Lower Bounds on Lattice Enumeration with Extreme Pruning

Yoshinori Aono[1], Phong Q. Nguyen[2,3], Takenobu Seito[4], and Junji Shikata[5]

[1] National Institute of Information and Communications Technology, Japan
[2] Inria Paris, France
[3] CNRS, JFLI, University of Tokyo, Japan
[4] Bank of Japan ⋆, Japan
[5] Yokohama National University, Japan

Abstract. At Eurocrypt ’10, Gama, Nguyen and Regev introduced lattice enumeration with extreme pruning: this algorithm is implemented in state-of-the-art lattice reduction software and used in challenge records. They showed that extreme pruning provided an exponential speed-up over full enumeration. However, no limit on its efficiency was known, which was problematic for long-term security estimates of lattice-based cryptosystems. We prove the first lower bounds on lattice enumeration with extreme pruning: if the success probability is lower bounded, we can lower bound the global running time taken by extreme pruning. Our results are based on geometric properties of cylinder intersections and some form of isoperimetry. We discuss their impact on lattice security estimates.

1 Introduction

Among all the candidates submitted in 2017 to the NIST standardization of post-quantum cryptography, the majority are based on hard lattice problems, such as LWE and NTRU problems. Unfortunately, security estimates for lattice problems are known to be difficult: many different assessments exist in the research literature, which is reflected in the wide range of security estimates in NIST submissions (see [2]), depending on the model used. One reason is that the performance of lattice algorithms depends on many parameters: we do not know how to select these parameters optimally, and we do not know how far from optimal are current parameter selections. The most sensitive issue is the evaluation of the cost of a subroutine to find shortest or nearly shortest lattice vectors in certain dimensions (typically the blocksize of blockwise reduction algorithms). In state-of-the-art lattice reduction software [11,7,9], this subroutine is implemented by lattice enumeration with extreme pruning, introduced at Eurocrypt ’10 by Gama, Nguyen and Regev [16] as a generalization of pruning methods introduced by Schnorr *et al* [34,35] in the 90s. Yet, most lattice-based

⋆ The views expressed in this paper are those of authors and do not necessarily reflect the official views of the Bank of Japan.

NIST submissions chose their parameters based on the assumption that sieving [1,28,22,20,8] (rather than enumeration) is the most efficient algorithm for this subroutine. This choice goes back to the analysis of NewHope [3, Sect. 6], which states that sieving is more efficient than enumeration in dimension ≥ 250 for both classical and quantum computers, based on a lower bound on the cost of sieving (ignoring subexponential terms) and an upper bound on the cost of of enumeration (either [11, Table 4] or [10, Table 5.2]). In dimensions around $140 - 150$, this upper bound is very close to actual running times for solving the largest record SVP challenges [32], which does not leave much margin for future progress; and for dimensions ≥ 250, a numerical extrapolation has been used, which is also debatable.

It would be more consistent to compare the sieving lower bound by a lower bound on lattice enumeration with extreme pruning. Unfortunately, no such lower bound is known: the performances of extreme pruning strongly depends on the choice of bounding function, and it is unknown how good can be such a function. There is only a partial lower bound on the cost of extreme pruning in [11], assuming that the choice of step bounding function analyzed in [16] is optimal. And this partial lower bound is much lower than the upper bound given in [11,10].

Our results. We study the limitations of lattice enumeration with extreme pruning. We prove the first lower bound on the cost of extreme pruning, given a lower bound on the global success probability. This is done by studying the case of a single enumeration with cylinder pruning, and generalizing it to the extreme pruning case of multiple enumerations, possibly infinitely many. Our results are based on geometric properties of cylinder intersections and a probabilistic form of isoperimetry: usually, isoperimetry refers to a geometric inequality involving the surface area of a set and its volume.

Our lower bounds are easy to compute and appear to be reasonably tight in practice, at least in the single enumeration case: we introduce a cross-entropy-based method which experimentally finds upper bounds somewhat close to our lower bounds.

Impact. By combining our lower bounds with models of strongly-reduced lattice bases introduced in [26,11,7] and quantum speed-ups for enumeration [6], we obtain more sound comparisons with sieving: see Fig. 1 for an overview. It suggests that enumeration is faster than sieving up to higher dimensions than previously considered by lattice-based submissions to NIST post-quantum cryptography standardization: the cost lower bound used by many NIST submissions is not as conservative as previously believed, especially in the quantum setting. Concretely, in the quantum setting, the lower bounds of enumeration and sieving cross in dimensions roughly 300-400 in the HKZ-basis model or beyond 500 in the Rankin-basis model, depending on how many enumerations are allowed. We note that in high dimension, our lower bound for enumeration with 10^{10} HKZ bases is somewhat close to the numerical extrapolation of [17, (2)], called Core-Enum+$O(1)$ in [2].

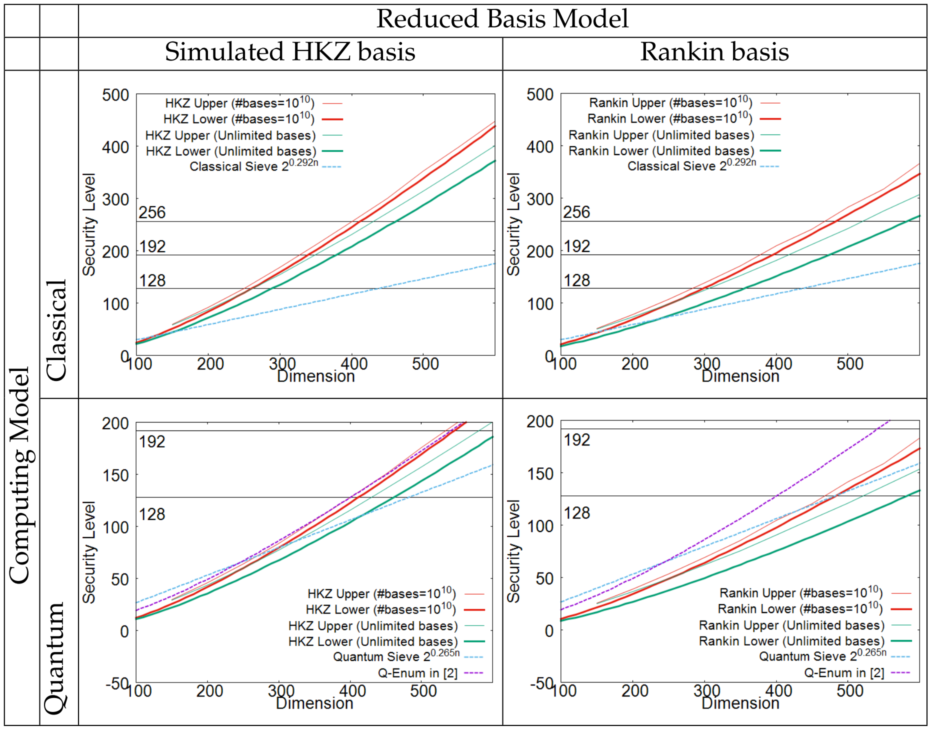

Fig. 1. Upper/lower bounds on the classical/quantum cost of enumeration with cylinder pruning, using strongly-reduced basis models. See Sect. 5 for the exact meaning of these curves: the lower bounds correspond to (16) and (17) and the upper bounds are found by the algorithm in Sect. 4. For comparison, we also displayed several curves from [2]: $2^{0.292n}$ and $2^{0.265n}$ as the simplified classical/quantum complexity of sieve algorithms, and the numerical extrapolation of enumeration cost of [17, (2)].

Technical overview. Enumeration is the simplest algorithm to solve hard lattice problems: it outputs $L \cap B$, given a lattice L and an n-dimensional ball $B \subseteq \mathbb{R}^n$. It dates back to the early 1980s [29,18,15] but has been significantly improved in practice in the past twenty years, thanks to pruning methods introduced by Schnorr *et al.* [34,35,33], and later revisited and generalized as respectively cylinder pruning by Gama, Nguyen and Regev [16] and discrete pruning by Aono and Nguyen [5]: pruning methods offer a trade-off by enumerating over a special subset $S \subseteq B$, at the expense of missing solutions. Gama *et al.* [16] introduced the idea of extreme pruning where one repeats pruned enumeration many times over different sets S: this can be globally faster than full enumeration, even if a single enumeration has a negligible probability of returning solutions. In the case of cylinder pruning, [16] showed that the speed-up can be asymptotically exponential for simple choices of the pruning subset S.

Cylinder pruning uses the intersection S of n cylinders defined by a lattice basis and a bounding function f: by using different lattice bases B, one obtains different sets S. The running time and the success probability of cylinder pruning depend on the quality of the basis, and the bounding function f. But when one uses different bases, these bases typically have approximately the same quality, which allows to focus on f, which determines the radii of S.

The probability of success of cylinder pruning is related to the volume of S, whereas its cost is related to the volumes of the 'canonical' projections of S. We show that if the success probability is lower bounded, that is, if S is sufficiently big (with respect to its volume, or its Gaussian measure for the case of solving LWE), then the function f defining S can be lower bounded: as a special case, if S occupies a large fraction of the ball, f is lower bounded by essentially the linear pruning function of [16]. This immediately gives lower bounds on the volumes of the projections of S, but we significantly improve these direct lower bounds using the following basic form of isoperimetry: for certain distributions such as the Gaussian distribution, among all Borel sets of a given volume, the ball centered at the origin has the largest probability. The extreme pruning case is obtained by a refinement of isoperimetry over finitely many sets: it is somewhat surprising that we obtain a lower bound even in the extreme case where we allow infinitely many sets S.

All our lower bounds are easy to compute. To evaluate their tightness, we introduce a method based on cross-entropy to compute good upper bounds in practice, i.e., good choices of f. This is based on earlier work by Chen [10].

Open problem. Our lower bounds are specific to cylinder pruning [16]. It would be interesting to obtain tight lower bounds for discrete pruning [5].

Roadmap. In Section 2, we introduce background and notation on lattices, enumeration and its cost estimations. Section 3 presents our lower bounds as geometric properties of cylinder intersections. Section 4 shows how to obtain good upper bounds in practice, by finding nice cylinder intersections using cross-entropy. Finally, in Section 5, we evaluate the tightness of our lower bounds and discuss security estimates for the hardness of finding nearly shortest lattice vectors. The appendix includes proofs of technical results. The full version of this paper on eprint also includes sage scripts to compute our lower bounds.

2 Background

2.1 Notation

Throughout the paper, we use row representations of matrices. The Euclidean norm of a vector $\mathbf{v} \in \mathbb{R}^n$ is denoted $\|\mathbf{v}\|$. The 'canonical' projection of $\mathbf{u} \in \mathbb{R}^n$ onto $\mathbb{R}^k$ for $1 \leq k \leq n$ is the truncation $\tau_k(\mathbf{u}) = (u_1, \ldots, u_k)$.

Measures. We denote by vol the standard Lebesgue measure over $\mathbb{R}^n$. We denote by $\rho_{n,\sigma}$ the centered Gaussian measure of variance σ^2, whose pdf over $\mathbb{R}^n$ is

$$(2\pi\sigma^2)^{-n/2}e^{-\|\mathbf{x}\|^2/(2\sigma^2)}.$$

The standard Gaussian measure is $\rho_n = \rho_{n,1}$.

