\right)$ (and the probability distribution $P_Q\left(z\right)$ if necessary).

The complexity of this algorithm is $O\left(n^2\right) + O\left(nQT\right) + O\left(nQ^2\right) + O\left(Q^{\frac{Q}{3}}\right)$ (the calculations of every $\left\|\boldsymbol{b}_i^*\right\|^2$, every $\mu_r\left(\zeta_i^2\right)$, every $\kappa_r\left(\zeta_i^2\right)$, and every c_r). If T is a finite number, $F_Q\left(z\right)$ converges to the true probability for $Q \to \infty$ [6, 21]. Unfortunately, no theoretical bound of Q has been achieved. However, the numerical experiments in Section 4 will show that $F_Q\left(z\right)$ can approximate the actual distributions if Q is larger than 50. It will be also shown that the calculation time can be within one minute even for $Q = 70$.

There is another (often more efficient) algorithm using a different calculation method of the coefficients. See Section 6.4 for the details.

4 Experiment

Here, it is numerically verified whether the Gram-Charlier approximation $F_Q\left(z\right)$ can estimate the actual success probability for various lattice bases $\boldsymbol{B}$ and various sampling distributions $\boldsymbol{\alpha}$. The following four lattice bases are used:

- **B128** was originally generated in TU Darmstadt SVP challenge [16] ($n = 128$, seed $= 0$) and was roughly reduced by the BKZ algorithm of the fplll package [20] with block size 20. The target function of FK was not reduced ($\Phi\left(\boldsymbol{B}\right) = 195.3$ where the Gaussian heuristic is normalized to 1).
- **B128reduced** was reduced largely from B128 by a variant algorithm of FK ($\Phi\left(\boldsymbol{B}\right) = 102.5$).
- **B100** was generated in the same way as B128 except for $n = 100$.
- **B150** was generated in the same way as B128 except for $n = 150$.

The following two sampling distributions are used, which are based on the standard RS and the extended RS:

- 2^{25}**-RS** consists of 2^{25} samples generated by the standard RS with $u = 25$.

– $2^{17}3^5$**-RS** consists of $2^{17} \times 3^5 \simeq 2^{25}$ samples, where T_i was set to 3 for $n-5 \leq i \leq n-1$ and 2 for $n-22 \leq i \leq n-6$ ($T_i = 1$ for the others). In other words, d_i for $i = n-5, \ldots, n-1$ is selected uniformly randomly from $0, 1, 2$ instead of $0, 1$.

Note that the above sampling distributions were estimated accurately (namely, without any sampling error) because they enumerated all the possible samples deterministically. The weighted RS was not employed in this experiment because of its inevitable sampling error. The actual probability of the squared length ℓ^2 over a sampling on a basis was estimated by a histogram of the squared lengths over all the generated lattice vectors. The logarithm with base 2 of the cumulative distribution was used for displaying the result because it clarifies the differences among the cumulative distribution functions when ℓ^2 is short.

Fig. 2 shows the results for B128 and B128reduced over 2^{25}-RS. The actual distributions over the sampling are displayed by the bumpy blue curves. The normal approximation is displayed by the uppermost orange curve. The corresponding Gram-Charlier approximations $F_Q(z)$ are also displayed ($Q = 10$, 50, 60, and 70). Note that the complete form of the approximation with $Q = 10$ can be described in Section 2.3. The approximation with the highest degree ($Q = 70$) is displayed by the light blue curve. The pre-estimated minimal squared length ℓ^2_{GH} (by the Gaussian heuristic) is displayed by the dashed vertical line. The squared lengths of the generated lattice vectors are normalized so that ℓ^2_{GH} is equal to 1. Moreover, the squared length of the current shortest vector (namely, the first index) of the basis (denoted by ℓ^2_{CSV}) is displayed by the dot-dashed vertical line. First, we can observe that $F_Q(z)$ estimated the actual distributions quite accurately if Q is larger than 50. It verifies that our method is useful at least when the cumulative probability is more than 2^{-25}. Next, we can observe that $F_Q(z)$ with $Q \geq 50$ approximately converged to a decreasing curve at least when the cumulative distribution is more than 2^{-50}. Though it is hard to guarantee that these converged curves can estimate the actual cumulative probability, it can be asserted that the curves are accurate under the randomness assumption because they are determined only by $\boldsymbol{\alpha}$ and $\boldsymbol{B}^*$. Third, $F_Q(z)$ with $Q \geq 50$ converged around ℓ^2_{CSV}. Therefore, it can be determined easily whether the current shortest vector is improved by a given sampling distribution or not. Fourth, the converged curves fell sharply below thresholds around the Gaussian heuristic in comparison with the gently-reducing normal approximations. It shows that the curve for B128 cannot achieve a very short lattice vector even if the number of samplings is near the infinity. On the other hand, the curve for B128reduced seems to achieve the Gaussian heuristic by a sufficiently large number of samplings. The results show that it is crucial for finding a very short lattice vector to reduce the lattice basis.

In order to verify the applicability of the Gram-Charlier approximations in various cases, different sizes of lattice bases (Fig. 3) and a different sampling distribution (Fig. 4) were employed. Fig. 3 shows the results for different sizes of lattice bases (B100 and B150) over 2^{25}-RS. In both cases, $F_Q(z)$ with $Q \geq 50$ converged to a curve, which was approximately equivalent to the actual cumulative

distribution. Fig. 4 shows the results of the bases with B128 and B128reduced over a different sampling distributions $2^{17}3^5$-RS. The results were quite similar to those in Fig. 2, where $F_Q(z)$ with $Q \geq 50$ converged to a curve and could estimate the actual cumulative distribution accurately. In summary, these results verify that the Gram-Charlier approximations are useful for various lattice bases and various sampling distributions.

Regarding the efficiency of our method, Fig. 5 shows the actual calculation time for the above experimental settings ({B128, B128reduced, B100, B150} over 2^{25}-RS and {B128, B128reduced} over $2^{17}3^5$-RS). for various degrees of approximation Q from 3 to 70. It shows that the calculation time mostly did not depend on the lattice bases nor the sampling distribution. However, the time is increased exponentially according to the degree of approximation Q. The time was within about one minute even for the largest $Q = 70$. The state of the art method proposed by Aono and Nguyen could estimate the success probability for one natural number representation within about 2 seconds (see Table 1 in Section 5.4 of [3]). Note also that their method estimates essentially only one point in the cumulative distribution curve because it is non-parametric. Though [3] proposed some acceleration techniques such as random sampling and parallel computing, it does not seem to be available for exponentially increasing number of samples. On the other hand, our parametric method can estimate the complete form of the curve over 2^{50} natural number representations. It shows that our method is much more efficient than the state of the art method [3]. In addition, our method could give sufficiently accurate results at least within the observable range.

