A Sieve Algorithm for the Shortest Lattice Vector Problem

Miklós Ajtai Ravi Kumar D. Sivakumar

IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120

{ajtai, ravi, siva}@almaden.ibm.com

ABSTRACT

We present a randomized $2^{O(n)}$ time algorithm to compute a shortest non-zero vector in an n-dimensional rational lattice. The best known time upper bound for this problem was $2^{O(n \log n)}$ first given by Kannan [7] in 1983. We obtain several consequences of this algorithm for related problems on lattices and codes, including an improvement for polynomial time approximations to the shortest vector problem. In this improvement we gain a factor of $\log \log n$ in the exponent of the approximating factor.

1. INTRODUCTION

A lattice in $\mathbf{R}^n$ is the set of all integer linear combinations of some n linearly independent basis vectors. The shortest lattice vector problem is the problem of finding a shortest non-zero vector (under some norm, usually L_2) in the lattice. The α-approximate version of this problem is to find a non-zero lattice vector whose length is at most α times the length of a shortest non-zero lattice vector.

The shortest lattice vector problem and its approximate versions have a rich history. In a celebrated result, Lenstra, Lenstra, and Lovász [10] gave an algorithm (the LLL or L^3 algorithm) that computes a $2^{n/2}$-approximate shortest vector in polynomial time. This was improved in a generalization of the LLL algorithm by Schnorr [13], who obtained a hierarchy of algorithms that provide a uniform trade-off between the running time and the approximation factor. This algorithm runs in $n^{O(1)}k^{O(k)}$ steps to obtain a $k^{O(n/k)}$-approximate shortest vector. For instance, a polynomial-time version of this algorithm improves the approximation factor obtained by the LLL algorithm to $2^{n(\log \log n)^2/\log n}$. Kannan [7] obtained a $2^{O(n \log n)}$ time algorithm for exactly computing a shortest vector. The constant in the exponent of Kannan's algorithm was improved to about $1/2$ by Helfrich [6].

On the hardness front, the shortest lattice vector problem for the L_∞ norm was shown to be NP-complete by van Emde Boas [4]. Recently, Ajtai [2] proved that the shortest lattice vector problem under the L_2 norm is NP-hard under randomized reductions. Micciancio [12] showed that the α-approximate version of this problem remains NP-hard for any $\alpha < \sqrt{2}$.

Given that a polynomial time algorithm that solves the shortest vector problem exactly or to within a small constant factor seems out of reach, it becomes interesting to finding faster algorithms for this problem, and to improve the approximation factor achieved by a polynomial time algorithm. These questions are the main subject of this paper.

We show that the shortest vector problem (in L_2 norm) for a lattice L in $\mathbf{R}^n$ can be solved in randomized $2^{O(n)}$ time. Specifically, there is a constant $c > 0$ and a randomized algorithm $\mathcal{A}$ such that for any lattice L in $\mathbf{R}^n$ presented by a basis with total bit length at most B, with probability exponentially close to one, $\mathcal{A}$ finds a shortest non-zero vector in L in 2^{cn} steps. Here each step works in time at most a polynomial in B on operands of bit length polynomial in B. The running time of our algorithm improves upon all the previous known algorithms for this problem. In fact, in $2^{O(n)}$ time, our algorithm can find all α-approximate shortest vectors for any constant $\alpha \geq 1$. Our algorithm is presented in Section 3. For polynomial time approximations, we present, as a consequence of our main algorithm, the first asymptotic improvement (in the exponent of the approximation factor) since Schnorr's [13] work: we show that in polynomial time, we can find a vector of length at most $2^{n \log \log n/\log n}$ times the length of the shortest non-zero vector. This result and other consequences of our algorithm, including results for the closest vector problem, basis reduction, minimum-weight codeword, are presented in Section 4.

A few words about our method. The starting point of our algorithm is a recent result of [9], who showed that in $2^{O(n)}$ time, one can approximate the length of shortest lattice vector to within a polynomial factor. Their algorithm was based on the idea behind a learning algorithm of Blum, Kalai, and Wasserman [3] and the worst-case/average-case reduction of lattice problems due to Ajtai [1].

Our algorithm works in the following way. Given a basis of the lattice, we uniformly choose $2^{O(n)}$ random lattice points, all inside a sufficiently large parallelepiped, and perturb the lattice points by small quantities. Then, we iteratively apply

STOC'01, July 6-8, 2001, Hersonissos, Crete, Greece.
Copyright 2001 ACM 1-58113-349-9/01/0007 ...$5.00.

a "sieving" procedure to these points. The basic property of this step is that given a set of perturbed lattice vectors with a bound on the longest vector in the set, we produce a new set of perturbed lattice vectors such that the longest vector is now about half of the original length while the set itself does not shrink by more than half. The iteration is applied until we obtain a set of vectors in a ball of constant radius. Finally, we argue that (due to the nature of the perturbations) there is a reasonable chance that a short lattice vector and its "nearest neighbor" in the lattice are both discovered by this process. Taking pairwise differences then gives the shortest vector.

2. PRELIMINARIES

As usual, $\mathbf{Z}$ and $\mathbf{R}$, respectively, will denote the set of integers and real numbers. For integers $n > 0$, $\mathbf{R}^n$ will be the set of all vectors $\langle x_1, ..., x_n \rangle$ where $x_i \in \mathbf{R}$ for $i = 1, ..., n$. $\mathbf{R}^n$ is a vector space over $\mathbf{R}$. The L_2 norm of a vector $x \in \mathbf{R}^n$ will be denoted by $\|x\|_2$ or by $\|x\|$ and the L_∞ norm of $x = \langle x_1, ..., x_n \rangle$, in other words $\max\{|x_1|, ..., |x_n|\}$, will be denoted by $\|x\|_\infty$. For $P \subseteq \mathbf{R}^n$, $\mathrm{vol}(P)$ denotes its volume, and for $x \in \mathbf{R}^n$, $B(x, r)$ denotes the open ball of radius r centered at x, i.e., $B(x, r) = \{y \in \mathbf{R}^n \mid \|x - y\| < r\}$.

Let $b_1, \ldots, b_n$ be n linearly independent vectors in $\mathbf{R}^n$. The set $L = L(b_1, \ldots, b_n) = \{\sum_{i=1}^n \alpha_i b_i \mid \alpha_1, \ldots, \alpha_n \in \mathbf{Z}\}$ is an n-dimensional *lattice* in $\mathbf{R}^n$. Let $\det L$ denote the determinant of L and $\mathrm{sh}(L)$ denote the length of the shortest non-zero vector in L. The SVP problem is the problem of finding a lattice vector of length $\mathrm{sh}(L)$. We assume that B is the bit length of the input to this problem. In this preliminary version we do not discuss the issue of precision of the real numbers used in the algorithm. It will be easy to see that $\mathrm{poly}(B)$ bits of precision are sufficient.

3. MAIN RESULT: SVP IN L_2

THEOREM 1. *There are absolute constants $c, c' > 0$ and a probabilistic algorithm $\mathcal{A}$ such that, on input $b_1, \ldots, b_n \in \mathbf{R}^n$ of total bit length B, $\mathcal{A}$ outputs in time $2^{cn}B^{c'}$ a non-zero vector $v \in L(b_1, \ldots, b_n)$ such that with probability exponentially close to 1, $\|v\| = sh(L)$.*

3.1 Overview of the proof

The key step in the proof the repeated application of a sieve procedure, Lemma 3, which has the following flavor. Given sufficiently many points in $\mathbf{R}^n$ of length at most R, identify a small set of "representatives" from the set of points and a large set of "survivors" such that for every survivor point, there is a representative at distance at most $R/2$. This lemma is used in the following way. After suitable pre-processing of the input lattice to ensure certain technical conditions (e.g., that $1 \le \mathrm{sh}(L) < 2$, $\max_i \|b_i\| \le 2^{\epsilon n}$), we create a large (sides of exponential length) parallelepiped $\mathcal{P}$ that is fairly close to being a cube. Then we uniformly sample a large number of lattice points $z_1, \ldots, z_N$, $N = 2^{O(n)}$, from $\mathcal{P} \cap L$, and to each sample lattice point z_i, we add a uniform perturbation vector y_i of expected length $O(1)$ to obtain a sequence of points $x_1, \ldots, x_N$. For each perturbed lattice point x_i, we will keep track of two lattice points: its "true identity" z_i, and an "approximator" a_i, initially set to 0. We repeatedly apply the sieve lemma to the vectors $x_i - a_i$; for each survivor $x_i - a_i$ with representative $x_j - a_j$, we know that the distance between $x_i - a_i$ and $x_j - a_j$ is about half the distance between x_i and a_i. Therefore, $a_i + x_j - a_j$ is a better approximation to x_i, and since x_j is close to its true identity z_j, we define the new approximator for x_i to be $a_i + z_j - a_j$. In these steps, once the true identity of a point is revealed, we will not use it in the future. We repeat this process until the distance between the points and their approximators are bounded by another constant. Finally, if x_i still survives and has an approximator a_i, output the lattice point $w_i = z_i - a_i$. Since both z_i and a_i are close to x_i, with high probability, the length of w_i is bounded by a constant c_5. We will denote this process as the basic algorithm.