Balls. We denote by $\mathrm{Ball}_n(R)$ the n-dimensional zero-centered ball of radius R. Let $V_n(R) = \mathrm{vol}(\mathrm{Ball}_n(R))$. Let $\mathbf{u} = (u_1, \ldots, u_n)$ be a point chosen uniformly at random from the unit sphere S^{n-1}, *e.g.* $u_i = x_i/\sqrt{\sum_{j=1}^n x_j^2}$, where $x_1, \ldots, x_n$ are independent, normally distributed random variables with mean 0 and variance 1. Then $\|\tau_k(\mathbf{u})\|^2 = \frac{\sum_{i=1}^k x_i^2}{\sum_{i=1}^k x_i^2 + \sum_{i=k+1}^n x_i^2} = \frac{X}{X+Y}$, where X and Y have distributions $\mathrm{Gamma}(k/2, \theta = 2)$ and $\mathrm{Gamma}((n-k)/2, \theta = 2)$ respectively. Here, we use the scale parametrization to represent Gamma distributions. Hence, $\|\tau_k(\mathbf{u})\|^2$ has distribution $\mathrm{Beta}(k/2, (n-k)/2)$. In particular, $\|\tau_{n-2}(\mathbf{u})\|^2$ has distribution $\mathrm{Beta}(n/2-1, 1)$, whose pdf is $x^{(n/2)-2}/B(n/2-1,1) = (n/2-1)x^{(n/2)-2}$. It follows that the truncation $\tau_{n-2}(\mathbf{u})$ is uniformly distributed over $\mathrm{Ball}_{n-2}(1)$, which allows to transfer our results to random points in balls.

Recall that the cumulative distribution function of the $\mathrm{Beta}(a, b)$ distribution is the regularized incomplete beta function $I_x(a, b)$ defined as:

$$I_x(a,b) = \frac{1}{B(a,b)} \int_0^x u^{a-1}(1-u)^{b-1} du, \tag{1}$$

where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ denotes the beta function. We have the following elementary bounds (by integrating by parts):

$$\frac{x^a(1-x)^{b-1}}{aB(a,b)} \le I_x(a,b) \qquad \forall a > 0, b \ge 1, 0 \le x \le 1 \tag{2}$$

$$I_x(a,b) \le \frac{x^a}{a \cdot B(a,b)} \qquad \forall a > 0, b \ge 1, 0 \le x \le 1 \tag{3}$$

For $z \in [0,1]$ and $a, b > 0$, $I_z^{-1}(a,b) + I_{1-z}^{-1}(b,a) = 1$ which is immediate from the relation $I_x(a,b) + I_{1-x}(b,a) = 1$.

Finally, $P(s,x) = \int_0^x t^{s-1}e^{-t}dt/\Gamma(s)$ is the regularized incomplete gamma function.

Lattices. A *lattice* L is a discrete subgroup of $\mathbb{R}^m$, or equivalently the set $L(\mathbf{b}_1, \ldots, \mathbf{b}_n) = \{\sum_{i=1}^n x_i\mathbf{b}_i : x_i \in \mathbb{Z}\}$ of all integer combinations of n linearly independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathbb{R}^m$. Such $\mathbf{b}_i$'s form a *basis* of L. All the bases of L have the same number n of elements, called the dimension or rank of L, and the same n-dimensional volume of the parallelepiped $\{\sum_{i=1}^n a_i\mathbf{b}_i : a_i \in [0,1)\}$ they generate. We call this volume the co-volume, or determinant, of L, and denote it by $\mathrm{covol}(L)$. The lattice L is said to be *full-rank* if $n = m$. The most famous lattice problem is the *shortest vector problem* (SVP), which asks to find a non-zero lattice vector of minimal Euclidean norm. The *closest vector problem* (CVP) asks to find a lattice vector closest to a target vector.

Orthogonalization. For a basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of a lattice L and $i \in \{1, \ldots, n\}$, we denote by π_i the orthogonal projection on $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{i-1})^\perp$. The *Gram-Schmidt orthogonalization* of the basis B is defined as the sequence of orthogonal vectors $B^\star = (\mathbf{b}_1^\star, \ldots, \mathbf{b}_n^\star)$, where $\mathbf{b}_i^\star := \pi_i(\mathbf{b}_i)$. We can write each $\mathbf{b}_i$ as $\mathbf{b}_i^\star + \sum_{j=1}^{i-1} \mu_{i,j}\mathbf{b}_j^\star$ for some unique $\mu_{i,1}, \ldots, \mu_{i,i-1} \in \mathbb{R}$. Thus, we may represent the $\mu_{i,j}$'s by a lower-triangular matrix μ with unit diagonal. The projection of a lattice may not be a lattice, but $\pi_i(L)$ is an $n+1-i$ dimensional lattice generated by $\pi_i(\mathbf{b}_i), \ldots, \pi_i(\mathbf{b}_n)$, with $\mathrm{covol}(\pi_i(L)) = \prod_{j=i}^n \|\mathbf{b}_j^\star\|$.

The Gaussian Heuristic. For a full-rank lattice L in $\mathbb{R}^n$ and a measurable set $C \subset \mathbb{R}^n$, the Gaussian heuristic estimates the number of lattice points inside of C to be approximately $\mathrm{vol}(C)/\mathrm{vol}(L)$. Accordingly, we would expect that $\lambda_1(L)$ might be close to $\mathrm{GH}(L) = V_n(1)^{-1/n}\mathrm{vol}(L)^{1/n}$, which holds for a random lattice L.

Cylinders. The performances of cylinder pruning are directly related to the following bodies. Define the (k-dimensional) cylinder-intersection of radii $R_1 \leq \cdots \leq R_k$ as the set

$$C_{R_1,\ldots,R_k} = \left\{ (x_1, \ldots, x_k) \in \mathbb{R}^k,\ \forall j \leq k,\ \sum_{\ell=1}^{j} x_\ell^2 \leq R_j^2 \right\} \subseteq \mathrm{Ball}_k(R_k).$$

Gama *et al.* [16] showed how to efficiently compute tight lower and upper bounds for $\mathrm{vol}(C_{R_1,\ldots,R_k})$, thanks to the Dirichlet distribution and special integrals.

2.2 Enumeration with Cylinder Pruning

To simplify notations, we assume that we focus on the SVP setting, *i.e.* to find short lattice vectors, rather than the more general CVP setting. Let L be a full-rank lattice in $\mathbb{R}^n$. Given a basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of L and a radius $R > 0$, Enumeration [29,18,15] outputs $L \cap S$ where $S = \mathrm{Ball}_n(R)$ by a depth-first tree search: by comparing all the norms of the vectors obtained, one extracts a shortest non-zero lattice vector.

We follow the general pruning framework of [5], which replaces S by a subset of S depending on B. Given a function $f : \{1, \ldots, n\} \to [0, 1]$, Gama *et al.* [16] introduced the following set to generalize the pruned enumeration of [34,35]:

$$P_f(B, R) = \{\mathbf{x} \in \mathbb{R}^n \text{ s.t. } \|\pi_{n+1-i}(\mathbf{x})\| \leq f(i)R \text{ for all } 1 \leq i \leq n\}, \qquad (4)$$

where the π_i is the projection over $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{i-1})^\perp$. The set $P_f(B, R)$ should be viewed as a random variable. Note that $P_f(B, R) \subseteq \mathrm{Ball}_n(R)$ and if g is the constant function equal to 1, then $P_g(B, R) = \mathrm{Ball}_n(R)$.

Gama *et al.* [16] noticed that the basic enumeration algorithm can actually compute $L \cap P_f(B, R)$ instead of $L \cap \mathrm{Ball}_n(R)$, just by changing its parameters. We call *cylinder pruning* this form of pruned enumeration, because $P_f(B, R)$ is an intersection of cylinders, since each equation $\|\pi_{n+1-i}(\mathbf{x})\| \leq f(i)R$ defines a cylinder. Cylinder pruning was historically introduced in the SVP setting, but its adaptation to CVP is straightforward, as was shown by Liu and Nguyen [21].

Complexity of Enumeration. The advantage is that for suitable choices of f, enumerating $L \cap P_f(B, R)$ is much cheaper than enumerating $L \cap \mathrm{Ball}_n(R)$: indeed, [16] shows that cylinder pruning runs in $\sum_{k=1}^n N_k$ poly-time operations, where N_k is the number of points of $\pi_{n+1-k}(L \cap P_f(B, R))$: this is because N_k is exactly the number of nodes at depth $n - k + 1$ of the enumeration tree which is searched by cylinder pruning. By the Gaussian heuristic, we have heuristically $N_k \approx H_k$ where:

$$H_k = \frac{\mathrm{vol}(\pi_{n+1-k}(P_f(B, R)))}{\mathrm{covol}(\pi_{n+1-k}(L))} = \frac{\mathrm{vol}(C_{Rf(1),\ldots,Rf(k)})}{\mathrm{covol}(\pi_{n+1-k}(L))}.$$

It follows that the complexity of cylinder pruning is heuristically:

$$N = \sum_{k=1}^{n} \frac{\mathrm{vol}(C_{Rf(1),\ldots,Rf(k)})}{\prod_{i=n-k+1}^{n} \|\mathbf{b}_i^\star\|} \tag{5}$$

This N is a heuristic estimate of the number of nodes in the tree searched by cylinder pruning. It depends on one hand on R and the bounding function f, but on the other hand on the quality of the basis B, because of the term $\prod_{i=n-k+1}^{n} \|\mathbf{b}_i^\star\|$. In the SVP setting, one can further divide (5) by two, because of symmetries in the enumeration tree.

Success Probability. We consider two settings:

Approximation Setting: The algorithm is successful if and only if we find at least one non-zero point of $L \cap P_f(B, R)$, that is $L \cap P_f(B, R) \not\subseteq \{0\}$. This is the situation studied in [5] and corresponds to the use of cylinder pruning in blockwise lattice reduction. By the Gaussian heuristic, the number of points of $L \cap P_f(B, R)$ is heuristically:

$$\frac{\mathrm{vol}(P_f(B, R))}{\mathrm{covol}(L)} = \frac{\mathrm{vol}(C_{Rf(1),\ldots,Rf(n)})}{\mathrm{covol}(L)}.$$

So we estimate the probability of success as:

$$\Pr_{\mathrm{succ}} = \min\left(1, \frac{\mathrm{vol}(C_{Rf(1),\ldots,Rf(n)})}{\mathrm{covol}(L)}\right). \tag{6}$$

Since $\mathrm{covol}(L) = V_n(\mathrm{GH}(L))$, if $R = \beta \mathrm{GH}(L)$, then (6) becomes

$$\Pr_{\mathrm{succ}} = \min\left(1, \beta^n \frac{\mathrm{vol}(C_{Rf(1),\ldots,Rf(n)})}{V_n(R)}\right). \tag{7}$$

Unique Setting: This corresponds to the situation studied in [16] and to bounded distance decoding (BDD). There is a secret vector $\mathbf{v} \in L$, whose distribution is assumed to be the Gaussian distribution over $\mathbb{R}^n$ of parameter σ. The algorithm is successful if and only if $\mathbf{v}$ is returned by the algorithm, *i.e.* if and only if $\mathbf{v} \in P_f(B,R)$. So we estimate the probability of success as:

$$\Pr_{\mathrm{succ}} = \rho_{n,\sigma}(P_f(B,R)) = \rho_{n,\sigma}(C_{f(1)R,\ldots,f(n)R}). \tag{8}$$

3 Lower Bounds for Cylinder Pruning

In this section, we prove novel geometric properties of cylinder intersections: if a cylinder intersection is sufficiently big (with respect to its volume or its Gaussian measure), we can lower bound the radii defining the intersection, as well as the volume of all its canonical projections, which are also cylinder intersections.