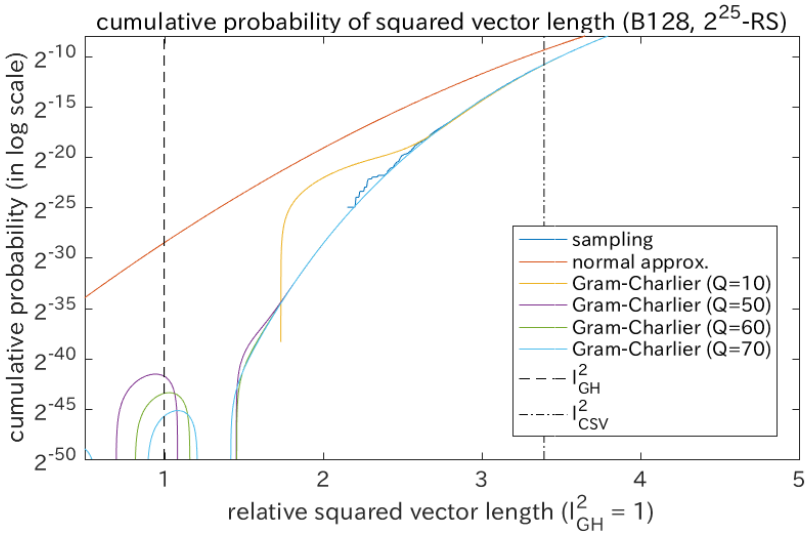
In summary, the numerical experiments verified the following points. First, if the cumulative probability is larger than 2^{-25}, the Gram-Charlier approximation $F_Q(z)$ with $Q \geq 50$ can estimate accurately the actual distribution of the squared length of generated lattice vectors for any lattice bases and any sampling distributions. Second, $F_Q(z)$ with $Q \geq 50$ converges to a curve at least when the cumulative probability is larger than 2^{-50}. Because the computation over 2^{50}-RS is expected to take hundreds of years [19], the limit 2^{-50} is sufficiently small in practice. Third, $F_Q(z)$ can be calculated efficiently within one minute even for the largest $Q = 70$.

5 Investigation of RSR and FK

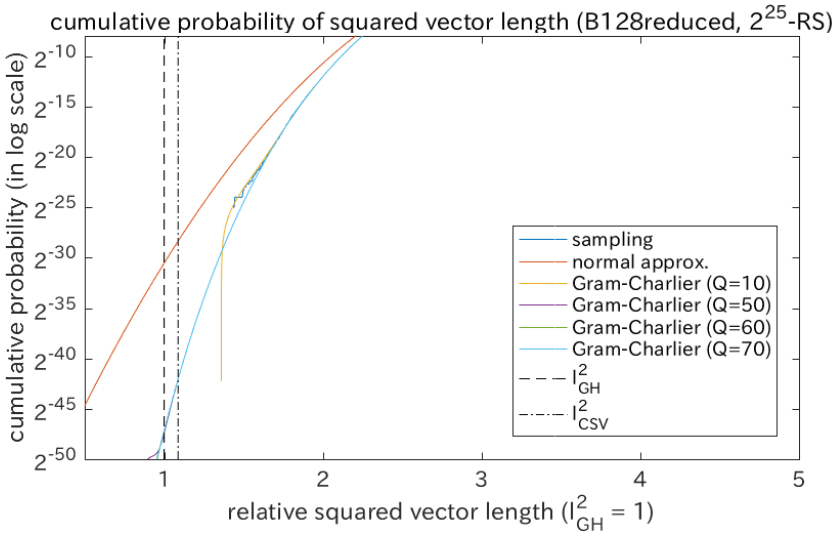
Here, we investigate RSR and its variant FK by the Gram-Charlier approximation in order to clarify why FK is superior to other methods in the SVP. Note that the following investigation is actualized only by the proposed Gram-Charlier approximation. It is quite hard for the non-parametric approach to investigate the detailed behavior of the algorithms over a huge size of samples such as 2^{50}.

5.1 Utilization of Weighted Sampling Distribution

One of the two important features of FK is that it employs the weighted RS (see Section 2.1). Here, it is verified experimentally by the Gram-Charlier approximation that the weighted RS is useful for finding very short lattice vectors.

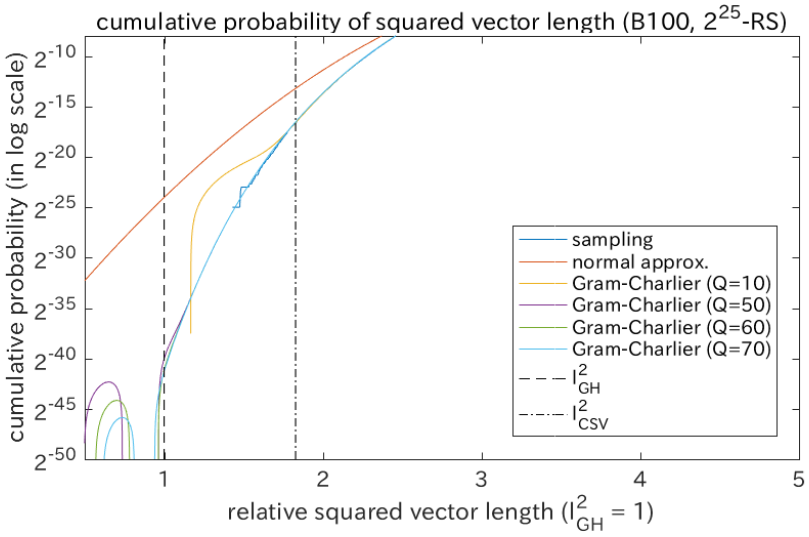


(a) B128 over 2^{25}-RS.

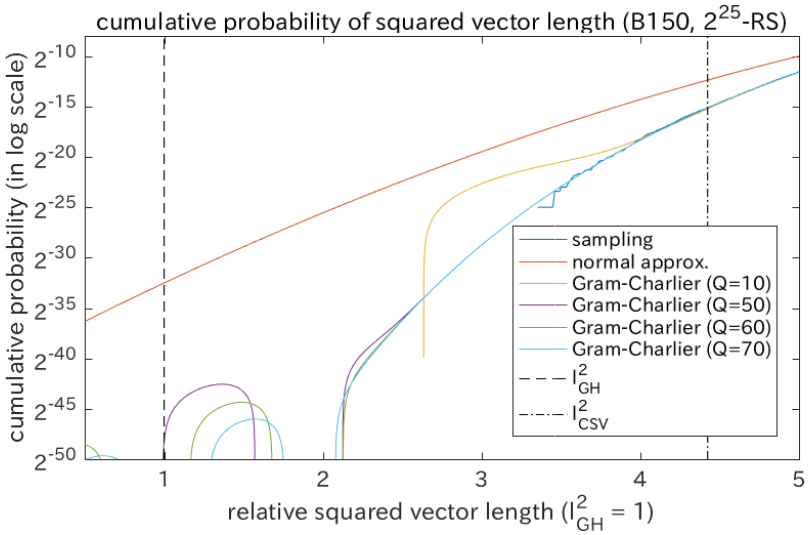


(b) B128reduced over 2^{25}-RS.

