Solving the Shortest Vector Problem in $2^{0.6039n}$ Time via Mid-point Hessian

Minki Hhan

KAIST, Daejeon, Korea
minkihhan@kaist.ac.kr

Abstract

We present randomized algorithms for the shortest vector problem (SVP). For the n-dimensional lattice $\mathcal{L}$, our algorithms solve SVP in time $2^{0.6039n+o(n)}$ classically and $2^{0.5411n+o(n)}$ quantumly and space $2^{0.5n+o(n)}$, improving the previous best algorithm running in $2^{n+o(n)}$ time and space of Aggarwal, Dadush, Regev, and Stephens-Davidowitz [STOC'15].

Our algorithms heavily use the property of the Hessian of the periodic Gaussian function at the half shortest vector: For a shortest vector $v \in \mathcal{L}$, the Hessian at $v/2$ has the eigenvector close to v, which can be used to recover v using the (preprocessing) bounded distance decoding algorithm. Given the periodicity modulo $\mathcal{L}$, the candidate midpoints are indexed by the parity classes in $\mathcal{L}/2\mathcal{L}$. Our algorithm searches for the class of a shortest vector by estimating the corresponding Hessians using discrete Gaussian samples.

We optimize the algorithm using random sublattice cosets and various sampling technique, achieving the final complexity. The optimization techniques may be of independent interest.

AI use disclosure. Most results of this paper are discovered with the assistance of ChatGPT 5.5 Pro and ChatGPT 5.6 Sol Ultra. While the author suggested several directions and optimizations, all the technical details are discovered by AI, and verified by the author. The manuscript is written by the author based on the initial draft generated by AI. The author takes full responsibility for the content. All errors in the manuscript are likely due to the author's human errors.

Contents

1 Introduction **3**
- 1.1 Technical overview . . . 4

2 Preliminaries **6**
- 2.1 Lattices and decoding . . . 6
- 2.2 Discrete Gaussians . . . 7
- 2.3 Useful inequalities . . . 9

3 Recovering Shortest Vector from Hessian **9**
- 3.1 Scale and Gaussian mass bounds . . . 9
- 3.2 The periodic Gaussian and the shortest parity class . . . 10
- 3.3 Estimating the Hessians and recovering a shortest vector . . . 12

4 Batch Hessian Estimation **14**

5 Random Sublattice Coset Hessian **15**
- 5.1 Gaussian and Hessian sums . . . 15
- 5.2 A random coset Hessian . . . 16
- 5.3 Estimating the coset Hessian . . . 17
- 5.4 Coset Hessians and shortest vectors . . . 18
- 5.5 The affine-coset algorithm . . . 20

6 Importance Sampling on an Affine Lattice Coset **21**
- 6.1 Lattice points in a shortest parity class . . . 21
- 6.2 Sampling (wider) discrete Gaussian from an affine lattice coset . . . 24
- 6.3 Estimating the Hessian by importance sampling . . . 25
- 6.4 Recovering a shortest vector . . . 28
- 6.5 The importance-sampling algorithm . . . 28
- 6.6 Reducing the space by sparsification . . . 29
- 6.7 Space-efficient implementation . . . 31
- 6.8 Quantum algorithm . . . 32

1 Introduction

An n-dimensional lattice $\mathcal{L} = \mathcal{L}(b_1, \ldots, b_n)$ is the set of all integer combinations of linearly independent vectors $b_1, \ldots, b_k \in \mathbb{R}^n$. For the sake of simplicity, we only consider the full rank lattice satisfies $k = n$. The (search) shortest vector problem (SVP) asks to find a nonzero vector of minimum norm, given a basis of $\mathcal{L}$.

There are two faces of lattice problems including SVP. Approximate variants of SVP have long served as algorithmic tools in computational number theory, integer programming, and cryptanalysis [25, 21, 26, 39, 24]. On the other hand, the exact problem and its small approximation factor version are known to be hard [7, 11, 4, 22, 19]. Given the close relation between the lattice problems and many recent cryptographic primitives [37, 12, 29, 36, 8, 17], lattice problems have become one of the most promising foundations for post-quantum cryptography.

SVP remains a fundamental benchmark for our understanding of the complexity of lattice problems. Researchers have discovered various worst-case algorithms [18, 30, 9, 21, 32, 35, 31, 13] and heuristic or average-case algorithms [32, 34, 10]. To date, however, the best provable worst-case running time was $2^{n+o(n)}$, with the same space complexity [2] using a new efficient discrete Gaussian sampling algorithm. Later, [1] gave several time-space tradeoffs and quantum speedup, including $2^{1.669n+o(n)}$ time and $2^{n/2+o(n)}$ space and $2^{0.950n+o(n)}$ and $2^{0.835n+o(n)}$ quantum time without and with quantum random access memory (QRAM), respectively.

We give a new class of worst-case algorithms for SVP, significantly improving the previous classical and quantum algorithms [2, 1]. Our main results are summarized as follows.

Theorem 1.1. *There are randomized classical and quantum algorithms that solve Search-*SVP *with probability at least* $2/3$ *in expected time* $2^{0.603867n+o(n)}$ *classically and* $2^{0.541051n+o(n)}$ *quantumly and space* $2^{0.5n+o(n)}$*, up to factors polynomial in the input length. The quantum algorithm requires a QRAM of size* $2^{0.36036n+o(n)}$*.*

Remark 1.2 (Concurrent works). During the preparation of this manuscript, OpenAI announced an improvement of the sphere packing bound [33]. Plugging that parameter, the time complexity of our algorithm is reduced to $2^{0.60221n+o(n)}$ classically and $2^{0.53925n+o(n)}$ quantumly, according to ChatGPT 5.6 Sol's calculation. We did not present these complexities because the result of OpenAI and the numerical calculation for those parameters are not verified (by the author, at least). We also noticed the concurrent works [16, 23] right before publicizing the manuscript, which obtains $2^{(0.5+\frac{\beta^2}{4e\ln 2})n+o(n)} \approx 2^{0.7134n+o(n)}$ time and $2^{0.5n+o(n)}$ space algorithm for $\beta = 2^{0.4014\cdots}$, matching Theorem 5.8, with a completely different approach.

We discuss some direct applications of Theorem 1.1. Obviously, our algorithm solves the approximate or unique SVP as well as its gap version in the same time and space complexity. Our algorithm beats the previous best time records [41, 3, 27] $2^{0.802n+o(n)}$ time and $2^{0.401n+o(n)}$ space for the approximate SVP with a constant factor, although they use less space.

This improvement can be readily used in other applications. For example, the best-known polynomial factor approximate SVP [3, Theorems 5.3] uses one call to a constant-factor SVP oracle in a small dimension. Originally the proof uses the oracle from [27] with an exponent $0.802n$. Replacing this to ours, we improve the time complexity for solving $\widetilde{O}(n^c)$-SVP from $2^{\frac{n}{2c+1.24}}$ to $2^{\frac{n}{2c+1.65}}$.

Another direct application is BDD or exact CVP with the distance promise. Using Kannan's embedding, CVP for $(t, \mathcal{L})$ with the promise $\mathrm{dist}(t, \mathcal{L}) < \sqrt{3}\lambda_1(\mathcal{L})/2$ can be reduced to SVP with one additional dimension (see e.g., [28]). Thus BDD and the CVP with the above promise can be solved in the same time and space using our algorithm. Previously, [14, 2] give $2^{n/2+o(n)}$ time algorithm for $\alpha < 0.422$.

Lastly, using the dimension-preserving reduction from centered DGS to SVP [40], we can sample one (or polynomially many) discrete Gaussian sample with an arbitrary parameter in the same time. This partly answers the open question in [2] for the centered discrete Gaussian sampling below smoothing, albeit with a worse time complexity than $2^{n/2+o(n)}$. A concurrent work [23] gives a $2^{n/2+o(n)}$ time discrete Gaussian sampling algorithm for arbitrary parameters.

1.1 Technical overview

Let $\mathcal{L}$ be a n-dimensional full-rank lattice with basis B and $s > 0$ be a width parameter. Its dual lattice $\mathcal{L}^* = \{y : \langle y, x\rangle \in \mathbb{Z} \text{ for every } x \in \mathcal{L}\}$ has the basis B^{-T}. Define $\rho_s(x) := \exp\left(-\pi \frac{\|x\|^2}{s^2}\right)$ and let $\rho_s(A) := \sum_{x\in A} \rho_s(x)$. We define the (centered) discrete Gaussian distribution $D_{\mathcal{L},s}(x) := \frac{\rho_s(x)}{\rho_s(\mathcal{L})}$. The periodic Gaussian function [5] $F_s : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$F_s(z) := \frac{\rho_s(\mathcal{L}+z)}{\rho_s(\mathcal{L})} = \mathbb{E}_{X\sim D_{\mathcal{L}^*,1/s}}\left[e^{2\pi i\langle X,z\rangle}\right]$$

where the last equality is due to the Poisson summation formula. It is clear that F_s is $\mathcal{L}$-periodic, i.e., $F_s(z) = F_s(z+v)$ for $v \in \mathcal{L}$.

At the midpoint, the Hessian points the way. Our algorithm starts with studying its Hessian [38, 14], which also has two representations

$$\nabla^2 F_s(z) + \frac{2\pi}{s^2} F_s(z) I_n = \frac{4\pi^2}{s^4 \rho_s(\mathcal{L})} \sum_{y\in\mathcal{L}} (y-z)(y-z)^T e^{-\pi\|y-z\|^2/s^2} \tag{1}$$

by differentiating the first representation, and by differentiating the second representation

$$\nabla^2 F_s(z) = -4\pi^2 \mathbb{E}_{X\sim D_{\mathcal{L}^*,1/s}}\left[XX^T e^{2\pi i\langle X,z\rangle}\right]. \tag{2}$$

From these two representations, we observe the following two properties. First, let v be a shortest vector with $\|v\| = \lambda$ and evaluate $\nabla^2 F_s(v/2)$ in Eq. (1), we can see that

$$\nabla^2 F_s(v/2) + aI = \frac{4\pi^2}{s^4\rho_s(\mathcal{L})}\left(\frac{vv^T}{2} e^{-\pi\lambda^2/4s^2} + \sum_{y\in\mathcal{L}\setminus\{0,v\}} (y-v/2)(y-v/2)^T e^{-\pi\|y-v/2\|^2/s^2}\right)$$

where $a := 2\pi F_s(v/2)/s^2$ which is unimportant in the spectral analysis. Noting that $\|y - v/2\|$ for $y \neq 0, v$ is much larger than $\|v/2\| = \lambda/2$, we can expect that $\nabla^2 F_s(v/2) + aI$ is close to a multiple of vv^T. For appropriately chosen s, this intuition can be formalized by the following statement for the normalized eigenvector $\tilde{v}$ of $\nabla^2 F_s(v/2)$ corresponding to the largest eigenvalue:

$$\tilde{v} \text{ is } \textit{inverse polynomially close} \text{ to } v/\|v\|.$$

Then, given $\tilde{v}$, we guess λ (which only takes polynomial iterations) and solve the bounded distance decoding (BDD) problem for the target vector $\lambda\tilde{v}$. This BDD step can be relatively quick, thanks to the preprocessing BDD algorithm in [1], which solves each BDD instance with inverse polynomial distances in time $2^{o(n)}$ after $2^{0.5n+o(n)}$ time preprocessing.

One Hessian is enough if we know where to look. This gives our very first result (Theorem 3.7) with time complexity $2^{1.463n+o(n)}$ and space complexity $2^{0.5n+o(n)}$. Suppose that for a shortest vector v, we know a representative w of the parity class $v + 2\mathcal{L}$. We *compute* the eigenvector $\tilde{w}$ of $\nabla^2 F_s(w/2)$, and solve the BDD problem for it using (preprocessing) BDD. But, how to compute $\nabla^2 F_s(w/2)$?

Here the second representation Eq. (2) comes in. We use the algorithm in [2] that samples $N = 2^{n/2}$ elements, say $X_1, ..., X_N$, from the discrete Gaussian distribution in time $2^{n/2+o(n)}$ on $\mathcal{L}^*$. Then, given $w/2$, we use the estimator

$$\widehat{\nabla^2 F_s}(w/2) = \frac{-4\pi^2}{N} \sum_{i=1}^{N} X_i X_i^T e^{2\pi i\langle X_i, w/2\rangle}.$$

Concentration inequalities and results from lattice literature show that about $2^{2t_0 n+o(n)}$ samples for $t_0 = 2^{0.802}/4e\ln 2 = 0.2314$ suffice to accurately estimate the actual $\nabla^2 F_s(w/2)$ as well as its eigenvectors. This leads to the time complexity $|\mathcal{L}/2\mathcal{L}| \cdot 2^{2t_0 n} = 2^{1.463n+o(n)}$ by enumerating all the parity class in $\mathcal{L}/2\mathcal{L}$. The $2^{n/2+o(n)}$ space is required in the discrete Gaussian sampling.

Walsh-Hadamard looks everywhere at once. The time can be reduced further. Write $w = Bu$. Since $B^T X \in \mathbb{Z}^n$,

$$e^{2\pi i\langle X, Bu/2\rangle} = e^{\pi i u^T B^T X} = (-1)^{u^T (B^T X \bmod 2)}.$$

With some work, the fast Walsh-Hadamard transform can be applied to compute $\widehat{\nabla^2 F_s}(Bu/2)$ simultaneously for different u's. Carefully choosing the coordinates leads to $2^{n+o(n)}$ time algorithm in Theorem 4.2 while preserving $2^{0.5n+o(n)}$ space.

A random coset narrows the view. The bottleneck 2^n of the above strategy is the size of the search space $\mathcal{L}/2\mathcal{L}$. Let $X \in \mathcal{L}^*$. Choose a random invertible linear map $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and divide

$$P(B^T X \bmod 2) \in \mathbb{F}_2^n$$

into h fixed coordinates and $\ell = n - h$ remaining coordinates. We then choose a random value $j \in \mathbb{F}_2^h$ for the first h coordinates, and use the discrete Gaussian samples X that agree to this value. These samples belong to an affine coset of an index-2^h sublattice of $\mathcal{L}^*$. It turns out that the discrete Gaussian samples (for an appropriate width) are almost uniformly distributed over different cosets, thus this collective sampling adds 2^h multiplicative factor in the time complexity in sampling. That is, the estimation can be anyway done.

At first sight, it may seem that choosing one random coset could lose the spectral property of Hessian. This is not the case. Conditioning the first h coordinates of the samples on the dual lattice side corresponds to *summing* over the corresponding h coordinates of u as shown in Eq. (21). Thus, the shortest-vector contribution still appears in one of the 2^ℓ coset Hessians! Given this intuition, we can indeed apply almost the same strategy to find a shortest vector. The overall time complexity becomes

$$2^{2t_0 n + h + o(n)} + 2^{\ell + o(n)}$$

where the first term is to sample $2^{2t_0 n}$ discrete Gaussian samples in the target coset (as we get roughly one coset Gaussian sample out of 2^h lattice Gaussian samples), and 2^ℓ is the number of outputs to be checked. Balancing these costs gives $2^{0.73147n+o(n)}$ algorithm as shown in Theorem 5.8.

May the fourth weigh what matters. The main loss in the above algorithm is due to the inefficient rejection step in the discrete Gaussian sampling over the sublattice coset. The last algorithm uses the following observation: For two probabilistic distributions D and E, where we know how to sample only from E and know their probabilistic density functions p and q, then we have

$$\mathbb{E}_{X\sim D}[f(X)] = \sum_X p(X) f(X) = \sum_X q(X) \cdot \frac{p(X)}{q(X)} f(X) = \mathbb{E}_{X\sim E}\left[\frac{p(X)}{q(X)} f(X)\right]$$

so that we can estimate the statistics regarding D only using samples from E. We use this idea.

We first show that for a slightly wider width s_R than original, we can sample discrete Gaussian directly on the target sublattice coset Λ. Of course, samples from the discrete Gaussian with a wider width s_R do not have exactly the distribution required for Hessian estimation. We therefore use the above idea and assign each sample X the importance weight for the narrower target width $s_r < s_R$

$$w(X) := \frac{\rho_{s_r}(X)}{\rho_{s_R}(X)} = \frac{\Pr_{D_{\Lambda,s_r}}[X]\rho_{s_r}(\Lambda)}{\Pr_{D_{\Lambda,s_R}}[X]\rho_{s_R}(\Lambda)} \propto \frac{\Pr_{D_{\Lambda,s_r}}[X]}{\Pr_{D_{\Lambda,s_R}}[X]},$$

which is easily computable. Similarly to the above, we have

$$\mathbb{E}_{X\sim D_{\Lambda,s_r}}[f(X)] = \frac{\mathbb{E}_{X\sim D_{\Lambda,s_R}}[w(X)f(X)]}{\mathbb{E}_{X\sim D_{\Lambda,s_R}}[w(X)]}$$

for a function f. Thus the wider Gaussian with width R is used only to generate the samples, while the weights recover the Hessian corresponding to the narrower target Gaussian with width r.

The loss caused by importance sampling is $\iota = \frac{1}{2}\log_2\left(\frac{R^2}{r(2R-r)}\right)$ in the exponent of the variance (Lemma 6.7), forcing us to sample $2^{\iota n}$ more samples. This turns $2^{2t_0 n+h}$ time in the previous algorithm to $2^{(2r+\iota)n}$, and balancing the parameters gives $2^{0.60387n+o(n)}$ time complexity in Theorem 6.11.[1]

Sparsity makes space for quantum. A careful reader may notice that the above algorithm requires $2^{(2r+\iota)n}$ Gaussian samples, thus its space complexity is also $2^{0.60387n+o(n)}$. We observe that the value $w(X)$ heavily depends on the norm of X, and in particular for most of X it is very close to 0. We can sparsify the Gaussian samples by employing a binary variable Z such that, conditioned on each sample X, choose $Z \in \{0,1\}$ such that $\Pr[Z=1 \mid X] = \pi(X)$ and $\Pr[Z=0 \mid X] = 1-\pi(X)$ for some $\pi(X) \propto w(X)$. We use the following: For every function f,

$$\mathbb{E}_{X,Z}\left[\frac{Zf(X)}{\pi(X)}\right] = \mathbb{E}_X\left[\mathbb{E}\left[\frac{Zf(X)}{\pi(X)}\middle| X\right]\right] = \mathbb{E}_X[f(X)]$$

thus we can compute the statistics with a smaller number of X's for which $Z=1$, which is about $\mathbb{E}[\pi(X)] \cdot 2^{(2r+\iota)n}$. This gives a $2^{0.5n+o(n)}$ space version of the above algorithm, showing Theorem 6.14. Note that we anyway need to sample $2^{(2r+\iota)n}$ Gaussian samples, thus the time complexity remains unchanged.

This reduced space actually allows us to use the quantum minimum finding algorithm [15]. The number of effective samples can be reduced to 2^{2rn} while ensuring the accuracy of the estimation at best. In this case, we choose $b \approx 2rn$ and write $\mathbb{F}_2^\ell = \mathbb{F}_2^b \times \mathbb{F}_2^{\ell-b}$. For each fixed value of the last $\ell - b$ coordinates, the fast Walsh-Hadamard transform computes the corresponding $2^b = 2^{2rn+o(n)}$ coset Hessians at once. Classically this is repeated $2^{\ell-b}$ times, while quantum minimum finding reduces the number of repetitions to $2^{(\ell-b)/2}$, which is $2^{\iota n+o(n)}$ for the optimized parameters. Balancing the parameter gives $2^{0.5411n+o(n)}$ quantum time and $2^{0.5n+o(n)}$ space in Theorem 6.15. The QRAM only stores the BDD preprocessing data and 2^{2rn} samples, which is $2^{0.3604n+o(n)}$ in the balanced parameters.

Acknowledgment. The author is grateful to Yixin Shen for helpful discussions about an early version of this work, and to Jiseung Kim for kindly sharing a preliminary draft of his ongoing work [23].

2 Preliminaries

The sets of integers, positive integers, and real numbers are denoted by $\mathbb{Z}$, $\mathbb{N}$, and $\mathbb{R}$. $\mathbb{F}_2$ denotes the field of two elements $\{0,1\}$. We write $[m] = \{1,\dots,m\}$ for $m \in \mathbb{N}$. The Euclidean, operator, and Frobenius norms are denoted by $\|\cdot\|$, $\|\cdot\|_{\mathrm{op}}$, and $\|\cdot\|_{\mathrm{F}}$, respectively. The closed Euclidean ball of radius R is $B_2^n(R) = \{x \in \mathbb{R}^n : \|x\| \le R\}$. For an event E, its indicator function is $\mathbf{1}_E$. All logarithms are natural unless a base is specified. We write $\mathrm{dist}(t,S) := \inf_{x\in S}\|t-x\|$ for a set S and a vector t.

The dimension n tends to infinity when discussing the asymptotic behavior. Unless specified otherwise, every $o(1)$ term is uniform over the lattices and the fixed parameters in the stated ranges. A negligible function, denoted by $\mathrm{negl}(n)$, is smaller than n^{-c} for every $c > 0$ for all sufficiently large n. We write $\langle B \rangle$ for the bit length of the input basis B and assume that it is polynomial in n.