A basic ingredient behind these properties is a special case of cylinder intersections, corresponding to the step-bounding functions used in [16]. More precisely, we consider the intersection of a ball with a cylinder, which we call a ball-cylinder:

$$D_{k,n}(R,R') = \left\{ (x_1,\ldots,x_n) \in \mathbb{R}^n,\ \sum_{l=1}^{k} x_l^2 \le R^2 \ \text{ and } \ \sum_{l=1}^{n} x_l^2 \le R'^2 \right\}.$$

In other words, $D_{k,n}(R,R') = C_{R,\ldots,R,R',\ldots,R'}$ where R is repeated k times, and R' is repeated $n-k$ times. The following result is trivial:

Lemma 1. *Let $R_1 \le R_2 \le \cdots \le R_n$ and $1 \le k \le n$. Then:*

$$C_{R_1,\ldots,R_n} \subseteq D_{k,n}(R_k,R_n).$$

Note that for fixed k,n and R', $\mathrm{vol}(D_{k,n}(R,R'))$ is an increasing function of R. The following lemma gives properties of the volume and Gaussian measures of ball-cylinders, based on the background:

Lemma 2. *Let $R \le R'$ and $1 \le k \le n$. Then:*

$$\begin{aligned}
&\mathrm{vol}(D_{k,n}(R,R')) = V_n(R') \times I_{(R/R')^2}(k/2, 1+(n-k)/2)\\
&\quad \rho_{k,\sigma}(\mathrm{Ball}_k(R)) \ge \rho_{n,\sigma}(D_{k,n}(R,R')) \ge \rho_{k,\sigma}(\mathrm{Ball}_k(R))\rho_{n,\sigma}(\mathrm{Ball}_n(R'))\\
&\quad \rho_{n,\sigma}(\mathrm{Ball}_n(R)) = P(n/2, R^2/(2\sigma^2))
\end{aligned}$$

Proof. Because $D_{k,n}(R,R') \subseteq \mathrm{Ball}_n(R')$, $\mathrm{vol}(D_{k,n}(R,R'))/V_n(R')$ is the probability that a random vector $(x_1,\ldots,x_n)$ (chosen uniformly at random from the n-dimensional ball of radius R') satisfies $\sum_{l=1}^{k} x_l^2 \leq R^2$, that is, $\sum_{l=1}^{k}(x_l/R')^2 \leq (R/R')^2$. It follows that this probability is also the probability that a random vector $(y_1,\ldots,y_n)$ (chosen uniformly at random from the n-dimensional unit ball) satisfies: $\sum_{l=1}^{k} y_l^2 \leq (R/R')^2$. From the background, we know that $\sum_{l=1}^{k} y_l^2$ has distribution $\mathrm{Beta}(k/2,(n+2-k)/2)$, which proves the first equality.

Note that $D_{k,n}(R,R') \subseteq D_{k,n}(R,+\infty)$, which proves that $\rho_{n,\sigma}(D_{k,n}(R,R')) \leq \rho_{k,\sigma}(\mathrm{Ball}_k(R))$. Furthermore, by the Gaussian correlation inequality on convex symmetric sets, we have:

$$\begin{aligned}\rho_{n,\sigma}(D_{k,n}(R,R')) &\geq \rho_{n,\sigma}(\mathrm{Ball}_n(R')) \times \rho_{n,\sigma}\left(\{(x_1,\ldots,x_n) \in \mathbb{R}^n : \sum_{i=1}^{k} x_i^2 \leq R^2\}\right)\\ &= \rho_{k,\sigma}(\mathrm{Ball}_k(R))\rho_{n,\sigma}(\mathrm{Ball}_n(R'))\end{aligned}$$

which proves that $\rho_{n,\sigma}(D_{k,n}(R,R')) \geq P(k/2, R^2/(2\sigma^2))P(n/2, R'^2/(2\sigma^2))$.

Finally, let $x_1,\ldots,x_n$ be independent, normally distributed random variables with mean 0 and variance 1. Then $X = \sum_{i=1}^{n} x_i^2$ has the distribution $\mathrm{Gamma}(n/2, \theta = 2)$ whose CDF is $P(n/2, x/2)$. Therefore $\rho_n(\mathrm{Ball}_n(R)) = P(n/2, R^2/2)$. □

3.1 Lower Bounds on Cylinder Radii

The following theorem lower bounds the radii of any cylinder intersection covering a fraction of the ball:

Theorem 1. *Let* $0 \leq R_1 \leq \cdots \leq R_n$ *be such that* $\mathrm{vol}(C_{R_1,\ldots,R_n}) \geq \alpha V_n(R_n)$, *where* $0 \leq \alpha \leq 1$. *If for all* $1 \leq k \leq n$, *we define* $\alpha_k > 0$ *by* $I_{\alpha_k}(k/2, 1+(n-k)/2) = \alpha$, *then* $\mathrm{vol}(D_{k,n}(\sqrt{\alpha_k}R_n, R_n)) \leq \mathrm{vol}(C_{R_1,\ldots,R_n})$ *and:*

$$R_k \geq \sqrt{\alpha_k}R_n.$$

Proof. Lemma 1 shows that:

$$\mathrm{vol}(C_{R_1,\ldots,R_n}) \leq \mathrm{vol}(D_{k,n}(R_k,R_n)).$$

On the other hand, Lemma 2 shows that by definition of α_k:

$$\begin{aligned}&\mathrm{vol}(D_{k,n}(\sqrt{\alpha_k}R_n,R_n))\\ &= V_n(R_n) \times I_{\alpha_k}\left(\frac{k}{2}, 1+\frac{n-k}{2}\right) = \alpha V_n(R_n) \leq \mathrm{vol}(C_{R_1,\ldots,R_n}),\end{aligned}$$

which proves the first statement. Hence:
$\mathrm{vol}(D_{k,n}(\sqrt{\alpha_k}R_n,R_n)) \leq \mathrm{vol}(D_{k,n}(R_k,R_n))$, which implies that $R_k \geq \sqrt{\alpha_k}R_n$. □

The parameter α in Th. 1 is directly related to our success probability (7) in the approximation setting: indeed, if $R_n = \beta GH(L)$ and $\Pr_{\text{succ}} \geq \gamma$, then $\alpha = \gamma/\beta^n$ satisfies the condition of Th. 1. We have the following Gaussian analogue of Th. 1, where the lower bound on the volume is replaced by a lower bound on the Gaussian measure:

Theorem 2. *Let $0 \leq R_1 \leq \cdots \leq R_n$ be such that $\rho_{n,\sigma}(C_{R_1,\ldots,R_n}) \geq \beta$, where $0 \leq \beta \leq 1$. If for all $1 \leq k \leq n$ we define $\beta_k > 0$ by $P(k/2, \beta_k/(2\sigma^2)) = \beta$, then $\rho_{n,\sigma}(D_{k,n}(\sqrt{\beta_k}, R_n)) \leq \rho_{n,\sigma}(C_{R_1,\ldots,R_n})$ and $R_k \geq \sqrt{\beta_k}$.*

Proof. On the one hand, Lemma 1 shows that:

$$\rho_{n,\sigma}(C_{R_1,\ldots,R_n}) \leq \rho_{n,\sigma}(D_{k,n}(R_k, R_n)).$$

On the other hand, Lemma 2 shows that by definition of β_k:

$$\rho_{n,\sigma}(D_{k,n}(\sqrt{\beta_k}, R_n)) \leq P(k/2, \beta_k/2(\sigma^2)) = \beta \leq \rho_{n,\sigma}(C_{R_1,\ldots,R_n}),$$

which proves the first statement. Hence:

$$\rho_{n,\sigma}(D_{k,n}(\sqrt{\beta_k}, R_n)) \leq \rho_{n,\sigma}(D_{k,n}(R_k, R_n)),$$

which implies that $R_k \geq \sqrt{\beta_k}$. □

In Th. 2, β can be chosen as any lower bound on the success probability in the unique setting (8).

Th. 1 allows to derive numerical lower bounds on the radii, from any lower bound on the success probability. However, there is a special case for which the lower bound has a simple algebraic form, thanks to the following technical lemma (proved in Appendix A):

Lemma 3. *If $1 \leq k \leq n$, then:*

$$1 - P(1/2, 1/2) \leq I_{k/n}(k/2, (n-k)/2) \leq P(1/2, 1/2) \tag{9}$$

By coupling Th. 1 and Lemma 3, we obtain that the squared radii of any high-volume cylinder intersection are lower bounded by linear functions:

Theorem 3. *Let $0 \leq R_1 \leq \cdots \leq R_n$ such that $\mathrm{vol}(C_{R_1,\ldots,R_n}) \geq P(1/2, 1/2) \times V_n(R_n)$. Then for all $1 \leq k \leq n$:*

$$R_k \geq \sqrt{\frac{k}{n+2}} R_n.$$

Proof. The assumption and (9) imply that

$$\mathrm{vol}(C_{R_1,\ldots,R_n}) \geq I_{k/(n+2)}(k/2, 1 + (n-k)/2) V_n(R_n).$$

Hence, we can apply Th. 1 with $\alpha_k = \sqrt{k/(n+2)}$. □

Note that $P(1/2, 1/2) \approx 0.683\ldots$, so any bounding function with high success probability must have a cost lower bounded by that of some linear pruning, which means that its speed-up (compared to full enumeration) is at most single-exponential (see [16]).

3.2 Lower Bounds on Cylinder Volumes from Isoperimetry

The lower bounds on radii given by Th. 1 and 2 provide lower bounds on $\mathrm{vol}(C_{R_1,\dots,R_k})$ for all $1 \leq k \leq n-1$. Indeed, if $R_k \geq \sqrt{\alpha_k} R_n$, then:

$$\mathrm{vol}(C_{R_1,\dots,R_k}) \geq \mathrm{vol}(C_{\sqrt{\alpha_1}R_n,\dots,\sqrt{\alpha_k}R_n}).$$

Such lower bounds immediately provide a lower bound on the cost of enumeration with cylinder pruning, because of (5).

In this subsection, we show that this direct lower bound can be significantly improved, namely it can be replaced by $V_k(\sqrt{\alpha_k}R_n)$. Our key ingredient is the following isoperimetric result, which says that among all Borel sets of given volume, the ball centered at the origin has the largest measure, for any isotropic measure which decays monotonically radially away :

Theorem 4 (Isoperimetry). *Let A be a Borel set of $\mathbb{R}^k$. Let $\mathcal{D}$ be a distribution over $\mathbb{R}^k$ such that its probability density function f is radial and decays monotonically radially away: $f(\mathbf{x}) \leq f(\mathbf{y})$ whenever $\|\mathbf{x}\| \geq \|\mathbf{y}\|$. If a random variable X has distribution $\mathcal{D}$, then:*

$$\Pr(X \in A) \leq \Pr(X \in B),$$

where B is the ball of $\mathbb{R}^k$ centered at the origin such that $\mathrm{vol}(B) = \mathrm{vol}(A)$.

Proof. The statement is proved in [38, p498-499] for the special case where $\mathcal{D}$ is the Gaussian distribution over $\mathbb{R}^k$. However, the proof actually works for any radial probability density function which decays monotonically radially away. □

It implies the following:

Lemma 4. *Let $1 \leq k \leq n$. Let $\pi = \tau_k$ be the canonical projection of $\mathbb{R}^n$ over $\mathbb{R}^k$. Let C be a subset of the n-dimensional ball of radius R' such that both C and $\pi(C)$ are measurable. If R is the radius of the k-dimensional ball of volume $\mathrm{vol}(\pi(C))$, then:*

$$\mathrm{vol}(C) \leq \mathrm{vol}(D_{k,n}(R,R')) \text{ and } \rho_{n,\sigma}(C) \leq \rho_{n,\sigma}(D_{k,n}(R,R')).$$

Proof. Let B' be the n-dimensional centered ball of radius R'. Let B be the k-dimensional centered ball of radius R. Let $\mathbf{x}$ be chosen uniformly at random from B'. Since $C \subseteq B'$, $\mathrm{vol}(C)/V_n(R')$ is exactly $\Pr(\mathbf{x} \in C)$, and we have:

$$\Pr(\mathbf{x} \in C) \leq \Pr(\pi(\mathbf{x}) \in \pi(C)).$$

Let $\mathcal{D}$ be the distribution of $\mathbf{y} = \pi(\mathbf{x}) \in \mathbb{R}^k$. Then by Th. 4,

$$\Pr(\mathbf{y} \in \pi(C)) \leq \Pr(\mathbf{y} \in B).$$

Hence:

$$\Pr(\mathbf{x} \in C) \leq \Pr(\mathbf{y} \in B) = \frac{\mathrm{vol}(D_{k,n}(R,R'))}{V_n(R')},$$

which proves the first statement. Similarly, if $\mathbf{x}$ is chosen from the Gaussian distribution corresponding to $\rho_{n,\sigma}$, then

$$\rho_{n,\sigma}(C)/\rho_{n,\sigma}(B') = \Pr(\mathbf{x} \in C) \leq \Pr(\pi(\mathbf{x}) \in \pi(C)).$$