Fig. 2. Cumulative distributions of the squared length from the bases with $n = 128$ over 2^{25}-RS and the Gram-Charlier approximations: The histograms of the squared lengths of the 2^{25} generated lattice vectors are displayed by blue (slightly bumpy) curves. The curves of the normal approximation (in orange) and the Gram-Charlier approximations $F_Q(z)$ ($Q = 10$, 50, 60, and 70) are also displayed. The Gaussian heuristic ℓ^2_{GH} (normalized to 1) is displayed by the dashed vertical line. The squared length of the current shortest vector ℓ^2_{CSV} is displayed by the dot-dashed vertical line. The two bases generated from the same basis are used (B128 and B128reduced).

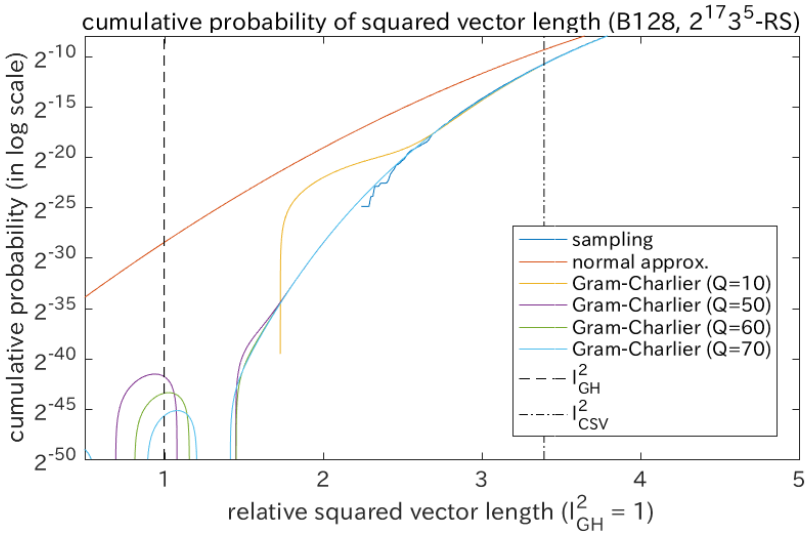


(a) B100 over 2^{25}-RS.

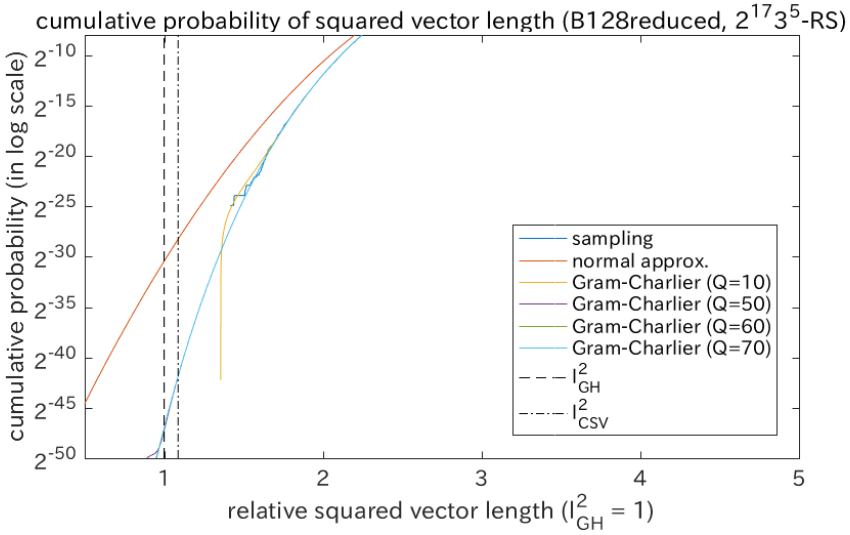


(b) B150 over 2^{25}-RS.

Fig. 3. Cumulative distributions of the squared length and the Gram-Charlier approximations over 2^{25}-RS from the bases of a small lattice (B100) and a large one (B150)



(a) B128 over $2^{17}3^5$-RS.



(b) B128reduced over $2^{17}3^5$-RS.

Fig. 4. Cumulative distributions of the squared length from the basis with $n = 128$ over $2^{17}3^5$-RS and the Gram-Charlier approximations.

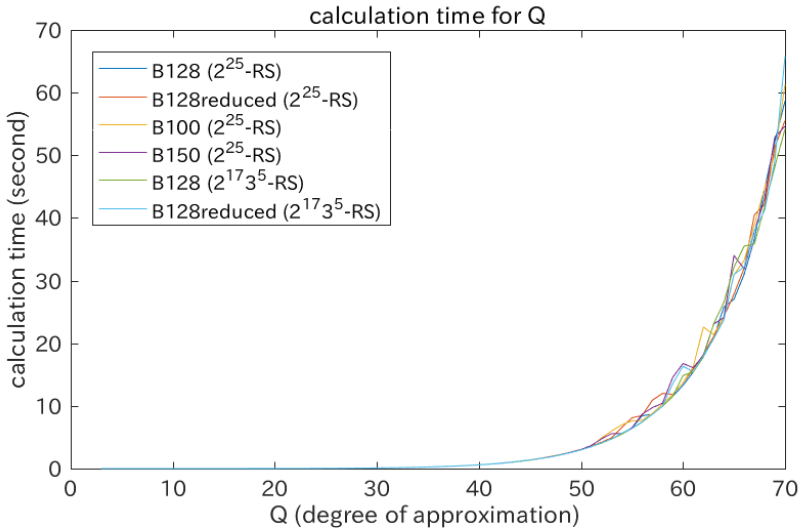


Fig. 5. Calculation time for various degrees of approximation: The actual calculation time for the Gram-Charlier approximation was measured for {B128, B128reduced, B100, B150} over 2^{25}-RS and {B128, B128reduced} over $2^{17}3^5$-RS. The degree of approximation Q was set from 3 to 70.

We used the two lattice bases B128 and B128reduced in Section 4. 2^{25}-RS and 3^{25}-RS were employed as an example of the standard RS and that of the extended RS, respectively. They consist of 2^{25} and $3^{25} \simeq 2^{40}$ possible samples. The natural number for each index occurs equally in $0, 1$ for 2^{25}-RS or $0, 1, 2$ for 3^{25}-RS. FK can give different weights to each natural number p. It is suggested in [10] that the inverse of β_{p1} of Eq. (38) gives an appropriate weight, where $\beta_{11} = \frac{1}{12}$, $\beta_{21} = \frac{7}{12}$, and $\beta_{31} = \frac{19}{12}$. Therefore, the appropriate weighted distribution is given as $(\alpha_{i1}, \alpha_{i2}) \propto \left(1, \frac{1}{7}\right)$ for $T_i = 2$ and $(\alpha_{i1}, \alpha_{i2}, \alpha_{i3}) \propto \left(1, \frac{1}{7}, \frac{1}{19}\right)$ for $T_i = 3$. Two sampling distributions were constructed by applying the appropriate weighted distribution to 2^{25}-RS and 3^{25}-RS, which are referred as to weighted-2^{25}-RS and weighted-3^{25}-RS. Each of the two sampling distributions corresponds to an example of the weighted RS.