Note that if the basic algorithm stops with a non-zero w_i, we already have a constant factor approximation algorithm for SVP. To ensure that w_i is non-zero and to obtain the shortest vector, we make the following argument.

Let u denote a shortest vector in L. Let $w = w_i$ be a lattice point of length at most c_5 that is output by the procedure above. Let x be a sample point from which w was obtained, and let $z \in L$ be the true identity of x. Since the perturbations are small, we will argue (Lemma 13) that the probability (conditioned on x being a sample) that one of $z \pm u$ is the true identity of x is at least 2^{-Cn} times the probability that z is the true identity of x, for some constant C. (A related idea is used in [5] in the context of AM proofs for SVP.) Furthermore — and this is the crucial point — the algorithm of Lemma 3, as well as the iterative process above, are oblivious to the true identity of x. Using this fact, we will argue that for some w, $w + u$ has at least 2^{-Cn} times the probability of w to be the output of the basic algorithm. By proving (Lemma 22) that the number of lattice points in the ball of radius c_5 around the origin is at most $2^{c_6 n}$ for some constant c_6, we obtain that there is at least one $w \in L$ whose probability of being output is at least $2^{-c_6 n}$, $w + u$ has a probability of at least $2^{-(c_6+3)n}$. Therefore, by repeating the basic algorithm $2^{c_7 n}$ times for a suitably large constant c_7, we can ensure that with high probability both w and $w + u$ are output. Thus, the final algorithm is the following: repeat the basic algorithm $2^{c_7 n}$ times, take all possible pairwise differences of the points output by the basic algorithm, and output the shortest of these vectors. In fact, it should be fairly obvious now that for any constant $d \ge 1$, the same algorithm can be modified to output all points of length at most $d \cdot \mathrm{sh}(L)$.

The rest of this Section is organized as follows. In Section 3.2, we state and prove Lemma 3, the sieve lemma. The proof of Lemma 3 uses a simpler version (Lemma 2) that works for the L_∞ norm; Lemma 2 is also used later in the algorithm prior to the application of Lemma 3. In Section 3.3, we describe the pre-processing steps, which establish that without loss of generality, we may assume certain properties of the lattice. In Section 3.4, we describe the initial sampling procedure, and prove some properties of the samples produced. In Section 3.5, we present the core of the algorithm, namely the the iterative sampling procedure and complete the analysis of the algorithm. Some technical lemmas that are used in the course of various proofs in Sections 3.2 through 3.5 are established in the Appendix. To obtain a high level understanding of the algorithm, the reader may accept the statements of Lemmas 2 and 3, 5, and proceed with Sections 3.4 and 3.5, consulting the Appendix for the statements of the technical lemmas used in the analysis.

3.2 The sieve lemma

Lemma 2. *There is a deterministic algorithm $\mathcal{A}$ so that for all positive integer q there exists a $\beta_2 > 0$ so that for all sufficiently large positive integers n and for all $R > 0$ the following holds. Suppose that $a_1, ..., a_t$ is a sequence of vectors in $\mathbf{R}^n$ so that $\|a_i\|_\infty \leq R$, for $i = 1, ..., t$.*

If $\mathcal{A}$ gets q, n, k, R, t and $a_1, ..., a_t$ as input, then in time t^{β_2} it returns a set $\Phi \subseteq \{1, .., t\}$ so that the following requirements are met:

(1). $|\Phi| \geq t - 2^{(q+1)n}$,

(2). $(\forall i \in \Phi)\ (\exists j \in \{1, ..., t\} - \Phi)\ \|a_i - a_j\|_\infty \leq 2^{-q}R.$

Proof. We cut the interval $[-R, R]$ into 2^{q+1} subintervals of equal length. Let J be the set of these 2^{q+1} intervals. Now we define a partition of $\{1, ..., t\}$ into at most $2^{(q+1)n}$ disjoint classes. $i, j \in \{1, ..., t\}$ will be in the same class iff for all $k = 1, ..., n$ the k-th components of the vectors a_i and a_j belong to the same interval from J. Let P be the partition of $\{1, ..., t\}$ defined this way. Form each class C of P we arbitrarily select a representative. Let Φ be the complement of the set of these representatives in $\{1, ..., t\}$. The number of representatives is at most $2^{(q+1)n}$ therefore $\Phi \geq t - 2^{(q+1)n}$ so (1) holds. For the proof of (2) let i be an arbitrary element of Φ. Let C be the class of P containing i and let j be the selected representative of this class. By the definition of Φ we have $j \notin \Phi$. Since for each fixed $k = 1, ..., n$ the k-th component of a_i and a_j are in an interval of length $2^{-q}R$ we have that $\|a_i - a_j\|_\infty \leq 2^{-q}R$, which completes the proof of the lemma. □

Our next goal is to formulate an analogue of Lemma 2 for the Euclidean norm. This L_2 version of the lemma will be somewhat weaker, namely: $\mathcal{A}$ will only be a probabilistic algorithm, and in the analogue of condition (1) we will have only $|\Phi| \geq t/2$, and the approximation will be also weaker.

Lemma 3. *There is a probabilistic algorithm $\mathcal{A}$ so that if $\beta_1 > 0$ is sufficiently large then there exists a $\beta_2 > 0$ so that for all sufficiently large positive integers n and for all $R > 0$ the following holds. Suppose that $t \geq 2^{\beta_1 n} \geq t/2$ and $a_1, ..., a_t$ is a sequence of vectors in $\mathbf{R}^n$ so that $\|a_i\|_2 \leq R$, for $i = 1, ..., t$.*

Then, if $\mathcal{A}$ gets n, R, t and $a_1, ..., a_t$ as input, then in time $2^{\beta_2 n}$ and with a probability of at least $1/2$ it returns a set $\Phi \subseteq \{1, .., t\}$ so that the following requirements are met:

(3). $|\Phi| \geq t/2$,

(4). $(\forall i \in \Phi)\ (\exists j \in \{1, ..., t\} - \Phi)\ \|a_i - a_j\|_2 \leq R/2.$

Remark: The following version of this lemma gives a a simpler and deterministic algorithm that is somewhat slower. The analysis of our SVP algorithm can be carried out using this version as well. We are grateful to Madhu Sudan and two anonymous referees for STOC who pointed it out to us.

Let $R > 0$ and let Ψ be a set of vectors in $\mathbf{R}^n$ such that $|\Psi| \geq 2 \cdot 8^n$, and $(\forall x \in \Psi)\ \|x\| \leq R$. Then there exists a subset $\Phi \subseteq \Psi$ such that $|\Phi| \geq |\Psi|/2$, and $(\forall x \in \Phi)\ (\exists x' \in \Psi \backslash \Phi)\ \|x - x'\| \leq R/2$. Moreover, given Ψ, such a Φ can be computed in deterministic time polynomial in $|\Psi|$.

For the proof, take a packing of $B(0, R)$ by balls of radius $R/8$; there are at most $\text{vol}(B(0, R))/\text{vol}(B(0, R/8)) = 8^n$ balls in the packing, and such a packing may be easily found by a greedy procedure. Double the radius of each ball in the packing to obtain a covering of $B(0, R)$ by balls of radius $R/4$. For each ball β in the covering, add all except one point of $\Psi \cap \beta$ to Φ. Let us call this one point the *representative for* β. The number of points not added to Φ is at most 8^n, so $|\Phi| \geq |\Psi| - 8^n \geq |\Psi|/2$. Also, for any point $x \in \Phi$ that came from a ball β, $\|x - \text{representative}(\beta)\| \leq \text{diameter}(\beta) = R/2$.
End of Remark.

We will prove the sieve lemma in the following way. We take a random orthonormal system e_i in $\mathbf{R}^n$. We consider the components of each a_i in this system. For a large constant M we will write each a_i in the form $b_i + r_i$ where all of the components of b_i are not larger than M in absolute value. (In fact we cut each coefficient of a_i into two parts: up to M and the remainder.) We show that with high probability for most of the possible values of i we will have $\|r_i\|_2 \leq R/8$. Now we consider the coefficient vectors of b_i written in the orthonormal basis e_i. We may now apply Lemma 2 to these vectors (the L_∞ norm is defined with respect to the e_i basis). We define Φ as in the lemma and get an approximation closer than $R/4$ for each b_i, $i \in \Phi$. The approximating vectors will also approximate the corresponding vectors a_i for a distance less than $R/2$.

For the actual proof, we need to prove some facts about random orthonormal systems.

If n is a positive integer we define a probability measure μ on the sphere S around 0 with radius 1. Namely if X is a Borel subset of S then $\mu(X)$ will be proportional to $\mathcal{L}_n(\{\alpha x \mid x \in X, 0 \leq \alpha \leq 1\})$ where $\mathcal{L}_n$ is the n dimensional Lebesgue measure. When we say that a random variable ξ takes its values with uniform distribution on S we will mean that the distribution of ξ is μ.

Assume that n is a positive integer. We define a probability distribution on the set $\mathcal{R}$ of all orthonormal bases of $\mathbf{R}^n$ in the following way. We will pick a random orthonormal basis $e = \langle e_1, ..., e_n \rangle$ by sequentially picking the vectors $e_1, ..., e_n$. Assume that $e_1, ..., e_{i-1}$ have been already selected for some $i = 1, ..., n$. Let V be the subspace of $\mathbf{R}^n$ orthogonal to $e_1, ..., e_{i-1}$ and let $S_V(0, 1)$ be the sphere in V centered around 0 and with radius 1. We pick e_i at random and with uniform distribution from $S_V(0, 1)$.