Let $\mathcal{D}'$ be the distribution of $\mathbf{y} = \pi(\mathbf{x}) \in \mathbb{R}^k$: this is a Gaussian distribution. Then by Th. 4,

$$\Pr(\mathbf{y} \in \pi(C)) \leq \Pr(\mathbf{y} \in B) = \frac{\rho_{n,\sigma}(D_{k,n}(R, R'))}{\rho_{n,\sigma}(B')}.$$

□

It has the following geometric consequence:

Corollary 1. *Let $R_1 \leq R_2 \leq \cdots \leq R_n$ and $1 \leq k \leq n$. Let $R > 0$ such that $\mathrm{vol}(C_{R_1,\ldots,R_n}) \geq \mathrm{vol}(D_{k,n}(R, R_n))$ or $\rho_{n,\sigma}(C_{R_1,\ldots,R_n}) \geq \rho_{n,\sigma}(D_{k,n}(R, R_n))$. Then:*

$$\mathrm{vol}(C_{R_1,\ldots,R_k}) \geq V_k(R).$$

Proof. Let $C = C_{R_1,\ldots,R_n}$ and $\pi = \tau_k$ be the canonical projection of $\mathbb{R}^n$ over $\mathbb{R}^k$. Then $\pi(C) = C_{R_1,\ldots,R_k}$. If r is the radius the k-dimensional ball of volume $\mathrm{vol}(\pi(C))$, Lemma 4 implies that: $\mathrm{vol}(C) \leq \mathrm{vol}(D_{k,n}(r, R_n))$ and $\rho_{n,\sigma}(C) \leq \rho_{n,\sigma}(D_{k,n}(r, R_n))$. Thus, by definition of R, we have either $\mathrm{vol}(D_{k,n}(R, R_n)) \leq \mathrm{vol}(C) \leq \mathrm{vol}(D_{k,n}(r, R_n))$ or $\rho_{n,\sigma}(D_{k,n}(R, R_n)) \leq \rho_{n,\sigma}(C) \leq \rho_{n,\sigma}(D_{k,n}(r, R_n))$, which each imply that $r \geq R$. □

Note that $C_{R_1,\ldots,R_k}$ and $\mathrm{Ball}_k(R)$ are the projections of respectively $C_{R_1,\ldots,R_n}$ and $D_{k,n}(R, R_n)$ over $\mathbb{R}^k$. So the corollary is a bit surprising: if one particular body is "bigger" than the other, then so are their projections. Obviously, this cannot hold for arbitrary bodies in the worst case.

This corollary implies the following lower bounds, which strengthens Theorem 1:

Corollary 2. *Under the same assumptions as Th. 1, we have:*

$$\mathrm{vol}(C_{R_1,\ldots,R_k}) \geq V_k(\sqrt{\alpha_k} R_n).$$

Proof. From Th. 1, we have: $\mathrm{vol}(C_{R_1,\ldots,R_n}) \geq \mathrm{vol}(D_{k,n}(\sqrt{\alpha_k} R_n, R_n))$. And we apply Cor. 1. □

Similarly, we obtain:

Corollary 3. *Under the same assumptions as Th. 2, we have:*

$$\mathrm{vol}(C_{R_1,\ldots,R_k}) \geq V_k(\sqrt{\beta_k} R_n).$$

It would be interesting to study if the lower bounds of the last two corollaries can be further improved.

3.3 Generalisation to Finitely Many Cylinder Intersections

In this section, we give an analogue of the results of Sect. 3.2 to finitely many cylinder intersections, which corresponds to the extreme pruning setting. The key ingredient is the following refinement of isoperimetry:

Theorem 5 (Isoperimetry). *Let $A_1, \dots, A_m$ be Borel sets of $\mathbb{R}^k$. Let $\mathcal{D}$ be a distribution over $\mathbb{R}^k$ such that its probability density function f is radial and decays monotonically radially away: $f(\mathbf{x}) \leq f(\mathbf{y})$ whenever $\|\mathbf{x}\| \geq \|\mathbf{y}\|$. If a random variable X has distribution $\mathcal{D}$, then:*

$$\frac{1}{m}\sum_{i=1}^{m} \Pr(X \in A_i) \leq \Pr(X \in B),$$

where B is the ball of $\mathbb{R}^k$ centered at the origin such that $\mathrm{vol}(B) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{vol}(A_i)$.

Proof. The statement is proved in [38, p499-500] for the special case where $\mathcal{D}$ is the Gaussian distribution over $\mathbb{R}^k$. However, the proof actually works for any radial probability density function which decays monotonically radially away. □

Lemma 5. *Let $1 \leq k \leq n$. Let $\pi = \tau_k$ be the canonical projection of $\mathbb{R}^n$ over $\mathbb{R}^k$. Let $C_1, \dots, C_m \subseteq \mathrm{Ball}_n(R')$ such that all the C_i's and $\pi(C_i)$'s are measurable. If R is the radius of the k-dimensional ball of volume $\frac{1}{m}\sum_{i=1}^{m} \mathrm{vol}(\pi(C_i))$, then:*

$$\frac{1}{m}\sum_{i=1}^{m} \mathrm{vol}(C_i) \leq \mathrm{vol}(D_{k,n}(R,R')) \textit{ and } \frac{1}{m}\sum_{i=1}^{m} \rho_{n,\sigma}(C_i) \leq \rho_{n,\sigma}(D_{k,n}(R,R')).$$

Proof. Let B' be the n-dimensional centered ball of radius R'. Let B be the k-dimensional centered ball of radius R such that $\mathrm{vol}(B) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{vol}(\pi(C_i))$. Let $\mathbf{x}$ be chosen uniformly at random from B'. Since $C_i \subseteq B'$, $\mathrm{vol}(C_i)/V_n(R')$ is exactly $\Pr(\mathbf{x} \in C_i)$, and we have:

$$\Pr(\mathbf{x} \in C_i) \leq \Pr(\pi(\mathbf{x}) \in \pi(C_i)).$$

Let $\mathcal{D}$ be the distribution of $\mathbf{y} = \pi(\mathbf{x}) \in \mathbb{R}^k$. Then by Th. 5,

$$\frac{1}{m}\sum_{i=1}^{m} \Pr(\mathbf{y} \in \pi(C_i)) \leq \Pr(\mathbf{y} \in B).$$

Hence:

$$\frac{1}{m}\sum_{i=1}^{m} \Pr(\mathbf{x} \in C_i) \leq \Pr(\mathbf{y} \in B) = \frac{\mathrm{vol}(D_{k,n}(R,R'))}{V_n(R')},$$

which proves the first statement. □

It has the following geometric consequence:

Corollary 4. *Let $C_1, \ldots, C_m \subseteq \mathrm{Ball}_n(R_n)$ be n-dimensional cylinder intersections. Let $1 \leq k \leq n$ and denote by $\pi = \tau_k$ the canonical projection of $\mathbb{R}^n$ over $\mathbb{R}^k$. Let $R > 0$ such that $\frac{1}{m}\sum_{i=1}^m \mathrm{vol}(C_i) \geq \mathrm{vol}(D_{k,n}(R, R_n))$ or $\frac{1}{m}\sum_{i=1}^m \rho_{n,\sigma}(C_i) \geq \rho_{n,\sigma}(D_{k,n}(R, R_n))$. Then:*

$$\frac{1}{m}\sum_{i=1}^m \mathrm{vol}(\pi(C_i)) \geq V_k(R).$$

Proof. If r is the radius of the k-dimensional ball of volume $\frac{1}{m}\sum_{i=1}^m \mathrm{vol}(\pi(C_i))$, the Lemma 5 implies that: $\frac{1}{m}\sum_{i=1}^m \mathrm{vol}(C_i) \leq \mathrm{vol}(D_{k,n}(r, R_n))$ and $\frac{1}{m}\sum_{i=1}^m \rho_{n,\sigma}(C_i) \leq \rho_{n,\sigma}(D_{k,n}(r, R_n))$. Thus, by definition of R, we have either $\mathrm{vol}(D_{k,n}(R, R_n)) \leq \mathrm{vol}(C) \leq \mathrm{vol}(D_{k,n}(r, R_n))$ or $\rho_{n,\sigma}(D_{k,n}(R, R_n)) \leq \rho_n(C) \leq \rho_{n,\sigma}(D_{k,n}(r, R_n))$, which each imply that $r \geq R$. □

Theorem 6. *Let $C_1, \ldots, C_m \subseteq \mathrm{Ball}_n(R_n)$ be n-dimensional cylinder intersections such that $\sum_{i=1}^m \mathrm{vol}(C_i) \geq m\alpha V_n(R_n)$, where $0 \leq \alpha \leq 1$. If for all $1 \leq k \leq n$, we define $\alpha_k > 0$ by $I_{\alpha_k}(k/2, 1 + (n-k)/2) = \alpha$, then $\mathrm{vol}(D_{k,n}(\sqrt{\alpha_k}R_n, R_n)) \leq \frac{1}{m}\sum_{i=1}^m \mathrm{vol}(C_i)$ and:*

$$\sum_{i=1}^m \mathrm{vol}(\pi(C_i)) \geq mV_k(\sqrt{\alpha_k}R_n),$$

where $\pi = \tau_k$ denotes the canonical projection of $\mathbb{R}^n$ over $\mathbb{R}^k$.

Proof. Lemma 2 shows that by definition of α_k:

$$\mathrm{vol}(D_{k,n}(\sqrt{\alpha_k}R_n, R_n)) = \alpha V_n(R_n) \leq \frac{1}{m}\sum_{i=1}^m \mathrm{vol}(C_i),$$

which proves the first statement. And the rest follows by Lemma 4. □

Again, the parameter α in Th. 6 is directly related to our global success probability (7) in the approximation setting: the global success probability is $\leq \sum_{i=1}^m \mathrm{vol}(C_i)/\mathrm{covol}(L)$ so if $R_n = \beta GH(L)$ and the global success probability is $\geq \gamma$, then $\alpha = \gamma/(m\beta^n)$ satisfies the condition of Th. 1.

We have the following Gaussian analogue of Th. 6:

Theorem 7. *Let $C_1, \ldots, C_m \subseteq \mathrm{Ball}_n(R_n)$ be n-dimensional cylinder intersections such that $\sum_{i=1}^m \rho_{n,\sigma}(C_i) \geq m\beta$, where $0 \leq \beta \leq 1/m$. If for all $1 \leq k \leq n$, we define $\beta_k > 0$ by $P(k/2, \beta_k/(2\sigma^2)) = \beta$, then $\rho_{n,\sigma}(D_{k,n}(\sqrt{\beta_k}R_n, R_n)) \leq \frac{1}{m}\sum_{i=1}^m \rho_{n,\sigma}(C_i)$ and:*

$$\sum_{i=1}^m \mathrm{vol}(\pi(C_i)) \geq mV_k(\beta_k),$$

where $\pi = \tau_k$ denotes the canonical projection of $\mathbb{R}^n$ over $\mathbb{R}^k$.

In the unique setting, the global success probability is $\leq \sum_{i=1}^m \rho_{n,\sigma}(C_i)$, so if the global success probability is $\geq \gamma$, then $\beta = \gamma/m$ satisfies the condition of Th. 7.

Surprisingly, we will show that Th. 6 and 7 imply that we can lower bound the cost of extreme pruning, independently of the number m of cylinder intersections:

Lemma 6. *Let the global probability $0 \leq \alpha' \leq 1$ and $1 \leq k \leq n$. Let $\alpha = \alpha'/m$ and $\alpha_k > 0$ such that $I_{\alpha_k}(k/2, 1+(n-k)/2) = \alpha$. Then, $mV_k(\sqrt{\alpha_k})$ is strictly decreasing w.r.t. m, yet lower bounded by some linear function of α':*

$$mV_k(\sqrt{\alpha_k}) > \alpha' \cdot \frac{kV_k(1)}{2} \cdot B\left(\frac{k}{2}, 1 + \frac{n-k}{2}\right).$$

Furthermore, for fixed α', k and n, the left-hand side converges to the right-hand side when m goes to infinity and α_k is defined as above.