2.1 Lattices and decoding

We only consider the full-rank lattice defined as follows.[2] Let $B = (b_1, ..., b_n) \in \mathbb{R}^{n\times n}$ be a matrix with independent columns. A full-rank lattice generated by B is a set $\mathcal{L} = \mathcal{L}(B) := B\mathbb{Z}^n$, i.e., integer linear combinations of column vectors of B. We define its dual lattice by

$$\mathcal{L}^* := \{y \in \mathrm{span}(\mathcal{L}) : \langle y, x\rangle \in \mathbb{Z} \text{ for every } x \in \mathcal{L}\}$$

[1] Note that since we do not even sample from the discrete Gaussian with width r, we can set r smaller than the limit of the original discrete Gaussian sampler [2]. We actually choose so, and extend several lemmas for $r < t_0$.

[2] It is well-known that the full-rank case implies to the general case.

which also satisfies $\mathcal{L}^* = B^{-T}\mathbb{Z}^n$ when $\mathcal{L}$ is full-rank.

A lattice $\mathcal{L}'$ is called a sublattice of $\mathcal{L}$ if $\mathcal{L}' \subseteq \mathcal{L}$. If $\mathcal{L}'$ is a full-rank sublattice, $\mathcal{L}/\mathcal{L}'$ is the finite abelian group of cosets $x + \mathcal{L}'$. In particular, it holds that

$$\mathcal{L}/2\mathcal{L} \cong \mathbb{F}_2^n, \qquad \mathcal{L}^*/2\mathcal{L}^* \cong \mathbb{F}_2^n.$$

For cosets $a + 2\mathcal{L}$ and $x + 2\mathcal{L}^*$, we can write $a = Bu \in \mathcal{L}$ and $x = B^{-T}z \in \mathcal{L}^*$ for $u, z \in \mathbb{F}_2^n$. We call u is the parity class. The following is well-defined

$$\chi_a(x) := (-1)^{\langle x,a\rangle} = (-1)^{z^T u}$$

and independent of the representatives. In the remainder of the paper, we use arbitrary representative of the coset which does not change any results.

The first minimum $\lambda = \lambda_1(\mathcal{L})$ is $\lambda_1(\mathcal{L}) := \min_{x\in\mathcal{L}\setminus\{0\}} \|x\|$. For $R \geq 0$, define the counting function for the lattice points inside a ball $N_{\mathcal{L}}(R) := |(\mathcal{L} \setminus \{0\}) \cap B_2^n(R)|$. We use the following bound [35, Lemma 3], based on the spherical-code bound of [20].

Lemma 2.1. *Uniformly for full-rank n-dimensional lattices $\mathcal{L}$ and $x \geq 1$, it holds for $\beta = 2^{0.4014\ldots}$*

$$N_{\mathcal{L}}(x\lambda_1(\mathcal{L})) \leq \beta^{n+o(n)}x^n.$$

Definition 2.2 (Search-SVP)**.** *Given a basis of a full-rank lattice $\mathcal{L}$, Search-*SVP *asks to find a nonzero $v \in \mathcal{L}$ with $\|v\| = \lambda_1(\mathcal{L})$.*

Definition 2.3 (BDD)**.** *For $0 < \alpha < 1/2$, α-*BDD *takes a lattice $\mathcal{L}$ and a target vector t promised to satisfy $\mathrm{dist}(t, \mathcal{L}) < \alpha\lambda_1(\mathcal{L})$ as input, and returns the unique closest lattice vector to t.*

We use the following result for preprocessing BDD from [1].

Theorem 2.4. *There is a randomized classical preprocessing algorithm for the $n^{-1/3}$-*BDD *problem such that:*

- *its preprocessing runs in expected time and worst-case space $2^{n/2+o(n)}$ and succeeds with a constant probability. After running, the size of advice is $2^{o(n)}$.*
- *given the successful preprocessing output, any instance of the $n^{-1/3}$-*BDD *problem can be deterministically solved in time $2^{o(n)}$ and polynomial additional space.*

Proof. Plug $\varepsilon = \exp(-\sqrt{n})$ in [1, Theorem 56] with $\alpha = 1/2$, which gives $(\phi(\mathcal{L})/\lambda_1(\mathcal{L}))$-BDD oracle in $m\,\mathrm{poly}(n)$ time and space after the DGS preprocessing $\phi(\mathcal{L}) = \frac{\sqrt{\ln(1/\varepsilon)/\pi - o(1)}}{2\eta_\varepsilon(\mathcal{L}^*)}$ and $m = O(n\ln(1/\varepsilon)/\sqrt{\varepsilon}) = 2^{o(n)}$. An asymptotic calculation gives $\phi(\mathcal{L})/\lambda_1(\mathcal{L}) = \Omega(n^{-1/4})$ using $\lambda_1(\mathcal{L})\eta_\varepsilon(\mathcal{L}^*) \leq \sqrt{\frac{\ln((1+\varepsilon)/\varepsilon)}{\pi}} + \sqrt{\frac{n}{2\pi}}$. This makes the query in $m\,\mathrm{poly}(n) = 2^{o(n)}$ time and space. The preprocessing is only the discrete Gaussian samples, which can be sampled in time $2^{n/2+o(n)}$ and space $2^{n/2+o(n)}$. □

We consider a time-truncated variant that returns $\perp$ and aborts if the time is exceeded. For an input that does not meet the promise of $n^{-1/3}$-BDD, there is no correctness guarantee; we use an output only after exact lattice-membership verification. Whenever an algorithm is repeated, every trial uses independent preprocessing and fresh randomness, and we return the shortest exactly verified nonzero vector over all trials. Every DGS call is time-truncated at its stated budget. If the call exceeds the budget or returns too few samples, the current scale is aborted.

2.2 Discrete Gaussians

For $s > 0$, $x \in \mathbb{R}^n$, and a countable set A, define

$$\rho_s(x) := \exp\left(-\pi\frac{\|x\|^2}{s^2}\right), \qquad \rho_s(A) := \sum_{x\in A}\rho_s(x)$$

For a full-rank lattice $\mathcal{L} \subset \mathbb{R}^n$ and a vector $t \in \mathbb{R}^n$, the discrete Gaussian on a lattice coset $\mathcal{L}+t$ is

$$D_{\mathcal{L}+t,s}(x) := \frac{\rho_s(x)}{\rho_s(\mathcal{L}+t)}, \qquad x \in \mathcal{L}+t.$$

In particular, $D_{\mathcal{L},s}$ denotes the central case $t=0$.

Definition 2.5 (Smoothing parameter [29, 37])**.** *For $\varepsilon \in (0,1)$, the smoothing parameter is the unique positive number $\eta_\varepsilon(\mathcal{L})$ satisfying $\rho_{1/\eta_\varepsilon(\mathcal{L})}(\mathcal{L}^* \setminus \{0\}) = \varepsilon$.*

Lemma 2.6 ([2, Lemma 6.1])**.** *Let β be the constant in Lemma 2.1. For any full-rank lattice $\mathcal{L} \subset \mathbb{R}^n$, it holds that*

$$\eta_{1/2}(\mathcal{L}^*) \le \left(\frac{\beta}{\sqrt{2\pi e}} + o(1)\right) \frac{\sqrt{n}}{\lambda_1(\mathcal{L})}.$$

Theorem 2.7 ([2, Theorem 5.11])**.** *Let $\mathcal{L} \subset \mathbb{R}^n$ be a full-rank lattice, $s > 0$, and $\kappa = \Omega(n)$. There is a classical algorithm DGS that outputs at most $M = 2^{n/2}$ lattice vectors in time $2^{n/2+\mathrm{polylog}(\kappa)+o(n)}$ and space $2^{n/2+o(n)}$. Its output distribution is $\exp(-\Omega(\kappa))$-close to the following distribution: choose $M' \in \{0,\dots,M\}$, draw $X_1,\dots,X_M$ independently from $D_{\mathcal{L},s}$ and independently of M', and output $X_1,\dots,X_{M'}$. If $s > \sqrt{2}\eta_{1/2}(\mathcal{L})$, then $M' = M$.*

Lemma 2.8. *For every $c > 0$, there is a constant $C_c > 0$ such that, for every lattice $\mathcal{L}$, every $s > 0$, and $X \sim D_{\mathcal{L},s}$,*

$$\mathbb{P}[\|X\| > C_c s\sqrt{n}] \le 2^{-2cn} \qquad \text{and} \qquad \mathbb{E}\left[\|X\|^2 \mathbf{1}_{\{\|X\| > C_c s\sqrt{n}\}}\right] \le s^2 n\, 2^{-cn}.$$

Proof. The discrete-Gaussian tail inequality [2, Lemma 2.4] gives, for all sufficiently large constants t,

$$\mathbb{P}[\|X\| > ts\sqrt{n}] \le \left(\sqrt{2\pi e} \cdot te^{-\pi t^2}\right)^n.$$

Choose $t = C_c$ sufficiently large (which only depends on c) so that

$$\sqrt{2\pi e} \cdot te^{-\pi t^2} \le e^{\frac{-\pi t^2}{2}} \le 2^{-2c}$$

holds, which proves the first part. For the second part, the tail integration formula for $Y = \|X\|$ gives

$$\mathbb{E}[Y^2 \mathbf{1}_{\{Y>u\}}] = u^2 \mathbb{P}[Y > u] + \int_u^\infty 2z\mathbb{P}[Y > z]dz.$$

Taking $u = C_c s\sqrt{n}$ and $z = ts\sqrt{n}$, we have

$$\begin{aligned}
\mathbb{E}[Y^2 \mathbf{1}_{\{Y>C_c s\sqrt{n}\}}] &\le C_c^2 s^2 n 2^{-2cn} + s^2 n \int_{C_c}^\infty 2t \cdot \mathbb{P}[\|X\| > ts\sqrt{n}]dt \\
&\le C_c^2 s^2 n 2^{-2cn} + s^2 n \int_{C_c}^\infty 2t \cdot \exp\left(\frac{-n\pi t^2}{2}\right) dt \\
&= C_c^2 s^2 n 2^{-2cn} + s^2 n \cdot \frac{2e^{\frac{-\pi n C_c^2}{2}}}{\pi n} \\
&\le s^2 n \left(C_c^2 + \frac{2}{\pi n}\right) 2^{-2cn}
\end{aligned}$$

where the last term is smaller than $s^2 n 2^{-cn}$ for sufficiently large n. □

2.3 Useful inequalities

We use the following standard probability bounds.

Lemma 2.9 (Markov's inequality)**.** *For a random variable $Y \geq 0$ and $a > 0$, $\mathbb{P}[Y \geq a] \leq \mathbb{E}[Y]/a$.*

Lemma 2.10 (Chebyshev's inequality)**.** *If Y has finite variance and $a > 0$, $\mathbb{P}[|Y - \mathbb{E}[Y]| \geq a] \leq \mathrm{Var}(Y)/a^2$.*

Lemma 2.11 (Hoeffding's inequality)**.** *If $Y_1, \ldots, Y_N$ are independent and $a_i \leq Y_i \leq b_i$ and $\delta > 0$, $\mathbb{P}\left[\left|\frac{1}{N}\sum_{i=1}^N (Y_i - \mathbb{E}[Y_i])\right| \geq \delta\right] \leq 2\exp\left(-\frac{2N^2\delta^2}{\sum_{i=1}^N (b_i - a_i)^2}\right)$.*

Lemma 2.12 (Chernoff bound)**.** *If $Y_1, \ldots, Y_N$ are independent Bernoulli random variables and $0 < \delta \leq 1$, then*

$$\mathbb{P}\left[\left|\sum_{i=1}^N Y_i - \sum_{i=1}^N \mathbb{E}[Y_i]\right| \geq \delta \sum_{i=1}^N \mathbb{E}[Y_i]\right] \leq 2\exp\left(-\frac{\delta^2}{3}\sum_{i=1}^N \mathbb{E}[Y_i]\right).$$

Lemma 2.13. *If $E_1, \ldots, E_m$ are independent events with $\mathbb{P}[E_i] \leq p < 1/4$, then*

$$\mathbb{P}\left[\sum_{i=1}^m \mathbf{1}_{E_i} \geq \frac{m}{2}\right] \leq 2^m p^{m/2}.$$

3 Recovering Shortest Vector from Hessian

We assume that an approximate shortest-vector length d satisfying

$$\lambda_1(\mathcal{L}) \leq d \leq (1 + 1/n)\lambda_1(\mathcal{L}) \tag{3}$$

is known. This can be assumed by applying LLL [25] and guessing d from $\{(1+1/n)^{-j}\|x\|\}_{j=0,\ldots,n^2}$, where x is the resulting nonzero vector. We formalize this in the final algorithm.

It will be convenient to define the following parameters:

$$\xi_t = \xi_t(d) := \sqrt{\frac{4nt\ln 2}{\pi d^2}}, \qquad t_0 := \frac{\beta^2}{4e\ln 2} = 0.23147\ldots \tag{4}$$

These choices give the Gaussian-mass estimates in Corollary 3.2. It will be convenient to note that for most of our interested case the following holds for most choices of t:

$$\xi_t^2 = \Theta(n/d^2) = \Theta(n/\lambda^2). \tag{5}$$

3.1 Scale and Gaussian mass bounds

Lemma 3.1. *Let $\mathcal{L}$ be an n-dimensional full-rank lattice and let $\lambda = \lambda_1(\mathcal{L})$. Let $s = \Theta(\lambda/\sqrt{n})$ and let $k \geq 0$ be a constant. Then*

$$\sum_{x \in \mathcal{L}\setminus\{0\}} \left(1 + \frac{\|x\|}{\lambda}\right)^k \rho_s(x) \leq 2^{o(n)} \left(\frac{\beta^2 n s^2}{2\pi e \lambda^2}\right)^{n/2}.$$

Proof. The case $k = 0$ is essentially [2, Lemma 4.2]. The same shell argument gives

$$\sum_{x \in \mathcal{L}\setminus\{0\}} \left(1 + \frac{\|x\|}{\lambda}\right)^k \rho_s(x) \leq 2^k \beta^{n+o(n)} \max_{y \geq 0} y^{n+k} e^{-\pi\lambda^2 y^2/s^2} \leq \beta^{n+o(n)} \max_{y > 0} y^{n+k} e^{-\pi\lambda^2 y^2/s^2}.$$

Here we used $\|x\| \geq \lambda$ and hence

$$\left(1 + \frac{\|x\|}{\lambda}\right)^k \leq 2^k \left(\frac{\|x\|}{\lambda}\right)^k.$$

The maximum is attained at $y^2 = \frac{(n+k)s^2}{2\pi\lambda^2}$, and its value is $\left(\frac{(n+k)s^2}{2\pi e\lambda^2}\right)^{(n+k)/2}$. Since $s = \Theta(\lambda/\sqrt{n})$ and k is fixed, replacing $n+k$ by n changes the expression by only $2^{o(n)}$. The fixed factor 2^k and the subexponential shell-summation factor are absorbed into the same term. □

Corollary 3.2. *Let $\lambda = \lambda_1(\mathcal{L})$ and $\lambda \le d \le (1+1/n)\lambda$. For every fixed $a > 0$ and $k \ge 0$,*

$$\sum_{x\in\mathcal{L}\setminus\{0\}} \left(1 + \frac{\|x\|}{\lambda_1(\mathcal{L})}\right)^k \rho_{a/\xi_t}(x) \le 2^{\left(\frac{1}{2}\log_2 \frac{a^2 t_0}{2t} + o(1)\right)n}.$$

Consequently, for every fixed $a > 0$ and $k > 0$,

$$\sum_{x\in\mathcal{L}} \|x\|^k \rho_{a/\xi_t}(x) \le \lambda^k 2^{(\frac{1}{2}\log_2 \frac{a^2 t_0}{2t} + o(1))n} \tag{6}$$

In particular, the following hold:

- *For a (possibly negative) constant c, it holds that $\rho_{\sqrt{2^{1+c}}/\xi_t}(\mathcal{L}) \le 1 + 2^{(c/2 - \frac{1}{2}\log_2 \frac{t}{t_0} + o(1))n}$.*
- *If $t > t_0/2$, then $\rho_{1/\xi_t}(\mathcal{L}) = 1 + 2^{-\Omega(n)}$.*
- *If $t > t_0$, then $\xi_t > \sqrt{2}\,\eta_{1/2}(\mathcal{L}^*)$, and Theorem 2.7 applies to $\mathcal{L}^*$ at parameter ξ_t.*

Proof. For $s = a/\xi_t$, we have $\frac{\beta^2 n s^2}{2\pi e\lambda^2} = \frac{a^2 t_0}{2t}\frac{d^2}{\lambda^2}$. The first claim follows from Lemma 3.1, since $d/\lambda \le 1 + 1/n$. The second claim follows from the first because its term at $x = 0$ vanishes and $\|x\|^k \le \lambda^k(1 + \|x\|/\lambda)^k$ for $x \ne 0$. The first bound follows by taking $a = \sqrt{2^{c+1}}$ for given c and $k = 0$. Plugging $t > t_0/2$ gives the second bound. If $t > t_0$, then the bound gives

$$\rho_{\sqrt{2}/\xi_t}(\mathcal{L}\setminus\{0\}) < \frac{1}{2}$$

for all sufficiently large n. The definition and monotonicity of the smoothing parameter give the last claim. □

3.2 The periodic Gaussian and the shortest parity class

For $s > 0$, define the periodic Gaussian function [6, 14] by

$$F_s(z) := \frac{\rho_s(\mathcal{L} + z)}{\rho_s(\mathcal{L})}, \qquad z \in \mathbb{R}^n. \tag{7}$$

The following is the Poisson representation of F_s.

$$F_s(z) = \mathbb{E}_{X\sim D_{\mathcal{L}^*,1/s}}\left[e^{2\pi i\langle X, z\rangle}\right]. \tag{8}$$

For $u \in \mathbb{F}_2^n$, define the Hessian

$$\mathcal{G}_{t,d}(u) := \nabla^2 F_{1/\xi_t}(Bu/2).$$

We also define

$$\mathcal{A}_{t,d}(u) := \mathcal{G}_{t,d}(u) + 2\pi\xi_t^2 F_{1/\xi_t}(Bu/2) I_n. \tag{9}$$

For fixed u, $\mathcal{A}_{t,d}(u)$ and $\mathcal{G}_{t,d}(u)$ differ by $2\pi\xi_t^2 F_{1/\xi_t}(Bu/2)I_n$, a constant multiple of I_n. Hence the two matrices have the same eigenvectors and eigengaps. We will use the following alternative formula.

Lemma 3.3. *For every $u \in \mathbb{F}_2^n$,*

$$\mathcal{G}_{t,d}(u) = -4\pi^2 \mathbb{E}_{X\sim D_{\mathcal{L}^*,\xi_t}}\left[XX^T(-1)^{u^T(B^T X \bmod 2)}\right]. \tag{10}$$

Moreover, the following equality holds thus $\mathcal{A}_{t,d}(u)$ is positive semidefinite:

$$\mathcal{A}_{t,d}(u) = \frac{\pi^2\xi_t^4}{\rho_{1/\xi_t}(\mathcal{L})} \sum_{w\in Bu+2\mathcal{L}} ww^T e^{-\pi\xi_t^2\|w\|^2/4}. \tag{11}$$

Proof. Differentiating the expression in Eq. (8) gives

$$\nabla^2 F_s(z) = -4\pi^2 \mathbb{E}_{X \sim D_{\mathcal{L}^*, 1/s}} \left[X X^T e^{2\pi i \langle X, z \rangle} \right].$$

Set $s = 1/\xi_t$ and $z = Bu/2$. Since $B^T X \in \mathbb{Z}^n$,

$$e^{2\pi i \langle X, Bu/2 \rangle} = e^{\pi i u^T B^T X} = (-1)^{u^T (B^T X \bmod 2)}.$$

This proves the first identity.

Differentiating the expression in Eq. (7) and using $\mathcal{L} = -\mathcal{L}$ gives

$$\nabla^2 F_s(z) = \frac{1}{\rho_s(\mathcal{L})} \sum_{y \in \mathcal{L}} \left(\frac{4\pi^2}{s^4} (y-z)(y-z)^T - \frac{2\pi}{s^2} I_n \right) \cdot e^{-\pi \|y-z\|^2 / s^2}.$$

It follows that

$$\nabla^2 F_s(z) + \frac{2\pi}{s^2} F_s(z) I_n = \frac{4\pi^2}{s^4 \rho_s(\mathcal{L})} \sum_{y \in \mathcal{L}} (y-z)(y-z)^T e^{-\pi \|y-z\|^2 / s^2}.$$

Again set $s = 1/\xi_t$, $z = Bu/2$, and reindex by $w = 2y - Bu$. The second identity follows from $-Bu + 2\mathcal{L} = Bu + 2\mathcal{L}$. □

The following lemma shows that, if a shortest vector v is included in the coset $Bu + 2\mathcal{L}$, the matrix $\mathcal{A}_{t,d}(u)$ is approximately proportional to the matrix vv^T.