Remark. It is easy to see that the defined probability measure on $\mathcal{R}$ in invariant under multiplication by unitary linear transformations of $\mathbf{R}^n$, that is, if U is a unitary linear transformation of $\mathbf{R}^n$ and Y is a Borel set of $\mathcal{R}$ then $\mu(Y) = \mu(UY)$. This is a consequence of the symmetry of the definition of the probability distribution. This property of the distribution implies that for each $i = 1, ..., n$, e_i has a uniform distribution on the unit sphere of $\mathbf{R}^n$.

For $x, M \in \mathbf{R}$, $M > 0$, we define a function $C(x, M)$ in the following way: if $x > M$ or $x < -M$ then $C(x, M) = \frac{x}{|x|}M$, otherwise $C(x, M) = x$.

Lemma 4. *There exists a $c > 0$ so that if n is a positive integer, $M > 1$, and $v \in \mathbf{R}^n$, then the following holds. Assume that $e_1, ..., e_n$ is a random orthonormal basis of $\mathbf{R}^n$, $v = \sum_{i=1}^n \alpha_i e_i$ and $w = \sum_{i=1}^n C(\alpha_i, n^{-1/2}M\|v\|)e_i$. Then* $\text{Pr}[\|v - w\| \leq 2^{-cM}\|v\|] \geq 1 - 2^{-cM}$.

PROOF. We may assume that $\|v\| = 1$. For all $x > 0$ let $N(x)$ be the number of integers $i \in [1, n]$ with the property $|\alpha_i| > xn^{-1/2}$. Each e_i has a uniform distribution on the unit sphere, so according to Lemma 20 there is a constant $c' > 0$ so that $\mathrm{E}[N(x)] \leq ne^{-c'x^2}$. Therefore Markov's inequality implies that $\Pr[N(x) \geq ne^{-c'x^2/2}] \leq e^{-c'x^2/2}$. Let $K(x)$ be the number of integers $i \in [1, n]$ with the property $(x + 1)n^{-1/2} \geq |\alpha_i| > xn^{-1/2}$. Clearly $K_i(x) \leq N_i(x)$ and $\|v - w\|^2 \leq \sum_{j=1}^{\infty} K(M + j)j^2n^{-1}$. Let B be the event that for all $j = 1, 2, ...,$ $K(M + j) \leq ne^{-c'(M+j)^2/2}$. We have $\Pr[B] \geq 1 - \sum_{j=1}^{\infty} e^{-c'(M+j)^2/2}$. Assume now that $c > 0$ is sufficiently small with respect to c'. Then $1 - \sum_{j=1}^{\infty} e^{-c'(M+j)^2/2} \geq 1 - e^{-cM}$ so $\Pr[B] \geq 1 - e^{-cM}$. If B holds then $\|v - w\|^2 \leq \sum_{j=1}^{\infty} K(M + j)j^2n^{-1} \leq \sum_{j=1}^{\infty} ne^{-c'(M+j)^2/2}j^2n^{-1} \leq e^{-2cM}$ and so $\|v - w\| \leq e^{-cM}$ which completes the proof of the Lemma. □

PROOF. (of Lemma 3) We will apply Lemma 2 where q will be a sufficiently large constant. Assume now that n is sufficiently large, t and the sequence a_i are fixed. Let e_i be a random orthonormal system in $\mathbf{R}^n$, and let $M = 3/c$ where c is the constant whose existence is stated in Lemma 4. We apply Lemma 4 with $v \to a_i$ for $i = 1, ..., t$. If $a_i = \sum_{j=1}^{n} \alpha_{ij}e_i$, then let $b_i = \sum_{j=1}^{n} C(\alpha_{ij}, M\|a_i\|)e_i$. According to Lemma 4 for each fixed $i = 1, ..., t$, we have $\Pr[\|a_i - b_i\|_2 \leq R/8] \geq 1 - 2^{-cM} = 7/8$. Therefore if Φ' is the set of all $i \in 1, ..., t$ with $\|a_i - b_i\|_2 \leq R/8$ then it is easy to see that $\Pr[|\Phi'| \geq 3t/4] > 1/2$. The algorithm $\mathcal{A}$ will provide the required answer if $|\Phi'| \geq 3t/4$. From now on we assume that this is the case.

We apply now Lemma 2 with $R \to n^{-1/2}MR$ so that the sequence $\langle b_i \mid i \in \Phi' \rangle$ takes the role of the sequence $\langle a_i \mid i = 1, ..., t \rangle$. Let $\Phi \subseteq \Phi'$ be the set whose existence is stated in the lemma. Since $|\Phi'| \geq 3t/4$ and $|\Phi| \geq |\Phi'| - 2^{(q+1)n}$ we have $|\Phi| \geq t/2$, provided $t \geq 4 \cdot 2^{(q+1)n}$, which holds if $\beta_1 > 0$ is chosen to be a sufficiently large constant. Assume now that $i \in \Phi$. (2) implies that there is a $j \in \Phi'$, $j \notin \Phi$ so that $\|b_i - b_j\|_\infty \leq 2^{-q}n^{-1/2}MR$. Thus $\|b_i - b_j\| \leq (n(2^{-q}n^{-1/2}MR)^2)^1/2 = 2^{-q}MR$. If $i, j \in \Phi'$, then $\|a_i - b_i\| \leq R/8$, and $\|a_j - b_j\| \leq R/8$, therefore $\|a_i - a_j\| \leq \|a_i - b_i\| + \|b_i - b_j\| + \|a_j - b_j\| \leq R/8 + 2^{-q}MR + R/8 \leq R/2$, provided that q is sufficiently large with respect to $M = 3/c$, specifically $2^q \geq 4M$. □

3.3 Preprocessing

To facilitate the analysis later, we make two simplifying assumptions: $1 \leq \mathrm{sh}(L) < 2$ and $\max_i \|b_i\| \leq 2^{\epsilon n}$ for some $\epsilon > 1$. The justification for these assumptions is given in following lemma.

LEMMA 5. *Without loss of generality, we may assume that there is an absolute constant $\epsilon < 1$ such that the instance L of the shortest lattice vector problem satisfies the following properties:*

(5). $1 \leq sh(L) \leq 2$,

(6). L is presented by a basis $b_1, \ldots, b_n$ that satisfies $\max_{i=1}^{n} \|b_i\| \leq 2^{\epsilon n}$.

PROOF. Let L denote the input lattice specified by a rational basis (numerator and denominator for each basis vector), and let B denote the bit length of the input.

First, we would like to assume that $1 \leq \mathrm{sh}(L) < 2$. To justify this, note that given L specified by B bits, $2^{-B^a} \leq \mathrm{sh}(L) \leq 2^{B^a}$ for some absolute constant a, independent of the lattice L or its presentation. The upper bound follows from the fact that one of the basis vectors has length at most 2^B. For the lower bound, we use the fact that for any lattice L and any basis $b_1, b_2, \ldots, b_n$ of L, $\mathrm{sh}(L) \geq \min_i \|b_i^*\|$ (see, for example, [11, Lemma 1.2.6]), where $b_1^*, b_2^*, \ldots, b_n^*$ denote the orthogonal basis of $\mathbf{R}^n$ obtained by applying the Gram–Schmidt orthogonalization process to $b_1, b_2, \ldots, b_n$. The Gram–Schmidt algorithm runs in polynomial time and produces a basis of $\mathbf{R}^n$ whose bit length is polynomially bounded by the bit length of the description of the b_i's. Therefore, $\min_i \|b_i^*\| \geq 2^{-B^a}$ for some absolute constant a. Therefore, if we consider the lattices 2^kL for each integer k satisfying $-B^a \leq k \leq B^a$, one of these lattices satisfies $1 \leq \mathrm{sh}(\cdot) < 2$. The bit length of the description of each of these lattices is clearly $O(B^a)$.

Secondly, if $L = L(b_1, \ldots, b_n)$, we need an upper bound on $\mu =_{\mathrm{def}} \max_i \|b_i\|$. Specifically, we will assume that $\mu \leq 2^{\epsilon n}$ for some fixe $\epsilon < 1$. We now show how this can be achieved. The L^3 basis reduction algorithm [10] finds, for any lattice X in d dimensions, a vector of length at most $2^{(d-1)/2}\lambda_1(X)$. We proceed inductively. For $j \geq 0$, assume that $b_1, \ldots, b_j$ have been found so that $\max_{i \leq j} \|b_i\| \leq 2^{n/2}$. Let L_j denote the orthogonal projection of L to the subspace orthogonal to the span of $b_1, \ldots, b_j$. Apply the L^3 algorithm to L_j to find a vector b. If $\|b\| > 2^{(n/2)+1}$, then the shortest vector of L must be in $L(b_1, \ldots, b_j)$, and we have a reduction in the dimension of the problem. Otherwise, we pick $b_{j+1} \in L$ such that the orthogonal projection of b_{j+1} is b and it is as close to b as possible. This leads to the closest lattice vector problem in the lattice generated by $b_1, \ldots, b_j$; by rounding the coefficients this can be solved to within an error of $\sum_{i \leq j} \|b_i\|$. Thus the construction either reduces the problem to a smaller dimensional problem or gives a basis of length at most $n^2 2^{n/2}$. To complete the argument that $\mu \leq 2^{\epsilon n}$, we simply choose an $\epsilon > 1/2$. □

3.4 The sampling procedure

After applying suitable pre-processing to ensure the conditions listed in Lemma 5, our SVP algorithm begins with a sampling step, where we create a sequence of $N = 2^{O(n)}$ points $x_1, x_2, \ldots, x_N$ of the form lattice point + small perturbation.