Lemma 6 implies that the cost of enumeration decreases as the number of cylinder intersections increases, if the global probability α' is fixed. However, there is a limit given by some linear function of α' which depends only on n.

To prove the lemma, we use the following two lemmas:

Lemma 7. *For $a \geq 0, b \geq 1, 0 < z \leq 1$:*

$$\frac{\partial}{\partial z} I_z^{-1}(a,b) \geq \frac{1}{az} I_z^{-1}(a,b)$$

Proof. Substituting $x = I_z^{-1}(a,b)$ in (3) we obtain:

$$\frac{(1 - I_z^{-1}(a,b))^{b-1}(I_z^{-1}(a,b))^a}{aB(a,b)} \leq z.$$

This implies that

$$\frac{\partial}{\partial z} I_z^{-1}(a,b) = B(a,b)(1 - I_z^{-1}(a,b))^{1-b}(I_z^{-1}(a,b))^{1-a} \geq \frac{1}{az} I_z^{-1}(a,b).$$

□

Lemma 8. *For $a \geq 0, b \geq 1$:*

$$\lim_{y \to 0+} \frac{y}{(I_y^{-1}(a,b))^a} = \frac{1}{a \cdot B(a,b)}$$

Proof. Bounding inequalities (2) and (3) from both sides implies that

$$\lim_{x \to 0+} \frac{I_x(a,b)}{x^a} = \frac{1}{a \cdot B(a,b)}.$$

Letting $x = I_y^{-1}(a,b)$, the claim holds. □

Proof of Lemma 6

We have $I_{\alpha_k}(k/2, 1+(n-k)/2) = \alpha'/m$ and $\alpha_k = I_{\alpha'/m}^{-1}(k/2, 1+(n-k)/2)$. This gives:

$$mV_k(\sqrt{\alpha_k}) = V_k(1)m \cdot \left(I_{\alpha'/m}^{-1}(k/2, 1+(n-k)/2)\right)^{k/2}.$$

Thus, to show the first claim, it suffices to prove that

$$g(y) = \frac{1}{y}\left(I^{-1}_{\alpha' y}(k/2, 1 + (n-k)/2)\right)^{k/2}$$

is strictly increasing over $0 < y \leq 1$.

For simplicity, we write $I := I^{-1}_{\alpha' y}(k/2, 1 + (n-k)/2)$ and we have:

$$g'(y) = \frac{\alpha' k}{2y} I^{k/2-1} \cdot \frac{\partial I}{\partial y} - \frac{I^{k/2}}{y^2}$$

By Lemma 7, we can see that $\frac{\partial I}{\partial y} \geq \frac{2}{\alpha' k y} > I$ and $g'(y) > 0$ which proves the first claim. The lower bound can be derived by the limit of the function. By the relationship

$$\lim_{m \to \infty} m V_k(\sqrt{\alpha_k}) = V_k(1) \cdot \lim_{y \to 0+} g(y),$$

and the straightforward consequence of Lemma 8,

$$\lim_{y \to 0+} g(y) = \alpha' \cdot \frac{k}{2} \cdot B\left(\frac{k}{2}, 1 + \frac{n-k}{2}\right),$$

we obtain the second claim. □

By a similar technique, we can show a similar result for the Gaussian case: the proof is postponed to Appendix A.3.

Lemma 9. *Let the global probability $0 \leq \beta' \leq 1$ and $1 \leq k \leq n$. Let $\beta = \beta'/m$ and $\beta_k > 0$ such that $P(k/2, \beta_k/(2\sigma^2)) = \beta$. Then, $mV_k(\sqrt{\beta_k})$ is strictly decreasing w.r.t. m, yet lower bounded by some linear function of β':*

$$m V_k(\sqrt{\beta_k}) > \beta' (2\pi\sigma^2)^{k/2}.$$

Moreover, for fixed β', k and σ, the left-hand side converges to the right-hand side when m goes to infinity and β_k is defined as above.

4 Efficient Upper Bounds based on Cross-Entropy

In order to guess how tight are our lower bounds in practice, we need to be able to find efficiently very good bounding functions for cylinder pruning. Different methods have been used over the years (see [16,10,7,9]). In this section, we present the method that we used to generate bounding functions that try to minimize the enumeration cost, under the constraint that the success probability is greater than a given $p_0 > 0$. From our experience, different methods usually give rise to close bounding functions, but their running time can vary significantly.

4.1 Our Formulation and Previous Algorithms

Usually, the problem to find optimal cost has two formulations and our algorithm targets the first one:

1. [11,7] for a given basis B, radius R, and target probability p_0, minimize the cost (5) subject to the constraint that the probability (6) is greater than p_0. The variables are $R_1, \ldots, R_n$. This kind of constrained optimization is known as *monotonic optimization* because the objective function and constraint functions are both monotonic, i.e., $f(x_1, \ldots, x_n) \leq f(x'_1, \ldots, x'_n)$ if $x_i \leq x'_i$ for all i. It is known that the optimal value is on the border (see, for example [12]). A heuristic random perturbation is implemented in the progressive BKZ library [7], and an outline of the cross-entropy method is mentioned in Chen's thesis [10].
2. [9] for a given basis B and radius, minimize the expected cost of extreme pruning [16]: $m \cdot EnumCost + (m-1) \cdot PreprocessCost$ where m is a variable defining the number of bases, and therefore the success probability of the enumeration. The variables are $R_1, \ldots, R_n$ and m. This is an unconstrained optimization problem. A heuristic gradient descent and the Nelder-Mead method are implemented in the fpLLL library [9].

We explain why we introduce a new approach. All the known approaches try to minimize an approximate upper bound of the enumeration cost: this approximation is the sum of n terms, where each term can be derived from the computation of a simplex volume (following [16]) which costs $O(n^2)$, where the unit is number of floating-point operations and the required precision might be linear in n. Although there exists an $O(n^2)$ algorithm to compute the approximate upper bound [4, Section 3.3], a naive random perturbation strategy is too slow to converge.

Besides, we think that the Nelder-Mead and gradient descent are not suitable for our optimization problem. If we want to apply such methods to the constrained problem, a usual approach converts the problem into a corresponding global optimization problem by introducing penalty functions. Then, we find a near-optimal solution to the original problem by using the optimized variable of the converted problem. However, we know that the optimal point is on the border at which the penalty functions must change drastically. It could make the optimal point of the new problem far from the original one. Hence, we need an algorithm to solve our constrained problem directly.

For this purpose, we revisit Chen's partial description [10] of the cross-entropy method to solve the problem (i). In Section 4.2, we give a brief overview of the cross-entropy method, and in Section 4.3, we explain how we modify it for our purpose.

4.2 A Brief Introduction to the Cross Entropy Method

The original motivation of the cross entropy method is to speed up Monte-Carlo simulation for approximating a probability. If the target probability is extremely small, the number of sampling points must be huge. To solve this issue, Rubinstein [30] introduced the cross entropy method and showed that the algorithm could be used for combinatorial optimization problems. This subsection gives a general presentation of the cross-entropy method: we will apply it to the optimization of pruning functions. For more information, see for example [30,14].

Let χ be the whole space of combinations and consider a cost function $S : \chi \to \mathbb{R}_{\geq 0}$ that we want to minimize. Assume that we have a probability distribution $D_{\chi,\mathbf{u}}$ defined over χ and parametrized by a vector $\mathbf{u}$. We fix the corresponding probability density function $f_{\mathbf{u}}(x)$. A cross-entropy algorithm to find the optimal combination $X^* := \text{argmin}_{X \in \chi} S(X)$ is outlined in Algorithm 1; here we use the description in the textbook [14, Algorithm 2.3].

Algorithm 1 A Generic Framework of the Cross-Entropy Method

```
Input: Searching space $\chi$,cost function $S : \chi \to \mathbb{R}_{\geq 0}$, initial parameter vector
    $\mathbf{v}_0$, algorithm parameter $\rho, N, d$;for example, $N = 1000$, $\rho = 0.1$ and $d = 10$.
Output: An approximation $S(x^*)$ of the minimal and corresponding $x^*$.
1: $t \leftarrow 1$
2: According to $D_{\chi,\mathbf{v}_{t-1}}$, sample $X_1, \ldots, X_N$ from $\chi$
3: Let the threshold $\gamma_t$ be the $\lceil \rho N \rceil$-th smallest value of $S(X_i)$
4: Solve the stochastic program (10) for the inputs $(X_1, \ldots, X_N, \gamma_t, \mathbf{v}_{t-1})$ and
   find the new parameter $\mathbf{v}_t$
5: if the found minimum $S(X^*)$ during the execution of the algorithm is not
   updated in the last $d$ loop then
6:     output the smallest $S(X^*)$ and $X^*$
7: else
8:     let $t \leftarrow t + 1$ and goto Step 2
9: end if
```

The stochastic program in Step 4 is the problem of finding the parameter vector $\mathbf{v}$ which optimizes

$$\arg\max_{\mathbf{v}} \sum_{i=1}^{N} I_{S(X_i) \leq \gamma_t} \log f_{\mathbf{v}}(X_i) \tag{10}$$

where

$$I_{S(X_i) \leq \gamma_t} = \begin{cases} 1 & \text{if } S(X_i) \leq \gamma_t \\ 0 & \text{if } S(X_i) > \gamma_t \end{cases}$$

is the characteristic function. It is known that the new distribution $D_{\chi,\mathbf{v}_t}$ derived from the solution is closer to the ideal distribution $D_{\chi,\mathbf{opt}}$ that outputs the optimal $X^{\mathbf{opt}} = \arg\min_X S(X)$ with probability 1, than the previous distribution $D_{\chi,\mathbf{v}_{t-1}}$. In other words, the cost of sampled elements from $D_{\chi,\mathbf{v}_t}$ are likely to smaller than that of samples from $D_{\chi,\mathbf{v}_{t-1}}$. This is quantified by the function to measure the distance between two probability distributions:

$$D(g, f_{\mathbf{v}}) := \int g(x) \log \frac{g(x)}{f_{\mathbf{v}}(x)} dx$$

which is known as the *cross-entropy*, or Kullback-Leibler distance. The above algorithm wants to minimize the distance from the optimal state g by changing the parameter vector $\mathbf{v}$.

The stochastic program (10) can be easily solved analytically if the family of distribution function $\{f_{\mathbf{v}}(x)\}_{\mathbf{v} \in V}$ is a natural exponential family (NEF) [31]. In particular, if the function $f_{\mathbf{v}}(x)$ is convex and differentiable with respect to $\mathbf{v}$, the solution of (10) is obtained by solving the simultaneous equations

$$\sum_{i=1}^{N} I_{S(X_i) \leq \gamma_t} \nabla \log f_{\mathbf{v}}(X_i) = \mathbf{0}. \tag{11}$$

The Gaussian product (12) used in the next section is one of the simplest examples of such functions.

4.3 Our Algorithm

For the generic algorithm (Algorithm 1), we substitute our cost function and constraints. Then, we modify the sampler and introduce the FACE strategy as explained in this section. Recall that the input is a lattice basis and its Gram-Schmidt lengths, a radius R and a target probability p_0. We mention that our algorithm follows [19, Algorithm 2] for optimization over a subset of $\mathbb{R}^m$ by Kroese, Porotsky and Rubinstein.

Modified sampler: The sampling parameter is $\mathbf{u} = (c_1, \ldots, c_{n-1}, \sigma_1, \ldots, \sigma_{n-1}) \in \mathbb{R}^{2n-2}_{\geq 0}$ where c and σ correspond to the center and deviation respectively.