Lemma 3.4. *Let $t > t_0$ and $\lambda = \lambda_1(\mathcal{L}) \le d \le (1 + 1/n)\lambda$. Let $v \in \mathcal{L}$ satisfy $\|v\| = \lambda$ and $u \in \mathbb{F}_2^n$ satisfy $v \in Bu + 2\mathcal{L}$. For the normalized vector $\widehat{v} = v/\lambda$, it holds that*

$$\mathcal{A}_{t,d}(u) = \mu_{t,d} \widehat{v}\widehat{v}^T + \mathcal{R}_{t,d}(u) \qquad \text{for} \quad \mu_{t,d} := \frac{2\pi^2 \xi_t^4 \lambda^2}{\rho_{1/\xi_t}(\mathcal{L})} e^{-\pi \xi_t^2 \lambda^2 / 4} = \xi_t^4 \lambda^2 2^{-tn + O(1)}, \tag{12}$$

where $\mathcal{R}_{t,d}(u)$ is positive semidefinite such that $\|\mathcal{R}_{t,d}(u)\|_{\mathrm{op}} \le \mu_{t,d} \cdot 2^{\left(-\frac{1}{2}\log_2 \frac{t}{t_0} + o(1)\right) n}$.

Proof. Since $Bu + 2\mathcal{L} = v + 2\mathcal{L}$, the terms $w = v$ and $w = -v$ in Eq. (11) contribute

$$\frac{2\pi^2 \xi_t^4}{\rho_{1/\xi_t}(\mathcal{L})} v v^T e^{-\pi \xi_t^2 \lambda^2 / 4} = \mu_{t,d} \widehat{v}\widehat{v}^T.$$

Let $\mathcal{R}_{t,d}(u)$ be the sum of the remaining terms. It is positive semidefinite.

Write every element of $v + 2\mathcal{L}$ as $w = -v + 2y$ for $y \in \mathcal{L}$. The two removed terms correspond to $y = 0$ and $y = v$. Since $\mathcal{R}_{t,d}(u)$ is positive semidefinite, we have $\|\mathcal{R}_{t,d}(u)\|_{\mathrm{op}} \le \mathrm{Tr}\left(\mathcal{R}_{t,d}(u)\right)$. This gives

$$\frac{\|\mathcal{R}_{t,d}(u)\|_{\mathrm{op}}}{\mu_{t,d}} \le 2 \sum_{y \in \mathcal{L} \setminus \{0, v\}} \frac{\|y - v/2\|^2}{\lambda^2} e^{-\pi \xi_t^2 \left(\|y - v/2\|^2 - \lambda^2/4 \right)}.$$

For $y \notin \{0, v\}$, both y and $y - v$ are nonzero lattice vectors. Hence

$$\|y - v/2\|^2 - \frac{\lambda^2}{4} = \frac{\|y\|^2 + \|y - v\|^2 - \lambda^2}{2} \ge \frac{\|y\|^2}{2}.$$

Moreover,

$$\frac{\|y - v/2\|}{\lambda} \le \frac{\|y\|}{\lambda} + \frac{\|v\|}{2\lambda} \le \frac{\|y\|}{\lambda} + 1$$

Combining these two bounds and Corollary 3.2 with $a = \sqrt{2}$ and $k = 2$ gives

$$\frac{\|\mathcal{R}_{t,d}(u)\|_{\mathrm{op}}}{\mu_{t,d}} \le 2 \sum_{y \in \mathcal{L} \setminus \{0\}} \left(1 + \frac{\|y\|}{\lambda} \right)^2 \rho_{\sqrt{2}/\xi_t}(y) \le 2^{\left(-\frac{1}{2}\log_2 \frac{t}{t_0} + o(1)\right) n}.$$

Finally, the asymptotic of $\mu_{t,d}$ follows from the choice of ξ_t in Eq. (4) and $\lambda \le d \le (1 + 1/n)\lambda$, as well as $\rho_{1/\xi_t}(\mathcal{L}) = 1 + 2^{-\Omega(n)}$ by Corollary 3.2. □

3.3 Estimating the Hessians and recovering a shortest vector

By Lemma 3.4, the exact Hessian at the parity class of a shortest vector v contains some information about v. In this section, we show how to actually recover v using this property. As discussed in the overview, this can be done by the three steps; 1) compute an estimation of Hessian, 2) compute a unit vector close to the direction of v, and 3) recover the vector v using (preprocessing) BDD oracle from Theorem 2.4.

Fix $0 < t < 1/4$ and let $N_t = \lceil n^5 \log n \cdot 2^{2tn} \rceil$.[3] For independent $X_1, \dots, X_{N_t} \sim D_{\mathcal{L}^*, \xi_t}$ and $u \in \mathbb{F}_2^n$, define

$$\widehat{\mathcal{G}}_{t,d}(u) := -\frac{4\pi^2}{N_t} \sum_{i=1}^{N_t} X_i X_i^T (-1)^{u^T (B^T X_i \bmod 2)} \mathbf{1}_{\{\|X_i\| \le C_1 \xi_t \sqrt{n}\}}, \tag{13}$$

where C_1 is the constant in Lemma 2.8 for $c = 1$. The parameters will be chosen so that we can use the DGS algorithm in Theorem 2.7.

The following lemma shows that $\widehat{\mathcal{G}}$ approximate $\mathcal{G}$ accurately.

Lemma 3.5. *For every fixed $0 < t < 1/4$, with probability $1 - 2^{-\Omega(n \log n)}$, it holds for all $u \in \mathbb{F}_2^n$ that*

$$\max_{u \in \mathbb{F}_2^n} \left\| \widehat{\mathcal{G}}_{t,d}(u) - \mathcal{G}_{t,d}(u) \right\|_{\mathrm{op}} \le 2\xi_t^2 2^{-tn}.$$

Proof. Let $X \sim D_{\mathcal{L}^*, \xi_t}$. Noting that $\|X\| > C_1 \xi_t \sqrt{n}$ is truncated in $\widehat{\mathcal{G}}$,

$$\mathbb{E}[\widehat{\mathcal{G}}_{t,d}(u)] - \mathcal{G}_{t,d}(u) = 4\pi^2 \mathbb{E}\left[X X^T (-1)^{u^T (B^T X \bmod 2)} \mathbf{1}_{\{\|X\| > C_1 \xi_t \sqrt{n}\}} \right].$$

Using $\|\mathbb{E}[M]\|_{\mathrm{op}} \le \mathbb{E}[\|M\|_{\mathrm{op}}]$ and $\left\|X X^T\right\|_{\mathrm{op}} = \|X\|^2$, Lemma 2.8 with $c = 1$ gives for sufficiently large n

$$\left\| \mathbb{E}[\widehat{\mathcal{G}}_{t,d}(u)] - \mathcal{G}_{t,d}(u) \right\|_{\mathrm{op}} \le 4\pi^2 \mathbb{E}\left[\|X\|^2 \mathbf{1}_{\{\|X\| > C_1 \xi_t \sqrt{n}\}} \right] \le 4\pi^2 \xi_t^2 n 2^{-n} \le \xi_t^2 2^{-tn}. \tag{14}$$

We next bound the probability that the estimation $\widehat{\mathcal{G}}$ is far from $\mathcal{G}$. Fix $u \in \mathbb{F}_2^n$ and $a, b \in \{1, \dots, n\}$. The absolute value of the (a,b)-entry of each summand in Eq. (13) is at most

$$4\pi^2 |(X_i)_a (X_i)_b| \mathbf{1}_{\{\|X_i\| \le C_1 \xi_t \sqrt{n}\}} \le 4\pi^2 C_1^2 \xi_t^2 n.$$

Therefore, Hoeffding's inequality gives, using $N_t = \lceil n^5 \log n \cdot 2^{2tn} \rceil$,

$$\mathbb{P}\left[\left| \left(\widehat{\mathcal{G}}_{t,d}(u) - \mathbb{E}[\widehat{\mathcal{G}}_{t,d}(u)] \right)_{a,b} \right| > \frac{\xi_t^2}{n} 2^{-tn} \right] \le 2 \exp\left(-\Omega\left(\frac{N_t}{n^4} 2^{-2tn} \right) \right) = 2e^{-\Omega(n \log n)}.$$

By the union bound over all $2^n n^2$ choices of $u \in \mathbb{F}_2^n$ and $a, b \in \{1, \dots, n\}$, with probability $1 - 2^{-\Omega(n \log n)}$, every entry of every difference $\widehat{\mathcal{G}}_{t,d}(u) - \mathbb{E}[\widehat{\mathcal{G}}_{t,d}(u)]$ has absolute value at most $\xi_t^2 2^{-tn}/n$. Since $\|M\|_{\mathrm{op}} \le n \max_{a,b} |M_{a,b}|$, the operator norm of every difference is at most $\xi_t^2 2^{-tn}$. Together with Eq. (14), this concludes the proof. □

Given that $\widehat{\mathcal{G}}_{t,d}(u) \approx \mathcal{G}_{t,d}(u) \approx \mu_{t,d} \widehat{v} \widehat{v}^T$ due to Lemmas 3.4 and 3.5, the eigenvector of $\widehat{\mathcal{G}}_{t,d}(u)$ must be close to the direction of v. The following lemma formalize this intuition.

Lemma 3.6. *Let $t_0 < t < 1/4$ be fixed and let $\lambda = \lambda_1(\mathcal{L}) \le d \le (1 + 1/n)\lambda$. Let $v \in \mathcal{L}$ satisfy $\|v\| = \lambda$, and let $u \in \mathbb{F}_2^n$ satisfy $v \in Bu + 2\mathcal{L}$. Suppose that the conclusion of Lemma 3.5 holds. If q is a unit eigenvector corresponding to the largest eigenvalue of $\widehat{\mathcal{G}}_{t,d}(u)$, then*

$$\min_{\sigma \in \{-1,1\}} \|q - \sigma v/\lambda\| = O(n^{-1/2}).$$

Consequently, the $n^{-1/3}$-BDD query at dq returns one of $\pm v$ for all sufficiently large n.

[3] Throughout this paper, we do not try to optimize the polynomial factors.

Proof. Let $\widehat{v} = v/\lambda$. By Lemma 3.4 and the definition of $\mathcal{A}_{t,d}(u)$ in Eq. (9),

$$\widehat{\mathcal{G}}_{t,d}(u) = -2\pi\xi_t^2 F_{1/\xi_t}(Bu/2)I_n + \mu_{t,d}\widehat{v}\widehat{v}^T + E, \tag{15}$$

where $E = \mathcal{R}_{t,d}(u) + \widehat{\mathcal{G}}_{t,d}(u) - \mathcal{G}_{t,d}(u)$ and $\|\mathcal{R}_{t,d}(u)\|_{\mathrm{op}}/\mu_{t,d} = 2^{-\Omega(n)}$.

Since $\lambda \le d \le (1+1/n)\lambda$, $\xi_t^2\lambda^2 = \Theta(n)$. Also, $\mu_{t,d} = \Theta(\xi_t^4\lambda^2 2^{-tn})$ by Lemma 3.4. Lemma 3.5 gives

$$\frac{\left\|\widehat{\mathcal{G}}_{t,d}(u) - \mathcal{G}_{t,d}(u)\right\|_{\mathrm{op}}}{\mu_{t,d}} = O\left(\frac{1}{\xi_t^2\lambda^2}\right) = O(n^{-1}).$$

Together with $\|\mathcal{R}_{t,d}(u)\|_{\mathrm{op}}/\mu_{t,d} = 2^{-\Omega(n)}$, this gives $\|E\|_{\mathrm{op}} = O(\mu_{t,d}n^{-1})$.

Since q is a eigenvector corresponding to the largest eigenvalue of $\widehat{\mathcal{G}}_{t,d}(u)$, we have $q^T\widehat{\mathcal{G}}_{t,d}(u)q \ge \widehat{v}^T\widehat{\mathcal{G}}_{t,d}(u)\widehat{v}$. Substituting the expression for $\widehat{\mathcal{G}}_{t,d}(u)$ using Eq. (15), the multiple of I_n has the same value on both sides because $\|q\| = \|\widehat{v}\| = 1$. The remaining inequality is $\mu_{t,d}\langle q, \widehat{v}\rangle^2 + q^TEq \ge \mu_{t,d} + \widehat{v}^TE\widehat{v}$. Since $|q^TEq| \le \|E\|_{\mathrm{op}}$ and $|\widehat{v}^TE\widehat{v}| \le \|E\|_{\mathrm{op}}$, it follows that

$$\mu_{t,d}\left(1 - \langle q, \widehat{v}\rangle^2\right) \le q^TEq - \widehat{v}^TE\widehat{v} \le 2\|E\|_{\mathrm{op}} = O(\mu_{t,d}n^{-1}).$$

Choose $\sigma \in \{-1, 1\}$ such that $\langle q, \sigma\widehat{v}\rangle \ge 0$. It follows that $\|q - \sigma\widehat{v}\|^2 \le 2(1 - \langle q, \widehat{v}\rangle^2) = O(n^{-1})$.

Finally, $\|dq - \sigma v\| \le d\|q - \sigma\widehat{v}\| + d - \lambda = O(\lambda n^{-1/2}) < \lambda n^{-1/3}$ for all sufficiently large n. The last claim follows from Theorem 2.4. □

Combining the estimate and the recovery argument gives the direct algorithm.

Algorithm 1: Direct Hessian SVP

1. Apply LLL reduction to obtain a nonzero $x \in \mathcal{L}$ and use the scales $d_j = (1+1/n)^{-j}\|x\|$ for $0 \le j \le n^2$. Construct the preprocessing $n^{-1/3}$-BDD data.
2. Fix $t_0 < t < 1/4$. At every scale d_j, call DGS from Theorem 2.7 with $\kappa = n^2$ and keep N_t samples from $D_{\mathcal{L}^*,\xi_t(d_j)}$. Abort the scale if fewer than N_t samples are returned.
3. For every $u \in \mathbb{F}_2^n$, evaluate $\widehat{\mathcal{G}}_{t,d_j}(u)$ directly and compute a unit eigenvector q_u corresponding to its largest eigenvalue. Query BDD at d_jq_u and retain the output only after exact lattice-membership verification.
4. Return the shortest verified nonzero vector over all scales and all u.

Theorem 3.7. *Algorithm 1 solves Search-*SVP *with constant probability in time* $2^{1.46295n+o(n)}$ *and space* $2^{n/2+o(n)}$.

Proof. The correctness follows from the previous lemmas. Lemma 3.5 shows that $\widehat{\mathcal{G}}$ estimate $\mathcal{G}$ for all inputs with overwhelming probability, and in that case Lemma 3.6 and Theorem 2.4 show the correctness of the final outcome.

We explain the time and space complexity below. There are polynomially many choices of d_j because of the property of LLL-reduced vectors. For each d_j, the algorithm considers all 2^n values of $u \in \mathbb{F}_2^n$, and evaluating $\widehat{\mathcal{G}}_{t,d_j}(u)$ takes $N_t\,\mathrm{poly}(n)$ time for $N_t = \mathrm{poly}(n)2^{2tn}$. Hence the total time is $2^{(1+2t)n+o(n)}$. Because of Lemma 3.4, we have $1 + 2t > 1 + 2t_0 = 1.4629\ldots$, thus choosing $t \to t_0$ gives the desired time complexity. All remaining steps take less time and do not affect the overall time complexity. The space complexity is due to DGS in Theorem 2.7. □

4 Batch Hessian Estimation

The factor 2^n in the $2^{1.4629n+o(n)}$ running time of the previous section comes from evaluating $\widehat{\mathcal{G}}_{t,d}(u)$ separately for every $u \in \mathbb{F}_2^n$. The definition of $\widehat{\mathcal{G}}_{t,d}(u)$ uses the same samples for all u, and Lemma 3.5 proves that all these estimators are simultaneously accurate. This section optimizes the algorithm by batching the Hessian estimations. The main idea is to use the Walsh-Hadamard transform in the following lemma.

Lemma 4.1. *Let $A : \mathbb{F}_2^\ell \to \mathbb{R}^{n\times n}$. Given $A(x)$ for every $x \in \mathbb{F}_2^\ell$, the matrices*

$$\sum_{x\in\mathbb{F}_2^\ell} (-1)^{\theta\cdot x} A(x), \qquad \theta \in \mathbb{F}_2^\ell,$$

can be computed in time $2^\ell \operatorname{poly}(n)$ and space $2^\ell \operatorname{poly}(n)$.

Proof. Store the matrices $A(x)$ in an array indexed by $\mathbb{F}_2^\ell$. For every $j \in \{1,\dots,\ell\}$ and every $x \in \mathbb{F}_2^\ell$ with $x_j = 0$, replace

$$\big(A(x), A(x+e_j)\big) \quad \text{by} \quad \big(A(x)+A(x+e_j), A(x)-A(x+e_j)\big),$$

where e_j is the jth standard basis vector. After processing all ℓ coordinates, the entry indexed by θ equals $\sum_x (-1)^{\theta\cdot x} A(x)$. There are $\ell 2^{\ell-1}$ replacements, and the array contains 2^ℓ matrices. □

Algorithm 2: Walsh-Hadamard Hessian SVP

1. Construct the preprocessing BDD data and the scale grid, choose t, and obtain the Discrete Gaussian samples at every scale d_j as in Algorithm 1. Set $\ell = \lfloor n/2 \rfloor$.
2. At each scale d_j that was not aborted, define, for every sample X_i,

   $$k_i := B^T X_i \bmod 2, \qquad W_i := -\frac{4\pi^2}{N_t} X_i X_i^T \mathbf{1}_{\{\|X_i\| \le C_1 \xi_t(d_j)\sqrt{n}\}}.$$

   Write $k_i = (k_i', k_i'') \in \mathbb{F}_2^\ell \times \mathbb{F}_2^{n-\ell}$.
3. For every $\theta'' \in \mathbb{F}_2^{n-\ell}$, construct the array $A_{\theta''} : \mathbb{F}_2^\ell \to \mathbb{R}^{n\times n}$ given by

   $$A_{\theta''}(x) := \sum_{\substack{1\le i\le N_t \\ k_i'=x}} (-1)^{\theta''\cdot k_i''} W_i.$$

   Apply Lemma 4.1 to $A_{\theta''}$. The matrix indexed by $\theta' \in \mathbb{F}_2^\ell$ is $\widehat{\mathcal{G}}_{t,d_j}(\theta', \theta'')$.

   For each such matrix, compute a unit eigenvector q corresponding to its largest eigenvalue, query BDD at $d_j q$, and retain the output if it is a nonzero lattice vector.
4. Return the shortest retained vector over all scales.

Theorem 4.2. *Algorithm 2 solves Search-SVP in time $2^{n+o(n)}$ and $2^{n/2+o(n)}$ space with constant success probability.*

Proof. For $u = (\theta', \theta'')$, the matrix produced by the transform is

$$\sum_{x\in\mathbb{F}_2^\ell} (-1)^{\theta'\cdot x} A_{\theta''}(x) = \sum_{i=1}^{N_t} (-1)^{\theta'\cdot k_i' + \theta''\cdot k_i''} W_i = \widehat{\mathcal{G}}_{t,d_j}(u).$$

Thus the algorithm evaluates exactly the estimators defined in Eq. (13). The correctness then follows the same argument as in Theorem 3.7.

Now we discuss the space and time complexity. In the algorithm, only one array $A_{\theta''}$ is stored at a time. The array contains 2^ℓ matrices, while the number of Discrete Gaussian samples is $N_t < 2^{n/2}$. Together with the preprocessing BDD data, this gives space $2^{n/2+o(n)}$.

For each fixed θ'', constructing the array $A_{\theta''}$ takes $N_t \operatorname{poly}(n)$ time and the following Hadamard-Walsh transform take $2^{\ell+o(n)}$ time by Lemma 4.1. Since there are $2^{n-\ell}$ choices of θ'', the total time for each d_j is $2^{n-\ell}(N_t + 2^\ell)2^{o(n)}$. Here $\ell = \lfloor n/2 \rfloor$ and $N_t = 2^{2tn+o(n)} < 2^{n/2}$ because $t < 1/4$. Therefore, the total time over the polynomially many scales is $2^{n+o(n)}$. □

5 Random Sublattice Coset Hessian

The algorithm in Section 4 computes $\widehat{\mathcal{G}}(u)$ for all $u \in \mathbb{F}_2^n$, resulting in the complexity $2^{n+o(n)}$. This section shows that the same strategy works by estimating the Hessians using vectors in a sublattice cosets.

As always, we mainly focus on the case $\lambda \le d \le (1+1/n)\lambda$ for $\lambda = \lambda_1(\mathcal{L})$. We write $g(t) := \frac{1}{2}\log_2(t/t_0)$ for $t > 0$. For our main interested parameter d, Corollary 3.2 gives, for every fixed c and positive integer k,

$$\sum_{x\in\mathcal{L}} \|x\|^k \rho_{\sqrt{2^{1+c}}/\xi_t}(x) \le \lambda^k 2^{(c/2-g(t)+o(1))n}, \qquad \text{and} \qquad \rho_{\sqrt{2^{1+c}}/\xi_t}(\mathcal{L}) \le 1 + 2^{(c/2-g(t)+o(1))n}. \tag{16}$$

The following term from Lemma 3.4 is convenient for us.

$$\mu_{t,d} = \xi_t^4 \lambda^2 2^{-tn+O(1)} = \xi_t^4 d^2 2^{-tn+O(1)} \qquad \text{for} \quad \lambda \le d \le (1+1/n)\lambda. \tag{17}$$

5.1 Gaussian and Hessian sums

The following periodic Gaussian and Hessian mass bounds will be used in this section.

Lemma 5.1. *For every $s > 0$,*

$$\sum_{u\in\mathbb{F}_2^n} F_s(Bu/2) = \frac{\rho_{2s}(\mathcal{L})}{\rho_s(\mathcal{L})}.$$

If $\lambda \le d \le (1+1/n)\lambda$, it holds that $\sum_{u\neq 0} F_{1/\xi_t}(Bu/2) \le 2^{(1/2-g(t)+o(1))n}$.