Fig. 6 shows the Gram-Charlier approximation of the cumulative distributions of the squared length of a generated lattice vector ($F_{70}(z)$) over 2^{25}-RS and 3^{25}-RS, weighted-2^{25}-RS, and weighted-3^{25}-RS. It shows that weighted-2^{25}-RS and weighted-3^{25}-RS are clearly superior to 2^{25}-RS and 3^{25}-RS. The weighted sampling distributions are always expected to find shorter lattice vectors than the non-weighted ones when the same number of samples are given. Though weighted-3^{25}-RS seems to be slightly inferior to weighted-2^{25}-RS, there is a significant advantage of weighted-3^{25}-RS. The maximum number of samples over weighted-2^{25}-RS is 2^{25}. Therefore, the lower bound of the cumulative probability is 2^{-25} (the upper horizontal line in Fig. 6) even if additional computational

resources are available. On the other hand, the maximum number of samples over weighted-3^{25}-RS is about 2^{40}. Therefore, the cumulative probability can be reduced to the lower horizontal line if available. It is expected to generate shorter lattice vectors than weighted-2^{25}-RS.

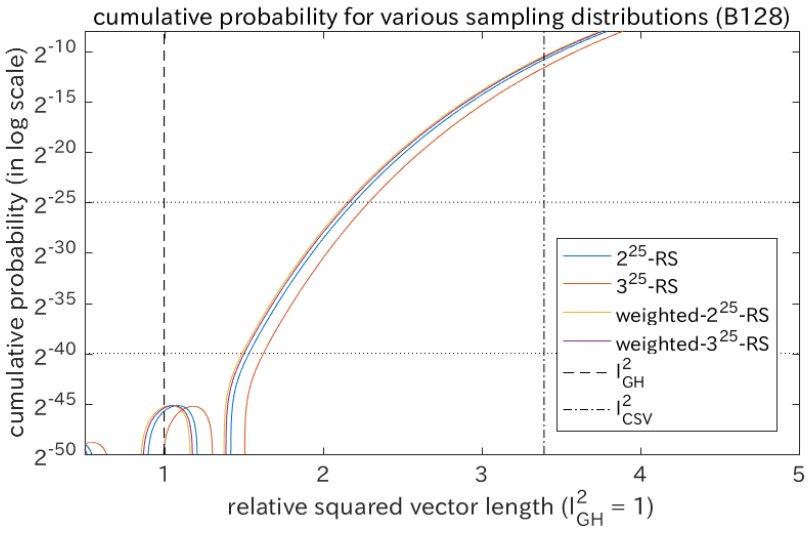
5.2 Periodical Improvement of Current Basis by Target Function

Another important feature of FK is that it periodically improves the basis so that the target function $\Phi(\boldsymbol{B})$ of Eq. (7) is reduced. Fukase and Kashiwabara [10] attempted to explain why this feature can accelerate the SVP by the normal approximation, where $\Phi(\boldsymbol{B})$ can be regarded as the mean of ℓ^2 over generated lattice vectors. They asserted that the success probability becomes higher as the mean of ℓ^2 is smaller. Though this tendency is also observed in the form of the Gram-Charlier approximation of $F_Q(z)$, it is not essential to utilize higher order cumulants. However, it cannot be explained by the normal approximation why the basis needs to be improved periodically. Although the normal approximation in Fig. 2 seems to achieve ℓ^2_{GH} by using about 2^{30} samples even for the initial basis B128, it is not true. Here, the investigation with the Gram-Charlier approximation is carried out for explaining why the periodical improvement of the basis is essential.

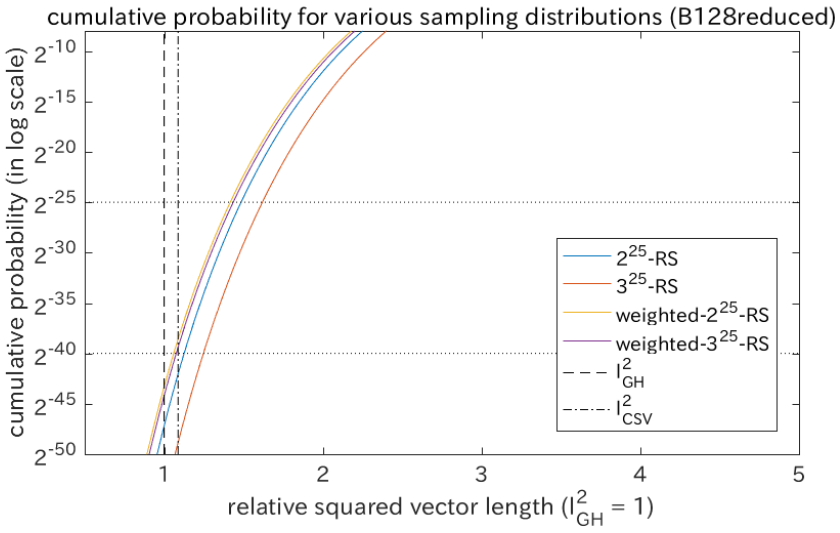
Fig. 7 shows the Gram-Charlier approximations over the 2^u-RS, which is denoted by $F_{70}(z; 2^u)$. The exponent parameter u determines the size of the search space over the random sampling distribution. u was set to 10, 15, 20, 25, 30, 35, 40, 45, and 50. B128 and B128reduced are employed as the basis. The horizontal lines correspond to the lower bounds of the cumulative probability 2^{-u}. Each black circle displays a cross-point between $F_{70}(z; 2^u)$ and the corresponding lower bound 2^{-u}. As u becomes larger, the cumulative distribution gets worse because the distribution moves to the right. On the other hand, as u becomes larger, the lower bound of the cumulative distribution becomes smaller (namely, better) because the search space becomes larger. Each cross-point represents an equilibrium point between the above two factors. We can regard the squared length at the cross-point as an estimated minimum for given u. The estimated minimum is calculated by solving the following equation with respect to z for given u:

$$F_{70}(z; 2^u) = \frac{1}{2^u}. \tag{43}$$

It is easily solved numerically because it is a continuous single-variable function. Fig. 8 shows the transitions of the estimated minimum of the squared length of generated lattice vectors for $u = 10, \ldots, 50$. It shows that the estimated minimum for B128 is saturated to about 1.5 even if 2^u is intractably large. In other words, unless the basis is improved periodically, the shortest length of generated lattice vectors cannot be reduced below a certain value even if the computational resources are provided sufficiently. It verifies that it is crucial for solving the SVP to improve the lattice basis periodically. It seems strange that the estimated minimum length for $u = 50$ is slightly larger than that for $u = 45$. In the deterministic search, the accurate minimum never increases as u is larger.



(a) B128.



(b) B128reduced.