Since the bounding radii must increase and the last coordinate is $R_n = 1$, the searching space is

$$\chi = \{(x_1, \ldots, x_{n-1}) \in (0, 1]^n : x_1 \leq x_2 \leq \cdots \leq x_{n-1}\} \subset \mathbb{R}^{n-1}.$$

To sample from the space following the parameter $\mathbf{u}$, define the corresponding probability distribution $D_{\chi,\mathbf{u}}$ as follows: sample each u_i from $N(c_i, \sigma_i^2)$ independently, if all $u_i \geq 0$, then let $(x_1, \ldots, x_n)$ be $(u_1, \ldots, u_n)$ sorted in increasing order and output it. We sort the output because because we do not know a suitable distribution from which the sampling from χ is easy. As we will see later, when the algorithm is about to converge, the Gaussian parameters σ_i become small, and the distributions of u_i's and x_i become close. Below we assume that the probability density function of $D_{\chi,\mathbf{u}}$ is sufficiently close to that of the Gaussian product

$$f_{\mathbf{u}}(X) = \frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{n-1} \left(\frac{1}{\sigma_i} \exp(-(x_i - c_i)^2/(2\sigma_i^2)) \right). \tag{12}$$

The gradients of log of the function are

$$\frac{\partial}{\partial c_i} \log f_{\mathbf{u}}(X) = \frac{x_i - c_i}{\sigma_i^2},$$

and

$$\frac{\partial}{\partial \sigma_i} \log f_{\mathbf{u}}(X) = -\frac{1}{\sigma_i} + \frac{(x_i - c_i)^2}{\sigma_i^3}.$$

Substituting them into (11), we obtain the formulas to update c_i and σ_i as follows

$$\begin{aligned} c_i^{new} &\leftarrow \frac{\sum_{j:S(X_j)\leq\gamma_t} x_{j,i}}{|\{j : S(X_j) \leq \gamma_t\}|} \\ \sigma_i^{new} &\leftarrow \sqrt{\frac{\sum_{j:S(X_j)\leq\gamma_t} (x_{j,i} - c_i)^2}{|\{j : S(X_j) \leq \gamma_t\}|}} \end{aligned} \tag{13}$$

where we denote $x_{j,i}$ for the i-th coordinate of X_j.

The FACE strategy: For practical speedup, we can employ the fully-automated cross-entropy (FACE) strategy described in [14, Section 4.2]. It simply replaces the full sampling in Step 2 in Figure 1 by a recycling strategy. Consider a list $L = \{X_1, \ldots, X_N\}$. If the cost of a new sample is less than $\max_{i\in[N]} S(X_i)$, replace the new sample to the maximum element in the list, and update the parameter vector by (13) using all items in the list, i.e., with $\gamma_t = +\infty$.

We did preliminary experiments on this strategy and found that our problem has a typical trend, *i.e.* if the size N of list is small (≈ 10), the minimum cost $\min_{i\in[N]} S(X_i)$ decreases very fast but seems to stay near a local minimum. On the other hand, if we choose a large N (≈ 1000), the speed of convergence is slow, but the pruning function found is better than in the small case if we use many loop iterations. Hence, we start with a small N and increase it little by little.

Integrating the above, we give the pseudocode of our optimizing algorithm in Algorithm 2. We used a heuristic parameter set $N_{init} = 10$ and $N_{max} = 50$, and terminate the computation if $\mathbf{v}$ is not updated in the last 10 loop iterations.

5 Tightness and Applications to Security Estimates

In this section, we study the heuristic cost N of (5) divided by two (SVP setting).

5.1 Modeling Strongly Reduced Bases

The cost (5) of cylinder pruning over $P_f(B, R)$ depends both on the quality of the basis B, the radius R and the pruning function f. The results of Sect. 3 allow to lower bound the numerator of each term of (5), but we also need to lower bound the part depending on the basis B. This was already discussed in [25,11,7] using two models of strongly reduced bases: the Rankin model

Algorithm 2 Cross-Entropy Method for Optimizing Pruning Radii

```
Input: Gram-Schmidt lengths $(\|\mathbf{b}_1^\star\|, \ldots, \|\mathbf{b}_n^\star\|)$, Radius of the ball $R$, Target
    probability $p_0$, initial and maximum size of list $N$, $N_{max}$, initial parameter
    vector $\mathbf{u} = (\mathbf{c}, \sigma)$, parameter to increase list size $d$.
Output: A near optimal cost and corresponding radii $(R_1, \ldots, R_n)$
 1: Sample new $X = (R_1, \ldots, R_m)$ from $D_{\chi,\mathbf{u}}$
 2: if $Pr(X) < p_0$ then
 3:     goto Step 1
 4: end if
 5: if $|L| < N$ then
 6:     $L \leftarrow L \cup X$
 7: else
 8:     $X_i \leftarrow \mathrm{argmax}_{X_i \in L} Cost(X_i)$
 9: end if
10: if $Cost(X) < Cost(X_i)$ then
11:     Replace $X_i$ by $X$
12:     Update $\mathbf{u}$ by using list $L$
13: end if
14: if $\mathbf{u}$ is not updated in the last $d$ loops then
15:     $N \leftarrow N + 1$
16: end if
17: if $N > N_{max}$ then
18:     output minimum among $X_1, \ldots, X_{N-1}$ and exit
19: end if
20: goto Step 1
```

used in [11,25] which provides conservative bounds by anticipating progress in lattice reduction, and the HKZ model used in [11,7] which is closer to the state-of-the-art. This part is more heuristic than Sect. 3.

The HKZ model. The BKZ algorithm tries to approximate HKZ-reduced bases, which are bases B such that $\|\mathbf{b}_i^\star\| = \lambda_1(\pi_i(L))$ for all $1 \leq i \leq n$. When running BKZ, an HKZ basis is the best output one can hope for. On the other hand, a BKZ-reduced basis with large blocksize will be close to an HKZ-basis, so this model is somewhat close to the state-of-the-art. It corresponds to an idealized Kannan's algorithm [18] where enumerations are only performed over HKZ-reduced bases (see [23] for more practical variants). Unfortunately, in theory, we do not know what the $\|\mathbf{b}_i^\star\|$'s of an HKZ basis will look like exactly, except for $i = 1$, but we can make a guess. Following [11,7], we assume that for $1 \leq i \leq n - 50$, $\|\mathbf{b}_i^\star\| \approx \mathrm{GH}(\pi_i(L)) = V_{n-i+1}(1)^{-1/(n-i+1)} \left(\prod_{k=i}^n \|\mathbf{b}_k^\star\|\right)^{1/(n-i+1)}$, which means that we assume that $\pi_i(L)$ behaves like a random lattice. Then we can simulate $\|\mathbf{b}_i^\star\|$ for $1 \leq i \leq n - 50$ by a simple recursive formula. We stop at $n - 50$, because Chen and Nguyen [11] reported that the last projected lattices do not behave like random lattices. For the remaining indices, they proposed to use a numerical table from experimental results in low dimension: we use the same table. Note that for a large dimension such as 200, errors in the last coordinates are not an issue because the contribution of the terms $k \leq 50$ in N is negligible.

The Rankin model. It is known that HKZ bases are not optimal for minimizing the running time of enumeration. For instance, Nguyen [27, Chapter 3] noticed a link between the cost of enumeration and the Rankin invariants of a lattice, which provides lower bounds on heuristic estimates of the number of nodes and identifies better bases than HKZ. However, finding these better bases is currently more expensive [13] than finding HKZ-reduced bases. Recall that the Rankin invariants $\gamma_{n,m}(L)$ of an n-rank lattice L satisfy:

$$\gamma_{n,m}(L) := \min_{\substack{S\text{: sublattice of } L \\ \mathrm{rank}(S)=m}} \left(\frac{\mathrm{vol}(S)}{\mathrm{covol}(L)^{m/n}} \right)^2 \leq \frac{\prod_{i=1}^{m} \|\mathbf{b}_i^\star\|^2}{\mathrm{covol}(L)^{2m/n}}, \tag{14}$$

for any basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of L. We have the following lower bound [37, Cor. 1] for Rankin's constant $\gamma_{n,m} := \max_L \gamma_{n,m}(L)$:

$$\gamma_{n,m} \geq \left(n \cdot \frac{\prod_{n-m+1}^{n} Z(j)}{\prod_{j=2}^{m} Z(j)} \right)^{2/n} \quad \text{where } Z(j) := \zeta(j)\Gamma(j/2)\pi^{-j/2}. \tag{15}$$

According to [36], it seems plausible that most lattices come close to realizing Rankin constants: for any $\varepsilon > 0$ and sufficiently large n, most lattices L "should" verify $\gamma_{n,m}(L)^{1/(2m)} \geq \gamma_{n,m}^{1/(2m)} - \varepsilon$ for all m.

Ignoring ε, if we lower bound any term of the form $\frac{\prod_{i=1}^{m} \|\mathbf{b}_i^\star\|^2}{\mathrm{covol}(L)^{2m/n}}$ in the simplified cost (5) by the right-hand side of (15), we obtain the following heuristic lower bound formula:

$$N = \frac{1}{2} \sum_{k=1}^{n} \frac{\mathrm{vol}(C_{R_1,\ldots,R_k}) \prod_{i=1}^{n-k} \|\mathbf{b}_i^\star\|}{\mathrm{vol}(L)} > \frac{1}{2} \sum_{k=1}^{n} \frac{\mathrm{vol}(C_{R_1,\ldots,R_k})}{\mathrm{vol}(L)^{k/n}} \left((n-k) \frac{\prod_{j=k+1}^{n} Z(j)}{\prod_{j=2}^{n-k} Z(j)} \right)^{\frac{1}{n-k}}$$

In both cases, substituting the volume lower bounds in Section 3.2 and 3.3, we obtain closed formulas to find the lower bound complexity which are suitable for numerical analyses.

On the other hand, for any n-rank lattice L, and any fixed $m \in \{1, \ldots, n-1\}$, there is a basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of L such that $\frac{\prod_{i=1}^{m} \|\mathbf{b}_i^\star\|^2}{\mathrm{covol}(L)^{2m/n}} = \gamma_{n,m}(L)$. This existence would only be guaranteed for fixed m, such as for the m maximizing the corresponding number N_{n+1-m} of nodes in the enumeration tree at depth m. By idealization, we call Rankin basis a basis such that for all $m \in \{1, \ldots, n-1\}$, $\frac{\prod_{i=1}^{m} \|\mathbf{b}_i^\star\|^2}{\mathrm{covol}(L)^{2m/n}}$ is approximately less than the right-hand side of (15): since such bases may not exist, this is an over-simplification to guess how much speed-up might be possible with the best bases. We use Rankin bases to compute speculative upper bounds, anticipating progress in lattice reduction.

5.2 Explicit Lower Bounds

We summarize the applications of the results of Section 3.2 and 3.3, to compute lower bounds on the number of nodes searched by cylinder pruning with lower bounded success probability.

Single Enumeration. By Corollary 2, if α is a lower bound on the success probability,

$$N \geq \frac{1}{2} \sum_{k=1}^{n} \frac{V_k(\sqrt{\alpha_k} R_n)}{\prod_{i=n-k+1}^{n} \|\mathbf{b}_i^\star\|} \tag{16}$$

where α_k is defined by $I_{\alpha_k}(k/2, 1+(n-k)/2) = \alpha$.

For the Gaussian case with success probability $\geq \beta$, from Corollary 3,

$$N \geq \frac{1}{2} \sum_{k=1}^{n} \frac{V_k(\sqrt{\beta_k})}{\prod_{i=n-k+1}^{n} \|\mathbf{b}_i^\star\|}$$

where β_k is defined by $P(k/2, \beta_k/(2\sigma^2)) = \beta$.

Multiple Enumerations. For the situation where one can use m bases, let α' be a lower bound on the global success probability. Then by Lemma 6,

$$N \geq \frac{\alpha'}{4} \sum_{k=1}^{n} \frac{k V_k(R_n) B(k/2, 1+(n-k)/2)}{\prod_{i=n-k+1}^{n} \|\mathbf{b}_i^\star\|} \tag{17}$$

where α' satisfies $\mathrm{vol}(\cup_{i=1}^{m} C_i) \geq \alpha' \mathrm{vol}(R_n)$.