Proof. The cosets $Bu/2 + \mathcal{L}$ for $u \in \mathbb{F}_2^n$ partition $\frac{1}{2}\mathcal{L}$. Hence,

$$\sum_{u\in\mathbb{F}_2^n} F_s(Bu/2) = \sum_{u\in\mathbb{F}_2^n} \frac{\rho_s(Bu/2+\mathcal{L})}{\rho_s(\mathcal{L})} = \frac{\rho_s(\frac{1}{2}\mathcal{L})}{\rho_s(\mathcal{L})} = \frac{\rho_{2s}(\mathcal{L})}{\rho_s(\mathcal{L})}.$$

The final statement follows from Eq. (16) with $c = 1$. □

Lemma 5.2. *Let $\lambda \le d \le (1+1/n)\lambda$ and $t > t_0/2$. Then*

$$\sum_{u\in\mathbb{F}_2^n} \|\mathcal{A}_{t,d}(u)\|_{\mathrm{F}}^2 \le \mu_{t,d}^2 2^{(2t-\min\{g(t),2g(t)\}+o(1))n}.$$

Proof. We use $\|X\|_{\mathrm{F}}^2 = \operatorname{Tr}(X^T X)$ for the real matrix X. Applying this to Eq. (11), the left-hand side becomes

$$\left(\frac{\pi^2\xi_t^4}{\rho_{1/\xi_t}(\mathcal{L})}\right)^2 \sum_{u\in\mathbb{F}_2^n} \sum_{z,z'\in Bu+2\mathcal{L}} \operatorname{Tr}(zz^T z'z'^T) e^{-\pi\xi_t^2(\|z\|^2+\|z'\|^2)/4}.$$

As u varies, the map from $(z,z') \in (Bu+2\mathcal{L})^2$ defined by $(z,z') \mapsto (p = \frac{z+z'}{2}, q = \frac{z-z'}{2})$ is a bijection to $\mathcal{L}^2$, and it satisfies

$$\operatorname{Tr}(zz^T z'z'^T) = (z^T z')^2 = (\|p\|^2 - \|q\|^2)^2, \qquad \|z\|^2 + \|z'\|^2 = 2(\|p\|^2 + \|q\|^2).$$

Therefore,

$$\sum_{u\in\mathbb{F}_2^n} \|\mathcal{A}_{t,d}(u)\|_{\mathrm{F}}^2 = \frac{\pi^4\xi_t^8}{\rho_{1/\xi_t}(\mathcal{L})^2} \sum_{p,q\in\mathcal{L}} (\|p\|^2 - \|q\|^2)^2 e^{-\pi\xi_t^2(\|p\|^2+\|q\|^2)/2}.$$

Let $w(x) = e^{-\pi\xi_t^2\|x\|^2/2} = \rho_{\sqrt{2}/\xi_t}(x)$. Using $(\|p\|^2 - \|q\|^2)^2 \le 2(\|p\|^4 + \|q\|^4)$, the sum over $p, q \in \mathcal{L}$ is at most

$$4\left(\sum_{p\in\mathcal{L}} \|p\|^4 w(p)\right)\left(\sum_{q\in\mathcal{L}} w(q)\right).$$

Indeed, the Gaussian weight factors as $w(p)w(q)$, and the two terms obtained from $\|p\|^4 + \|q\|^4$ are equal. Applying the first inequality with $k = 4$ and $c = 0$ of Eq. (16) to the first term and the second inequality with $c = 0$ to the second term, we have an upper bound

$$\lambda^4 2^{(-g(t)+o(1))n} \cdot (1 + 2^{(-g(t)+o(1))n}) = \lambda^4 2^{(-\min\{g(t),2g(t)\}+o(1))n}.$$

The resulting exponent is $-g(t)$ if $g(t) \ge 0$ and $-2g(t)$ otherwise. The fixed constants and $\rho_{1/\xi_t}(\mathcal{L})^{-2} \le 1$ are absorbed into $2^{o(n)}$. The final result is obtained using Eq. (17). □

5.2 A random coset Hessian

Fix $0 \le \chi \le 1/2$. We split each parity class $u \in \mathbb{F}_2^n$ into $h = \lfloor \chi n \rfloor$ and $\ell = n - h$ bits in random basis. For $P \in \mathrm{GL}_n(\mathbb{F}_2)$, $u \in \mathbb{F}_2^n$, and $X \in \mathcal{L}^*$, define

$$Pu = (\alpha, \theta) \in \mathbb{F}_2^h \times \mathbb{F}_2^\ell, \qquad P^{-T}B^T X \bmod 2 = (J_P(X), V_P(X)) \in \mathbb{F}_2^h \times \mathbb{F}_2^\ell.$$

Define $\Lambda_j := \{X \in \mathcal{L}^* : J_P(X) = j\}$. The set Λ_0 is a sublattice of $\mathcal{L}^*$, and each Λ_j for $j \in \mathbb{F}_2^h$ is one of its cosets. Note that every Λ_j is nonempty because the map $X \mapsto P^{-T}(B^T X \bmod 2)$ from $\mathcal{L}^*$ to $\mathbb{F}_2^n$ is surjective. Therefore, Λ_j can be written as $\Lambda_0 + v_j$ for some $v_j \in \Lambda_j$ thus the discrete Gaussian $D_{\Lambda_j,s}$ is well-defined.

For $\theta \in \mathbb{F}_2^\ell$, we define the (sublattice) coset Hessian along Λ_j analogous to Lemma 3.3 by

$$\mathcal{G}_{t,d,j}(\theta) := -4\pi^2 \mathbb{E}_{X\sim D_{\Lambda_j,\xi_t}}\left[XX^T(-1)^{\theta^T V_P(X)}\right] \tag{18}$$

based on the identity for $X \in \Lambda_j$

$$u^T(B^T X \bmod 2) = (Pu)^T(P^{-T}B^T X \bmod 2) = \alpha^T J_P(X) + \theta^T V_P(X) = \alpha^T j + \theta^T V_P(X) \pmod 2 \tag{19}$$

For $u = P^{-1}(\alpha, \theta)$, Eq. (18) omits the factor $(-1)^{\alpha^T j}$. This omission allows $\mathcal{G}_{t,d,j}$ is defined independent of α, yet we should be careful about that factor.

Let $\zeta_{t,j} := 2^h \mathbb{P}_{X\sim D_{\mathcal{L}^*,\xi_t}}[J_P(X) = j]$ be the probability, normalized by the uniform probability 2^{-h}, that a discrete Gaussian sample of width ξ_t belongs to the affine lattice coset Λ_j. The following lemma shows that $\zeta_{t,j}$ is near 1, i.e., the distribution of $J_P(X)$ for the Discrete Gaussian sample X is pointwise exponentially close to uniform. The parameter condition of this lemma is intentionally general for the later application.

Lemma 5.3. *If $\lambda = \lambda_1(\mathcal{L}) \le d \le (1 + 1/n)\lambda$, $t > t_0/2$, and $0 \le \chi \le 1/2$ and $\chi < 1/2 + g(t)$, then, with probability $1 - 2^{-\Omega(n)}$ over the uniform random $P \in \mathrm{GL}_n(\mathbb{F}_2)$, $\zeta_{t,j} = 1 + O(2^{-\Omega(n)})$ for every $j \in \mathbb{F}_2^h$.*

Proof. Recall $2^h \mathbf{1}_{\{y=j\}} = \sum_{\alpha\in\mathbb{F}_2^h} (-1)^{\alpha\cdot(y+j)}$ for $y, j \in \mathbb{F}_2^h$. Applying the identity with $y = J_P(X)$ and taking the expectation over $X \sim D_{\mathcal{L}^*,\xi_t}$ gives

$$\zeta_{t,j} = \sum_{\alpha\in\mathbb{F}_2^h} (-1)^{\alpha\cdot j} \mathbb{E}[(-1)^{\alpha\cdot J_P(X)}].$$

Let $\mathcal{K}_P := P^{-1}(\mathbb{F}_2^h \times \{0\})$ and $\kappa = P^{-1}(\alpha, 0) \in \mathcal{K}_P$ for $\alpha \in \mathbb{F}_2^h$. The Poisson representation in Eq. (8) gives

$$F_{1/\xi_t}(B\kappa/2) = \mathbb{E}_{X\sim D_{\mathcal{L}^*,\xi_t}}[e^{2\pi i\langle X, B\kappa/2\rangle}] = \mathbb{E}_{X\sim D_{\mathcal{L}^*,\xi_t}}[(-1)^{\kappa^T(B^T X \bmod 2)}] = \mathbb{E}_{X\sim D_{\mathcal{L}^*,\xi_t}}[(-1)^{\alpha\cdot J_P(X)}].$$

Plugging this equation to the summand in $\zeta_{t,j}$ above, we have

$$\zeta_{t,j} = \sum_{\alpha\in\mathbb{F}_2^h} (-1)^{\alpha\cdot j} F_{1/\xi_t}\left(BP^{-1}(\alpha,0)/2\right) = 1 + \sum_{\alpha\neq 0} (-1)^{\alpha\cdot j} F_{1/\xi_t}\left(BP^{-1}(\alpha,0)/2\right).$$

By substituting $\kappa = P^{-1}(\alpha, 0)$, we have $|\zeta_{t,j} - 1| \le \sum_{\substack{\kappa\in\mathcal{K}_P\\ \kappa\neq 0}} F_{1/\xi_t}(B\kappa/2)$ for all j. We bound this term using Lemma 5.1. For each nonzero $\kappa \in \mathbb{F}_2^n$, the space $\mathcal{K}_P$ for random P is a uniform random h-dimensional subspace so that $\mathbb{P}_P[\kappa \in \mathcal{K}_P] = (2^h - 1)/(2^n - 1) = 2^{(\chi-1+o(1))n}$. This gives

$$\mathbb{E}_P \sum_{\kappa\in\mathcal{K}_P,\kappa\neq 0} F_{1/\xi_t}(B\kappa/2) \le 2^{(\chi-1/2-g(t)+o(1))n}.$$

The exponent is a negative constant by the assumption $\chi < 1/2 + g(t)$. Markov's inequality shows that, except with probability $2^{-\Omega(n)}$ over P,

$$\sum_{\kappa\in\mathcal{K}_P,\kappa\neq 0} F_{1/\xi_t}(B\kappa/2) \le 2^{-\Omega(n)}.$$

This proves the lemma, as the upper bound of $|\zeta_{t,j} - 1|$ holds simultaneously for all j. □

The following lemma shows the sampling from D_{Λ_j,ξ_t} can be done by sampling from the entire $D_{\mathcal{L}^*,\xi_t}$ and then collecting the samples with $J_P(X) = j$ thanks to the almost uniformity.

Lemma 5.4. *If $\lambda \le d \le (1 + 1/n)\lambda$, $t_0 < t < 1/4$, and $0 \le \chi \le 1/2$, there is a classical algorithm that for independent uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$, outputs $N_t = \lceil n^5 \log n \cdot 2^{2tn}\rceil$ lattice vectors or aborts in time $2^{(\max\{1/2,\chi+2t\}+o(1))n}$ and space $2^{n/2+o(n)}$. It does not abort with probability $1 - 2^{-\Omega(n)}$ over P and the sampler randomness, and in that case, the joint distribution of its outputs has statistical distance at most $\exp(-\Omega(n^2))$ from $D_{\Lambda_j,\xi_t}^{N_t}$.*

Proof. We use Theorem 2.7 $\lceil 4N_t 2^{h-n/2}\rceil$ times sequentially, collect the samples satisfying $J_P(X) = j$, and discard the others. By Lemma 5.3, except with probability $2^{-\Omega(n)}$ over P, $\mathbb{P}_{X\sim D_{\mathcal{L}^*,\xi_t}}[J_P(X) = j] = 2^{-h}\left(1 + O(2^{-\Omega(n)})\right)$ simultaneously for every $j \in \mathbb{F}_2^h$. For every such P and every j, the Chernoff bound in Lemma 2.12 shows that fewer than N_t samples are accepted among $4N_t 2^h$ independent samples with probability $2^{-\Omega(N_t)}$. The statistical distance from independent $D_{\mathcal{L}^*,\xi_t}$ samples changes this probability by at most $\exp(-\Omega(n^2))$. Therefore, the procedure aborts with probability $2^{-\Omega(n)}$. The time and space complexity is clear. □

5.3 Estimating the coset Hessian

Fix $t_0 = 0.23147\ldots < t < 1/4$ and $N_t = \lceil n^5 \log n \cdot 2^{2tn}\rceil$. For independent samples from the coset discrete Gaussian distribution $X_1, \ldots, X_{N_t} \sim D_{\Lambda_j,\xi_t}$, define an estimation of Eq. (18) analogous to Eq. (13) by

$$\widehat{\mathcal{G}}_{t,d,j}(\theta) := -\frac{4\pi^2}{N_t}\sum_{i=1}^{N_t} X_i X_i^T (-1)^{\theta\cdot V_P(X_i)} \mathbf{1}_{\{\|X_i\|\le C_1\xi_t\sqrt{n}\}} \tag{20}$$

which can be computed using Lemma 5.4. We show that $\widehat{\mathcal{G}}_{t,d,j}$ approximates $\mathcal{G}_{t,d,j}$ similar to Lemma 3.5.

Lemma 5.5. *Let $\lambda = \lambda_1(\mathcal{L})$ and $\lambda \le d \le (1+1/n)\lambda$. Choose independent uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$. For independent $X_1, \ldots, X_{N_t} \sim D_{\Lambda_j,\xi_t}$, with probability $1 - 2^{-\Omega(n)}$ over $P, j, X_1, \ldots, X_{N_t}$,*

$$\max_{\theta\in\mathbb{F}_2^\ell} \left\|\widehat{\mathcal{G}}_{t,d,j}(\theta) - \mathcal{G}_{t,d,j}(\theta)\right\|_{\mathrm{op}} = O\left(\frac{\mu_{t,d}}{n}\right).$$

Proof. By Lemma 5.3, except with probability $2^{-\Omega(n)}$ over P, we have $\zeta_{t,j} \geq 1/2$ for every $j \in \mathbb{F}_2^h$. Fix such a matrix P. Let $Y \sim D_{\mathcal{L}^*,\xi_t}$ and $X \sim D_{\Lambda_j,\xi_t}$. By $\mathbb{P}[J_P(Y) = j] = 2^{-h}\zeta_{t,j}$ and Lemma 2.8 with $c = 1$, we have

$$\mathbb{E}\left[\|X\|^2 \mathbf{1}_{\{\|X\|>C_1\xi_t\sqrt{n}\}}\right] \leq \frac{2^h}{\zeta_{t,j}}\mathbb{E}\left[\|Y\|^2 \mathbf{1}_{\{\|Y\|>C_1\xi_t\sqrt{n}\}}\right] \leq 2^{h+1}\xi_t^2 n 2^{-n}.$$

With the same argument for Eq. (14), we have

$$\left\|\mathbb{E}[\widehat{\mathcal{G}}_{t,d,j}(\theta)] - \mathcal{G}_{t,d,j}(\theta)\right\|_{\mathrm{op}} \leq 8\pi^2\xi_t^2 n 2^{-\ell} \leq \frac{\xi_t^4 d^2}{n} 2^{-tn}$$

for all sufficiently large n, where we used $\ell \geq n/2$, $t < 1/4$, and $\xi_t^2 d^2 = 4nt \ln 2/\pi$.

Each entry of the random matrix in Eq. (20) has absolute value at most $4\pi^2 C_1^2 \xi_t^2 n$. Hence, for fixed $\theta \in \mathbb{F}_2^\ell$ and $a, b \in \{1, \ldots, n\}$, Hoeffding's inequality gives

$$\mathbb{P}\left[\left|\left(\widehat{\mathcal{G}}_{t,d,j}(\theta) - \mathbb{E}[\widehat{\mathcal{G}}_{t,d,j}(\theta)]\right)_{a,b}\right| > \frac{\xi_t^4 d^2}{n^2} 2^{-tn}\right] \leq 2\exp\left(-\Omega\left(\frac{N_t \xi_t^4 d^4}{n^6} 2^{-2tn}\right)\right) = 2e^{-\Omega(n\log n)}.$$

Here we used $N_t = \lceil n^5 \log n \cdot 2^{2tn} \rceil$ and $\xi_t^2 d^2 = 4nt \ln 2/\pi$. A union bound over the $2^\ell n^2$ choices of θ, a, b shows that, with probability $1 - 2^{-\Omega(n\log n)}$ over the samples,

$$\max_{\theta \in \mathbb{F}_2^\ell} \left\|\widehat{\mathcal{G}}_{t,d,j}(\theta) - \mathbb{E}[\widehat{\mathcal{G}}_{t,d,j}(\theta)]\right\|_{\mathrm{op}} \leq \frac{\xi_t^4 d^2}{n} 2^{-tn}$$

The triangle inequality applied to $\widehat{\mathcal{G}}_{t,d,j}(\theta) - \mathbb{E}[\widehat{\mathcal{G}}_{t,d,j}(\theta)]$ and $\mathbb{E}[\widehat{\mathcal{G}}_{t,d,j}(\theta)] - \mathcal{G}_{t,d,j}(\theta)$ proves the lemma, after using $\mu_{t,d} = \xi_t^4 d^2 2^{-tn+O(1)}$. □

5.4 Coset Hessians and shortest vectors

It remains to show how to relate the coset Hessians and the (direction) of shortest vectors as in Lemma 3.4. The following lemma extends Eq. (9) to the coset Hessian.

Lemma 5.6. $\mathcal{G}_{t,d,j}(\theta) - \frac{1}{\zeta_{t,j}} \sum_{\alpha \in \mathbb{F}_2^h} (-1)^{\alpha \cdot j} \mathcal{A}_{t,d}\left(P^{-1}(\alpha, \theta)\right) \in \mathrm{span}\{I_n\}$.

Proof. Since $\mathbb{P}[J_P(X) = j] = 2^{-h}\zeta_{t,j}$, the definition of $\mathcal{G}_{t,d,j}(\theta)$ gives

$$\mathcal{G}_{t,d,j}(\theta) = -\frac{4\pi^2 2^h}{\zeta_{t,j}} \mathbb{E}_{X \sim D_{\mathcal{L}^*,\xi_t}}\left[XX^T (-1)^{\theta \cdot V_P(X)} \mathbf{1}_{\{J_P(X)=j\}}\right].$$

Observe that $2^h \mathbf{1}_{\{y=j\}} = \sum_{\alpha \in \mathbb{F}_2^h} (-1)^{\alpha \cdot (y+j)}$ for every $y, j \in \mathbb{F}_2^h$. Indeed, if $y = j$, every summand is one; otherwise, the summands can be paired with opposite signs. Applying this identity with $y = J_P(X)$ and interchanging the finite sum and expectation gives

$$\mathcal{G}_{t,d,j}(\theta) = \frac{1}{\zeta_{t,j}} \sum_{\alpha \in \mathbb{F}_2^h} (-1)^{\alpha \cdot j} \mathcal{G}_{t,d}\left(P^{-1}(\alpha, \theta)\right) \tag{21}$$

where we put $u = P^{-1}(\alpha, \theta)$ and use Eq. (19) to derive

$$(-1)^{\theta \cdot V_P(X)} \cdot (-1)^{\alpha \cdot J_P(X) + \alpha \cdot j} = (-1)^{\alpha \cdot j} \cdot (-1)^{\alpha \cdot J_P(X) + \theta \cdot V_P(X)} = (-1)^{\alpha \cdot j} \cdot (-1)^{u^T (B^T X \bmod 2)}.$$

The term $(-1)^{u^T (B^T X \bmod 2)}$ is exactly the term in Lemma 3.3, proving Eq. (21). The final inclusion is proven by using Lemma 3.3 on $\mathcal{G}_{t,d}(P^{-1}(\alpha, \theta)) - \mathcal{A}_{t,d}(P^{-1}(\alpha, \theta)) \in \mathrm{span}\{I_n\}$ for each α. □

This lemma gives some intuition why the approach in this section works. If one of $P^{-1}(\alpha,\theta) = u$ is such that $Bu + 2\mathcal{L}$ contains a shortest vector v, the corresponding $\mathcal{A}_{t,d}(P^{-1}(\alpha,\theta))$ must be approximately proportional to vv^T. We will prove that the sum of the remaining terms is exponentially smaller with high probability over P and j. Therefore, some eigenvector of $\mathcal{G}_{t,d,j}(\theta)$ must be approximately proportional to v, as the multiple of I_n does not affect the eigen vectors and the order of eigenvalues.

The coefficient of the first term is $(-1)^{\alpha_* \cdot j}$, so it may determine either the largest or the smallest eigenvalue. We therefore consider both extreme eigenvalues. Similarly to Lemma 3.6, we obtain the following lemma.