Fig. 6. Cumulative distributions of the squared length from B128 and B128reduced over the weighted RS: The Gram-Charlier approximations $F_{70}(z)$ are displayed over the four sampling distributions (2^{25}-RS, 3^{25}-RS, weighted-2^{25}-RS, and weighted-3^{25}-RS). The two horizontal lines are also displayed, which correspond to 2^{-25} and $2^{-40} \simeq 3^{-25}$. See text for the details.

Though it may be because the estimated minimum is based on the probabilistic search which allows duplicate samples, it is beyond the scope of this paper to clarify this phenomenon.

6 Discussions

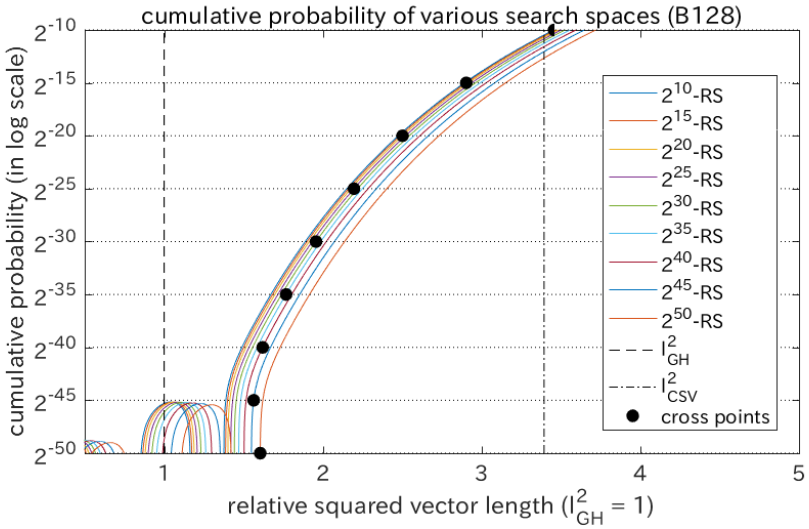
6.1 Validity of Randomness Assumption

The accuracy of our method heavily relies on the randomness assumption. Reversely, the converged Gram-Charlier approximation with sufficiently large Q is rigorously accurate if the randomness assumption holds. Though it is hard currently to verify the validity of the randomness assumption, the assumptions on the Gaussian heuristic [12] suggest that its validity depends on the relative squared vector length $\ell^2/\ell^2_{\mathrm{GH}}$. If the relative length is 1 (namely, ℓ^2 is equal to ℓ_{GH}), the randomness assumption does not hold because the distribution includes only one sample. On the other hand, if the relative length is sufficiently large, the randomness assumption seems to be valid because the number of possible lattice vectors within the corresponding volume is expected to be high. Letting $\ell^2/\ell^2_{\mathrm{GH}}$ be $1+\epsilon$, the number is approximately estimated as $(1+\epsilon)^{\frac{n}{2}}$. For example, it is $1.1^{64} \simeq 445$ for $n = 128$ and $\epsilon = 0.1$. The estimated number seems to be sufficiently high to approximate a continuous distribution even for such a small $\epsilon = 0.1$.

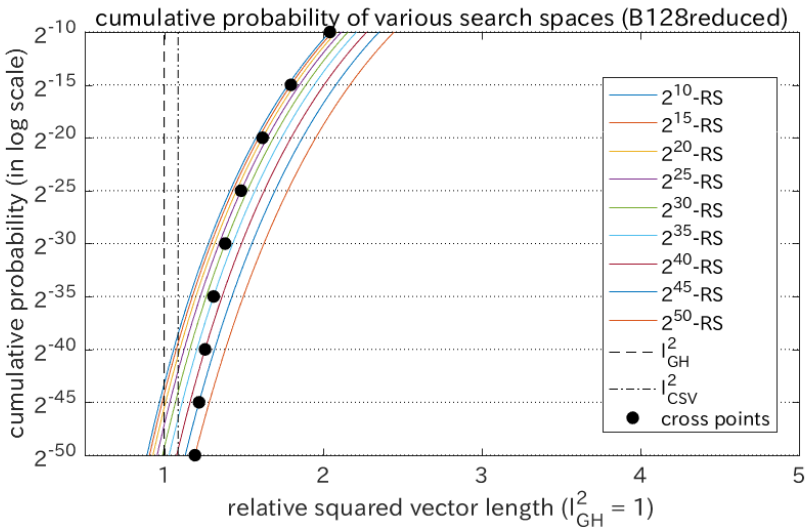
6.2 Dependence among Indices of Natural Number Representation

One of the disadvantages of our proposed model is that it cannot allow the dependence among the indices of the natural number representation (see Section 2.1). In addition, our model seems to be based on an inefficient non-deterministic search. On the other hand, almost all of the state of the art algorithms employ a deterministic search enumerating a set of candidate natural number representations, where the dependence among the indices is often allowed. Here, we show that our model can give a probabilistic approximation of such a deterministic search and the optimal $\boldsymbol{\alpha}$ can be calculated by minimizing the Kullback-Leibler divergence. In other words, by regarding any set of candidate natural number representations in any algorithm as a probability distribution, our model can approximate this distribution as accurately as possible. Though there are various metrics such as the Euclidean distance, we employed the Kullback-Leibler divergence here. It is because its optimum can be derived easily and its mathematical properties have been extensively investigated [5].

When only a single natural number representation $\boldsymbol{d} = (d_i)$ is given, we can estimate the Gram-Charlier approximation of the conditional probability $P_Q(z|\boldsymbol{d})$ by letting α_{ip} be 1 only if $p = d_i$ (otherwise $\alpha_{ip} = 0$). Let Ω be the set of candidate (namely, enumerated) natural number representations in a deterministic search. This deterministic search is equivalent to the uniformly distributed search over Ω if the number of samplings is sufficiently large. Therefore,

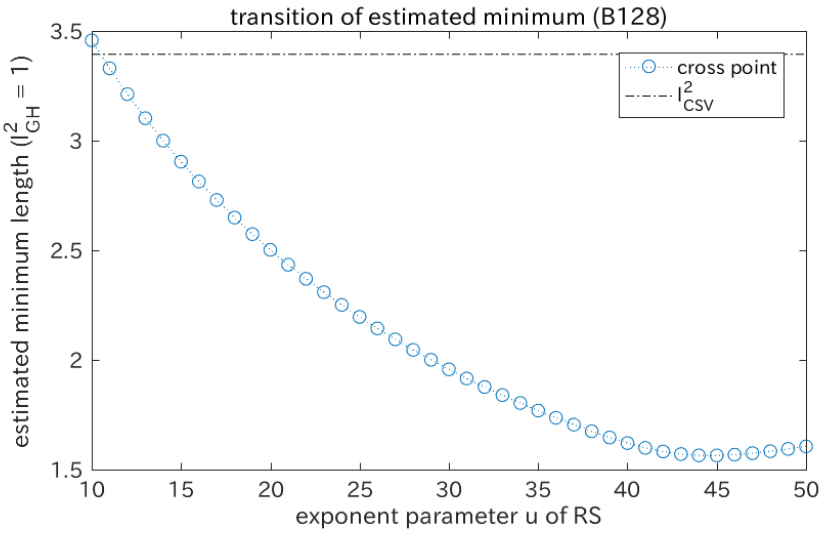


(a) B128.