Let $M = 2^{\epsilon n}$ (where ϵ is the constant from Lemma 5). Let $D = 2^{c_0 n}, N = 2^{c_1 n}$, and let $K > 0$, where c_0, c_1 and K are constants to be specified later. (The constant c_1 needs to be large enough so that $(N - O(n2^{2n}))2^{-O(\log n)} > 0$ (see Lemmas 10 and 11 below); c_0 is from Lemma 7, and needs to be large enough with respect to ϵ, c_1, and with respect to two constants c_2, C, which are specified in Lemma 13 later. The role of K is described in Lemma 6, and the choice of K is made in the proof of Lemma 7 to be large enough with respect to c_1 and C (from Lemma 13).)

Let $L = L(b_1, \ldots, b_n)$; by Lemma 5, we may assume that $\max_i \|b_i\| \leq 2^{\epsilon n}$. Let $e_1, e_2, \ldots, e_n$ denote the standard orthonormal basis of $\mathbf{R}^n$, and for $1 \leq i \leq n$, let $f_i = De_i$. For each i, $1 \leq i \leq n$, express f_i in the basis $b_1, \ldots, b_n$, and round each coefficient to the nearest integer. Let a_i be the lattice point obtained this way. It is not hard to see that for each i, $\|f_i - a_i\| \leq M\sqrt{n}$. Let $\mathcal{P}$ denote the half-closed parallelepiped determined by $a_1, \ldots, a_n$, that is, the set $\{\sum_{i=1}^{n} \alpha_i a_i \mid 0 \leq \alpha_i < 1, i = 1, ..., n\}$.

Let $\xi = \xi^{(L,\mathcal{P})}$ be a random variable defined in the following way. Let ζ be a random variable whose values are the points of $\mathcal{P} \cap L$ with uniform distribution on $\mathcal{P} \cap L$. Let η be a random variable on $\mathbf{R}^n$ whose components are i.i.d. random variables of normal distribution with mean 0 and standard deviation $1/(\sqrt{Kn})$. Finally we take independent copies of η and ζ and let $\xi = \eta + \zeta$. (See [1, Lemma 8] for details on how to sample almost uniformly from the set $\mathcal{P} \cap L$ in polynomial time. The error in sampling, that is, the variational distance between the uniform distribution and the distribution induced by the sampling algorithm, can be at most $2^{-\alpha n}$ for any $\alpha > 0$.)

Let $x_1, x_2, \ldots, x_N$ be independent random values of the random variable ξ, with the decomposition $x_i = z_i + y_i$, where z_i and y_i are the corresponding values, respectively, of ζ and η.

We begin by stating some useful properties of the random samples produced.

Lemma 6. *For every $c' > 1$ and $C' > 0$, there exists a constant $K > 0$ such that if the above sampling process is executed with this choice of K, then with probability at least $1 - 2^{-C'n}$, the following holds: for every $i \in \{1, \ldots, N\}$, $\|y_i\| \leq c'$. In particular, for any $C' > 2$, K can be chosen so that with probability at least $1 - 2^{-C'n}$, every y_i satisfies $\|y_i\| \leq c_3 = 2$.*

Proof. By Lemma 21, for any fixed i, $\Pr[\|y_i\| > c'] = \Pr[\|y_i\|^2 > c'^2] < e^{-n((K/4)-(1/2c'^2))} < 2^{-(C'+c_1)n}$ by choosing K large enough. By the union bound, the probability that any of the $N = 2^{c_1 n}$ y_i's has norm more than c' is at most $2^{c_1 n} \cdot 2^{-(C'+c_1)n} < 2^{-C'n}$. □

Lemma 7. *For every constant $c_1 > 0, c_2 > 0, C > 0$, there exist constants $c_0', c_0 > 0$ such that if the above sampling process is executed with $D = 2^{c_0 n}$, then with probability at least $1 - 2^{-Cn}$, the following holds: for every $i \in \{1, \ldots, N\}$, where $N = 2^{c_1 n}$,*

(7). $\|x_i\|_\infty \leq 2^{c_0' n}$,

(8). the distance from x_i to the complement of $\mathcal{P}$ is at least c_2.

Proof. The proof uses several ideas from [1, Lemma 3]. First recall that for $i = 1, \ldots, n$, $f_i = De_i$ (where the e_i's denote a standard orthonormal basis of $\mathbf{R}^n$), and that $\mathcal{P} = \mathcal{P}(a_1, \ldots, a_n)$ is the parallelepiped formed by the lattice points a_i such that for $i = 1, \ldots, n$, $\|f_i - a_i\| \leq M\sqrt{n}/2$. (Recall that $L = L(b_1, \ldots, b_n)$ and $\max_{i=1}^{n} \|b_i\| \leq M$.) Let $\mathcal{Q} = \mathcal{P}(f_1, \ldots, f_n)$ denote the parallelepiped formed by the points f_i. The distance of each vertex of $\mathcal{P}$ from the corresponding vertex of $\mathcal{Q}$ is at most $n \cdot \max_{i=1}^{n} \|f_i - a_i\| \leq Mn^{3/2}/2$. Therefore, if we enlarge the cube $\mathcal{Q}$ from its center by a factor of $1 + Mn^{3/2}/(2D)$, it will contain $\mathcal{P}$. Let $\mathcal{Q}^+$ denote the enlarged cube. Similarly, if we obtain a cube $\mathcal{Q}^-$ by shrinking $\mathcal{Q}$ about its center by the factor $1 - Mn^{3/2}/(2D)$, then $\mathcal{Q}^-$ will be contained in $\mathcal{P}$. Since $\mathcal{Q}^- \subseteq \mathcal{P}$, $\mathcal{P}$ contains a sphere of diameter at least $D(1 - Mn^{3/2}/(2D)) \geq 3D/4$, provided c_0 is sufficiently large. Therefore, the minimal height of $\mathcal{P}$ is at least $3D/4$.

Condition (7) is now easily seen to be true: since $\mathcal{P}$ is contained in $\mathcal{Q}^+$, every lattice point in $\mathcal{P}$ has L_∞ norm at most $D + Mn^{3/2}$. Furthermore, by Lemma 6, with probability at least $1 - 2^{-2n}$, every perturbation vector y_i has L_2 norm, and hence L_∞ norm at most c_3. Therefore, for each i, $\|x_i\|_\infty \leq \|z_i\|_\infty + \|y_i\|_\infty \leq D + Mn^{3/2} + c_3 \leq 2D$, again assuming that c_0 is sufficiently large. Take $c_0' = c_0 + 1$.

For condition (8), we proceed as follows. Let $H = 3D/4$ denote the minimal height of $\mathcal{P}$. Let $c' = c_2 + 2$.

Let $\mathcal{P}'$ denote the parallelepiped obtained by shrinking $\mathcal{P}$ about its center by a factor of $1 - c'/H$. Then every point that is within distance c' from the complement of $\mathcal{P}$ is outside $\mathcal{P}'$. We will give a lower bound on the number of lattice points inside $\mathcal{P}'$ and upper bound on the number of lattice points inside $\mathcal{P}$. This will give us a lower bound on the fraction of lattice points that are far from the complement of $\mathcal{P}$, from which we will derive a lower bound on the probability that all the x_i's are far from the complement of $\mathcal{P}$.

For the lower bound on $|\mathcal{P}' \cap L|$, we further shrink $\mathcal{P}'$ by a factor of $1 - 2Mn/H'$, where H' is the minimal height of $\mathcal{P}'$. Clearly, $H' = H - c'$. Let $\mathcal{P}^-$ be the parallelepiped thus obtained. Recall that $b_1, \ldots, b_n$ are a basis of L. Let W denote the set of parallelepipeds of the form $v + \mathcal{P}(b_1, \ldots, b_n)$, such that $v \in L$ and $\mathcal{P} \cap (v + \mathcal{P}(b_1, \ldots, b_n))$ is non-empty. Now every element of W that intersects $\mathcal{P}^-$ is completely contained in $\mathcal{P}'$, therefore,

$$\begin{aligned}
&\operatorname{vol}(\mathcal{P}^-)(\det L)^{-1} \\
&= \left(1 - \frac{2Mn}{H - c'}\right) \operatorname{vol}(\mathcal{P}')(\det L)^{-1} \\
&= \left(1 - \frac{2Mn}{H - c'}\right)\left(1 - \frac{c'}{H}\right) \operatorname{vol}(\mathcal{P})(\det L)^{-1}
\end{aligned}$$

is a lower bound on the number of lattice points inside $\mathcal{P}'$.

