New Lattice Based Cryptographic Constructions

Oded Regev *

August 17, 2004

Abstract

We introduce the use of Fourier analysis on lattices as an integral part of a lattice based construction. The tools we develop provide an elegant description of certain Gaussian distributions around lattice points. Our results include two cryptographic constructions that are based on the worst-case hardness of the unique shortest vector problem. The main result is a new public key cryptosystem whose security guarantee is considerably stronger than previous results ($O(n^{1.5})$ instead of $O(n^7)$). This provides the first alternative to Ajtai and Dwork's original 1996 cryptosystem. Our second result is a family of collision resistant hash functions with an improved security guarantee in terms of the unique shortest vector problem. Surprisingly, both results are derived from one theorem that presents two indistinguishable distributions on the segment $[0, 1)$. It seems that this theorem can have further applications; as an example, we use it to solve an open problem in quantum computation related to the dihedral hidden subgroup problem.

1 Introduction

Cryptographic constructions based on lattices have attracted considerable interest in recent years. The main reason is that, unlike many other cryptographic constructions, lattice based constructions can be based on the *worst-case* hardness of a problem. That is, breaking them would imply a solution to *any* instance of a certain lattice problem. In this paper we will be interested in the unique shortest vector problem (uSVP), a lattice problem that is believed to be hard. For a constant c, the n^c-uSVP is defined as follows: we are asked to find the shortest non-zero vector in an n-dimensional lattice with the promise that it is shorter by a factor of n^c than all other non-parallel vectors. Hence, the problem becomes harder as c decreases. The results in this field can be divided into two types. The first includes public key cryptosystems and the second includes families of collision resistant hash functions.

The only previously known public key cryptosystem based on a worst-case lattice problem is the one due to Ajtai and Dwork [2], which appeared in 1996. They presented a public key cryptosystem based on the worst-case hardness of $O(n^8)$-uSVP. Then, in [10], Goldreich, Goldwasser and Halevi showed how to eliminate decryption errors that existed in the original scheme. They also improved the security to $O(n^7)$-uSVP. Although there are other lattice based cryptosystems (see, e.g., [11,

*Department of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Most of this work was done while the author was at the Institute for Advanced Study, Princeton, NJ. Work supported by the Army Research Office grant DAAD19-03-1-0082 and NSF grant CCR-9987845.

15, 20]), none of them is based on the worst-case hardness of a lattice problem. Our main result is a new public key cryptosystem whose security is based on $O(n^{1.5})$-uSVP.

In [1], Ajtai presented a family of one-way hash functions based on the worst-case hardness of several lattice problems. In terms of the uSVP, it was based on the hardness of $O(n^c)$-uSVP. The constant c was not explicitly specified but later it was noted to be $c = 19$ [4]. In [9], it was shown that under the same assumptions one can obtain a family of collision resistant hash functions. This is a stronger primitive than a one-way function with many uses in cryptography. Cai and Nerurkar [5] improved the exponent to $c = 9 + \epsilon$ and later, by providing an improved analysis, Cai [4] obtained $c = 4 + \epsilon$. These papers also showed how to base the security of the hash function on other lattice problems that are potentially harder than uSVP (such as the shortest vector problem and the shortest independent vectors problem). In [22], Micciancio constructed a family of hash functions with the best known constant c for several important lattice problems (but not for uSVP). In another related paper [21], Micciancio improved the efficiency of the hash function by using cyclic lattices.

Our contribution

The main contribution of this paper is the introduction of Fourier analysis on lattices as an integral part of a lattice based construction. Fourier analysis was previously used indirectly through transference theorems, i.e., theorems that relate properties of a lattice and its dual (see, e.g., [4]). Our constructions are the first to use Fourier analysis directly.

Our main theorem is a reduction from the $O(n^{1.5})$-uSVP to the problem of distinguishing between two types of distributions on the segment $[0, 1)$. We believe that this theorem will find other uses in the future.

Using the main theorem, we present three results. The main one is a new public key cryptosystem that is based on the worst-case hardness of $O(n^{1.5})$-uSVP. This is a major improvement to the 1996 cryptosystem by Ajtai and Dwork. Its description is surprising in that it essentially consists only of numbers modulo some large number N.