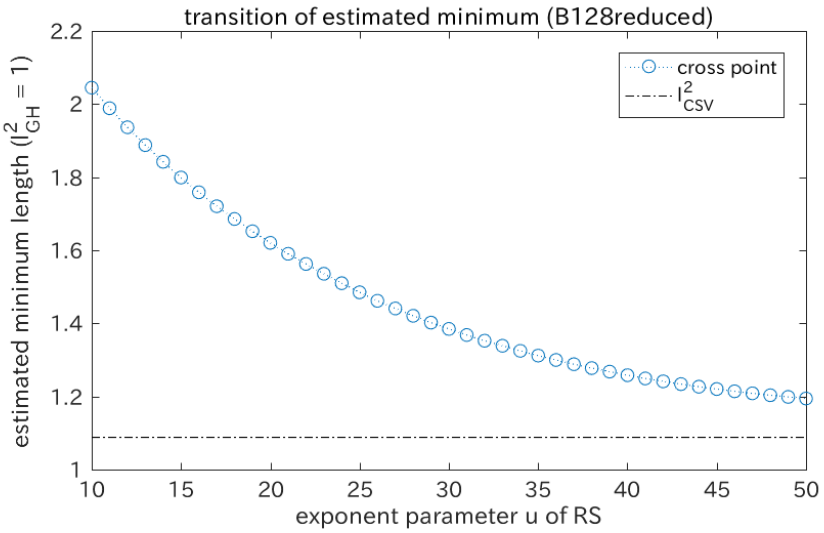


(b) B128reduced.

Fig. 7. Cumulative distributions of the squared length from B128 and B128reduced over the various search spaces 2^u of RS: The Gram-Charlier approximations $F_{70}\left(z; 2^u\right)$ are displayed over the various RS distributions with different search spaces ($u = 10, 15, 20, 25, 30, 35, 40, 45, 50$). The horizontal lines are also displayed by 2^{-u}. Each black circle displays the cross-point of the Gram-Charlier approximation and the horizontal line which corresponds to the search space. See text for the details.



(a) B128.



(b) B128reduced.

Fig. 8. Estimated minimum of the squared length from B128 and B128reduced for the exponent parameter u of RS: The parameter u of 2^u-RS determines the size of the search space. Each estimated minimum corresponds to a cross-point in Fig. 7. The squared length of the current shortest vector ℓ^2_{CSV} is displayed by the dot-dashed horizontal line.

$P_Q(z)$ is given as

$$P_Q(z) = \sum_{\boldsymbol{d}} P_Q(z|\boldsymbol{d}) P(\boldsymbol{d}) = \frac{\sum_{\boldsymbol{d}\in\Omega} P_Q(z|\boldsymbol{d})}{|\Omega|}, \tag{44}$$

where $|\Omega|$ is the cardinality (the number of members) of Ω. $P(\boldsymbol{d})$ is the probability that $\boldsymbol{d}$ occurs in the search, which is defined as

$$P(\boldsymbol{d}) = \begin{cases} \frac{1}{|\Omega|} & \boldsymbol{d} \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

We can estimate $P_Q(z)$ over any Ω by these equations. However, the calculation is time-consuming because it requires the estimation of $P_Q(z|\boldsymbol{d})$ for every $\boldsymbol{d} \in \Omega$. Our proposed model is essentially equivalent to employing the following approximation:

$$P(\boldsymbol{d}) \simeq \prod_i P(d_i) = \prod_i \alpha_{id_i}. \tag{46}$$

$P_Q(z)$ is efficiently calculated by our model, because the summation over Ω can be divided into the estimation of an independent probability distribution for each index i (see Section 3). Now, the Kullback-Leibler divergence between $P(\boldsymbol{d})$ and $\prod_i \alpha_{id_i}$ is given as

$$\begin{aligned} &D_{\mathrm{KL}}\left(P(\boldsymbol{d}) \mid \prod_i \alpha_{id_i}\right) = \sum_{\boldsymbol{d}} P(\boldsymbol{d}) \log\left(\frac{P(\boldsymbol{d})}{\prod_i \alpha_{id_i}}\right) \\ &= \sum_{\boldsymbol{d}\in\Omega} \frac{1}{|\Omega|} \log\left(\frac{1}{|\Omega| \prod_i \alpha_{id_i}}\right) = -\log(|\Omega|) - \frac{\sum_{\boldsymbol{d}\in\Omega} \sum_i \log(\alpha_{id_i})}{|\Omega|}. \end{aligned} \tag{47}$$

This divergence is equal to 0 only if the two distributions are completely the same, and otherwise always positive. Therefore, $\prod_i \alpha_{id_i}$ can approximate $P(\boldsymbol{d})$ as accurately as possible by minimizing D_{KL}. Then, the optimum of $\boldsymbol{\alpha}$ (denoted $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_{ip})$) is given as

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \sum_{\boldsymbol{d}\in\Omega} \sum_i \log(\alpha_{id_i}) \text{ subject to } \sum_p \alpha_{ip} = 1. \tag{48}$$

By the method of Lagrange multipliers, $\hat{\alpha}_{ip}$ is given as

$$\hat{\alpha}_{ip} = \frac{\sum_{\boldsymbol{d}\in\Omega} \delta_{pd_i}}{|\Omega|}, \tag{49}$$

where δ_{pd_i} is the Kronecker delta. In other words, $\hat{\alpha}_{ip}$ is proportional to the number of occurrences of p in the index i of Ω. It is easily calculated for any Ω.

Reversely, there is also a promising approach from our proposed model to a deterministic search framework. For a given natural number representation $\boldsymbol{d} = (d_i)$, its occurrence probability is given as $\prod_i \alpha_{id_i}$. Then, $\prod_i \alpha_{id_i}$ can be

regarded as a score measuring the "goodness" of $\boldsymbol{d}$. Using the logarithm, the score $\varphi(\boldsymbol{d})$ is given as

$$\varphi(\boldsymbol{d}) = \sum_i \log \alpha_{id_i}. \tag{50}$$

φ is the sum of weights, each of which depends on only the index i and d_i. If the natural number representations with high scores can be collected, we can construct deterministically a set Ω including only the "better" candidates. This principle is similar to the enumeration of the lattice vectors with the smaller lengths in the previous lattice basis reduction algorithms. Therefore, our proposed model is promising for improving the previous algorithms.