For the upper bound on $|\mathcal{P} \cap L|$, we enlarge $\mathcal{P}$ by a factor of $1 - 2Mn/H$. Let $\mathcal{P}^+$ be the parallelepiped thus obtained. Now every element of W that intersects $\mathcal{P}$ is completely contained in $\mathcal{P}^+$, therefore, $\operatorname{vol}(\mathcal{P}^+)(\det L)^{-1} = (1 - 2Mn/H)\operatorname{vol}(\mathcal{P})(\det L)^{-1}$ is an upper bound on the number of lattice points inside $\mathcal{P}$.

Therefore, when a lattice point is picked uniformly from inside $\mathcal{P}$, the probability that it is inside $\mathcal{P}'$ is at least

$$\begin{aligned}
\frac{(1 - \frac{2Mn}{H - c'})(1 - \frac{c'}{H})}{(1 - \frac{2Mn}{H})} &\geq \left(1 - \frac{2Mn}{H}\right)\left(1 - \frac{c'}{H}\right) \\
&\geq 1 - \frac{3Mn}{H} \geq 1 - \frac{4Mn}{D},
\end{aligned}$$

which is at least $1 - 2^{-Cn-1}$, provided $D = 2^{c_0 n}$ is sufficiently large with respect to $M = 2^{\epsilon n}$. In addition, from Lemma 6 with $C'n = Cn + 1$, we have that with probability at least $1 - 2^{-Cn-1}$, every perturbation vector has norm at most 2. Finally, for each i, the distance of x_i from the complement of $\mathcal{P}$, is at most the distance of z_i from the complement of $\mathcal{P}$, plus the length of the perturbation y_i. Therefore, with probability at least $1 - 2^{-Cn}$, for every i, the distance of x_i from the complement of $\mathcal{P}$ is at least $c' - 2 = c_2$. □

The next lemma bounds the ratio of the probability that a sample x was derived from a lattice point z to the probability that it was derived from $z \pm u$, the nearest neighbors of z.

LEMMA 8. *Assume that $B(a, 2)$ is completely contained in the parallelepiped $\mathcal{P}$, and let u be a shortest non-zero vector of L satisfying $1 \leq \|u\| < 2$. Then $\Pr[\zeta = a \pm u \mid \zeta + \eta = x] \geq 2^{-2Kn} \Pr[\zeta = a \mid \zeta + \eta = x]$.*

PROOF. If $B(a, 2)$ is completely contained in $\mathcal{P}$, then both points $a \pm u$ are in $\mathcal{P}$. For $b = a \pm u$, $\|x - b\| - \|x - a\| = \|x - (a \pm u)\| - \|x - a\| \leq \|u\| < 2$. By Lemma 9 (proved below), we have $\Pr[\zeta = b \mid \zeta + \eta = x] = (g(x - a)/g(x - b)) \cdot \Pr[\zeta = a \mid \zeta + \eta = x]$. Using the fact that the density function of the perturbation vector (i.i.d. normally random variables with mean 0 and variance $1/(Kn)$) is given by

$$g(x) = \gamma(n, K) \exp\left(-Kn\|x\|^2/2\right),$$

where γ is some fixed function, we have

$$\begin{aligned} \frac{g(x-b)}{g(x-a)} &= \exp\left(\frac{-Kn(\|x-b\|^2 - \|x-a\|^2)}{2}\right) \\ &\geq \exp(-2Kn) > 2^{-2Kn}. \end{aligned}$$

□

LEMMA 9. *Assume that $x \in \mathbf{R}^n$, and $a, b \in (L \cap \mathcal{P})$. Then*

$$\frac{\Pr[\zeta = a \mid \zeta + \eta = x]}{\Pr[\zeta = b \mid \zeta + \eta = x]} = \frac{g(x-a)}{g(x-b)}.$$

Here g is the density function of η and , $\Pr[\zeta = a \mid \zeta + \eta = x]$ is defined to be $\lim_{\epsilon \to 0} \Pr[\zeta = a \mid \zeta + \eta \in B(x, \epsilon)]$, where $B(x, \epsilon)$ is the ball with radius ϵ and center x.

PROOF. It is easy to see that even though the conditions in the conditional probabilities are events of zero probability, a finite non-zero limit always exists for the the conditional probabilities. For $\epsilon > 0$, let X_ϵ denote the event that $\zeta + \eta \in B(x, \epsilon)$. Using Bayes' rule, it is easy to see that

$$\Pr[\zeta = a \mid X_\epsilon] \cdot \Pr[X_\epsilon] = \Pr[\zeta = a] \cdot \Pr[X_\epsilon \mid \zeta = a]$$

and

$$\Pr[\zeta = b \mid X_\epsilon] \cdot \Pr[X_\epsilon] = \Pr[\zeta = b] \cdot \Pr[X_\epsilon \mid \zeta = b].$$

Since a, b are uniform choices in $L \cap \mathcal{P}$, $\Pr[\zeta = a] = \Pr[\zeta = b]$. Since all the probabilities are non-zero, we get

$$\frac{\Pr[\zeta = a \mid X_\epsilon]}{\Pr[\zeta = b \mid X_\epsilon]} = \frac{\Pr[X_\epsilon \mid \zeta = a]}{\Pr[X_\epsilon \mid \zeta = b]}.$$

Now, take the limit $\epsilon \to 0$ on both sides. Since the limit exists and is non-zero,

$$\begin{aligned} \lim_{\epsilon \to 0} \frac{\Pr[\zeta = a \mid X_\epsilon]}{\Pr[\zeta = b \mid X_\epsilon]} &= \frac{\lim_{\epsilon \to 0} \Pr[\zeta = a \mid X_\epsilon]}{\lim_{\epsilon \to 0} \Pr[\zeta = b \mid X_\epsilon]} \\ &= \frac{\Pr[\zeta = a \mid \zeta + \eta = x]}{\Pr[\zeta = b \mid \zeta + \eta = x]}. \end{aligned}$$

Similarly,

$$\lim_{\epsilon \to 0} \frac{\Pr[X_\epsilon \mid \zeta = a]}{\Pr[X_\epsilon \mid \zeta = b]} = \frac{\lim_{\epsilon \to 0} \Pr[\zeta + \eta \in B(x, \epsilon) \mid \zeta = a]}{\lim_{\epsilon \to 0} \Pr[\zeta + \eta \in B(x, \epsilon) \mid \zeta = b]}.$$

The lemma follows by noting that the continuity of g implies that for any $y \in \mathbf{R}^n$, $\lim_{\epsilon \to 0} \Pr[\zeta + \eta \in B(x, \epsilon) \mid \zeta = y] = g(x - y)$, where $g(\cdot)$ is the density function of η. □

3.5 The iterative step and analysis

We shall assume that we now have a lattice L satisfying the properties listed in 5, and we also have N samples of the random variable $\xi = \xi^{(L, \mathcal{P})}$ as defined in Section 3.4, with the properties stated in Lemmas 6 and 7. Note that by Lemma 7(7), each x_i satisfies $\|x_i\|_\infty \leq 2^{c_0' n}$. Define $F = 2^{c_0' n}$, and let c_3 denote the constant from Lemma 6.

LEMMA 10. *There exists a constant $\alpha > 0$ so that for each $j = 1, 2, \ldots,$ we can construct in time less than $j 2^{\alpha n}$, a subset S_j of $\{1, \ldots, N\}$ and for each $r \in S_j$ an $a_r^{(j)} \in L$, such that $\|x_r - a_r^{(j)}\|_\infty \leq F 2^{-j} + 2c_3$ and $|S_j| \geq N - j 2^{2n}$. For this construction we use only x_r (and not y_r or z_r) if $r \in S_j$. For $r \notin S_j$ we use y_r and z_r as well.*

PROOF. The proof is by induction on j, using Lemma 2. Let $S_0 = \{1, \ldots, N\}$, and for each $r \in S_0$, define $a_r^{(0)} = 0$. Clearly, $a_r^{(0)} \in L$ and $\|x_r - a_r^{(0)}\|_\infty \leq F + 2c_3$ for each $r \in S_0$. For $j > 0$, the set S_j is constructed from S_{j-1} as follows: define the set $X_{j-1} = \{x_r - a_r^{(j-1)} \mid r \in S_{j-1}\}$. Inductively, we have that for each $r \in S_{j-1}$, $\|x_r - a_r^{(j-1)}\|_\infty \leq F 2^{-(j-1)} + 2c_3$, so every point in X_{j-1} has L_∞ norm at most $F 2^{-(j-1)} + 2c_3$. Apply Lemma 2 with X_{j-1} playing the role of the sequence $\langle a_i \mid i = 1, \ldots, t\rangle$ and setting $q = 1$. Let S_j be the Φ returned by the algorithm of Lemma 2. Lemma 2(1) ensures that $|S_j| \geq |X_{j-1}| - 2^{2n} = |S_{j-1}| - 2^{2n} \geq (N - (j - 1)2^{2n}) - 2^{2n} = N - j 2^{2n}$. For $r \in S_j$, let $s \in S_{j-1} - S_j$ be the element that satisfies $\|(x_r - a_r^{(j-1)}) - (x_s - a_s^{(j-1)})\|_\infty \leq (F 2^{-(j-1)} + 2c_3)/2$. For $r \in S_j$, define $a_r^{(j)} = a_r^{(j-1)} + z_s - a_s^{(j-1)}$ (recall that x_s is defined to be $z_s + y_s$, where $z_s \in L$ and y_s is the small perturbation). Clearly, $a_r^{(j)} \in L$. Moreover, $\|x_r - a_r^{(j)}\|_\infty = \|x_r - (a_r^{(j-1)} + z_s - a_s^{(j-1)})\|_\infty = \|x_r - (a_r^{(j-1)} + x_s - y_s - a_s^{(j-1)})\|_\infty \leq \|(x_r - a_r^{(j-1)}) - (x_s - a_s^{(j-1)})\|_\infty + \|y_s\|_\infty \leq (F 2^{-j} + c_3) + \|y_s\|_2 \leq (F 2^{-j} + c_3) + c_3$ (by Lemma 6) $= F 2^{-j} + 2c_3$. The running time for each iterative step is the running time of the algorithm of Lemma 2, which is bounded by $2^{\alpha n}$ for a suitably large constant α since the maximum number of points sent to this algorithm is at most $N = 2^{c_1 n}$. Finally, it is also clear that as long as $r \in S_j$, the decomposition of x_r as $y_r + z_r$ is never used by the algorithm. □