Our second result is a family of collision resistant hash functions whose security is based on the worst-case hardness of $O(n^{1.5})$-uSVP. In terms of the uSVP, this improves all the previous results mentioned above. However, we should mention that previous results were not based only on uSVP and are therefore incomparable with our result. The hash function that we consider is simple and is known as the modular subset sum function[1]. This function already appeared in previous papers; for example, one of the results in [16] is an average-case to average-case reduction for the function. Finally, let us mention that a recent paper [24] greatly improves on our result as well as on all previous results. There, a family of collision resistant hash functions is shown whose security is based on the worst-case hardness of three important lattice problems (shortest vector problem, shortest independent vectors problem, and covering radius problem) all with an approximation factor of $\tilde{O}(n)$ (i.e., linear up to logarithmic terms). The proof of this strong result is based on ideas from this paper. Our third result is related to an open question in quantum computation and will be discussed in Section 7.

[1]Previous constructions of hash functions were usually presented as functions on random lattices. However, most of these results can be easily extended to the modular subset sum function. This was already noted in Ajtai's original paper ([1]).

Intuitive overview

In the following we provide an informal overview of the results in this paper. Many of the details are omitted for the sake of clarity.

Main theorem: Our main theorem is a reduction from n-dimensional instances of $n^{1.5}$-uSVP (and in general, n^c-uSVP) to the problem of distinguishing between two types of distributions on $[0, 1)$. One distribution is the uniform distribution U while the other, T_h, is concentrated around integer multiples of $1/h$ for some *unknown* large integer $h \leq 2^{O(n^2)}$. Notice that if we knew h, we could easily distinguish between the two by multiplying by h and checking if the result is close to an integer. The sharpness of the concentration in this 'wavy' distribution depends on the exponent c in the n^c-uSVP problem. For example, $n^{1.5}$-uSVP translates to a concentration of around $1/n$, that is, the difference between two adjacent peaks is roughly n times the width of a peak (see Figure 1).

The reduction is a Cook reduction. It works by producing distributions that are known to be either U or T_h for some h and then it uses the oracle to find out which of the two is the case. The value h in the T_h distributions produced by the reduction depends on the shortest vector in the lattice; this explains why h is unknown. Still, we can guarantee that h is not greater than $2^{O(n^2)}$ (this is done by working with an LLL-reduced basis).

Notice that the reduction is to a worst-case problem in the sense that one has to distinguish between U and T_h for *all* values $h \leq 2^{O(n^2)}$. Nevertheless, T_h has the property that if one can distinguish it from uniform for some *non-negligible fraction* of all possible h in some range then one can also distinguish it from uniform for *all* values of h. This average-case to worst-case property will be described in more detail later. In the following we describe four reductions that together form the proof of the main theorem.

The first reduction is from the search problem $n^{1.5}$-uSVP to a certain decision problem on lattices. The input to the decision problem is an $n^{1.5}$-unique lattice given by a basis $v_1, \ldots, v_n$. Assume that the shortest vector is $\sum_{i=1}^{n} a_i v_i$ where $a_i \in \mathbb{Z}$. The decision problem asks whether $p \mid a_i$ where p is some prime number chosen to be slightly more than $n^{1.5}$. The reduction is a Cook reduction and the idea is to make the lattice sparser and sparser without losing the shortest vector. At the end, the lattice is so sparse that we can easily find the shortest vector. For example, if we find some i such that $p \mid a_i$ then we can replace v_i with $p \cdot v_i$ without losing the shortest vector. We also need to handle the case where for all i, $p \nmid a_i$. Here we perform a slightly more complicated procedure. The idea is to add a multiple of one basis vector to another in such a way that $p \mid a_i$ for some i and then, as before, we can replace v_i with $p \cdot v_i$.