Lemma 5.7. *Let $t_0 < t < 1/4$ be fixed, let $0 \le \chi \le 1/2$, and let $\lambda = \lambda_1(\mathcal{L}) \le d \le (1+1/n)\lambda$. Let $v \in \mathcal{L}$ satisfy $\|v\| = \lambda$, and let $u_* \in \mathbb{F}_2^n$ satisfy $v \in Bu_* + 2\mathcal{L}$. Choose independent uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$, and write $Pu_* = (\alpha_*, \theta_*)$. With probability $1 - 2^{-\Omega(n)}$ over P and j, the following holds. Suppose that the conclusion of Lemma 5.5 holds. Let q_+ and q_- be unit eigenvectors corresponding to the largest and smallest eigenvalues of $\widehat{\mathcal{G}}_{t,d,j}(\theta_*)$, respectively. Then there are $s \in \{+,-\}$ and $\tau \in \{-1,1\}$ such that*

$$\|q_s - \tau v/\lambda\| = O(n^{-1/2}).$$

Consequently, the $n^{-1/3}$-BDD query at dq_s returns τv for all sufficiently large n.

Proof. Let $\widehat{v} = v/\lambda$ and recall Lemma 3.4 stating that $\mathcal{A}_{t,d}(u_*) = \mu_{t,d}\widehat{v}\widehat{v}^T + \mathcal{R}_{t,d}(u_*)$ for $\|\mathcal{R}_{t,d}(u_*)\|_{\mathrm{op}}/\mu_{t,d} = 2^{-\Omega(n)}$, where $\mu_{t,d} = \xi_t^4 \lambda^2 2^{-tn+O(1)}$.

We first bound the contribution from $\alpha \neq \alpha_*$. For fixed P, expanding the squared Frobenius norm using $\|X\|_{\mathrm{F}}^2 = \mathrm{Tr}(X^T X)$ for the real matrix X,

$$\left\| \sum_{\alpha \neq \alpha_*} (-1)^{j\cdot(\alpha-\alpha_*)} \mathcal{A}_{t,d}\left(P^{-1}(\alpha,\theta_*)\right) \right\|_{\mathrm{F}}^2 = \sum_{\alpha,\alpha' \neq \alpha_*} (-1)^{j\cdot(\alpha-\alpha')} \mathrm{Tr}\left(\mathcal{A}_{t,d}\left(P^{-1}(\alpha',\theta_*)\right)^T \mathcal{A}_{t,d}\left(P^{-1}(\alpha,\theta_*)\right)\right)$$

and taking the expectation over uniform j using $\mathbb{E}_j[(-1)^{j\cdot(\alpha-\alpha')}] = 0$ for $\alpha \neq \alpha'$ gives

$$\mathbb{E}_j \left\| \sum_{\alpha \neq \alpha_*} (-1)^{j\cdot(\alpha-\alpha_*)} \mathcal{A}_{t,d}\left(P^{-1}(\alpha,\theta_*)\right) \right\|_{\mathrm{F}}^2 = \sum_{\alpha \neq \alpha_*} \left\| \mathcal{A}_{t,d}\left(P^{-1}(\alpha,\theta_*)\right)\right\|_{\mathrm{F}}^2.$$

Now we are taking the expectation over P. For each fixed $u \neq u_*$, the matrix $\mathcal{A}_{t,d}(u)$ occurs in the sum on exactly when $P(u-u_*) \in \mathbb{F}_2^h \times \{0\}$. Since $P(u-u_*)$ is uniform over $\mathbb{F}_2^n \setminus \{0\}$, the probability that $P(u-u_*)$ is included in $\mathbb{F}_2^h \times \{0\}$ is exactly $(2^h-1)/(2^n-1)$. Consequently, Lemma 5.2 gives

$$\mathbb{E}_{P,j} \left\| \sum_{\alpha \neq \alpha_*} (-1)^{j\cdot(\alpha-\alpha_*)} \mathcal{A}_{t,d}\left(P^{-1}(\alpha,\theta_*)\right) \right\|_{\mathrm{F}}^2 \le 2^{h-n} \cdot (\mu_{t,d}^2 2^{2tn - g(t)n + o(n)}) = \mu_{t,d}^2 2^{(\chi - 1 - g(t) + 2t + o(1))n}$$

where we use $h = \lfloor \chi n \rfloor$. Since $\chi \le 1/2$, $t < 1/4$, and $g(t) > 0$, the exponent satisfies $\chi - 1 - g(t) + 2t < 0$. Markov's inequality shows that

$$\left\| \sum_{\alpha \neq \alpha_*} (-1)^{j\cdot(\alpha-\alpha_*)} \mathcal{A}_{t,d}\left(P^{-1}(\alpha,\theta_*)\right) \right\|_{\mathrm{F}} \le \mu_{t,d} 2^{-\Omega(n)} \tag{22}$$

holds except with probability $2^{-\Omega(n)}$ over P and j, and the same bound holds for the operator norm as well.

For the hidden sign $\sigma = (-1)^{\alpha_* \cdot j}$, we can decompose $E := \sigma\widehat{\mathcal{G}}_{t,d,j}(\theta_*) - \frac{\mu_{t,d}}{\zeta_{t,j}}\widehat{v}\widehat{v}^T - aI_n$ using Lemma 5.6 for some a and bound its operator norm by

$$\left\|\widehat{\mathcal{G}}_{t,d,j}(\theta_*) - \mathcal{G}_{t,d,j}(\theta_*)\right\|_{\mathrm{op}} + \frac{1}{\zeta_{t,j}} \left(\left\| \sum_{\alpha \neq \alpha_*} (-1)^{j\cdot(\alpha-\alpha_*)} \mathcal{A}_{t,d}\left(P^{-1}(\alpha,\theta_*)\right) \right\|_{\mathrm{op}} + \|\mathcal{R}_{t,d}(u_*)\|_{\mathrm{op}} \right) = O\left(\mu_{t,d} n^{-1}\right)$$

where we use Lemma 5.5 and Eq. (22) together with $\zeta_{t,j} = 1 + O(2^{-\Omega(n)})$ from Lemma 5.3.

Let q be the unit eigenvector corresponding to the largest eigenvalue of $\sigma\widehat{\mathcal{G}}_{t,d,j}(\theta_*)$, which must be one of q_+ (if $\sigma = 1$) or q_- (if $\sigma = -1$). Hence $q^T\sigma\widehat{\mathcal{G}}_{t,d,j}(\theta_*)q \geq \widehat{v}^T\sigma\widehat{\mathcal{G}}_{t,d,j}(\theta_*)\widehat{v}$. Plugging $\sigma\widehat{\mathcal{G}}_{t,d,j}(\theta_*) = \frac{\mu_{t,d}}{\zeta_{t,j}}\widehat{v}\widehat{v}^T + aI_n + E$, we obtain the following inequality

$$\frac{\mu_{t,d}}{\zeta_{t,j}}\left(1 - \langle q, \widehat{v}\rangle^2\right) \leq 2\left\|E\right\|_{\mathrm{op}} = O\left(\mu_{t,d}n^{-1}\right).$$

Thus $1 - \langle q, \widehat{v}\rangle^2 = O(n^{-1})$. Choosing $\tau \in \{-1, 1\}$ such that $\langle q, \tau\widehat{v}\rangle \geq 0$ gives $\|q - \tau\widehat{v}\| = O(n^{-1/2})$.

Finally, $\|dq - \tau v\| \leq d\|q - \tau\widehat{v}\| + d - \lambda = O(\lambda n^{-1/2}) < \lambda n^{-1/3}$ for all sufficiently large n. The final claim follows from Theorem 2.4. □

5.5 The affine-coset algorithm

This section presents the algorithm using the affine cosets. The overall time complexity is $2^{0.7314n} \approx 2^{(0.5+t)n+o(n)}$ by balancing $2^{\max(0.5,\chi+2t)n+o(n)}$ of the DGS sampling from Lemma 5.4 and $2^{\ell+o(n)} = 2^{(1-\chi)n+o(n)}$ of the number of $\theta \in \mathbb{F}_2^\ell$, where $h = \lfloor \chi n \rfloor$, $t_0 = 0.2314\ldots < t < 1/4$ and $N_t = 2^{(2t+o(1))n}$. This can be obtained using the Walsh-Hadamard transform as before, and the other steps take much smaller time.

Algorithm 3: Affine-coset Hessian SVP

1. Construct the preprocessing $n^{-1/3}$-BDD data and the scale grid from Algorithm 1. Choose $t_0 < t < 1/4$, and put $\chi = 1/2 - t$.
2. At every scale d, choose independent uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$. Apply the sampling algorithm from Lemma 5.4. Abort the scale d unless N_t samples from D_{Λ_j,ξ_t} are obtained.
3. Let $b = \lfloor n/2 \rfloor$. For every accepted sample X_i, define

$$V_P(X_i) = (k_i', k_i'') \in \mathbb{F}_2^b \times \mathbb{F}_2^{\ell-b}, \qquad W_i := -\frac{4\pi^2}{N_t}X_iX_i^T\mathbf{1}_{\{\|X_i\| \leq C_1\xi_t\sqrt{n}\}}.$$

   For each $\theta'' \in \mathbb{F}_2^{\ell-b}$, do the following:

   (a) Construct the array $A_{\theta''}(x) := \sum_{\substack{1\leq i\leq N_t \\ k_i'=x}}(-1)^{\theta''\cdot k_i''}W_i$ for all $x \in \mathbb{F}_2^b$.

   (b) Apply the matrix-valued Walsh-Hadamard transform to obtain for every $\theta' \in \mathbb{F}_2^b$:

$$\sum_{x\in\mathbb{F}_2^b}(-1)^{\theta'\cdot x}A_{\theta''}(x) = \sum_{i=1}^{N_t}(-1)^{\theta'\cdot k_i' + \theta''\cdot k_i''}W_i = \widehat{\mathcal{G}}_{t,d,j}(\theta', \theta''). \tag{23}$$

   (c) For each output in Eq. (23), compute unit eigenvectors q_+ and q_- corresponding to its largest and smallest eigenvalues and query BDD at $dq_\pm$. Store a nonzero output in $\mathcal{L}$. Keep only the shortest vector found, and discard the other vectors and temporal data.

4. Return the shortest vector over all scales.

Theorem 5.8. *Algorithm 3 solves Search-SVP with constant success probability in time* $2^{0.7314n+o(n)}$ *and space* $2^{n/2+o(n)}$.

Proof. Suppose $\lambda_1(\mathcal{L}) \leq d \leq (1 + 1/n)\lambda_1(\mathcal{L})$ and let v a shortest vector. Choose $u_* \in \mathbb{F}_2^n$ such that $v \in Bu_* + 2\mathcal{L}$, write $Pu_* = (\alpha_*, \theta_*)$, and write $\theta_* = (\theta_*', \theta_*'') \in \mathbb{F}_2^b \times \mathbb{F}_2^{\ell-b}$.

By Lemma 5.4, the algorithm obtains the specified samples from D_{Λ_j,ξ_t}. By Eq. (23), the iteration indexed by θ_*'' computes $\widehat{\mathcal{G}}_{t,d,j}(\theta_*', \theta_*'') = \widehat{\mathcal{G}}_{t,d,j}(\theta_*)$. The conclusion of Lemma 5.5 holds simultaneously for

every $\theta \in \mathbb{F}_2^\ell$, and Lemma 5.7 shows the correctness. The statistical distance in Lemma 5.4 only changes the success probability by $\exp\bigl(-\Omega(n^2)\bigr)$.

We analyze the time and space complexity. Sampling takes $2^{(\max\{1/2,\chi+2t\}+o(1))n}$ time by Lemma 5.4. There are $2^{\ell-b}$ sequential iterations. In each iteration, constructing the array takes $N_t\,\mathrm{poly}(n)$ time, and its Walsh-Hadamard transform, eigenvector computations, and 2^{b+1} BDD queries take $2^{b+o(n)}$ time. Since $N_t = 2^{(2t+o(1))n} < 2^{b+o(n)} = 2^{0.5n+o(n)}$, their total time is $2^{\ell+o(n)}$. Since $\ell = (1-\chi)n + O(1)$ and $\chi = 1/2 - t$, the time exponent is $\max\{\chi+2t, 1-\chi\} = 1/2 + t \to 0.7314$ for $t \to t_0$. The space complexity is dominated by the number of accepted samples and the preprocessing BDD data, which is $2^{n/2+o(n)}$. □

6 Importance Sampling on an Affine Lattice Coset

This section improves the algorithm in Section 5 by sampling from D_{Λ_j,ξ_R} and estimating the Hessian defined using the smaller width ξ_r.

Fix constants $0 < r < R$ with $r < 1/4$ and $0 < \chi < 1/2$. Let $h = \lfloor \chi n \rfloor$ and $\ell = n - h$ as in Section 5.2. Recall $g(t) = \frac{1}{2}\log_2(t/t_0)$. We use the parameters R and r for the source and target widths, respectively.

6.1 Lattice points in a shortest parity class

We first study the number of lattice points in the parity classes to improve some inequalities and conditions in some lemmas. For $0 < \varphi < \pi/2$, define the Kabatiansky-Levenshtein constant

$$B_{\mathrm{KL}}(\varphi) := \frac{1+\sin\varphi}{2\sin\varphi}\log_2\frac{1+\sin\varphi}{2\sin\varphi} - \frac{1-\sin\varphi}{2\sin\varphi}\log_2\frac{1-\sin\varphi}{2\sin\varphi}.$$

Note that β in Lemma 2.1 is such that $\log_2\beta = B_{\mathrm{KL}}(\pi/3)$ as shown in [35]. We use the following bound [20].

Theorem 6.1. *Let $0 < \varphi < \pi/2$ be fixed. If $S \subseteq \{x \in \mathbb{R}^n : \|x\| = 1\}$ and the angle between every two distinct vectors in S is at least φ, then $|S| \le 2^{(B_{\mathrm{KL}}(\varphi)+o(1))n}$. The $o(1)$ term is uniform when φ ranges over a compact subinterval of $(0, \pi/2)$.*

The counting argument in the following lemma is a variant of [35, Lemma 3]. The restriction to $v + 2\mathcal{L}$ gives the stronger separation used below.

Lemma 6.2. *Let v be a shortest vector and put $\lambda = \|v\|$. Every $w \in v + 2\mathcal{L} \setminus \{\pm v\}$ satisfies $\|w\| \ge \sqrt{3}\lambda$. Moreover, for every fixed $C \ge \sqrt{3}$, uniformly for $\sqrt{3} \le x \le C$,*

$$|\{w \in v + 2\mathcal{L} \setminus \{\pm v\} : \|w\| \le x\lambda\}| \le 2^{(B_{\mathrm{KL}}(\arccos(1-2/x^2))+o(1))n}.$$

Proof. Write $w = 2y - v$. If $w \ne \pm v$, then y and $y - v$ are nonzero lattice vectors. The parallelogram identity gives $\|w\|^2 + \lambda^2 = 2\|y\|^2 + 2\|y - v\|^2 \ge 4\lambda^2$.

For distinct $w, w' \in v + 2\mathcal{L}$, the vector $(w - w')/2$ is a nonzero lattice vector, and hence $\|w - w'\| \ge 2\lambda$. For each vector with $\|w\| \le x\lambda$, append the coordinate $\sqrt{x^2\lambda^2 - \|w\|^2}$. The resulting vectors in $\mathbb{R}^{n+1}$ have norm $x\lambda$ and pairwise distance at least 2λ. After normalization, their pairwise angles are at least $\arccos(1 - 2/x^2)$. The result follows from Theorem 6.1. □

We improve the bound on $\rho_{2\sqrt{2}/\xi_t}(\mathcal{L}) \le 1 + 2^{(1-g(t)+o(1))n}$ in Corollary 3.2 (i.e. Eq. (16)) using a more fine-grained counting argument with the vectors in each coset modulo $2\mathcal{L}$. For $x \ge 1$, define

$$K_2(x) := \begin{cases} 0, & 1 \le x \le \sqrt{2}, \\ B_{\mathrm{KL}}(\arccos(1 - 2/x^2)), & x > \sqrt{2}, \end{cases}$$

Combining Theorem 6.1 with Lemma 2.1, the number of nonzero vectors in $\mathcal{L}$ of norm at most $x\lambda$ is

$$N_{\mathcal{L}}(x\lambda_1(\mathcal{L})) \le 2^{(\min\{\log_2\beta + \log_2 x, 1 + K_2(x)\}+o(1))n}.$$

Recall $g(t) = \frac{1}{2}\log_2(t/t_0)$ for $t_0 = 0.23147\ldots$. Define[4]

$$g_2(R) := 1 - \sup_{x \ge 1}\left(\min\left\{\log_2 \beta + \log_2 x, 1 + K_2(x)\right\} - \frac{R}{2}x^2\right) \ge g(R). \tag{24}$$

Lemma 6.3. *Let $\lambda = \lambda_1(\mathcal{L}) \le d \le (1+1/n)\lambda$. For every fixed $R > 0$, $\rho_{2\sqrt{2}/\xi_R}(\mathcal{L} \setminus \{0\}) \le 2^{(1-g_2(R)+o(1))n}$.*

Note that the proof of this lemma is rather complicated and somewhat independent to the other arguments. The readers may use a simpler bound with g. All arguments in this section works well by using g in place of g_2, which gives a only slightly bad complexity of $2^{0.6040n+o(n)}$ time and space. See Remark 6.10.

Proof. We first bound the number of short vectors in one coset modulo $2\mathcal{L}$. Fix a coset $C \in \mathcal{L}/2\mathcal{L}$ and put

$$S_C(x) := \{w \in C \setminus \{0\} : \|w\| \le x\lambda\}, \qquad m := |S_C(x)|.$$

For every $w \in S_C(x)$, define $u_w := \frac{1}{x\lambda}\left(w, \sqrt{x^2\lambda^2 - \|w\|^2}\right) \in \mathbb{R}^{n+1}$. Then $\|u_w\| = 1$. If $w, w' \in S_C(x)$ are distinct, then $(w - w')/2$ is a nonzero vector in $\mathcal{L}$, and hence $\|w - w'\| \ge 2\lambda$. It follows that

$$\|u_w - u_{w'}\| \ge \frac{\|w - w'\|}{x\lambda} \ge \frac{2}{x} \quad \Longrightarrow \quad \langle u_w, u_{w'}\rangle = 1 - \frac{1}{2}\|u_w - u_{w'}\|^2 \le 1 - \frac{2}{x^2}.$$

Suppose first that $1 \le x \le \sqrt{2}$. Then $\langle u_w, u_{w'}\rangle \le 0$ for distinct w, w'. Let $G = (\langle u_w, u_{w'}\rangle)_{w,w' \in S_C(x)}$ be the Gram matrix of these unit vectors. Since $\operatorname{rank}(G) \le n+1$ and $\operatorname{tr}(G) = m$, the Cauchy-Schwarz inequality for the eigenvalues of G gives

$$\|G\|_{\mathrm{F}}^2 \ge \frac{\operatorname{tr}(G)^2}{\operatorname{rank}(G)} \ge \frac{m^2}{n+1}.$$

On the other hand, $-1 \le \langle u_w, u_{w'}\rangle \le 0$ for $w \ne w'$, and hence $\langle u_w, u_{w'}\rangle^2 \le -\langle u_w, u_{w'}\rangle$. Moreover,

$$0 \le \left\| \sum_{w \in S_C(x)} u_w \right\|^2 = m + \sum_{\substack{w,w' \in S_C(x) \\ w \ne w'}} \langle u_w, u_{w'}\rangle.$$

Consequently,

$$\|G\|_{\mathrm{F}}^2 = m + \sum_{\substack{w,w' \in S_C(x) \\ w \ne w'}} \langle u_w, u_{w'}\rangle^2 \le m - \sum_{\substack{w,w' \in S_C(x) \\ w \ne w'}} \langle u_w, u_{w'}\rangle \le 2m.$$

Combining the two bounds on $\|G\|_{\mathrm{F}}^2$ gives $m \le 2(n+1) = 2^{o(n)}$.

Now consider $x > \sqrt{2}$. The preceding inner-product bound shows that the angle between every two distinct vectors $u_w, u_{w'}$ is at least $\arccos(1 - 2/x^2)$. Applying Theorem 6.1 in dimension $n+1$ gives

$$|S_C(x)| \le 2^{(K_2(x)+o(1))n}.$$

Combining two cases, each coset modulo $2\mathcal{L}$ contains $2^{(K_2(x)+o(1))n}$ vectors of norm $\le x\lambda$ for every fixed $x \ge 1$. Summing over 2^n cosets modulo $2\mathcal{L}$, giving $2^{(1+K_2(x)+o(1))n}$. Together with Lemma 2.1, we have

$$N_{\mathcal{L}}(x\lambda) \le 2^{(\min\{\log_2\beta + \log_2 x, 1+K_2(x)\}+o(1))n} = 2^{A(x)n+o(n)}$$

[4]The inequality is not obvious. We sketch the proof for $R < 1/\ln 2$ which is only relevant to our analysis. In this case, dropping the second term from the minimum gives $g_2(R) \ge 1 - \sup_{x\ge1}(\log_2\beta + \log_2 x - Rx^2/2)$. The supremum is attained at $x = 1/\sqrt{R\ln 2}$, and the resulting lower bound equals $g(R)$.

for $A(x) = \min\{\log_2 \beta + \log_2 x, 1 + K_2(x)\}$. We now estimate the Gaussian mass. Put $x_k = (1+1/n)^k$ and $S_k := \{w \in \mathcal{L} : x_k\lambda \le \|w\| < x_{k+1}\lambda\}$. Using $\rho_{2\sqrt{2}/\xi_R}(w) = 2^{-(R/2)n\|w\|^2/d^2}$, we have

$$\rho_{2\sqrt{2}/\xi_R}(\mathcal{L}\setminus\{0\}) \le \sum_{k\ge 0} N_{\mathcal{L}}(x_{k+1}\lambda) 2^{-(R/2)n(\lambda^2/d^2)x_k^2}.$$

Fix $\delta > 0$ and a constant $X_0 > 1$. Partition $[1, X_0]$ into finitely many intervals $[a, b]$ such that $R(b^2-a^2)/2 \le \delta$. For each interval, consider $w \in \mathcal{L}$ satisfying $a\lambda \le \|w\| < b\lambda$. Their number is at most $2^{(A(b)+o(1))n}$, while each of their Gaussian weights is at most $2^{-(R/2)n(\lambda^2/d^2)a^2}$. Hence their total contribution is at most

$$2^{\left(A(b)-Ra^2/2+o(1)\right)n} \le 2^{\left(\sup_{x\ge 1} F_R(x)+\delta+o(1)\right)n}$$

where $F_R(x) := A(x) - \frac{R}{2}x^2$. The last upper bound is independent of the interval, and the number of intervals in the decomposition of $[1, X_0]$ is independent of n, so their total contribution satisfies the same bound.