6.3 Convergence Property of Gram-Charlier Approximation

Here, we investigate experimentally the convergence property of the Gram-Charlier approximation of the cumulative distribution function with the degree of approximation. One of the important factors is the convergence of c_r (the coefficients of the Hermite polynomials $H_r(x)$). Another important factor is the convergence of the value of $c_r H_{r-1}(x)$. The bound of $|H_r(x)|$ is given as

$$|H_r(x)| < 1.09\pi^{\frac{1}{4}} e^{\frac{x^2}{4}} \sqrt{r!}, \tag{51}$$

which holds for any r and x (see 22.14.17 and 22.5.19 in [1].) Therefore, $c_r H_{r-1}(x)$ converges if $\sqrt{(r-1)!}c_r$ converges. Fig. 9 shows the transitions of $|c_r|$ and $|\sqrt{(r-1)!}c_r|$ for $3 \leq r \leq 70$. Note that c_r is determined only by the basis and the sampling distribution. We used {B128, B128reduced, B100, B150} over 2^{25}-RS and {B128, B128reduced} over $2^{17}3^5$-RS. Fig. 9 experimentally verified that the coefficients and the bounds converges to 0 roughly exponentially. It is interesting that the transitions are similar irrespective of the basis and the sampling distribution. It suggests that some theoretical approximations may be available.

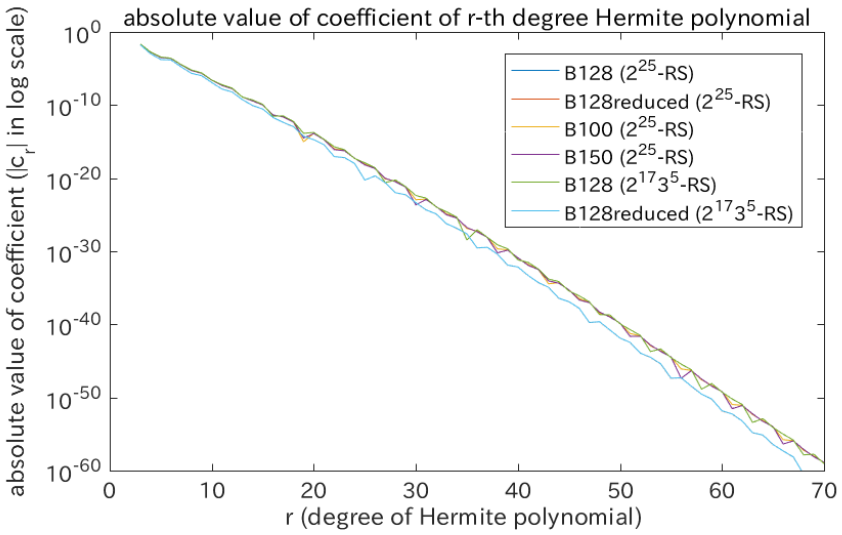
6.4 Accelerated Algorithm for Estimating Coefficients

The standard formulation of the coefficients of the Gram-Charlier approximation is given in Section 2.3. Here, another formulation is utilized, which can accelerate Algorithm 2. The new formulation is based on the orthogonality property of the Hermite polynomials, which is given as

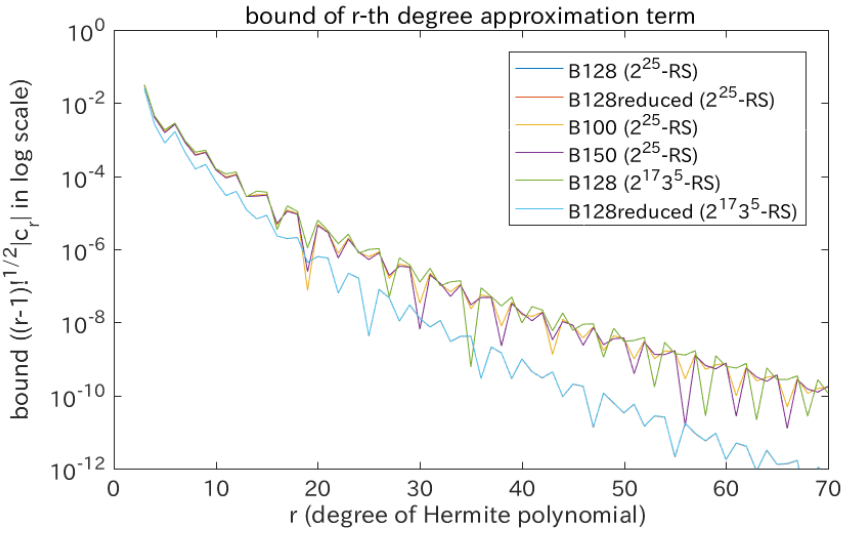
$$\int_{-\infty}^{\infty} H_p(x) H_q(x) e^{-\frac{x^2}{2}} \mathrm{d}x = \sqrt{2\pi} p! \delta_{pq}. \tag{52}$$

Using this property, the expectation of $H_r(x)$ $(r \geq 3)$ over a normalized probability distribution function $P(x)$ is given as

$$\int_{-\infty}^{\infty} H_r(x) P(x) \mathrm{d}x = c_r r!, \tag{53}$$



(a) Coefficients.



(b) Bounds.

Fig. 9. Transitions of the coefficients and the bounds of the Hermite polynomials: $|c_r|$ and $\sqrt{(r-1)!}|c_r|$ $(3 \leq r \leq 70)$ are calculated for {B128, B128reduced, B100, B150} over 2^{25}-RS and {B128, B128reduced} over $2^{17}3^5$-RS.

where c_r is the r-th coefficient of the Gram-Charlier approximation. Now, the explicit expression of $H_r(x)$ is given as

$$H_r(x) = r! \sum_{p=0}^{\lfloor \frac{r}{2} \rfloor} \frac{(-1)^p x^{(r-2p)}}{2^p p! (r-2p)!}. \tag{54}$$

Then, c_r is given as

$$c_r = \sum_{p=0}^{\lfloor \frac{r}{2} \rfloor} \frac{(-1)^p \mu_{r-2p}(x)}{2^p p! (r-2p)!}, \tag{55}$$

where $\mu_{r-2p}(x)$ is the $(r-2p)$-th order moment of a random variable x over $P(x)$. In the similar way as in Eq. (39), $\mu_r(x)$ can be calculated recursively from the normalized cumulants λ_r by

$$\mu_r = \lambda_r + \sum_{m=1}^{r-1} \binom{r-1}{m-1} \lambda_m \mu_{r-m}. \tag{56}$$

Consequently, the new algorithm calculating the coefficients is given as follows.

Algorithm 3 The accelerated Gram-Charlier approximation of the cumulative distribution of the squared lengths of generated lattice vectors.

Require: $\boldsymbol{\alpha}$, $\boldsymbol{B}^*$, and Q.
Calculate the moments $\mu_r\left(\zeta_i^2\right)$ by $\boldsymbol{\alpha}$ ($r \leq Q$, the same hereinafter).
Calculate the cumulants $\kappa_r\left(\zeta_i^2\right)$ by $\mu_r\left(\zeta_i^2\right)$.
Calculate the cumulants $\kappa_r\left(\ell^2\right)$ and the normalized ones $\lambda_r\left(\ell^2\right)$ by $\kappa_r\left(\zeta_i^2\right)$ and $\boldsymbol{B}^*$.
Calculate the moments $\mu_r\left(\bar{\ell}^2\right)$ by $\lambda_r\left(\ell^2\right)$.
Calculate the coefficients c_r by $\mu_r\left(\bar{\ell}^2\right)$.
return the approximate cumulative distribution $F_Q(z)$ (and the probability distribution $P_Q(z)$ if necessary).