The second reduction is from the above decision problem to a promise problem. In this promise problem, we are given a lattice that either has one short vector of length $1/n$ and all other non-parallel vectors are of length more than $\sqrt{n}$, or all vectors are of length more than $\sqrt{n}$. The goal is to distinguish between these two cases with the promise that one of them is the case. The input to the reduction is a $n^{1.5}$-unique lattice and we should decide whether $p \mid a_i$. We do this by first scaling the lattice so that the length of the shortest vector is $1/n$ and therefore all non-parallel vectors are of length more than $n^{1.5}/n = \sqrt{n}$. We then replace the basis vector v_i with pv_i and call the oracle for the promise problem with the resulting lattice. If $p \mid a_i$ then the shortest vector remains in the lattice and therefore we end up with a lattice whose shortest vector is of length $1/n$ and all other non-parallel vectors are of length more than $\sqrt{n}$. If $p \nmid a_i$ then the shortest vector

disappears and so do all of its multiples up to the pth multiple. Since $p > n^{1.5}$, all the vectors in the resulting lattice are of length more than $\sqrt{n}$. Therefore, we managed to solve the decision problem by one call to the oracle for the promise problem. Let us mention that the idea of multiplying a basis vector by a prime was already used by [12] in a different context.

The third reduction is the core of the proof. Here, we reduce the above promise problem to a problem of distinguishing between two n-dimensional distributions. Namely, one distribution is U_{L^*}, an n-dimensional uniform distribution, and the other is T_{L^*}, an n-dimensional distribution that has the form of fuzzy $n-1$-dimensional hyperplanes (see Figure 3 for a two-dimensional illustration). The reason we have L^* in the subscript should become clear later in the paper. The main intuitive idea in this reduction is the following. Imagine looking at a grid of points with your vision blurred – see the left image in Figure 2 for an illustration. If the grid is dense enough then all the blurs merge together and the picture you see is *uniform*. In other words, by adding noise to a lattice, we can essentially 'erase' its fine structure. This is also the underlying idea in the work of Ajtai and Dwork [2].

The reduction itself is as follows: given an input lattice L, the output distribution is obtained by choosing a 'random' lattice point from the dual lattice L^* and perturbing it by a Gaussian of radius $\sqrt{n}$. We analyze this reduction by applying an important lemma of Banaszczyk [3]. This lemma makes precise the intuitive idea mentioned above. It says that the distribution obtained in the reduction can be closely approximated by a function that depends only on points in L (the primal lattice) that are within distance $\sqrt{n}$ of the origin. Using this lemma, we can now prove the correctness of the reduction. First consider the case that all nonzero vectors in L are of length more than $\sqrt{n}$. Intuitively, since L has no short vectors, L^* is quite dense. Indeed, in this case the only vector in L that is within distance $\sqrt{n}$ of the origin is the origin itself and the lemma tells us that the distribution given by the reduction is very close to U_{L^*}, i.e., the n-dimensional uniform distribution. Now consider the case that L is a lattice with one short vector u of length $1/n$ and all other non-parallel vectors of length more than $\sqrt{n}$. By the definition of the dual lattice, L^* is aligned on $n-1$-dimensional hyperplanes orthogonal to u. The orthogonal distance between two adjacent hyperplanes is $1/\|u\| = n$ and the structure of L^* on each hyperplane is quite dense. Intuitively, we would expect the distribution given by the reduction to be concentrated around these hyperplanes and uniform on them. This is the distribution T_{L^*} mentioned above. As we will see later, this is exactly what Banaszczyk's lemma gives us.

The fourth and final reduction transforms the n-dimensional distributions described above into one-dimensional distributions. The reduction is based on a mapping from the support of our n-dimensional distributions to the segment $[0,1)$. The mapping has the property that U_{L^*} is mapped to U, the uniform distribution on $[0,1)$, and that T_{L^*} is mapped to T_h, a wavy distribution on $[0,1)$, for some h that depends on the hyperplane structure (and hence on the shortest vector in the original lattice). The mapping is performed by partitioning the support of our n-dimensional distributions, which turns out to be a parallelepiped, into W parallelepipeds for some very large W. Each of these parallelepipeds is 'tall and narrow', i.e., it is long in one dimension and very short in all other dimensions. Consider our n-dimensional distributions restricted to such a parallelepiped. The crucial observation is that since it is so narrow, the distribution that we see is essentially a one-dimensional cross section of the n-dimensional distribution: either uniform in the case of U_{L^*} or wavy in the case of T_{L^*}. This leaves us with W one-dimensional distributions – one for each narrow