It remains to consider $x_k > X_0$. Here we use only Lemma 2.1. Apart from its uniform $2^{o(n)}$ factor, the upper bound for the k-th shell is

$$U_k := \beta^n x_{k+1}^n 2^{-(R/2)n(\lambda^2/d^2)x_k^2} \quad\Longrightarrow\quad \frac{U_{k+1}}{U_k} \le e\, 2^{-R(\lambda^2/d^2)x_k^2(1+1/(2n))}$$

since $x_{k+1} = (1+1/n)x_k$. Choose X_0 sufficiently so large that this ratio is at most $1/2$ for all sufficiently large n and simultaneously the first such shell is bounded by $2^{(\sup_{x\ge1} F_R(x)+\delta+o(1))n}$, which is possible because $\log_2\beta + \log_2 x - Rx^2/2$ tends to $-\infty$ as $x\to\infty$. The bounds for the remaining shells form a geometric series and hence have the same total upper bound.

Letting δ tend to zero gives

$$\rho_{2\sqrt{2}/\xi_R}(\mathcal{L}\setminus\{0\}) \le 2^{(\sup_{x\ge1} F_R(x)+o(1))n} = 2^{(1-g_2(R)+o(1))n},$$

where the last equality follows from Eq. (24). □

The following lemma shows a spectral phenomena for the Hessian, or more precisely its translated version $\mathcal{A}$ in Eq. (11), similar to Lemma 3.4 even if $r < t_0 = 0.23147\ldots$. We choose the lower bound $9/50 = 0.18$ only to ensure

$$\sup_{x\ge\sqrt{3}} \left\{B_{\mathrm{KL}}(\arccos(1-2/x^2)) - r(x^2-1)\right\} < -\frac{1}{10}, \tag{25}$$

which follows by standard calculus. Also note that $B_{\mathrm{KL}}(\arccos(1-2/x^2))$ is Lipschitz in the interval $[\sqrt{3}, 7]$. We define $\mu_{r,d} = \frac{2\pi^2\xi_r^4\lambda^2}{\rho_{1/\xi_r}(\mathcal{L})} e^{-\pi\xi_r^2\lambda^2/4}$ for any r.

Lemma 6.4. *Let $\lambda = \lambda_1(\mathcal{L}) \le d \le (1+1/n)\lambda$. Let $v \in \mathcal{L}$ satisfy $\|v\| = \lambda$, let $u \in \mathbb{F}_2^n$ satisfy $v \in Bu + 2\mathcal{L}$, and put $\widehat{v} = v/\lambda$. If $r \ge 9/50$, then*

$$\left\|\mathcal{A}_{r,d}(u) - \mu_{r,d}\widehat{v}\widehat{v}^T\right\|_{\mathrm{F}} \le \mu_{r,d} 2^{-\Omega(n)}.$$

Proof. The terms $w = \pm v$ in Eq. (11) give $\mu_{r,d}\widehat{v}\widehat{v}^T$. The sum of the other terms is positive semidefinite, so its Frobenius norm is at most its trace. Consequently,

$$\frac{\left\|\mathcal{A}_{r,d}(u) - \mu_{r,d}\widehat{v}\widehat{v}^T\right\|_{\mathrm{F}}}{\mu_{r,d}} \le \frac{1}{2}\sum_{\substack{w\in v+2\mathcal{L}\\ w\ne\pm v}} \left(\frac{\|w\|}{\lambda}\right)^2 2^{-rn(\lambda^2/d^2)((\|w\|/\lambda)^2-1)}.$$

Put $x_k := \sqrt{3}(1+1/n)^k$ and let $S_k := \left\{w \in v+2\mathcal{L}\setminus\{\pm v\} : x_k \le \frac{\|w\|}{\lambda} < x_{k+1}\right\}$. Then

$$\frac{1}{2}\sum_{\substack{w\in v+2\mathcal{L}\\ w\ne\pm v}} \left(\frac{\|w\|}{\lambda}\right)^2 2^{-rn(\lambda^2/d^2)((\|w\|/\lambda)^2-1)} \le \frac{1}{2}\sum_{k\ge0} |S_k| x_{k+1}^2 2^{-rn(\lambda^2/d^2)(x_k^2-1)}.$$

We will show that the summations for $x_k < 7$ and $x_k \ge 7$ both are bounded above by $2^{-\Omega(n)}$.

First, if $x_k < 7$, Lemma 6.2 gives the bound

$$\frac{1}{n}\log_2\left(|S_k|x_{k+1}^2 2^{-rn(\lambda^2/d^2)(x_k^2-1)}\right) \le (B_{\mathrm{KL}}(\arccos(1-2/x_{k+1}^2)) + o(1)) - \frac{r\lambda^2(x_k^2-1)}{d^2}$$

and using the fact that $B_{\mathrm{KL}}(\arccos(1-2/x^2))$ is Lipschitz, $x_{k+1} = x_k + x_k/n = x_k + O(1/n)$ for $x_k < 7$, and $d/\lambda = 1 + O(1/n)$ gives an upper bound

$$B_{\mathrm{KL}}(\arccos(1-2/x_k^2)) - r(x_k^2-1) + o(1) \le -\frac{1}{7} + o(1)$$

where we use Eq. (25), proving $\frac{1}{2}|S_k|x_{k+1}^2 2^{-rn(\lambda^2/d^2)(x_k^2-1)} \le 2^{(-1/7+o(1))n}$. Since there are at most $O(n)$ k's such that $x_k < 7$, it gives the desire bound for the first summand over $x_k < 7$.

Now consider $x_k \ge 7$. Since $d \le (1+1/n)\lambda$, $r\lambda^2/d^2 \ge 1/6$ for sufficiently large n. Lemma 2.1 on $|S_k|$ gives

$$\frac{1}{n}\log_2\left(|S_k|2^{-rn(\lambda^2/d^2)(x_k^2-1)}\right) \le \log_2\beta + \log_2 x_{k+1} - \frac{x_k^2-1}{6} + o(1)$$

Since $x_{k+1} = (1+1/n)x_k \le 2x_k$, the exponent is at most

$$\log_2\beta + \log_2(2x_k) - \frac{1}{6}(x_k^2-1) + o(1) \le -\frac{x_k^2}{14} + o(1)$$

which holds for $x_k \ge 7$. Applying this inequality with $x = x_k$ proves

$$\frac{1}{2}|S_k|x_{k+1}^2 2^{-rn(\lambda^2/d^2)(x_k^2-1)} \le \frac{x_{k+1}^2}{2}2^{-nx_k^2/14+o(n)}.$$

This term (ignoring the uniform $o(1)$ factor) decays (super-)exponentially fast,[5] thus the asymptotic bound of the summation of the above term for k such that $x_k \ge 7$ becomes $2^{-\Omega(n)}$. □

6.2 Sampling (wider) discrete Gaussian from an affine lattice coset

This section shows that sampling from the discrete Gaussian D_{Λ_j,ξ_R} for a large ξ_R is more efficient than the one in Lemma 5.4 by removing the multiplicative factor 2^χ in the time complexity.

For the fixed d and R, recall that $\xi_R = \sqrt{4nR\ln 2/(\pi d^2)}$. For $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$, recall from Section 5.2 that $\Lambda_j := \{X \in \mathcal{L}^* : J_P(X) = j\}$. In particular, Λ_0 is a sublattice of $\mathcal{L}^*$ and Λ_j is a coset of Λ_0. The following lemma shows some Gaussian mass bound from Corollary 3.2 works for Λ_0 in expectation.

Lemma 6.5. *Let $\lambda = \lambda_1(\mathcal{L})$ and $\lambda \le d \le (1+1/n)\lambda$. If $R > t_0$ and $\chi < g_2(R)$, then $\rho_{\sqrt{2}/\xi_R}(\Lambda_0^* \setminus \{0\}) \le 2^{-\Omega(n)}$ with probability $1 - 2^{-\Omega(n)}$ over uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$. Consequently, $\xi_R > \sqrt{2}\,\eta_{1/2}(\Lambda_0)$.*

Proof. Let $\mathcal{K}_P = P^{-1}(\mathbb{F}_2^h \times \{0\})$. We first observe that $\Lambda_0^* = \mathcal{L} + B\mathcal{K}_P/2$. For every nonzero $\kappa \in \mathbb{F}_2^n$, it is easy to see that $\mathbb{P}_P[\kappa \in \mathcal{K}_P] = (2^h-1)/(2^n-1)$. Since the cosets $\mathcal{L} + B\kappa/2$ for $\kappa \in \mathbb{F}_2^n$ partition $\mathcal{L}/2$, we have

$$\begin{aligned}\mathbb{E}_P\rho_{\sqrt{2}/\xi_R}(\Lambda_0^* \setminus \{0\}) &\le \rho_{\sqrt{2}/\xi_R}(\mathcal{L} \setminus \{0\}) + 2^{h-n+o(n)}\left(\rho_{2\sqrt{2}/\xi_R}(\mathcal{L}) - \rho_{\sqrt{2}/\xi_R}(\mathcal{L})\right)\\ &\le 2^{(-g(R)+o(1))n} + 2^{(\chi-g_2(R)+o(1))n}.\end{aligned}$$

The last inequality follows from Eq. (16) with $c = 0$ and Lemma 6.3. Since $\chi < g_2(R)$ and $R > t_0$, Markov's inequality in Lemma 2.9 gives $\rho_{\sqrt{2}/\xi_R}(\Lambda_0^* \setminus \{0\}) = 2^{-\Omega(n)} < 1/2$ except with probability $2^{-\Omega(n)}$. The last claim follows from the definition and monotonicity of the smoothing parameter. □

[5] In particular, for $x_k \ge 7$, $(x_{k+2}^2 2^{-nx_{k+1}^2/14})/(x_{k+1}^2 2^{-nx_k^2/14}) = (1+1/n)^2 2^{-nx_{k+1}^2(2/n-1/n^2)/14} \le 1/2$ holds.

Given this lemma, we can apply Theorem 2.7 to sample the discrete Gaussian on Λ_j. More precisely, we define $\Gamma := \Lambda_0 + \mathbb{Z}x = \Lambda_0 \sqcup \Lambda_j$ for $j \neq 0$ and $x \in \Lambda_j$, and run the algorithm from Theorem 2.7 on Γ. Collecting the samples in Λ_j works well. This gives the following lemma.

Lemma 6.6. *Let $\lambda = \lambda_1(\mathcal{L})$, $\lambda \le d \le (1+1/n)\lambda$, $\chi < g_2(R)$, $R > t_0$, and $M = 2^{O(n)}$. With probability $1 - 2^{-\Omega(n)}$ over uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$, the following holds for every $j \in \mathbb{F}_2^h$. There is an algorithm that outputs M vectors or aborts, and its abort probability is at most $2^{-\Omega(n)}$. If it does not abort, the joint distribution of its outputs has statistical distance at most $\exp\bigl(-\Omega(n^2)\bigr)$ from $D^M_{\Lambda_j,\xi_R}$. The algorithm takes time $(M + 2^{n/2})2^{o(n)}$ and space $2^{n/2+o(n)}$.*

Proof. Note that $\rho_{\sqrt{2}/\xi_R}(\Lambda_0^* \setminus \{0\}) \le 2^{-\Omega(n)}$ and $\xi_R > \sqrt{2}\,\eta_{1/2}(\Lambda_0)$ holds except with probability $2^{-\Omega(n)}$ over P because of Lemma 6.5. Fix any such P.

For $j = 0$, we can directly use Theorem 2.7 $\lceil M/2^{n/2} \rceil$ times to Λ_0 with parameter ξ_R and $\kappa = n^2$. Suppose that $j \neq 0$. Solve the linear equations $J_P(x) = j$ over $\mathbb{F}_2$ and lift a solution to obtain $x \in \mathcal{L}^*$ satisfying $x \in \Lambda_j$. Let $\Gamma := \Lambda_0 + \mathbb{Z}x = \Lambda_0 \sqcup \Lambda_j$. Since $\Gamma^* \subseteq \Lambda_0^*$, we have $\rho_{\sqrt{2}/\xi_R}(\Gamma^* \setminus \{0\}) \le 2^{-\Omega(n)}$, and hence $\xi_R > \sqrt{2}\,\eta_{1/2}(\Gamma)$ for large n.

Choose arbitrary $y \in \Lambda_0^* \setminus \Gamma^*$. Poisson summation formula gives

$$\left|\mathbb{E}_{X\sim D_{\Gamma,\xi_R}} e^{2\pi i\langle X,y\rangle}\right| = \frac{\rho_{1/\xi_R}(\Gamma^* - y)}{\rho_{1/\xi_R}(\Gamma^*)} \le \rho_{\sqrt{2}/\xi_R}(\Lambda_0^* \setminus \{0\}) = 2^{-\Omega(n)}.$$

The character $e^{2\pi i\langle X,y\rangle}$ equals 1 for every $X \in \Lambda_0$. Moreover, $2x \in \Lambda_0$, while $y \notin \Gamma^*$, so $\langle x, y\rangle \in \mathbb{Z} + 1/2$. Hence the character equals -1 for every $X \in x + \Lambda_0 = \Lambda_j$. The above inequality gives $|1 - 2p| \le 2^{-\Omega(n)}$, where $p = \mathbb{P}_{X\sim D_{\Gamma,\xi_R}}[X \in \Lambda_j]$. It follows that $p = 1/2 + O(2^{-\Omega(n)})$.

We obtain the following sampling algorithm: Use Theorem 2.7 on Γ to generate $4(M+n)$ candidates, with parameter ξ_R and $\kappa = n^2$. Collect the first M candidates in Λ_j and abort if fewer than M are obtained. The Chernoff bound in Lemma 2.12 gives abort probability $2^{-\Omega(n)}$.

For exact samples from D_{Γ,ξ_R}, conditioning on $X \in \Lambda_j$ gives exactly D_{Λ_j,ξ_R}. There are at most $2^{O(n)}$ calls to Theorem 2.7, so their total statistical distance is still $\exp\bigl(-\Omega(n^2)\bigr)$. Applying the collection algorithm does not increase statistical distance, and conditioning on its non-abort event changes it by at most a factor $1 + 2^{-\Omega(n)}$. Finally, the calls can be made sequentially, giving time $(M + 2^{n/2})2^{o(n)}$ and space $2^{n/2+o(n)}$. □

6.3 Estimating the Hessian by importance sampling

Importance sampling estimates an expectation under a target distribution using samples from another distribution. Here we sample from D_{Λ_j,ξ_R} for large ξ_R and estimate the Hessian defined using D_{Λ_j,ξ_r} for the smaller target ξ_r. For $X \in \Lambda_j$, let $w(X) := \rho_{\xi_r}(X)/\rho_{\xi_R}(X)$ and $\overline{w}_j := \mathbb{E}_{X\sim D_{\Lambda_j,\xi_R}}[w(X)] = \rho_{\xi_r}(\Lambda_j)/\rho_{\xi_R}(\Lambda_j)$. For every matrix-valued function H for which the expectations exist,

$$\mathbb{E}_{X\sim D_{\Lambda_j,\xi_R}}[w(X)H(X)] = \sum_{X\in\Lambda_j} \frac{\rho_{\xi_R}(X)}{\rho_{\xi_R}(\Lambda_j)} \frac{\rho_{\xi_r}(X)}{\rho_{\xi_R}(X)} H(X) = \overline{w}_j \mathbb{E}_{X\sim D_{\Lambda_j,\xi_r}}[H(X)]. \tag{26}$$

Thus the weighted expectation is $\overline{w}_j \mathcal{G}_{r,d,j}(\theta)$ when $H(X) = -4\pi^2 XX^T(-1)^{\theta\cdot V_P(X)}$.

Let M be the number of sample to be determined later. Consider n independent families $\{X_{a,1}, \dots, X_{a,M}\}$ of samples from D_{Λ_j,ξ_R} for $a \in [n]$. Define the importance-sampling Hessian estimator

$$\widehat{\mathcal{N}}_{a,j}(\theta) := -\frac{4\pi^2}{M} \sum_{i=1}^{M} w(X_{a,i}) X_{a,i} X_{a,i}^T (-1)^{\theta\cdot V_P(X_{a,i})} \mathbf{1}_{\{\|X_{a,i}\| \le C_1 \xi_R \sqrt{n}\}}. \tag{27}$$

For $p, q \in [n]$, define the (p,q)-entry of $\widehat{\mathcal{N}}_j(\theta)$ to be the median (i.e., $(\lfloor n/2 \rfloor + 1)$-st smallest) among the corresponding entries of $\widehat{\mathcal{N}}_{1,j}(\theta), \dots, \widehat{\mathcal{N}}_{n,j}(\theta)$. We will show the median estimator $\widehat{\mathcal{N}}_j(\theta)$ approximates $\overline{w}_j \mathcal{G}_{r,d,j}(\theta)$ well for uniform random P and j.

We need some notations. Define $s > 0$ by $1/s = 2/r - 1/R$. Then

$$\frac{1}{\xi_s^2} = \frac{2}{\xi_r^2} - \frac{1}{\xi_R^2}, \qquad \frac{\rho_{\xi_r}(X)^2}{\rho_{\xi_R}(X)} = \rho_{\xi_s}(X), \qquad s = \frac{rR}{2R-r} < r < R. \tag{28}$$

We also recall μ from Eq. (12), and choose the number of samples M using ι defined below:

$$\mu_{r,d} = \frac{2\pi^2\xi_r^4\lambda^2}{\rho_{1/\xi_r}(\mathcal{L})} e^{-\pi\xi_r^2\lambda^2/4}, \qquad M = \lceil n^6 2^{(\iota(r,R)+2r)n} \rceil \quad \text{for} \quad \iota(r,R) := \frac{1}{2}\log_2 \frac{R^2}{r(2R-r)} \tag{29}$$

where $\mu_{r,d} = \Theta(\xi_r^4\lambda^2 2^{-rn})$ if $\lambda \le d \le (1+1/n)\lambda$ and $s > t_0/2$.

Lemma 6.7. *Let $\lambda = \lambda_1(\mathcal{L})$ and $\lambda \le d \le (1+1/n)\lambda$. If $s > t_0/2$ and $\chi < 1/2 + g(r)$, then, the following hold with probability $1 - O(n^{-1})$ over uniform random $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$:*

$$\zeta_{r,j} \ge \frac{1}{2}, \qquad \overline{w}_j = \left(\frac{r}{R}\right)^{n/2}\left(1 + O(2^{-\Omega(n)})\right), \qquad \frac{\mathbb{E}_{X\sim D_{\Lambda_j,\xi_R}}[w(X)^2]}{\overline{w}_j^2} \le 2n2^{\iota(r,R)n}.$$

Proof. Since $s > t_0/2$ and $s < r < R$, Corollary 3.2 gives $\rho_{1/\xi_s}(\mathcal{L}), \rho_{1/\xi_r}(\mathcal{L}), \rho_{1/\xi_R}(\mathcal{L}) = 1 + 2^{-\Omega(n)}$. Since g is increasing, Lemma 5.3 gives $\zeta_{r,j}, \zeta_{R,j} = 1 + O(2^{-\Omega(n)})$ simultaneously for every j, except with probability $2^{-\Omega(n)}$ over P. In particular, $\zeta_{r,j} \ge 1/2$ for all sufficiently large n.

By the definitions of $\overline{w}_j$ and $\zeta_{t,j}$, and by applying the Poisson summation formula to $\mathcal{L}^*$, we have

$$\overline{w}_j = \frac{\zeta_{r,j}}{\zeta_{R,j}}\frac{\rho_{\xi_r}(\mathcal{L}^*)}{\rho_{\xi_R}(\mathcal{L}^*)} = \left(\frac{r}{R}\right)^{n/2}\frac{\rho_{1/\xi_r}(\mathcal{L})}{\rho_{1/\xi_R}(\mathcal{L})}\frac{\zeta_{r,j}}{\zeta_{R,j}} = \left(\frac{r}{R}\right)^{n/2}\left(1 + O(2^{-\Omega(n)})\right)$$

where we use the approximations of ρ and ζ above. This proves the estimate of $\overline{w}_j$ in the statement.