Let j be the smallest integer such that $F 2^{-j} \leq c_3$. Apply the algorithm of Lemma 10 until this value of j. Since $F = 2^{c_0' n}$, $j = O(n)$. Note that $|S_j| \geq N - j 2^{2n}$. Furthermore, for every $r \in S_j$, we have $\|x_r - a_r^{(j)}\|_2 \leq \sqrt{n}\|x_r - a_r^{(j)}\|_\infty \leq 3c_3\sqrt{n}$. Let $G = c_3\sqrt{n}$, $N' = N - j 2^{2n}$.

LEMMA 11. *There exists a constant $\beta > 0$ so that for each $k = 1, 2, \ldots,$ we can construct in time less than $k 2^{\beta n}$, a subset T_k of $\{1, \ldots, N\}$ and for each $r \in T_k$ a $b_r^{(k)} \in L$, such that $\|x_r - b_r^{(k)}\|_2 \leq G 2^{-k} + 2c_3$ and $|T_k| \geq N' 2^{-k}$. For this construction we use only x_r (and not y_r or z_r) if $r \in T_k$. For $r \notin T_k$ we use y_r and z_r as well. The constructions succeed with probability at least $1 - 2^{-n}$.*

PROOF. The proof mimics the proof of Lemma 10, using Lemma 3 instead of Lemma 2; consequently, the number of points lost in each stage is much larger, and there is a small error probability to deal with. In the applications of

Lemma 3 below, we will assume that the algorithm outlined in Lemma 3 is repeated many times so that all the applications succeed with probability at least $1-2^{-2n}$. Other than this error probability, the only other source of error here is from the application of Lemma 6, which is bounded by 2^{-2n}. In the sequel, we will therefore assume that all applications of Lemma 3 are successful.

The proof is by induction on k, using Lemma 3. Let $T_0 = S_j$, and for each $r \in T_0$, define $b_r^{(0)} = a_r^{(j)}$. Clearly, $b_r^{(0)} \in L$ and $\|x_r - b_r^{(0)}\| \leq G + 2c_3$ for each $r \in T_0$. For $k > 0$, the set T_k is constructed from T_{k-1} as follows: define the set $X'_{k-1} = \{x_r - b_r^{(k-1)} \mid r \in T_{k-1}\}$. Inductively, we have that for each $r \in T_{k-1}$, $\|x_r - b_r^{(k-1)}\| \leq G2^{-(k-1)} + 2c_3$, so every point in X'_{k-1} has L_2 norm at most $G2^{-(k-1)} + 2c_3$. Apply Lemma 3 with X'_{k-1} playing the role of the sequence $\langle a_i \mid i = 1, \ldots, t\rangle$. Let T_k be the Φ returned by the algorithm of Lemma 3. Lemma 3(3) ensures that $|T_k| \geq |X'_{k-1}|/2 = |T_{k-1}|/2 \geq (N'2^{-(k-1)})/2 = N'2^{-k}$. For $r \in T_k$, let $s \in T_{k-1} - T_j$ be the element that satisfies $\|(x_r - b_r^{(k-1)}) - (x_s - b_s^{(k-1)})\| \leq (G2^{-(k-1)} + 2c_3)/2$. For $r \in T_k$, define $b_r^{(k)} = b_r^{(k-1)} + z_s - b_s^{(k-1)}$ (recall that x_s is defined to be $z_s + y_s$, where $z_s \in L$ and y_s is the small perturbation). Clearly, $b_r^{(k)} \in L$. Moreover, $\|x_r - b_r^{(j)}\| = \|x_r - (b_r^{(k-1)} + z_s - b_s^{(k-1)})\| = \|x_r - (b_r^{(k-1)} + x_s - y_s - b_s^{(k-1)})\| \leq \|(x_r - b_r^{(k-1)}) - (x_s - b_s^{(k-1)})\| + \|y_s\| \leq (G2^{-k} + c_3) + c_3$ (by Lemma 6) $= G2^{-k} + 2c_3$. The running time for each iterative step is the running time of the algorithm of Lemma 3, which is bounded by $2^{\beta n}$ for a suitably large constant β since the maximum number of points sent to this algorithm is at most $N = 2^{c_1 n}$. Finally, it is also clear that as long as $r \in S_k$, the decomposition of x_r as $y_r + z_r$ is never used by the algorithm. □

Let k be the smallest integer such that $G2^{-k} \leq c_3$. Apply the algorithm of Lemma 11 until this value of k. Since $G = c_3\sqrt{n}$, $k = O(\log n)$; also recall that $j = O(n)$. Note that $|T_k| \geq N'2^{-k} = (N - j2^{2n})2^{-k} \geq 2^{(c_1-1)n}$ for sufficiently large n, provided $c_1 > 2$. Furthermore, for every $r \in T_k$, we have $\|x_r - b_r^{(k)}\|_2 \leq 3c_3$. Denote $b_r^{(k)}$ by b_r. Finally, for each $r \in T_k$, let $w_r = z_r - b_r$. Clearly, $w_r \in L$. Furthermore, $\|w_r\| = \|z_r - b_r\| = \|(x_r + y_r) - b_r\| = \|(x_r - b_r) + y_r\| \leq \|x_r - b_r\| + \|y_r\| \leq 3c_3 + c_3 = 4c_3$, using Lemma 6.

For ease of further analysis, we will just pick one $r \in T_k$, and let $x = x_r, z = z_r, y = y_r, b = b_r$, and $w = w_r$. Note that $x = z + y$, $w = z - b$. The output of this algorithm is the lattice point w.

At this point, we have a procedure that outputs a lattice point w of length at most $4c_3$, a constant. This would already be a useful algorithm, except for two problems. First, while we have assumed that $\mathrm{sh}(L) < 2$, this algorithm only finds a vector of length at most $4c_3$. Secondly, and more importantly, the point it produces could be zero, in which case we obtain nothing non-trivial. To solve these problems, we will take the following approach. The algorithm of Lemma 11 is completely oblivious to the decomposition of x as $z+y$; indeed, it would work just as well if $x = z' + y'$ for some other lattice point z' with corresponding perturbation y'. Furthermore, since $\mathrm{sh}(L) < 2$ and since the perturbations have (constant) expected length $1/K$, there is a pretty good probability that $x = (z+u)+y'$, where u is a shortest vector in L. We proceed to the details.

We have proved that if we pick $N = 2^{c_1 n}$ samples of the random variable $\xi = \xi^{(L,\mathcal{P})}$ and apply the procedures of Lemmas 10 and 11 suitable number of times (after appropriate pre-processing outlined in Lemma 5), then we will produce, with probability exponentially close to one, a point $w \in L$ such that $\|w\| \leq 4c_3$. This probability is with respect to the randomization of the samples $x_1, \ldots, x_N$, as well as with respect to the randomness in the algorithm of Lemma 3. The next lemma summarizes some crucial properties of the w produced by the entire algorithm.

For $w \in L$, let p_w denote the probability that the algorithm outputs w, where the probability is taken over all the random choices made by the algorithm.

LEMMA 12. *There exist constants $c_5, c_6 > 0$ such that with probability at least $1-2^{-n}$ the following conditions hold:*

(9). $\|w\| \leq c_5$,

(10). $p_w \geq 2^{-c_6 n}$.

PROOF. Part (9) has already been proved above with $c_5 = 4c_3$; Part (10) is based on Lemma 22, which states that there is a c_6 depending only on c_5 such that if $\mathrm{sh}(L) \geq 1$, then the number of lattice points of length c_5 is at most $2^{c_6 n}$. Therefore, conditioned on the fact that the algorithm produces a lattice point of length at most c_5, at least one of these points has a probability of at least $2^{-c_6 n}$. □

To complete the description and the analysis of the algorithm, we will use the following lemma, whose proof we present after the proof of the main theorem.