parallelepiped. We then describe a certain ordering of these narrow parallelepipeds. Consider the distribution on $[0,1)$ obtained by concatenating all the one-dimensional distributions together. In other words, the ith parallelepiped is assigned to the segment $[(i-1)/W, i/W)$. In case we started with U_{L^*}, the resulting distribution is clearly uniform. In case we started with T_{L^*}, we show that our ordering is such that the individual distributions glue together nicely and that the resulting distribution is T_h. This completes the description of the main theorem.

Worst-case to average-case: The problem shown hard by the main theorem is, in a way, a worst-case problem: to solve all instances of $n^{1.5}$-uSVP, we need to distinguish between U and T_h for *all* h. We therefore provide another theorem that shows that even the average-case problem is hard. This theorem will be used in the proof of security of the public key cryptosystem. We prove the theorem by showing that a distinguisher between U and T_h for some non-negligible fraction of all possible values for h in some range leads to a distinguisher between U and T_h for all h (and hence to a solution to $n^{1.5}$-uSVP). The idea of this proof is the following. Assume we can distinguish between U and T_{2h}. Let us show how to distinguish between U and T_h. We sample a value $x \in [0,1)$ from the unknown distribution. We then let y be either $x/2$ or $(1+x)/2$ with equal probability. Notice that if the unknown distribution is uniform, then y is also uniform. If the unknown distribution is T_h then y is distributed according to T_{2h}. Notice that we did not use the (unknown) value of h. Hence, this operation transforms any T_h into T_{2h}. We can extend this idea and transform T_h into $T_{\eta h}$ for any $\eta \geq 1$. Now, given a distinguisher that works for some non-negligible fraction of all possible values for h we construct a distinguisher that works for any h. This is done by applying the above transformation with many random values η. With high probability, in one of these attempts, ηh will be such that the given distinguisher can distinguish between U and $T_{\eta h}$. Hence, with high probability, we can distinguish between U and T_h.

Public key cryptosystem: Let N be some large integer. The private key consists of a single integer h chosen randomly in the range (say) $[\sqrt{N}, 2\sqrt{N})$. The public key consists of $m = O(\log N)$ numbers $a_1, \ldots, a_m$ in $\{0, 1, \ldots, N-1\}$ that are 'close' to integer multiples of N/h (notice that h doesn't necessarily divide N). We also include in the public key an index $i_0 \in [m]$ such that a_{i_0} is close to an *odd* multiple of N/h. We encrypt one bit at a time. An encryption of the bit 0 is the sum of a random subset of $\{a_1, \ldots, a_m\}$ modulo N. An encryption of the bit 1 is done in the same way except we add $\lfloor a_{i_0}/2 \rfloor$ to the result. On receiving an encrypted word w we consider its remainder on division by N/h. If it is small, we decrypt 0 and otherwise we decrypt 1. Correctness is established as follows. Since $a_1, \ldots, a_m$ are all close to integer multiples of N/h, any sum of a subset of them is also close to a multiple of N/h and hence encryptions of 0 are decrypted correctly. Similarly, since $\lfloor a_{i_0}/2 \rfloor$ is far from a multiple of N/h, encryptions of 1 are also far from multiples of N/h and the decryption is 1.