Fix a choice of P satisfying the preceding conclusions. Since $2^{-h}\sum_j \zeta_{s,j} = 1$ and $\zeta_{R,j}/\zeta_{r,j}^2 = 1 + 2^{-\Omega(n)}$ uniformly in j, the expectation over uniform j of $\zeta_{s,j}\zeta_{R,j}/\zeta_{r,j}^2$ is $1 + 2^{-\Omega(n)}$. Markov's inequality gives

$$\frac{\zeta_{s,j}\zeta_{R,j}}{\zeta_{r,j}^2} \le n$$

except with probability $O(n^{-1})$ over uniform j.

The definitions of $w(X)$ and D_{Λ_j,ξ_R}, together with Eq. (28), give

$$\mathbb{E}_{X\sim D_{\Lambda_j,\xi_R}}[w(X)^2] = \frac{1}{\rho_{\xi_R}(\Lambda_j)}\sum_{X\in\Lambda_j}\frac{\rho_{\xi_r}(X)^2}{\rho_{\xi_R}(X)} = \frac{\rho_{\xi_s}(\Lambda_j)}{\rho_{\xi_R}(\Lambda_j)}.$$

Also, $\overline{w}_j = \rho_{\xi_r}(\Lambda_j)/\rho_{\xi_R}(\Lambda_j)$. Consequently, we can compute $\mathbb{E}_{X\sim D_{\Lambda_j,\xi_R}}[w(X)^2]/\overline{w}_j^2$ as follows:

$$\frac{\rho_{\xi_s}(\Lambda_j)\rho_{\xi_R}(\Lambda_j)}{\rho_{\xi_r}(\Lambda_j)^2} = \frac{\rho_{\xi_s}(\mathcal{L}^*)\rho_{\xi_R}(\mathcal{L}^*)}{\rho_{\xi_r}(\mathcal{L}^*)^2}\frac{\zeta_{s,j}\zeta_{R,j}}{\zeta_{r,j}^2} = \left(\frac{sR}{r^2}\right)^{n/2}\frac{\rho_{1/\xi_s}(\mathcal{L})\rho_{1/\xi_R}(\mathcal{L})}{\rho_{1/\xi_r}(\mathcal{L})^2}\frac{\zeta_{s,j}\zeta_{R,j}}{\zeta_{r,j}^2}.$$

Here the first equality follows by writing $\rho_{\xi_t}(\Lambda_j) = 2^{-h}\zeta_{t,j}\rho_{\xi_t}(\mathcal{L}^*)$, and the second equality is obtained by applying Poisson summation to $\mathcal{L}^*$. Plugging $s = rR/(2R-r)$ and $R/\sqrt{r(2R-r)} = 2^{\iota(r,R)}$ gives the multiplicative factor $2^{\iota(r,R)n}$. The last two factors are at most 2 and n, respectively, for all sufficiently large n. This proves the result. □

Lemma 6.8. *Let $\lambda = \lambda_1(\mathcal{L})$ and $\lambda \le d \le (1+1/n)\lambda$. If $s > t_0/2$ and $\chi < 1/2 + g(r)$, then it holds simultaneously for every $\theta \in \mathbb{F}_2^\ell$:*

$$\max_{\theta\in\mathbb{F}_2^\ell}\left\|\widehat{\mathcal{N}}_j(\theta) - \overline{w}_j\mathcal{G}_{r,d,j}(\theta)\right\|_{\mathrm{op}} = O\left(\frac{\overline{w}_j\mu_{r,d}}{n}\right)$$

with probability $1 - O(n^{-1})$ over P, j, and the choices of $X_{a,i}$.

Proof. Suppose that the conclusions of Lemma 6.7 hold, and fix the corresponding pair P, j. For every a, the expectation of $\widehat{\mathcal{N}}_{a,j}(\theta)$ is independent of a. From Eqs. (26) and (27), we have

$$\mathbb{E}[\widehat{\mathcal{N}}_{a,j}(\theta)] = -4\pi^2 \overline{w}_j \mathbb{E}_{X\sim D_{\Lambda_j,\xi_r}}\left[XX^T(-1)^{\theta\cdot V_P(X)}\mathbf{1}_{\{\|X\|\le C_1\xi_R\sqrt{n}\}}\right].$$

Consequently,

$$\widehat{\mathcal{N}}_j(\theta) - \overline{w}_j\mathcal{G}_{r,d,j}(\theta) = \left(\widehat{\mathcal{N}}_j(\theta) - \mathbb{E}[\widehat{\mathcal{N}}_{a,j}(\theta)]\right) + 4\pi^2\overline{w}_j\mathbb{E}_{X\sim D_{\Lambda_j,\xi_r}}\left[XX^T(-1)^{\theta\cdot V_P(X)}\mathbf{1}_{\{\|X\|>C_1\xi_R\sqrt{n}\}}\right].$$

We bound the two terms separately.

We first bound the first term by applying Chebyshev's inequality to each entry. Fix a, θ and the entry indices p, q. Write $(\widehat{\mathcal{N}}_{a,j}(\theta))_{p,q} = M^{-1}\sum_{i=1}^M Y_i$ by defining

$$Y_i := -4\pi^2 w(X_{a,i})(X_{a,i})_p(X_{a,i})_q(-1)^{\theta\cdot V_P(X_{a,i})}\mathbf{1}_{\{\|X_{a,i}\|\le C_1\xi_R\sqrt{n}\}}.$$

This gives the following upper bound

$$\mathbb{E}\left[\left(\widehat{\mathcal{N}}_{a,j}(\theta) - \mathbb{E}[\widehat{\mathcal{N}}_{a,j}(\theta)]\right)^2_{p,q}\right] = \frac{1}{M^2}\sum_{i=1}^M \mathbb{E}\left[(Y_i - \mathbb{E}[Y_i])^2\right] \le \frac{1}{M}\mathbb{E}[Y_1^2].$$

where the cross terms between $Y_i - \mathbb{E}[Y_i]$ and $Y_k - \mathbb{E}[Y_k]$ vanish for $i \ne k$ since the Y_i are independent. Consequently,

$$\mathbb{E}\left[\left(\widehat{\mathcal{N}}_{a,j}(\theta) - \mathbb{E}[\widehat{\mathcal{N}}_{a,j}(\theta)]\right)^2_{p,q}\right] \le \frac{1}{M}\mathbb{E}_{X\sim D_{\Lambda_j,\xi_R}}\left[w(X)^2(4\pi^2X_pX_q)^2\mathbf{1}_{\{\|X\|\le C_1\xi_R\sqrt{n}\}}\right].$$

When the indicator is nonzero, $(4\pi^2X_pX_q)^2 = O(\xi_R^4 n^2) = O(\xi_r^8\lambda^4)$. Hence

$$\mathbb{E}\left[\left(\widehat{\mathcal{N}}_{a,j}(\theta) - \mathbb{E}[\widehat{\mathcal{N}}_{a,j}(\theta)]\right)^2_{p,q}\right] = O\left(\frac{\xi_r^8\lambda^4}{M}\mathbb{E}_{X\sim D_{\Lambda_j,\xi_R}}[w(X)^2]\right) = O\left(\frac{n\overline{w}_j^2\xi_r^8\lambda^4 2^{\iota(r,R)n}}{M}\right) = O\left(\frac{\overline{w}_j^2\mu_{r,d}^2}{n^5}\right)$$

where we use $\mathbb{E}[w(X)^2] \le 2n\overline{w}_j^2 2^{\iota(r,R)n}$ from Lemma 6.7 and the definition of $\mu_{r,d}$ and M from Eq. (29).

Chebyshev's inequality now gives

$$\mathbb{P}\left[\left|\left(\widehat{\mathcal{N}}_{a,j}(\theta) - \mathbb{E}[\widehat{\mathcal{N}}_{a,j}(\theta)]\right)_{p,q}\right| > \frac{\overline{w}_j\mu_{r,d}}{n^2}\right] = O(n^{-1}).$$

Thus each of the n independent estimates of the (p,q)-entry has the required accuracy with probability $1 - O(n^{-1})$. If more than half of them have this accuracy, then their median has the same accuracy. By Lemma 2.13, the probability that the median does not have this accuracy is $2^{-\Omega(n\log n)}$. A union bound over $\theta \in \mathbb{F}_2^\ell$ and $p, q \in [n]$, together with $\|A\|_{\mathrm{op}} \le n\max_{p,q}|A_{p,q}|$, therefore gives

$$\max_{\theta\in\mathbb{F}_2^\ell}\left\|\widehat{\mathcal{N}}_j(\theta) - \mathbb{E}[\widehat{\mathcal{N}}_{a,j}(\theta)]\right\|_{\mathrm{op}} = O\left(\frac{\overline{w}_j\mu_{r,d}}{n}\right)$$

except with probability $2^{-\Omega(n\log n)}$.

It remains to bound the second term. Since $\zeta_{r,j} \ge 1/2$ and its definition $\zeta_{r,j} = 2^h\rho_{\xi_r}(\Lambda_j)/\rho_{\xi_r}(\mathcal{L}^*)$, we have $\rho_{\xi_r}(\Lambda_j) \ge 2^{-h-1}\rho_{\xi_r}(\mathcal{L}^*)$. Since $\xi_R > \xi_r$, Lemma 2.8 shows that the operator norm of the second term in the decomposition is at most $4\pi^2\overline{w}_j 2^{h+1}\xi_r^2 n 2^{-n}$. Dividing this bound by $\overline{w}_j\mu_{r,d}/n$ gives an upper bound $O(n2^{h-(1-r)n}) = 2^{-\Omega(n)}$, because $h \le n/2$ and $r < 1/4$. Therefore, the second term is much smaller:

$$4\pi^2\overline{w}_j\left\|\mathbb{E}_{X\sim D_{\Lambda_j,\xi_r}}\left[XX^T(-1)^{\theta\cdot V_P(X)}\mathbf{1}_{\{\|X\|>C_1\xi_R\sqrt{n}\}}\right]\right\|_{\mathrm{op}} = o(\overline{w}_j\mu_{r,d}/n).$$

Combining the two terms proves the lemma. The total failure probability is the summand of $O(n^{-1})$ from Lemma 6.7 and the additional failure probability $2^{-\Omega(n\log n)}$. □

6.4 Recovering a shortest vector

The following lemma is almost identical to Lemma 5.7, except that we have a normalizing factor $\overline{w}_j$ and we use Lemma 6.4 instead of Lemma 3.4.

Lemma 6.9. *Let $\lambda = \lambda_1(\mathcal{L}) \le d \le (1+1/n)\lambda$. Let $v \in \mathcal{L}$ satisfy $\|v\| = \lambda$, and let $u_* \in \mathbb{F}_2^n$ satisfy $v \in Bu_* + 2\mathcal{L}$. Choose independent uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$, and write $Pu_* = (\alpha_*, \theta_*)$. Assume that $r \ge 9/50$, $\chi < 1/2 + g(r)$, $\chi - 1 - \min\{g(r), 2g(r)\} + 2r < 0$ and that the conclusion of Lemma 6.8 holds. The following holds with probability $1 - O(n^{-1})$ over P and j: Let q_+ and q_- be unit eigenvectors corresponding to the largest and smallest eigenvalues of $\widehat{\mathcal{N}}_j(\theta_*)$, respectively. Then there are $s \in \{+,-\}$ and $\tau \in \{-1,1\}$ such that*

$$\|q_s - \tau v/\lambda\| = O(n^{-1/2}).$$

Consequently, the $n^{-1/3}$-BDD query at dq_s returns τv for all sufficiently large n.

Proof. Put $\widehat{v} = v/\lambda$ and $\sigma = (-1)^{\alpha_* \cdot j}$. By Lemma 6.4, $\mathcal{A}_{r,d}(u_*) = \mu_{r,d}\widehat{v}\widehat{v}^T + \mathcal{R}$, where $\|\mathcal{R}\|_{\mathrm{F}} \le \mu_{r,d} 2^{-\Omega(n)}$. Let $S_{P,j} := \sum_{\alpha \ne \alpha_*} (-1)^{j\cdot(\alpha-\alpha_*)} \mathcal{A}_{r,d}(P^{-1}(\alpha, \theta_*))$. The calculation in the proof of Lemma 5.7, with t replaced by r, together with Lemma 5.2, gives

$$\mathbb{E}_{P,j} \|S_{P,j}\|_{\mathrm{F}}^2 \le \mu_{r,d}^2 2^{(\chi - 1 - \min\{g(r), 2g(r)\} + 2r + o(1))n}.$$

The exponent is negative by assumption. Hence Markov's inequality gives $\|S_{P,j}\|_{\mathrm{F}} \le \mu_{r,d} 2^{-\Omega(n)}$, except with probability $2^{-\Omega(n)}$ over P, j.

By Lemma 5.3, $\zeta_{r,j} = 1 + O(2^{-\Omega(n)})$, except with probability $2^{-\Omega(n)}$ over P. Applying Lemma 5.6 and substituting the expressions for $\mathcal{A}_{r,d}(u_*)$ and $S_{P,j}$, and then applying Lemma 6.8, gives, for some $a \in \mathbb{R}$,

$$\sigma \widehat{\mathcal{N}}_j(\theta_*) = aI_n + \frac{\overline{w}_j \mu_{r,d}}{\zeta_{r,j}} \widehat{v}\widehat{v}^T + E, \qquad \|E\|_{\mathrm{op}} = O\left(\frac{\overline{w}_j \mu_{r,d}}{n}\right).$$

If $\sigma = 1$, let $q = q_+$, and otherwise let $q = q_-$. The remainder proof is identical to Lemma 5.7 except that we use $\overline{w}_j \mu_{r,d}/\zeta_{r,j}$ instead of $\mu_{t,d}/\zeta_{t,j}$. □

6.5 The importance-sampling algorithm

This section presents the algorithm using importance sampling. The time complexity is determined by the $2^{(\iota(r,R)+2r)n+o(n)}$ samples used in Lemma 6.8 and the $2^{\ell+o(n)} = 2^{(1-\chi)n+o(n)}$ matrices, e.g., $\widehat{\mathcal{N}}_j(\theta)$, indexed by $\theta \in \mathbb{F}_2^\ell$. The space complexity is determined by the samples and these matrices.

We choose the following parameters

$$r = 0.2222355, \qquad R = 0.400613, \qquad \chi = 0.3961331,$$

which gives $s = \frac{rR}{2R-r} = 0.1537683785\ldots$ and $\iota(r,R) = \frac{1}{2}\log_2 \frac{R^2}{r(2R-r)} = 0.1593947534\ldots$. A direct numerical calculation shows that all the conditions of Lemmas 6.4, 6.6, 6.8 and 6.9, and

$$\max\left\{\frac{1}{2}, \iota(r,R) + 2r, 1 - \chi\right\} < 0.603867. \tag{30}$$

Remark 6.10. Without the refined bound in Lemma 6.3, the source-coset sampling argument requires $\chi < g(R)$. Choose $r = 0.222275$, $R = 0.400765$, and $\chi = 0.395965$. Then

$$g(R) = 0.395965\ldots > \chi, \qquad \iota(r,R) + 2r = 0.604035\ldots, \qquad 1 - \chi = 0.604035.$$

These values satisfy the other assumptions of the algorithm. Therefore, using $g(R)$ instead of $g_2(R)$ gives time and space $2^{0.60404n+o(n)}$.

Algorithm 4: Importance-sampling affine-coset Hessian SVP

1. Construct the preprocessing $n^{-1/3}$-BDD data and the scale grid from Algorithm 1.
2. At every scale d, choose independent uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$. Put $M = \lceil n^6 2^{(\iota(r,R)+2r)n} \rceil$ and apply the sampling algorithm from Lemma 6.6. Abort the scale unless nM samples from D_{Λ_j,ξ_R} are obtained, and divide them into n families $\{X_{a,1},\dots,X_{a,M}\}$, $a \in [n]$.
3. For every $a \in [n]$, construct the matrix-valued array
$$A_a(y) := -\frac{4\pi^2}{M} \sum_{\substack{1\le i\le M \\ V_P(X_{a,i})=y}} w(X_{a,i}) X_{a,i} X_{a,i}^T \mathbf{1}_{\{\|X_{a,i}\|\le C_1\xi_R\sqrt{n}\}}, \qquad y \in \mathbb{F}_2^\ell.$$
4. Apply the matrix-valued Walsh-Hadamard transform to each A_a. For every $\theta \in \mathbb{F}_2^\ell$, the output satisfies
$$\sum_{y\in\mathbb{F}_2^\ell} (-1)^{\theta\cdot y} A_a(y) = \widehat{\mathcal{N}}_{a,j}(\theta). \tag{31}$$
Define $\widehat{\mathcal{N}}_j(\theta)$ by taking the median of the corresponding entries of $\widehat{\mathcal{N}}_{1,j}(\theta),\dots,\widehat{\mathcal{N}}_{n,j}(\theta)$.
5. For every $\theta \in \mathbb{F}_2^\ell$, compute unit eigenvectors q_+ and q_- corresponding to the largest and smallest eigenvalues of $\widehat{\mathcal{N}}_j(\theta)$ and query BDD at $dq_\pm$. Store a nonzero output in $\mathcal{L}$. Keep only the shortest vector found and discard the other vectors and temporary data.
6. Return the shortest vector over all scales.

Theorem 6.11. *Algorithm 4 solves Search-SVP with constant success probability in time $2^{0.60387n+o(n)}$ and space $2^{0.60387n+o(n)}$.*

Proof. Suppose $\lambda_1(\mathcal{L}) \le d \le (1+1/n)\lambda_1(\mathcal{L})$ and let v be a shortest vector. We can ignore the other d's as before. Choose $u_* \in \mathbb{F}_2^n$ such that $v \in Bu_* + 2\mathcal{L}$, and write $Pu_* = (\alpha_*, \theta_*)$.

By Lemma 6.6, the algorithm obtains samples whose joint distribution is $\exp\bigl(-\Omega(n^2)\bigr)$-close to independent samples from D_{Λ_j,ξ_R}. By Eq. (31), the Walsh-Hadamard transforms compute $\widehat{\mathcal{N}}_{a,j}(\theta)$ for every $a \in [n]$ and $\theta \in \mathbb{F}_2^\ell$. Their entrywise medians are therefore the matrices $\widehat{\mathcal{N}}_j(\theta)$ defined in Section 6.3.

The conclusion of Lemma 6.8 holds simultaneously for every $\theta \in \mathbb{F}_2^\ell$, and Lemma 6.9 shows that one of the two BDD queries corresponding to θ_* returns one of $\pm v$. The statistical distance in Lemma 6.6 changes the success probability by at most $\exp\bigl(-\Omega(n^2)\bigr)$. Since every stored vector is verified to be a nonzero vector in $\mathcal{L}$, the shortest stored vector has norm $\lambda_1(\mathcal{L})$.

We analyze the complexity. Generating and processing the samples takes $2^{(\max\{1/2,\iota(r,R)+2r\}+o(1))n}$ time by Lemma 6.6. Constructing the arrays takes $nM\,\mathrm{poly}(n)$ time. Their Walsh-Hadamard transforms (Lemma 4.1), the entrywise medians, the eigenvector computations, and the BDD queries take $2^{\ell+o(n)} = 2^{(1-\chi)n+o(n)}$ time. Hence the time exponent is 0.60387 as in Eq. (30).

The Gaussian samples use $2^{(\iota(r,R)+2r)n+o(n)}$ space, the arrays and their transforms use $2^{(1-\chi)n+o(n)}$ space, and the sampling algorithm uses $2^{n/2+o(n)}$ space. Thus the space exponent is bounded by the same maximum in Eq. (30). This proves the claimed time and space bounds. □

6.6 Reducing the space by sparsification

We use the same parameters as in the previous section. Choose independent uniform $P \in \mathrm{GL}_n(\mathbb{F}_2)$ and $j \in \mathbb{F}_2^h$. The algorithm in Section 6.5 stores n arrays indexed by $\mathbb{F}_2^\ell$ of size $2^\ell = 2^{0.60387n+o(n)}$. To reduce the space, we randomly and sparsely select some of the samples before applying the Walsh-Hadamard transform, and they are only used in the later steps.

More precisely, we do the following. For $X \in \Lambda_j$, define

$$\pi(X) := \min\left\{1, \frac{w(X)}{2^{\iota(r,R)n}(r/R)^{n/2}}\right\}. \tag{32}$$

Note that all terms are efficiently computable. Conditional on each sampled X, independently choose $Z \in \{0,1\}$ such that $\Pr[Z = 1 \mid X] = \pi(X)$ and $\Pr[Z = 0 \mid X] = 1 - \pi(X)$. We use the following: For every function f for which the expectation exists,

$$\mathbb{E}_{X,Z}\left[\frac{Zf(X)}{\pi(X)}\right] = \mathbb{E}_X\left[\mathbb{E}\left[\frac{Zf(X)}{\pi(X)}\middle| X\right]\right] = \mathbb{E}_X[f(X)]. \tag{33}$$

Let $M = \lceil n^6 2^{(\iota(r,R)+2r)n}\rceil$, and take n independent families $\{X_{a,1}, \ldots, X_{a,M}\}$, $a \in [n]$, of samples from D_{Λ_j,ξ_R}. Conditional on these samples, let the $Z_{a,i}$ be independent random variables such that $Z_{a,i} = 1$ with probability $\pi(X_{a,i})$ and $Z_{a,i} = 0$ otherwise. Define the sparsified estimator

$$\widehat{\mathcal{N}}^{\mathrm{sp}}_{a,j}(\theta) := -\frac{4\pi^2}{M}\sum_{i=1}^{M}\frac{Z_{a,i}w(X_{a,i})}{\pi(X_{a,i})}X_{a,i}X_{a,i}^T(-1)^{\theta\cdot V_P(X_{a,i})}\mathbf{1}_{\{\|X_{a,i}\|\le C_1\xi_R\sqrt{n}\}}. \tag{34}$$

Because of Eq. (33), the expectation of this sparsification is the same as the expectation of the original Eq. (27). For each matrix entry, define $\widehat{\mathcal{N}}^{\mathrm{sp}}_j(\theta)$ by taking the median among the corresponding entries of $\widehat{\mathcal{N}}^{\mathrm{sp}}_{1,j}(\theta), \ldots, \widehat{\mathcal{N}}^{\mathrm{sp}}_{n,j}(\theta)$.