LEMMA 13. *Assume that $u \in L$ be such that $\|u\| = sh(L)$. Then for any $w \in L$ with $\|w\| \leq c_5$ and $p_w \geq 2^{-c_6 n}$, we have $p_{w \pm u} \geq 2^{-3Kn} p_w$. Here K is the parameter described in Section 3.4.*

Let $w \in L$ be a lattice point whose existence is promised by Lemma 12, and let u be a shortest non-zero vector in L. Then by Lemma 13, $w + u$ and $w - u$ have probability at least $2^{-3Kn} p_w \geq 2^{-(c_6+3K)n}$. Therefore, if we repeat the entire algorithm $2^{c_7 n}$ times (with independent random choices each time) for a suitably large constant $c_7 > 0$, we can ensure that with probability at least $1 - 2^{-n}$, both w and $w + u$ must be produced as an output. Thus, our final algorithm is the following: repeat the entire algorithm $2^{c_7 n}$ times; for every pair of vectors $a, b \in L$ produced by the algorithm, compute $a + b$ and $a - b$, and output the shortest non-zero vector from all such sums and differences. ■

QED Theorem 1

Remark. If we do not assume that u is a shortest vector but assume only that $\|u\| \leq \mu$ then the statement of Lemma 13 remains true if we replace the conclusion by "either $p_{w+u} \geq 2^{-\nu n} p_w$ or $p_{w-u} \geq 2^{-\nu n} p_w$, where $\nu > 0$ depends only on μ". Using this, we can compute all vectors of length at most $d \cdot \mathrm{sh}(L)$ for any constant $d \geq 1$, in time $2^{d' n}$, where d' depends only on d.

Let $g(x)$ denote the probability density function of the distribution of η.

PROOF. (of Lemma 13) Let $w \in L$ such that $\|w\| \leq c_5$ and $p_w \geq 2^{-c_6 n}$. The proof uses conditional probabilities. To avoid using conditions with 0 probability we will partition the whole space $\mathbf{R}^n$ into small cubes with sides of length ϑ where ϑ is small enough with respect to n. We will denote by Z the set of all cubes of this type.

Our algorithm gives at the end a perturbed vector x and a lattice point b so that $\|x - b\| < 3c_3$. At this point we randomize the way x was derived and we get $x = z + y$ where $z \in L$ and y is the perturbation.

Let G be the set of all pairs $(S, h) \in Z \times (L \cap \mathcal{P})$, so that $\Pr[x \in S \wedge b = h] \neq 0$. For $(S, h) \in G$, and $w \in L$, we define $p_w^{(S,h)}$ to be the probability p_w, conditioned on the event "$x \in S$ and $b = h$." Note that the probability in $p_w^{(S,h)}$ is only with respect to the randomization of z, indeed, $p_w^{(S,h)} = \Pr[z - h = w \mid x \in S \wedge b = h]$.

Now $p_w = \sum_{(S,h) \in Z \times (L \cap \mathcal{P})} p_w^{(S,h)} \Pr[x \in S \wedge g = h]$. Let $I \subseteq G$ be the set of all pairs (S, h) such that both S and h are at a distance of at least c_2 from the complement of $\mathcal{P}$. Here c_2 is the constant from Lemma 7, which we will take to be at least 2. Note that $\sum_{(S,h) \in G - I} \Pr[x \in S \wedge b = h] \leq \sum_{S \mid (\exists h)[(S,h) \in G \setminus I]} \Pr[x \in S] \leq 2^{-Cn}$.

Next we prove that if $(S, h) \in I$, $p_{w+u}^{(S,h)} \geq 2^{-2Kn} p_w^{(S,h)}$. This follows since h is at a distance of at least $c_2 \geq 2$ from the complement of $\mathcal{P}$, by Lemma 8,

$$\begin{aligned} p_{w+u}^{(S,h)} &= \Pr[z - h = w + u \mid x \in S \wedge b = h] \\ &= \Pr[\zeta = w + h + u \mid x \in S \wedge b = h] \\ &\geq 2^{-2Kn} \Pr[\zeta = w + h \mid x \in S \wedge b = h] \\ &= p_w^{(S,h)}. \end{aligned}$$

Thus

$$\begin{aligned} p_w &= \sum_{(S,h) \in I} p_w^{(S,h)} \Pr[x \in S \wedge b = h] \\ &\quad + \sum_{(S,h) \in G \setminus I} p_w^{(S,h)} \Pr[x \in S \wedge b = h] \\ &\leq \sum_{(S,h) \in I} 2^{2Kn} p_{w+u}^{(S,h)} \Pr[x \in S \wedge b = h] + 2^{-Cn} \\ &\leq 2^{2Kn} p_{w+u} + 2^{-Cn}. \end{aligned}$$

Using the fact that $2^{c_6 n} p_w \geq 1$, we have

$$\begin{aligned} p_{w+u} &\geq 2^{-2Kn} p_w - 2^{-(C+2K)n} \\ &\geq 2^{-2Kn} p_w - 2^{-(C+2K)n} 2^{c_6 n} p_w \\ &= (2^{-2Kn} - 2^{-(2K+1)n}) p_w \geq 2^{-3Kn} p_w. \end{aligned}$$

Here we assume that $C - c_6 \geq 1$; recall that c_6 is the constant from Lemma 22, and it depends only on $c_5 = 4c_3$, and c_3 is the constant from Lemma 6 that bounds the length of all perturbation vectors. Thus the choice of C is (so far) independent of c_6. Now we choose $C \geq c_6 + 3$, and apply Lemma 7 with this choice of C. This also defines the constants c_0', c_0 in Lemma 7, and the constant K in Lemma 6 (via Lemma 7). □

Remark. If we want not only the shortest vector but all the vectors shorter than a constant, then we have to use the same lemma once for each required vector u', replacing u by u'. Naturally the size of the perturbations should be modified to derive similar conclusions from Lemma 8.

4. SOME APPLICATIONS

4.1 Basis reductions

COROLLARY 14. *There is a randomized algorithm that, given an n dimensional lattice L, runs in time $2^{O(n)}$ to obtain the Korkine-Zolotareff basis for L.*

PROOF. Let the lattice L be given by its basis $b_1, \ldots, b_n$. First, we find the shortest vector a_1 in L. Then, we inductively find the Korkine-Zolotareff basis for the lattice defined by $b_2(2), \ldots, b_n(2)$, where $b_i(j)$ denotes the projection of b_i onto the space which is the orthogonal complement of $b_1, \ldots, b_i$. Then, for $i = 2, \ldots, n$, we output a_i, the shortest lattice vector whose projection onto the space orthogonal to a_1 is b_i, as in [6], to satisfy $|\mu_{1,i}| \leq 1/2$, where $\mu_{1,i} = \langle a_1, a_i(i) \rangle / \|b_i(i)\|$. Clearly, the running time of this algorithm is at most $2^{O(n)}$. □

COROLLARY 15. *There is a randomized algorithm that, given an integer $k > 0$ and an n dimensional lattice L with $1 \leq sh(L) \leq 2$, finds a non-zero lattice vector u such that $\|u\| \leq O((k^2)^{n/k})$ in time $O(n^2(2^{O(k)} + n^2))$.*

PROOF. The corollary follows from [13], with the enhancement that Korkine-Zolotareff bases can be found in $2^{O(n)}$ time by Corollary 14. □

By choosing $k = \epsilon n$, we can obtain an $O(n^{1/\epsilon})$-approximate short vector in $2^{O(\epsilon n)}$ time. By choosing $k = \log n$, we can obtain a $2^{n \log \log n / \log n}$-approximate short vector in polynomial time; this is an improvement over the LLL algorithm which produces an approximation of $2^{n/2}$ in polynomial time and its subsequent enhancement [13] which produces an approximation of $2^{n \log \log^2 n / \log n}$ in polynomial time. All the applications of the LLL algorithm can be adapted to this improved approximation to the shortest lattice vector.

4.2 Closest vector problem

The closest vector problem is the inhomogeneous version of the shortest vector problem. In this problem, given an n-dimensional lattice L and a vector $u \in \mathbf{R}^n$, find a vector $v \in L$ that minimizes $\|u - v\|$. If one can find $v \in L$ such that $\|u - v\| \leq \epsilon(n) \|u - v'\|, \forall v' \in L$, then we say that the closest vector problem can be approximated to within factor $\epsilon(n)$. As a consequence of our $2^{O(n)}$ algorithm for the shortest vector problem, we obtain the following two applications:

COROLLARY 16. *There is a randomized $2^{O(n)}$-time algorithm to approximate the closest vector problem to within a factor $\sqrt{n/2}$.*

PROOF. We apply Theorem 6.8 of [7]. This theorem shows that the problem of finding an approximate closest vector to within a factor of $\sqrt{n/2}$ is polynomial-time Cook reducible to the decision version of the shortest vector problem. Since the latter can be solved in time $2^{O(n)}$, the corollary follows. □

COROLLARY 17. *Given an n dimensional lattice L such that the shortest vector of L is a constant and a vector $u \in \mathbf{R}^n$, there is a randomized algorithm that runs in time $2^{O(n)}$ such that if there is a $v \in L$ such that $\|u - v\| \leq \sqrt{n / \log n}$, then the algorithm outputs v.*

PROOF. We apply the main theorem of [8]. This theorem gives an algorithm that for any $k > 0$, runs in time $n^{k^2+O(1)}$ and finds the closest vector to u if the distance of v to the lattice is at most k time the length of the shortest Gram-Schmidt vector. Now, given any basis for L, we first construct in time $2^{O(n)}$, a Gram-Schmidt basis with the shortest vector of L included. Now, the corollary follows by setting $k = \sqrt{n/\log n}$. □

Using ideas similar to the proof of the main theorem, we can also obtain a $2^{O(n)}$ time algorithm to find a non-zero codeword of minimum weight in linear codes of block length n over fields of size poly(n). The details of this, and further consequences for the integer programming problem, following [7], will be given in the full paper.