Lemma 9 also implies a lower bound for the Gaussian setting with global success probability $\rho_{n,\sigma}(\cup_{i=1}^{m} C_i) \geq \beta'$:

$$N \geq \frac{\beta'}{2} \sum_{k=1}^{n} \frac{(2\pi\sigma^2)^{k/2}}{\prod_{i=n-k+1}^{n} \|\mathbf{b}_i^\star\|}.$$

5.3 Radii Tightness

To check tightness, we give two figures (Figure 2) that compare the lower bound of radii from Corollary 2, and the best radii generated by our cross entropy method. The comparison is for two regimes: high and low success probability. Note that the left probability 0.6827 is an approximation of $P(\frac{1}{2}, \frac{1}{2})$ for which the linear pruning is the best known proved lower bound.

We see that the radii bounds are reasonably tight in both cases. We deduce that in these examples, the enumeration cost bounds will also be tight, because the cost is dominated by what happens around $k \approx n/2$.

We note that it is to easier to compute lower bounds than upper bounds.

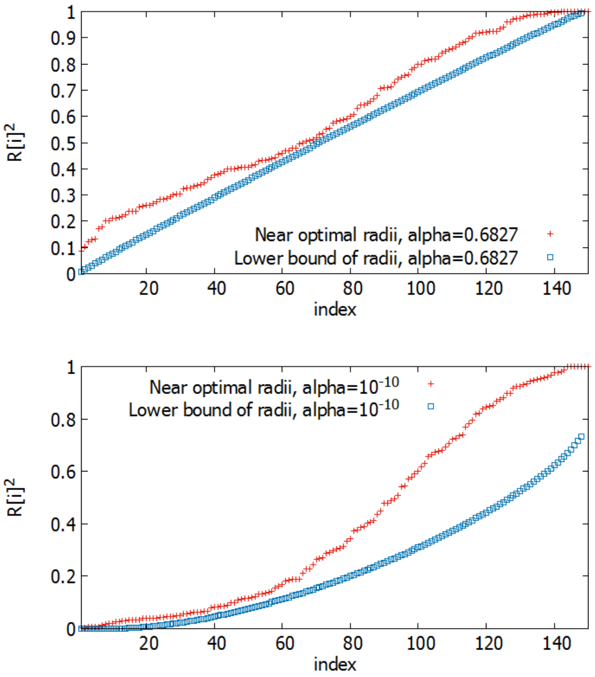

Fig. 2. Comparison of lower bound and near optimal radii; for the 150-dimensional simulated HKZ basis, compute near optimal radii and lower bound radii for $\alpha = 0.6827 \gtrsim P(\frac{1}{2}, \frac{1}{2})$ (Top) and $\alpha = 10^{-10}$ (Bottom).

5.4 Security Estimates for Enumeration

Fig. 1 (in the introduction) displays four bounds on the cost of enumeration in several situations, for varying dimension and simulated HKZ bases and Rankin bases:

- The thin red curve is an upper bound of the enumeration cost using $M = 10^{10}$ bases with single success probability $\alpha = 10^{-10}$ computed by the cross-entropy method.
- The bold red curve is a lower bound of the enumeration cost using $M = 10^{10}$ bases with single success probability $\alpha = 10^{-10}$ computed by M times (16).

- The thin green curve is an upper bound of the enumeration cost w.r.t. infinitely many bases with global success probability $\alpha' = 1$. This is computed by M times an upper bound of the enumeration cost with single success probability $1/M$ for a very large M where the single cost is greater than lattice dimension.
- The bold green curve is a lower bound of the enumeration cost w.r.t. infinitely many bases with a large global success probability. This is computed by (17) with $\alpha' = 1$.

In all experiments, we take the radius by $R_n = GH(L)$. The cost is the number of nodes of of the enumeration tree in the classical computing model. The security level is the base-2 logarithm of the cost, which is divided by two in the quantum computing model [6,24].

We also draw the curve of $2^{0.292n}$ and $2^{0.265n}$ which are simplified lower bounds of the cost for solving SVP-n used in [2] for classical and quantum computers, respectively.

In all the situations where we use 10^{10} bases, the upper bounds (thin red curve) and the lower bounds (bold red curve) are close to each other, which demonstrates the tightness of our lower bound.

In the classical setting, our lower bounds for enumeration are higher than sieve lower bounds. On the other hand, in the quantum setting, there are cases where enumeration is faster than quantum sieving. For instance, if an attacker could find many quasi-Rankin bases by some new lattice reduction algorithm, the claimed 2^{128} quantum security might be dropped to about 2^{96} security. In such a situation, the required blocksize would increase from about 480 to 580.

5.5 Experimental Environments

All experiments were performed by a standard server with two Intel Xeon E5-2660 CPUs and 256-GB RAM. We used the boost library version 1.56.0, which has efficient subroutines to compute (incomplete) beta, (incomplete) gamma and zeta functions with high precision.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 16H02780, 16H02830 and 18H03238, and JST CREST JPMJCR168A.

References

1. M. Ajtai, R. Kumar, and D. Sivakumar. A sieve algorithm for the shortest lattice vector problem. In *Proc. of 33rd STOC*, pages 601–610. ACM, 2001.
2. M. R. Albrecht, B. R. Curtis, A. Deo, A. Davidson, R. Player, E. Postlethwaite, F. Virdia, and T. Wunderer. Estimate all the {LWE, NTRU} schemes! Posted on the pqc-forum on Feb. 1, 2018. Available at `https://estimate-all-the-lwe-ntru-schemes.github.io/paper.pdf`.

3. E. Alkim, L. Ducas, T. Pöppelmann, and P. Schwabe. Post-quantum key exchange - A new hope. In *Proc. 25th USENIX Security Symposium*, pages 327–343. USENIX Association, 2016.
4. Y. Aono. A faster method for computing Gama-Nguyen-Regev's extreme pruning coefficients. *CoRR*, abs/1406.0342, 2014.
5. Y. Aono and P. Q. Nguyen. Random sampling revisited: Lattice enumeration with discrete pruning. In *Advances in cryptology—EUROCRYPT 2017 Part II*, volume 10211 of *LNCS*, pages 65–102. Springer, 2017. Full version on `https://eprint.iacr.org/2017/155`.
6. Y. Aono, P. Q. Nguyen, and Y. Shen. Quantum lattice enumeration and tweaking discrete pruning. `https://eprint.iacr.org/2018/546`, 2018.
7. Y. Aono, Y. Wang, T. Hayashi, and T. Takagi. Improved progressive BKZ algorithms and their precise cost estimation by sharp simulator. *IACR Cryptology ePrint Archive*, 2016:146, 2016. Full version of EUROCRYPT 2016.
8. A. Becker, L. Ducas, N. Gama, and T. Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In *Proc. 27th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 10–24, 2016.
9. D. Cadé, X. Pujol, and D. Stehlé. FPLLL library, version 3.0. Available from `http://perso.ens-lyon.fr/damien.stehle`, Sep 2008.
10. Y. Chen. *Réduction de réseau et sécurité concrète du chiffrement complètement homomorphe*. PhD thesis, Univ. Paris 7, 2013.
11. Y. Chen and P. Q. Nguyen. BKZ 2.0: better lattice security estimates. In *Proc. ASIACRYPT 2011*, volume 7073 of *LNCS*, pages 1–20. Springer, 2011.
12. M.-S. Cheon. *Global Optimization of Monotonic Programs: Applications in Polynomial and Stochastic Programming*. PhD thesis, Georgia Institute of Technology, 2005.
13. D. Dadush and D. Micciancio. Algorithms for the densest sub-lattice problem. In *Proc. 24th ACM-SIAM Symposium on Discrete Algorithms, SODA 2013*, pages 1103–1122, 2013.
14. P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of Operations Research*, 134(1):19–67, 2005.
15. U. Fincke and M. Pohst. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Mathematics of Computation*, 44(170):463–471, 1985.
16. N. Gama, P. Q. Nguyen, and O. Regev. Lattice enumeration using extreme pruning. In *Advances in cryptology—EUROCRYPT 2010*, volume 6110 of *LNCS*, pages 257–278. Springer, 2010.
17. A. Hülsing, J. Rijneveld, J. M. Schanck, and P. Schwabe. NTRU-HRSS-KEM: Algorithm specifications and supporting documentation. NIST submission of Nov. 30, 2017.
18. R. Kannan. Improved algorithms for integer programming and related lattice problems. In *Proc. 15th ACM STOC*, pages 193–206, 1983.
19. D. P. Kroese, S. Porotsky, and R. Y. Rubinstein. The cross-entropy method for continuous multi-extremal optimization. *Methodology and Computing in Applied Probability*, V8(3):383–407, 2006.
20. T. Laarhoven. Sieving for shortest vectors in lattices using angular locality-sensitive hashing. In *Advances in Cryptology - Proc. CRYPTO 2015 - Part I*, volume 9215 of *LNCS*, pages 3–22. Springer, 2015.
21. M. Liu and P. Q. Nguyen. Solving BDD by enumeration: An update. In *Topics in Cryptology - Proc. CT-RSA 2013*, volume 7779 of *LNCS*, pages 293–309. Springer, 2013.

22. D. Micciancio and P. Voulgaris. Faster exponential time algorithms for the shortest vector problem. In *Proc. ACM-SIAM SODA*, pages 1468–1480, 2010.
23. D. Micciancio and M. Walter. Fast lattice point enumeration with minimal overhead. In *Proc. SODA '15*, pages 276–294, 2015.
24. A. Montanaro. Quantum walk speedup of backtracking algorithms. *ArXiv e-prints*, 2015.
25. P. Q. Nguyen. Public-key cryptanalysis. In I. Luengo, editor, *Recent Trends in Cryptography*, volume 477 of *Contemporary Mathematics*. AMS–RSME, 2009.
26. P. Q. Nguyen. Hermite's constant and lattice algorithms. In *The LLL Algorithm: Survey and Applications*. Springer, 2010. In [27].
27. P. Q. Nguyen and B. Vallée, editors. *The LLL Algorithm: Survey and Applications*. Information Security and Cryptography. Springer, 2009.
28. P. Q. Nguyen and T. Vidick. Sieve algorithms for the shortest vector problem are practical. *J. of Mathematical Cryptology*, 2008.
29. M. Pohst. On the computation of lattice vectors of minimal length, successive minima and reduced bases with applications. *SIGSAM Bull.*, 15(1):37–44, 1981.
30. R. Y. Rubinstein. Optimization of computer simulation models with rare events. *European Journal of Operations Research*, 99:89–112, 1996.
31. R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method, A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer-Verlag New York, 2004.
32. M. Schneider and N. Gama. SVP challenge. Available at `http://www.latticechallenge.org/svp-challenge/`.
33. C. P. Schnorr. Lattice reduction by random sampling and birthday methods. In *Proc. STACS 2003*, volume 2607 of *LNCS*, pages 145–156. Springer, 2003.
34. C.-P. Schnorr and M. Euchner. Lattice basis reduction: improved practical algorithms and solving subset sum problems. *Math. Programming*, 66:181–199, 1994.
35. C.-P. Schnorr and H. H. Hörner. Attacking the Chor-Rivest cryptosystem by improved lattice reduction. In *Proc. of Eurocrypt '95*, volume 921 of *LNCS*, pages 1–12. IACR, Springer-Verlag, 1995.
36. U. Shapira and B. Weiss. A volume estimate for the set of stable lattices. *Comptes Rendus Mathématique*, 352(11):875 – 879, 2014.
37. J. L. Thunder. Higher-dimensional analogs of Hermite's constant. *Michigan Math. J.*, 45(2), 1998.
38. S. S. Venkatesh. *The Theory of Probability: Explorations and Applications*. Cambridge University Press, 2012.