The factor $1/\pi(X_{a,i})$ compensates for selecting a sample with probability $\pi(X_{a,i})$: conditional on $X_{a,i}$, the expectation of $Z_{a,i}w(X_{a,i})/\pi(X_{a,i})$ is $w(X_{a,i})$.

Lemma 6.12. *Let $\lambda = \lambda_1(\mathcal{L})$ and $\lambda \le d \le (1+1/n)\lambda$. Then, the following holds for all $\theta \in \mathbb{F}_2^\ell$:*

$$\max_{\theta\in\mathbb{F}_2^\ell}\left\|\widehat{\mathcal{N}}^{\mathrm{sp}}_j(\theta) - \overline{w}_j\mathcal{G}_{r,d,j}(\theta)\right\|_{\mathrm{op}} = O\left(\frac{\overline{w}_j\mu_{r,d}}{n}\right)$$

with probability $1 - O(n^{-1})$ over P, j, the samples $X_{a,i}$, and the variables $Z_{a,i}$.

Proof. Fix P, j for which the conclusion of Lemma 6.7 holds, which happens with probability $1 - O(n^{-1})$. For fixed a, let $\mathbb{E}_{X_a}$ denote expectation over the independent samples $X_{a,1}, \ldots, X_{a,M} \sim D_{\Lambda_j,\xi_R}$. Let $\mathbb{E}_{X_a,Z_a}$ additionally include the conditionally independent choices $Z_{a,i} \mid X_{a,i} \sim \mathrm{Bernoulli}(\pi(X_{a,i}))$.

By Eq. (33), $\mathbb{E}_{X_a,Z_a}[\widehat{\mathcal{N}}^{\mathrm{sp}}_{a,j}(\theta)] = \mathbb{E}_{X_a}[\widehat{\mathcal{N}}_{a,j}(\theta)]$. Consequently,

$$\widehat{\mathcal{N}}^{\mathrm{sp}}_j(\theta) - \overline{w}_j\mathcal{G}_{r,d,j}(\theta) = \widehat{\mathcal{N}}^{\mathrm{sp}}_j(\theta) - \mathbb{E}_{X_a,Z_a}[\widehat{\mathcal{N}}^{\mathrm{sp}}_{a,j}(\theta)] + \mathbb{E}_{X_a}[\widehat{\mathcal{N}}_{a,j}(\theta)] - \overline{w}_j\mathcal{G}_{r,d,j}(\theta).$$

The proof of Lemma 6.8 shows that the operator norm of $\mathbb{E}_{X_a}[\widehat{\mathcal{N}}_{a,j}(\theta)] - \overline{w}_j\mathcal{G}_{r,d,j}(\theta)$ is $o(\overline{w}_j\mu_{r,d}/n)$. We bound the first difference using the same concentration argument.

By the definition of $\pi(X)$ and $1/\pi(X) = \max(1, 2^{\iota(r,R)n}(r/R)^{n/2}/w(X)) \le 1 + 2^{\iota(r,R)n}(r/R)^{n/2}/w(X)$,

$$\mathbb{E}_{Z|X}\left[\left(\frac{Zw(X)}{\pi(X)}\right)^2\right] = \frac{w(X)^2}{\pi(X)} \le w(X)^2 + 2^{\iota(r,R)n}(r/R)^{n/2}w(X),$$

where $Z \mid X \sim \mathrm{Bernoulli}(\pi(X))$. In the proof of Lemma 6.8, we proved that $\left\|4\pi^2XX^T\right\|_{\mathrm{F}} = O(\xi_r^4\lambda^2)$ if $\{\|X\| \le C_1\xi_R\sqrt{n}\}$. Hence Lemma 6.7, together with $\mathbb{E}_{X\sim D_{\Lambda_j,\xi_R}}[w(X)] = \overline{w}_j = (r/R)^{n/2}(1 + O(2^{-\Omega(n)}))$, gives

$$\mathop{\mathbb{E}}_{\substack{X\sim D_{\Lambda_j,\xi_R}\\ Z|X\sim\mathrm{Bernoulli}(\pi(X))}}\left[\left(\frac{Zw(X)}{\pi(X)}\right)^2\left\|4\pi^2XX^T\right\|_{\mathrm{F}}^2\mathbf{1}_{\{\|X\|\le C_1\xi_R\sqrt{n}\}}\right] = O\left(n\overline{w}_j^2\xi_r^8\lambda^4 2^{\iota(r,R)n}\right).$$

For fixed a, θ, p, q, the estimator entry is the average of M independent summands. Therefore,

$$\mathbb{E}_{X_a, Z_a}\left[\left(\widehat{\mathcal{N}}^{\mathrm{sp}}_{a,j}(\theta) - \mathbb{E}_{X_a, Z_a}[\widehat{\mathcal{N}}^{\mathrm{sp}}_{a,j}(\theta)]\right)^2_{p,q}\right] = O\left(\frac{\overline{w}_j^2 \mu_{r,d}^2}{n^5}\right)$$

by the definitions of M and $\mu_{r,d}$ in Eq. (29). This is the same bound used in the proof of Lemma 6.8. Applying its Chebyshev, median, and union-bound arguments proves the required operator-norm bound simultaneously for every $\theta \in \mathbb{F}_2^\ell$. The failure probability in Lemma 6.7 is $O(n^{-1})$. □

The next lemma bounds the number of samples for which $Z_{a,i} = 1$.

Lemma 6.13. *Under the hypotheses of Lemma 6.12, with probability $1 - O(n^{-1})$ over P, j, the samples $X_{a,i}$, and the variables $Z_{a,i}$, the number of pairs (a, i) satisfying $Z_{a,i} = 1$ and $\|X_{a,i}\| \le C_1 \xi_R \sqrt{n}$ is at most $n^8 2^{2rn}$.*

Proof. Since $\pi(X) \le w(X)/(2^{\iota(r,R)n}(r/R)^{n/2})$, Lemma 6.7 gives

$$\mathbb{E}_{X \sim D_{\Lambda_j, \xi_R}}[\pi(X)] \le \frac{\overline{w}_j}{2^{\iota(r,R)n}(r/R)^{n/2}} = 2^{-\iota(r,R)n}\left(1 + O(2^{-\Omega(n)})\right).$$

The expected number of pairs satisfying $Z_{a,i} = 1$ is therefore $O(nM2^{-\iota(r,R)n}) = O(n^7 2^{2rn})$. The norm condition can only decrease this number. Markov's inequality proves the lemma. □

6.7 Space-efficient implementation

Now we introduce an algorithm with the reduced space complexity. For every importance-weighted sample, we choose the random variable Z using Eq. (32) and store the sample only if $Z = 1$. The division by $\pi(X)$ preserves the expectation of its contribution. By Lemma 6.13, the number of stored samples is $2^{2rn+o(n)}$.

We choose the same parameters r, R, χ as before, satisfying the conditions of all the lemmas and Eq. (30). The time complexity remains $2^{0.60387n+o(n)}$. Generating and selecting the samples takes $2^{(\iota(r,R)+2r)n+o(n)}$ time, and all the Walsh-Hadamard transforms and the examination of all matrices take $2^{\ell+o(n)} = 2^{(1-\chi)n+o(n)}$ time. The space complexity becomes $2^{n/2+o(n)}$, because the stored samples use $2^{2rn+o(n)}$ space for $2r < 1/2$, and only $2^{\lfloor n/2 \rfloor}$ matrices are processed at a time.

Algorithm 5: Space-efficient importance-sampling SVP

1. Follow the first two steps of Algorithm 4, but process each sample $X_{a,i}$ immediately and then discard it. If $\|X_{a,i}\| \le C_1 \xi_R \sqrt{n}$, choose $Z_{a,i}$ as in Eq. (32). When $Z_{a,i} = 1$, write

   $$V_P(X_{a,i}) = (k'_{a,i}, k''_{a,i}) \in \mathbb{F}_2^b \times \mathbb{F}_2^{\ell-b}, \qquad b = \lfloor n/2 \rfloor,$$

   and store $(a, k'_{a,i}, k''_{a,i}, W_{a,i})$, where $W_{a,i} := -\frac{4\pi^2 w(X_{a,i})}{M\pi(X_{a,i})} X_{a,i} X_{a,i}^T$. Abort the scale if more than $n^8 2^{2rn}$ tuples are stored.

2. For every $\theta'' \in \mathbb{F}_2^{\ell-b}$ and $a \in [n]$, form

   $$A_{a,\theta''}(x) := \sum_{\substack{(a,k',k'',W) \text{ stored} \\ k'=x}} (-1)^{\theta'' \cdot k''} W, \qquad x \in \mathbb{F}_2^b.$$

   Apply the matrix-valued Walsh-Hadamard transform and take the entrywise median of the n outputs for each $\theta' \in \mathbb{F}_2^b$. Perform the eigenvector and BDD steps of Algorithm 4, and discard the arrays before proceeding to the next θ''.

3. Keep the shortest verified vector over all scales. Repeat a sufficiently large constant number of times and return the shortest vector found.

Theorem 6.14. *Algorithm 5 solves Search-*SVP *with success probability at least* $2/3$ *in expected time* $2^{0.60387n+o(n)}$ *and space* $2^{n/2+o(n)}$.

Proof. The correctness follows from the proof of Theorem 6.11, replacing Lemma 6.8 by Lemma 6.12. The additional abort probability is $O(n^{-1})$ by Lemma 6.13.

The samples are discarded after the selection step, so only $2^{2rn+o(n)}$ tuples are stored. For each fixed θ'', the algorithm takes $\left(2^{2rn}+2^b\right)2^{o(n)}$ time and $2^{b+o(n)}$ space. Therefore, since $2r<1/2$ and $b=\lfloor n/2\rfloor$, the total time is $2^{\ell-b}\left(2^{2rn}+2^b\right)2^{o(n)}=2^{(1-\chi)n+o(n)}$. Together with the sampling time, this is $2^{0.60387n+o(n)}$. The selected tuples, the arrays for one fixed θ'', the discrete Gaussian sampler, and the preprocessing BDD data use $2^{n/2+o(n)}$ space. A constant number of repetitions gives success probability at least $2/3$. □

6.8 Quantum algorithm

We describe the quantum version of Algorithm 5 based on the quantum minimum finding algorithm [15]. The application is rather straightforward: We search for $\theta''\in\mathbb{F}_2^{\ell-b}$ that minimizes the shortest vector found by the preprocessing BDD algorithm. This reduces the complexity from $2^{\ell-b+o(n)}(2^{2rn}+2^b)$ to $2^{(\ell-b)/2+o(n)}(2^{2rn}+2^b)$.

In the quantum version, the collected tuples (of size $2^{2rn+o(n)}$) and the preprocessing BDD advice (of size $2^{o(n)}$) are stored in QRAM, which will be of size $2^{2rn+o(n)}$ at total. The overall steps for the Hessian estimation, eigenvector evaluations, and the BDD queries can be done reversibly. This gives the coherent oracle required by quantum minimum finding with only a $2^{o(n)}$ multiplicative overhead.

To maximize the speedup, we also choose $b=\lfloor 2rn\rfloor$ and other parameters by

$$r=0.180182,\qquad R=0.340429,\qquad \chi=0.278262.$$

giving $2^{0.54106n+o(n)}$ time and $2^{0.5n+o(n)}$ space.

Theorem 6.15. *There is a quantum algorithm that solves Search-*SVP *with success probability at least* $2/3$ *in expected time* $2^{0.54106n+o(n)}$ *and space* $2^{n/2+o(n)}$. *The QRAM size is* $2^{0.36036n+o(n)}$ *and* $2^{0.36036n+o(n)}$ *qubits.*

References

[1] Divesh Aggarwal, Yanlin Chen, Rajendra Kumar, and Yixin Shen. Improved classical and quantum algorithms for the shortest vector problem via bounded distance decoding. *SIAM Journal on Computing*, 54(2):233–278, 2025. 3, 4, 7

[2] Divesh Aggarwal, Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. Solving the shortest vector problem in 2^n time using discrete Gaussian sampling: Extended abstract. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *47th ACM STOC*, pages 733–742, Portland, OR, USA, June 14–17, 2015. ACM Press. 3, 4, 6, 8, 9

[3] Divesh Aggarwal, Zeyong Li, and Noah Stephens-Davidowitz. A $2^{n/2}$-time algorithm for $\sqrt{n}$-SVP and $\sqrt{n}$-Hermite SVP, and an improved time-approximation tradeoff for (H)SVP. In Anne Canteaut and François-Xavier Standaert, editors, *EUROCRYPT 2021, Part I*, volume 12696 of *LNCS*, pages 467–497, Zagreb, Croatia, October 17–21, 2021. Springer, Cham, Switzerland. 3

[4] Divesh Aggarwal and Noah Stephens-Davidowitz. (Gap/S)ETH hardness of SVP. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *50th ACM STOC*, pages 228–238, Los Angeles, CA, USA, June 25–29, 2018. ACM Press. 3

[5] Dorit Aharonov and Oded Regev. Lattice problems in np ∩ conp. *J. ACM*, 52(5):749–765, September 2005. 4

[6] Dorit Aharonov and Oded Regev. Lattice problems in NP intersect coNP. *Journal of the ACM*, 52(5):749–765, 2005. 10

[7] Miklós Ajtai. The shortest vector problem in L2 is NP-hard for randomized reductions (extended abstract). In *30th ACM STOC*, pages 10–19, Dallas, TX, USA, May 23–26, 1998. ACM Press. 3

[8] Miklós Ajtai and Cynthia Dwork. A public-key cryptosystem with worst-case/average-case equivalence. In *29th ACM STOC*, pages 284–293, El Paso, TX, USA, May 4–6, 1997. ACM Press. 3

[9] Miklós Ajtai, Ravi Kumar, and D. Sivakumar. A sieve algorithm for the shortest lattice vector problem. In *33rd ACM STOC*, pages 601–610, Crete, Greece, July 6–8, 2001. ACM Press. 3

[10] Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In Robert Krauthgamer, editor, *27th SODA*, pages 10–24, Arlington, VA, USA, January 10–12, 2016. ACM-SIAM. 3

[11] Huck Bennett, Mahdi Cheraghchi, Venkatesan Guruswami, and João Ribeiro. Parameterized inapproximability of the minimum distance problem over all fields and the shortest vector problem in all ℓ_p norms. In Barna Saha and Rocco A. Servedio, editors, *55th ACM STOC*, pages 553–566, Orlando, FL, USA, June 20–23, 2023. ACM Press. 3

[12] Zvika Brakerski, Adeline Langlois, Chris Peikert, Oded Regev, and Damien Stehlé. Classical hardness of learning with errors. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *45th ACM STOC*, pages 575–584, Palo Alto, CA, USA, June 1–4, 2013. ACM Press. 3

[13] Yanlin Chen, Kai-Min Chung, and Ching-Yi Lai. Space-efficient classical and quantum algorithms for the shortest vector problem. *Quantum Info. Comput.*, 18(3–4):283–305, March 2018. 3

[14] Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. On the closest vector problem with a distance guarantee. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 98–109, 2014. 3, 4, 10

[15] Christoph Durr and Peter Hoyer. A quantum algorithm for finding the minimum, 1999. 6, 32

[16] Yiming Gao, Yansong Feng, and Honggang Hu. Solving the shortest vector problem in $2^{0.7314n+o(n)}$ time via discrete gaussian sampling on superlattices. Cryptology ePrint Archive, Paper 2026/1587, 2026. 3

[17] Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In Richard E. Ladner and Cynthia Dwork, editors, *40th ACM STOC*, pages 197–206, Victoria, BC, Canada, May 17–20, 2008. ACM Press. 3

[18] Guillaume Hanrot and Damien Stehlé. Improved analysis of kannan's shortest lattice vector algorithm. In Alfred Menezes, editor, *CRYPTO 2007*, volume 4622 of *LNCS*, pages 170–186, Santa Barbara, CA, USA, August 19–23, 2007. Springer Berlin Heidelberg, Germany. 3

[19] Ishay Haviv and Oded Regev. Tensor-based hardness of the shortest vector problem to within almost polynomial factors. In David S. Johnson and Uriel Feige, editors, *39th ACM STOC*, pages 469–477, San Diego, CA, USA, June 11–13, 2007. ACM Press. 3

[20] Grigorii Anatolévich Kabatiansky and Vladimir Iosifovich Levenshtein. On bounds for packings on a sphere and in space. *Problemy peredachi informatsii*, 14(1):3–25, 1978. 7, 21

[21] Ravi Kannan. Improved algorithms for integer programming and related lattice problems. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, pages 193–206, 1983. 3

[22] Subhash Khot. Hardness of approximating the shortest vector problem in lattices. *J. ACM*, 52(5):789–808, September 2005. 3

[23] Jiseung Kim, 2026. Personal communication. 3, 6

[24] J. C. Lagarias and A. M. Odlyzko. Solving low-density subset sum problems. *J. ACM*, 32(1):229–246, January 1985. 3

[25] Arjen K. Lenstra, Jr. Lenstra, Hendrik W., and László Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261:515–534, 1982. 3, 9

[26] Hendrik W Lenstra Jr. Integer programming with a fixed number of variables. *Mathematics of operations research*, 8(4):538–548, 1983. 3

[27] Mingjie Liu, Xiaoyun Wang, Guangwu Xu, and Xuexin Zheng. Shortest lattice vectors in the presence of gaps. Cryptology ePrint Archive, Paper 2011/139, 2011. 3

[28] Mingjie Liu, Xiaoyun Wang, Guangwu Xu, and Xuexin Zheng. A note on BDD problems with λ_2-gap. *Information Processing Letters*, 114(1–2):9–12, 2014. 3

[29] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on Gaussian measures. In *45th FOCS*, pages 372–381, Rome, Italy, October 17–19, 2004. IEEE Computer Society Press. 3, 8

[30] Daniele Micciancio and Panagiotis Voulgaris. A deterministic single exponential time algorithm for most lattice problems based on voronoi cell computations. In Leonard J. Schulman, editor, *42nd ACM STOC*, pages 351–358, Cambridge, MA, USA, June 5–8, 2010. ACM Press. 3

[31] Daniele Micciancio and Panagiotis Voulgaris. Faster exponential time algorithms for the shortest vector problem. In Moses Charika, editor, *21st SODA*, pages 1468–1480, Austin, TX, USA, January 17–19, 2010. ACM-SIAM. 3

[32] Phong Q. Nguyen and Thomas Vidick. Sieve algorithms for the shortest vector problem are practical. *Journal of Mathematical Cryptology*, 2(2):181–207, 2008. 3

[33] OpenAI. Ten advances in mathematics and theoretical computer science. Technical report, OpenAI, August 2026. Published August 1, 2026. 3

[34] Amaury Pouly and Yixin Shen. Solving the shortest vector problem in $2^{0.63269n+o(n)}$ time on random lattices. In Joan Daemen and Emmanuel Thomé, editors, *Advances in Cryptology – EUROCRYPT 2026*, pages 92–123, Cham, 2026. Springer Nature Switzerland. 3

[35] Xavier Pujol and Damien Stehlé. Solving the shortest lattice vector problem in time $2^{2.465n}$. Cryptology ePrint Archive, Report 2009/605, 2009. 3, 7, 21

[36] Oded Regev. Lattice-based cryptography (invited talk). In Cynthia Dwork, editor, *CRYPTO 2006*, volume 4117 of *LNCS*, pages 131–141, Santa Barbara, CA, USA, August 20–24, 2006. Springer Berlin Heidelberg, Germany. 3

[37] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6), September 2009. 3, 8

[38] Oded Regev and Noah Stephens-Davidowitz. An inequality for gaussians on lattices. *SIAM Journal on Discrete Mathematics*, 31(2):749–757, 2017. 4

[39] Adi Shamir. A polynomial time algorithm for breaking the basic Merkle-Hellman cryptosystem. In *23rd FOCS*, pages 145–152, Chicago, Illinois, November 3–5, 1982. IEEE Computer Society Press. 3

[40] Noah Stephens-Davidowitz. Discrete Gaussian sampling reduces to CVP and SVP. In Robert Krauthgamer, editor, *27th SODA*, pages 1748–1764, Arlington, VA, USA, January 10–12, 2016. ACM-SIAM. 3

[41] Wei Wei, Mingjie Liu, and Xiaoyun Wang. Finding shortest lattice vectors in the presence of gaps. In Kaisa Nyberg, editor, *CT-RSA 2015*, volume 9048 of *LNCS*, pages 239–257, San Francisco, CA, USA, April 20–24, 2015. Springer, Cham, Switzerland. 3