The following is a rough description of how we establish the security of the cryptosystem. Assume that there exists a distinguisher $\mathcal{A}$ that given the public key can distinguish encryptions of 0 from encryptions of 1. In other words, the difference between the acceptance probabilities p_0 on encryptions of 0 and the acceptance probability p_1 on encryptions of 1 is non-negligible. Therefore, if p_u is the acceptance probability on random words w, it must be the case that either $|p_u - p_0|$ or $|p_u - p_1|$ is non-negligible. Assume that the former case holds (the latter case is similar). Then we construct a distinguisher between the distributions U and T_h. Let R be the unknown distribution

on $[0,1)$. We choose m values from R, multiply them by N and round the result. Let $a_1,\ldots,a_m$ be the result. We then estimate $\mathcal{A}$'s acceptance probability when the public key $a_1,\ldots,a_m$ (for simplicity we ignore i_0) is fixed and the word w is chosen randomly as an encryption of 0. This is done by simply calling $\mathcal{A}$ many times, each time with a new w computed according to the encryption algorithm. We also estimate $\mathcal{A}$'s acceptance probability when w is chosen uniformly from $\{0,1,\ldots,N-1\}$ and not according to the encryption algorithm. If there is a non-negligible difference between the two estimates, we decide that R is T_h and otherwise we say that R is U. We claim that this distinguishes between U and T_h. If R is U then $a_1,\ldots,a_m$ are uniform in $\{0,1,\ldots,N-1\}$. One can show that this implies that the distribution of encryptions of 0 is very close to the uniform distribution and therefore $\mathcal{A}$ (as well as any other algorithm) cannot have different acceptance probabilities for the two distributions. Otherwise, R is T_h and the distribution that we obtain on $a_1,\ldots,a_m$ is the same one that is used in the public key algorithm. Therefore, according to our hypothesis, $\mathcal{A}$ should have a non-negligible difference between the two cases.

A family of collision resistant hash functions: We choose $m = O(\log N)$ random numbers $a_1,\ldots,a_m$ uniformly from $\{0,1,\ldots,N-1\}$ and define the hash function $f(b) = \sum_{i=1}^{m} b_i a_i \bmod N$ where $b \in \{0,1\}^m$. A collision finding algorithm in this case means an algorithm $\mathcal{A}$ that given random $a_1,\ldots,a_m$ finds with non-negligible probability a nonzero vector $b \in \{-1,0,1\}^m$ such that $\sum b_i a_i \equiv 0 (\bmod N)$. Using $\mathcal{A}$ we show how to build a distinguisher between U and T_h. By trying many values of the form $(1+1/\mathrm{poly}(m))^i$ we can have an estimate $\tilde{h}$ of h up to some small $1/\mathrm{poly}(m)$ error. We would like to use $\tilde{h}$ to check if the distribution is concentrated around multiples of $1/h$. Sampling values from the unknown distribution R and reducing modulo $1/\tilde{h}$ does not help because the difference between i/h and $i/\tilde{h}$ is much larger than $1/\tilde{h}$ for almost all $0 \le i < h$, since h is exponential in m. The idea is to use the collision finding algorithm to create from T_h a distribution that is also concentrated around the peaks i/h but only for small i, namely $i \le m$.

The distinguisher first samples m values $x_1,\ldots,x_m$ from the unknown distribution R. We then add small perturbations $y_1,\ldots,y_m$ chosen uniformly in $[0,1/\tilde{h})$ to each $x_1,\ldots,x_m$ respectively. We denote the result by $z_1,\ldots,z_m$. We now call $\mathcal{A}$ with $\lfloor N\cdot z_1\rfloor,\ldots,\lfloor N\cdot z_m\rfloor$ and we get a subset S such that $\sum_{i\in S} z_i \bmod 1$ is very close to zero. For simplicity assume that it is exactly zero. If $\sum_{i\in S} x_i \bmod 1 = -\sum_{i\in S} y_i \bmod 1$ is close to an integer multiple of $1/\tilde{h}$ we say that R is T_h; otherwise, we say that R is U.