A Proof of Lemma 3

Let

$$p(k,n) := I_{k/n}\left(\frac{k}{2},\frac{n-k}{2}\right) = \frac{\int_0^{k/n} z^{\frac{k}{2}-1}(1-z)^{\frac{n-k}{2}-1}dz}{B(\frac{k}{2},\frac{n-k}{2})} \tag{18}$$

To prove Lemma 3, it suffices to show that: for any integers $1 \leq k < n$,

$$p(n,k) \leq P(\frac{1}{2},\frac{1}{2}) = \frac{\int_0^{1/2} t^{-1/2}e^{-t}}{\Gamma(\frac{1}{2})} \approx 0.682689...$$

A.1 Formulas and Lemmas

We have

$$\Gamma(a+1) = a\Gamma(a) \quad \text{and} \quad B(a,b+1) = B(a,b)\frac{b}{a+b}. \tag{19}$$

The following recurrence formulas hold (see `8.17.18` and `8.17.21` of *NIST Digital Library of Mathematical Functions* `http://dlmf.nist.gov/8.17` respectively):

$$I_x(a,b) = I_x(a+1,b-1) + \frac{x^a(1-x)^{b-1}}{aB(a,b)}. \tag{20}$$

$$I_x(a,b) = I_x(a,b+1) - \frac{x^a(1-x)^b}{bB(a,b)}. \tag{21}$$

We recall:

Theorem 8. *(Chebyshev integral inequality) For any nonnegative, monotonically increasing function $f(x)$ and monotonically decreasing function $g(x)$, we have*

$$\int_a^b f(x)g(x)dx \le \frac{1}{b-a}\left(\int_a^b f(x)dx\right)\cdot\left(\int_a^b g(x)dx\right).$$

Lemma 10. *For $a,b > 1$, the function $z^a(1-z)^b$ is maximized at $z_{max} = \frac{a}{a+b}$. Furthermore, it is strictly increasing over $z \in [0, z_{max}]$ and strictly decreasing over $z \in [z_{max}, 1]$.*

A.2 Proof Body

The proof of Lemma 3 can be derived from the following three lemmas.

Lemma 11. *If $n \ge 2$, then: $p(1,n) < p(1,n+2)$.*

Proof. By (21), we have

$$\begin{aligned}
p(1,n) &= I_{\frac{1}{n}}\left(\frac{1}{2},\frac{n+1}{2}\right) - \frac{n^{-1/2}(1-1/n)^{\frac{n-1}{2}}}{\frac{n-1}{2}\cdot B\left(\frac{1}{2},\frac{n-1}{2}\right)} \\
&= I_{\frac{1}{n+2}}\left(\frac{1}{2},\frac{n+1}{2}\right) + \frac{\int_{\frac{1}{n+2}}^{\frac{1}{n}} z^{-1/2}(1-z)^{\frac{n-1}{2}}dz}{B\left(\frac{1}{2},\frac{n+1}{2}\right)} - \frac{n^{-1/2}(1-1/n)^{\frac{n-1}{2}}}{\frac{n-1}{2}\cdot B\left(\frac{1}{2},\frac{n-1}{2}\right)}.
\end{aligned}$$

Then $J = p(1,n) - p(1,n+2)$ is equal to the last two terms. We will show that $J < 0$. From (19), we have $B\left(\frac{1}{2},\frac{n+1}{2}\right) = \frac{n-1}{n}B\left(\frac{1}{2},\frac{n-1}{2}\right)$ and we get

$$J' = J\cdot(n-1)B\left(\frac{1}{2},\frac{n-1}{2}\right) = n\int_{\frac{1}{n+2}}^{\frac{1}{n}} z^{-1/2}(1-z)^{\frac{n-1}{2}}dz - 2n^{-1/2}(1-1/n)^{\frac{n-1}{2}}$$

of which we want to show negativeness.

Since the integral function $z^{-1/2}(1-z)^{\frac{n-1}{2}}$ is strictly decreasing, the trivial bound

$$\begin{aligned} &n \int_{\frac{1}{n+2}}^{\frac{1}{n}} z^{-1/2}(1-z)^{\frac{n-1}{2}} dz \\ &< n\left(\frac{1}{n}-\frac{1}{n+2}\right)\left(\frac{1}{n+2}\right)^{-1/2}\left(1-\frac{1}{n+2}\right)^{\frac{n-1}{2}} = \frac{2}{\sqrt{n+2}}\left(1-\frac{1}{n+2}\right)^{\frac{n-1}{2}} \end{aligned}$$

holds. Thus, letting $f(x) = \frac{1}{\sqrt{x}}(1-1/x)^{\frac{n-1}{2}}$, we have $J' < 2(f(n+2)-f(n))$ and it suffices to show that $f(x)$ is strictly decreasing over the range $x \in (n, n+2)$. It is equivalent to check that the derivative of $g(x) = f(1/x) = \sqrt{x}(1-x)^{\frac{n-1}{2}}$ is > 0 for $\frac{1}{n+2} < x < \frac{1}{n}$. We have:

$$(\log g(x))' = \frac{g'(x)}{g(x)} = \frac{1}{2x} + \frac{n-1}{2}\frac{1}{x-1} = \frac{nx-1}{2x(x-1)}$$

which is > 0 if $0 < x < \frac{1}{n}$. Hence, $g(1/(n+2)) < g(1/n)$, $f(n+2) < f(n)$, and

$$J' = J \cdot (n-1)B\left(\frac{1}{2}, \frac{n-1}{2}\right) = 2(f(n+2)-f(n)) < 0.$$

Therefore, $p(1,n) = p(1,n+2) + J < p(1,n+2)$ for any $n \geq 2$. □

Corollary 5. *If $n \geq 2$, then $p(1,n) < P(\frac{1}{2}, \frac{1}{2})$.*

Proof. With $p(1,2) = \frac{1}{2}$ and $p(1,3) = \frac{1}{\sqrt{3}} \approx 0.5773$ and the known result $p(1,n) \to P(\frac{1}{2}, \frac{1}{2})$ $(n \to \infty)$, we obtain that $p(1,n) < P(\frac{1}{2}, \frac{1}{2})$ for $n \geq 2$. □

Lemma 12. *$p(2,n) < P(\frac{1}{2}, \frac{1}{2})$ for any $n \geq 2$*

Proof. By definition,

$$\begin{aligned} p(2,n) &= \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma(1)\Gamma\left(\frac{n}{2}-1\right)} \int_0^{\frac{2}{n}} (1-z)^{\frac{n-4}{2}} dz = \frac{n-2}{2} \int_0^{\frac{2}{n}} (1-z)^{\frac{n-4}{2}} dz \\ &= 1 - \left(1-\frac{2}{n}\right)^{\frac{n}{2}-1}. \end{aligned}$$

For $2 \leq n \leq 8$, we can check it is smaller than 0.68 numerically, Also, for $n \geq 9$, since the function $(1-1/x)^x$ is monotonically increasing with x, we have

$$1 - \left(1-\frac{2}{n}\right)^{\frac{n}{2}-1} < 1 - \left(1-\frac{2}{n}\right)^{\frac{n}{2}} \leq 1 - (1-2/9)^{9/2} < 0.68 < P\left(\frac{1}{2}, \frac{1}{2}\right).$$

□

Lemma 13. *$p(k+2,n) < p(k,n)$ for any $1 \le k < n$.*

Proof. By definition and (20)

$$\begin{aligned}
&p(k+2,n)\\
&= I_{\frac{k+2}{n}}\left(\frac{k}{2}+1, \frac{n-k}{2}-1\right) = I_{\frac{k+2}{n}}\left(\frac{k}{2}, \frac{n-k}{2}\right) - \frac{2}{k}\frac{(\frac{k+2}{n})^{\frac{k}{2}}(\frac{n-k-2}{n})^{\frac{n-k-2}{2}}}{B(\frac{k}{2},\frac{n-k}{2})}\\
&= p(k,n) + \frac{\int_{\frac{k}{n}}^{\frac{k+2}{n}} z^{\frac{k}{2}-1}(1-z)^{\frac{n-k}{2}-1}dz}{B(\frac{k}{2},\frac{n-k}{2})} - \frac{2}{k}\frac{(\frac{k+2}{n})^{\frac{k}{2}}(\frac{n-k-2}{n})^{\frac{n-k-2}{2}}}{B(\frac{k}{2},\frac{n-k}{2})}.
\end{aligned}$$

Thus, it suffices to show

$$\int_{\frac{k}{n}}^{\frac{k+2}{n}} z^{\frac{k}{2}-1}(1-z)^{\frac{n-k}{2}-1}dz - \frac{2}{k}\left(\frac{k+2}{n}\right)^{\frac{k}{2}}\left(\frac{n-k-2}{n}\right)^{\frac{n-k-2}{2}} := I - J < 0.$$

Let us define $g(z) = z^{\frac{k}{2}+1}(1-z)^{\frac{n-k}{2}-1}$ which is strictly increasing over $[0, \frac{k+2}{n}]$. Since z^{-2} is strictly decreasing, Chebyshev's integral inequality implies that

$$\begin{aligned}
I &= \int_{\frac{k}{n}}^{\frac{k+2}{n}} z^{-2}g(z)dz < \frac{n}{2}\int_{\frac{k}{n}}^{\frac{k+2}{n}} z^{-2}dz \int_{\frac{k}{n}}^{\frac{k+2}{n}} g(z)dz = \frac{n^2}{k(k+2)}\int_{\frac{k}{n}}^{\frac{k+2}{n}} g(z)dz\\
&< \frac{n^2}{k(k+2)}\cdot\frac{2}{n}g\left(\frac{k+2}{n}\right) = J.
\end{aligned}$$

Therefore, we have $I < J$ and it derives $p(k+2,n) < p(k,n)$. □

A.3 Proof of Lemma 9

Recall that $P(a,z) := \frac{\int_0^z x^{a-1}e^{-x}dx}{\Gamma(a)}$ which implies:

$$\frac{e^{-z}z^a}{\Gamma(a+1)} < P(a,z) < \frac{z^a}{\Gamma(a+1)} \quad \text{for } a > 0, z > 0. \tag{22}$$

Also, they imply the bound

$$P^{-1}(a,x) > (\Gamma(a+1)x)^{1/a} \text{ for } a > 0, 0 < x < 1 \tag{23}$$

and the limits

$$\lim_{z\to 0+}\frac{P(a,z)}{z^a} = \frac{1}{\Gamma(a+1)} \text{ and } \lim_{x\to 0+}\frac{x}{(P^{-1}(a,x))^a} = \frac{1}{\Gamma(a+1)}. \tag{24}$$

Hence, we have

$$\begin{aligned}
&\lim_{m\to\infty} mV_k(\sqrt{\beta_k}) = V_k(1)\lim_{m\to\infty} m\cdot\left(2\sigma^2 P^{-1}(k/2,\beta)\right)^{k/2}\\
&= V_k(1)\cdot\beta'\cdot\Gamma(k/2+1)(2\sigma^2)^{k/2} = \beta'\cdot(2\pi\sigma^2)^{k/2}.
\end{aligned}$$

To show the decreasing property, it suffices to show that $g(y) = \frac{1}{y} \cdot (P^{-1}(k/2, \beta' y))^{k/2}$ is strictly increasing over $0 < y \leq 1$.

We use the inequality

$$\frac{\partial}{\partial x} P^{-1}(a,x) = \Gamma(a) e^{P^{-1}(a,x)} P^{-1}(a,x)^{1-a} \geq \frac{P^{-1}(a,x)}{ax}$$

which is immediate from the left hand side of (22) with $z = P^{-1}(a,x)$.

Hence, denoting $P := P^{-1}(k/2, \beta' y)$ for simplicity,

$$g'(y) = \frac{\beta' k}{2y} P^{k/2-1} \cdot \frac{\partial P}{\partial y} - \frac{P^{k/2}}{y^2} > \frac{\beta' k}{2y} P^{k/2-1} \cdot \frac{P}{(k/2)\beta' y} P^{k/2-1} - \frac{P^{k/2}}{y^2} = 0$$

This completes the proof. □