5. REFERENCES

[1] M. Ajtai. Generating hard instances of lattice problems. *Proc. 28th ACM Symposium on Theory of Computing*, pp. 99–108, 1996. Full version available as ECCC Technical Report TR96-007 at www.eccc.uni-trier.de/eccc/.

[2] M. Ajtai. The shortest vector problem in L_2 is NP-hard for randomized reductions. *Proc. 30th ACM Symposium on Theory of Computing*, pp. 10–19, 1998.

[3] A. Blum, A. Kalai, and H. Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *Proc. 32nd ACM Symposium on Theory of Computing*, pp. 435–440, 2000.

[4] P. van Emde Boas. Another NP-complete partition problem and the complexity of computing short vectors in lattices. *Mathematics Department, University of Amsterdam*, TR 81-04, 1981.

[5] O. Goldreich and S. Goldwasser. On the limits of nonapproximability of lattice problems. *Journal of Computer and System Sciences*, 60(3):540–563, 2000.

[6] B. Helfrich. Algorithms to construct Minkowski reduced and Hermite reduced bases. *Theoretical Computer Science*, 41:125–139, 1985.

[7] R. Kannan. Minkowski's convex body theorem and integer programming. *Mathematics of Operations Research*, 12:415–440, 1987. Preliminary version in *ACM Symposium on Theory of Computing* 1983.

[8] P. Klein. Finding the closest lattice vector when it's unusually close. *Proc. 11th Symposium on Discrete Algorithms*, 2000.

[9] R. Kumar and D. Sivakumar. On polynomial approximations to the shortest lattice vector length. *Proc. 12th Symposium on Discrete Algorithms*, 2001. To appear.

[10] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261:515–534, 1982.

[11] L. Lovász. *An Algorithmic Theory of Numbers, Graphs and Convexity*. CBMS-NSF Regional Conference Series on Applied Mathematics, SIAM, 1986.

[12] D. Micciancio. The shortest vector in a lattice is hard to approximate to within some constant. *Proc. 39th IEEE Symposium on Foundations of Computer Science*, pp. 92–98, 1998.

[13] C. P. Schnorr. A hierarchy of polynomial time basis reduction algorithms. *Theoretical Computer Science*, 53:201–224, 1987.

APPENDIX

LEMMA 18. *There is a $c > 0$ so that if n is a positive integer, and $\xi = \langle \xi_1, ..., \xi_n \rangle$ is a random variable with values in $\mathbf{R}^n$ so that the random variables $\xi_1, ..., \xi_n$ are independent and each have normal distribution with expected value 0 and variance $n^{-1/2}$, then*

$$\Pr\left[1/2 \le \|\xi\|_2\right] \ge 1 - 2^{-nc}$$

PROOF. We have to prove that $\Pr[1/4 \le \sum_{i=1}^n \xi_i^2] \ge 1 - 2^{-cn}$ for some constant $c > 0$. If $y > 0$ let $\xi_{i,y} = \min\{\xi_i, y\}$. Since $n^{1/2}\xi$ has normal distribution with mean 0 and variance 1 there is a constant $\alpha > 0$ so that $\mathrm{E}[n\xi_{i,\alpha}^2] > 3/4$ and so $\mathrm{E}[\xi_{i,\alpha}^2] \ge (3/4)n^{-1}$. Since $\xi_{i,\alpha} \le \xi$ it is enough to show that $\Pr[1/4 \le \sum_{i=1}^n \xi_{i,\alpha}^2] \ge 1 - 2^{-cn}$. $\sum_{i=1}^n \xi_{i,\alpha}^2$ is a sum of independent random variables with identical distributions each taking their values in the interval $[0, \alpha]$ where α does not depend on n. By the definition of α the expected value of the sum is $3/4$. So our statement follows from Lemma 19 below. □

LEMMA 19. *For all $\alpha > 0$, $c_1 > 0$ there is a c_2 so that for all sufficiently large n if $\eta_1, ..., \eta_n$ are independent random variables with identical distributions so that $0 \le \eta_i \le \alpha$ for $i = 1, ..., t$ then $\Pr[|(\sum_{i=1}^n \eta_i) - E[\sum_{i=1}^n \eta_i]| > c_1] \le 2^{-c_2 n}$.*

PROOF. The statement of the lemma is a consequence of Chernoff's inequality. □

LEMMA 20. *There is a $c > 0$ so that for all $x > 1$ if n is a sufficiently large positive integer, $u \in \mathbf{R}^n$ and v is a random element of the unit sphere of $\mathbf{R}^n$ then $\Pr[\langle u, v \rangle \ge xn^{-1/2}\|u\|] \le e^{-cx^2}$.*

PROOF. We randomize w in the following way. Let ξ be a random variable with values in $\mathbf{R}^n$ whose components are independent and each have a normal distribution with expected value 0 and variance $n^{-1/2}$. Let $v = \frac{\xi}{\|\xi\|}$. With probability 1, v is defined and it is an element of the unit sphere. It is easy to see that its distribution is uniform. $\langle u, v \rangle = \frac{1}{\|\xi\|}\langle u, \xi \rangle$. Lemma 18 implies that with a probability of at least $1 - 2^{-c'n}$ we have $n^{1/2}/2 \le \|\xi\|_2$. Therefore $\Pr[\langle u, v \rangle \ge xn^{-1/2}\|u\|] \le 2^{-c'n} + \Pr[\langle u, \xi \rangle \ge 1/2xn^{-1/2}\|u\|]$. We may assume that $\|u\| = 1$. $\langle u, \xi \rangle$ has a normal distribution with mean 0 and variance $n^{-1/2}$. So $\Pr[\langle u, \xi \rangle \ge 1/2xn^{-1/2}] \le (2\pi)^{-1/2}e^{-x^2/8}$. We obtain $\Pr[\langle u, v \rangle \ge xn^{-1/2}\|u\|] \le 2^{-c'n} + (2\pi)^{-1/2}e^{-x^2/8} \le e^{-cx^2}$ if $c > 0$ is sufficiently large. □

LEMMA 21. *If $\xi_i, i = 1, \ldots, n$ are identical independent normally distributed random variables with mean 0 and variance $1/(Kn)$, then for any constant $C > 1$,*

$$\Pr\left[\sum_{j=1}^{n} \xi_j^2 > C\right] \le \exp\left(-n\left(\frac{K}{4} - \frac{1}{2C}\right)\right).$$

Proof. Let $Z = \sum_{j=1}^{n} \xi_j^2$. Since ξ_j's are identically distributed normal random variables with mean 0 and variance $1/(Kn)$, then for any $\delta > 0$,

$$\mathrm{E}[e^{\delta Z}] = \mathrm{E}\left[e^{\delta \sum_j \xi_j^2}\right] = \mathrm{E}\left[\prod_j e^{\delta \xi_j^2}\right] = \left(\mathrm{E}[e^{\delta \xi^2}]\right)^n,$$

where the third equality is by independence of the ξ_j's, and where ξ is a normally distributed random variable with mean 0 and variance $1/(Kn)$. It can be shown that

$$\mathrm{E}[\exp(\delta\xi^2)] = \left(1 - \frac{2\delta}{Kn}\right)^{-1/2},$$

and hence

$$E[\exp(\delta Z)] = \left(1 - \frac{2\delta}{Kn}\right)^{-n/2}.$$

Therefore, for any $\delta > 0$, using Markov's inequality,

$$\begin{aligned} \Pr[Z > C] &= \Pr[\exp(\delta Z) > \exp(\delta C)] \\ &< \exp(-\delta C)\, E[\exp(\delta Z)] \\ &= \exp(-\delta C)\left(1 - \frac{2\delta}{Kn}\right)^{-n/2}. \end{aligned}$$

Let $\epsilon = 1/(4C)$ and $\delta = \epsilon K n$. Note that since $C > 1$, $\epsilon < 1/4$. Also, since $1 - 2\epsilon > \exp(-4\epsilon)$ for $\epsilon < 1/4$,

$$\begin{aligned} \Pr[Z > C] &\leq \exp(-\epsilon C K n)\,(1 - 2\epsilon)^{-n/2} \\ &< \exp(-\epsilon C K n) \cdot \exp(4\epsilon n/2) \\ &= \exp\left(-\left(\frac{K}{4} - \frac{1}{2C}\right) n\right). \end{aligned}$$

□

Lemma 22. *For any $q > 0$ and any lattice L in $\mathbf{R}^n$ such that $sh(L) \geq 1$, $|L \cap B(0,q)| \leq (2q+1)^n$.*

Proof. If we place balls of radius $1/2$ with centers at all points in $L \cap B(0,q)$, the balls don't intersect, and every such ball is contained in $B(0, q+1/2)$. Therefore, we may upper bound $|L \cap B(0,q)|$ by the maximum number of non-intersecting balls of radius $1/2$ completely contained inside a ball of radius $q + 1/2$. This number is at most $\mathrm{vol}(B(0, q + 1/2))/\,\mathrm{vol}(B(0, 1/2))$, which is at most $(2q+1)^n$. □