Here, $\bar{\ell}^2$ denotes the normalized squared length. This algorithm avoids the combinatorial problem in the direct estimation of the coefficients from the cumulants by utilizing an additional step calculating the moments from the cumulants. Because the bottleneck process is the recursive interconversion between the moments and the cumulants, the complexity of this algorithm is $O\left(n^2\right) + O(nQT) + O\left(nQ^2\right)$, where the term of $O\left(Q^{\frac{Q}{3}}\right)$ is removed. Therefore, Algorithm 3 is much more rapid than Algorithm 2 if Q is large. However, there is one practical problem in Algorithm 3. Because the recursive calculation accumulates small rounding errors repeatedly, the estimation error of a high degree coefficient is not negligible. We are now constructing an implementation with keeping a high accuracy.

7 Conclusion

In this paper, we proposed a new method for estimating the success probability of finding very short lattice vectors in the lattice basis reduction algorithm. The proposed method is based on a parametric approach using the Gram-Charlier approximation and gives a closed-form expression with a few parameters. It could estimate the actual distribution quite accurately and quite efficiently. The investigations with the proposed method discovered some important properties of RSR and FK. The most significant advantage of the proposed method is that it can estimate the success probability quite efficiently with keeping the accuracy. The investigation in Section 5 is intractable for the non-parametric approach because it needs to manage 2^{50}-RS. On the other hand, the Gram-Charlier approximation estimated all the distributions within at most one minute. The experimental results showed that the calculation time and the accuracy of our proposed method do not depend on the lattice basis and the sampling distribution. In other words, they verified that our method is useful for a high-dimensional lattice over a large number of samplings.

An unsolved problem of the proposed method is that we have not clarified yet the theoretical relationship between the degree of approximation Q and the convergence of the estimation. We are planning to investigate the relationship furthermore in order to set Q to an appropriate value. In addition, we are planning to investigate furthermore the validity of the randomness assumption by the Gram-Charlier approximation. Another problem is that the Gram-Charlier approximation was used only for verifying the superiority of the previous algorithms in this paper. It is promising to use this method for finding the optimal settings and parameters adaptively and for solving the SVP much more efficiently. For example, we are planning to estimate the optimal weighted sampling distribution according to a given basis and to set the optimal size of the search space by u. Moreover, we are planning to investigate the state of the art algorithms (for example, [3], [19], and [8]) by the approximation method in Section 6.2.

References

1. Abramowitz, M., Stegun, I.A.: Handbook of mathematical functions with formulas, graphs, and mathematical table. Dover New York (1965)
2. Ajtai, M., Kumar, R., Sivakumar, D.: A sieve algorithm for the shortest lattice vector problem. In: Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing. pp. 601–610. STOC '01, ACM, New York, NY, USA (2001), http://doi.acm.org/10.1145/380752.380857
3. Aono, Y., Nguyen, P.Q.: Random sampling revisited: Lattice enumeration with discrete pruning. In: EUROCRYPT (2). pp. 65–102. Springer (2017)
4. Buchmann, J., Ludwig, C.: Practical lattice basis sampling reduction. In: Hess, F., Pauli, S., Pohst, M. (eds.) Algorithmic Number Theory: 7th International Symposium, ANTS-VII, Berlin, Germany, July 23-28, 2006. Proceedings. pp. 222–237. Springer Berlin Heidelberg, Berlin, Heidelberg (2006), https://doi.org/10.1007/11792086_17

5. Burnham, K.P., Anderson, D.R.: Model selection and multimodel inference: A practical-theoretic approach. Springer-Verlag, Berlin; New York, 2 edn. (2002)
6. Cramér, H.: On some classes of series used in mathematical statistics. In: Proceedings of the Sixth Scandinavian Congress of Mathematicians. pp. 399–425 (1925)
7. Cramér, H.: Mathematical Methods of Statistics. Princeton mathematical series, Princeton University Press (1946)
8. Ducas, L.: Shortest vector from lattice sieving: A few dimensions for free. In: Nielsen, J.B., Rijmen, V. (eds.) Advances in Cryptology – EUROCRYPT 2018. pp. 125–145. Springer International Publishing, Cham (2018)
9. Fincke, U., Pohst, M.: Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. Mathematics of computation 44(170), 463–471 (1985)
10. Fukase, M., Kashiwabara, K.: An accelerated algorithm for solving SVP based on statistical analysis. JIP 23, 67–80 (2015)
11. Gama, N., Nguyen, P.Q., Regev, O.: Lattice enumeration using extreme pruning. In: Gilbert, H. (ed.) Advances in Cryptology – EUROCRYPT 2010: 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, French Riviera, May 30 – June 3, 2010. Proceedings. pp. 257–278. Springer Berlin Heidelberg, Berlin, Heidelberg (2010), https://doi.org/10.1007/978-3-642-13190-5_13
12. Hoffstein, J., Pipher, J., Silverman, J.: An Introduction to Mathematical Cryptography. Springer Publishing Company, Incorporated, 2 edn. (2014)
13. Kannan, R.: Improved algorithms for integer programming and related lattice problems. In: Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing. pp. 193–206. STOC '83, ACM, New York, NY, USA (1983), http://doi.acm.org/10.1145/800061.808749
14. Lenstra, A., Lenstra, H., Lovász, L.: Factoring polynomials with rational coefficients. Mathematische Annalen 261(4), 515–534 (12 1982)
15. Ludwig, C.: Practical lattice basis sampling reduction. Ph.D. thesis, Technische Universität Darmstadt (2006)
16. Schneider, M., Gama, N.: SVP challenge, available at https://www.latticechallenge.org/svp-challenge/
17. Schnorr, C.P., Euchner, M.: Lattice basis reduction: Improved practical algorithms and solving subset sum problems. Math. Program. 66(2), 181–199 (Sep 1994), http://dx.doi.org/10.1007/BF01581144
18. Schnorr, C.: Lattice reduction by random sampling and birthday methods. In: Alt, H., Habib, M. (eds.) STACS 2003, 20th Annual Symposium on Theoretical Aspects of Computer Science, Berlin, Germany, February 27 - March 1, 2003, Proceedings. Lecture Notes in Computer Science, vol. 2607, pp. 145–156. Springer (2003)
19. Teruya, T., Kashiwabara, K., Hanaoka, G.: Fast lattice basis reduction suitable for massive parallelization and its application to the shortest vector problem. In: Public-Key Cryptography - PKC 2018 - 21st IACR International Conference on Practice and Theory of Public-Key Cryptography, Rio de Janeiro, Brazil, March 25-29, 2018, Proceedings, Part I. pp. 437–460 (2018)
20. The FPLLL development team: fplll, a lattice reduction library (2016), available at https://github.com/fplll/fplll
21. Wallace, D.L.: Asymptotic approximations to distributions. Ann. Math. Statist. 29(3), 635–654 (09 1958), http://dx.doi.org/10.1214/aoms/1177706528