An important aspect of the analysis is that we condition on any values of $z_1,\ldots,z_m,S$ and analyze the remaining randomness in $x_1,\ldots,x_m$ and $y_1,\ldots,y_m$. Intuitively, this allows us to argue that our distinguisher works no matter what the collision finding algorithm does. If R is the uniform distribution on $[0,1)$ then conditioned on any values of $z_1,\ldots,z_m$ the distribution of $y_1,\ldots,y_m$ is still uniform in $[0,1/\tilde{h})$ and hence, with high probability, $\sum_{i\in S} y_i$ is not close to an integer multiple of $1/\tilde{h}$. If R is T_h then conditioned on any values of $z_1,\ldots,z_m$, each x_i is distributed on the interval $(z_i - 1/\tilde{h}, z_i]$. More precisely, its distribution is the one obtained by restricting T_h to $(z_i - 1/\tilde{h}, z_i]$. This distribution can be closely approximated by the distribution obtained by restricting T_h to $(z_i - 1/h, z_i]$. Because the length of this interval is $1/h$, this distribution includes exactly one peak of T_h. Hence, x_i is distributed around some i/h. Therefore, $\sum_{i\in S} x_i \bmod 1$ is close to a multiple of $1/h$. Moreover, since the y_i's are at most $1/\tilde{h}$, $\sum_{i\in S} y_i$ is at most $m/\tilde{h}$. Since the estimate $\tilde{h}$ satisfies that for $1 \le i \le m$, i/h is very close to $i/\tilde{h}$, $\sum_{i\in S} x_i \bmod 1$ is close to a multiple of $1/\tilde{h}$ and the distinguisher decides that R is T_h, as required.

One last issue that we have to address is that $\mathcal{A}$ might not find collisions on inputs of the form $\lfloor N \cdot z_1 \rfloor, \ldots, \lfloor N \cdot z_m \rfloor$ when R is not the uniform distribution. This is because our assumption was that $\mathcal{A}$ finds collisions on inputs chosen uniformly. But if $\mathcal{A}$ does not find collisions we know that R has to be T_h and hence we can still distinguish between U and T_h.

Outline

In Section 2 we list several definitions and some properties of lattices that will be needed in this paper (for an introduction to lattices see [23]). The distributions that appear in our main theorem are defined in Section 2.1. The main theorem is developed in Section 3. In Section 4 we obtain an average-case version of the main theorem. This version is then used in Section 5 where we describe the public key cryptosystem and its analysis. The hash function and its analysis are given in Section 6. In Section 7 we present a solution to an open problem related to quantum computation. Sections 5, 6 and 7 are independent. The average-case formulation of Section 4 is used only in Section 5.

2 Preliminaries

A lattice in $\mathbb{R}^n$ is defined as the set of all integer combinations of n linearly independent vectors. This set of vectors is known as a basis of the lattice and is not unique. Given a basis $(v_1, \ldots, v_n)$ of a lattice L, the fundamental parallelepiped is defined as

$$\mathcal{P}(v_1, \ldots, v_n) = \left\{ \sum_{i=1}^{n} x_i v_i \;\middle|\; x_i \in [0, 1) \right\}.$$

When the basis is clear from the context we will use the notation $\mathcal{P}(L)$ instead of $\mathcal{P}(v_1, \ldots, v_n)$. Note that a lattice has a different fundamental parallelepiped for each possible basis. We denote by $d(L)$ the volume of the fundamental parallelepiped of L or equivalently, the determinant of the matrix whose columns are the basis vectors of the lattice. The point $x \in \mathbb{R}^n$ reduced modulo the parallelepiped $\mathcal{P}(v_1, \ldots, v_n)$ is the unique point $y \in \mathcal{P}(v_1, \ldots, v_n)$ such that $y - x$ is an integer combination of $v_1, \ldots, v_n$. The dual of a lattice L in $\mathbb{R}^n$, denoted L^*, is the set of all vectors $y \in \mathbb{R}^n$ such that $\langle x, y \rangle \in \mathbb{Z}$ for all vectors $x \in L$. Similarly, given a basis $(v_1, \ldots, v_n)$ of a lattice, we define the dual basis as the set of vectors $(v_1^*, \ldots, v_n^*)$ such that $\left\langle v_i, v_j^* \right\rangle = \delta_{ij}$ for all $i, j \in [n]$ where δ_{ij} denotes the Kronecker delta, i.e., 1 if $i = j$ and 0 otherwise. Note that if $B = (v_1, \ldots, v_n)$ is the $n \times n$ matrix whose columns are the basis vectors then the columns of $(B^T)^{-1}$ are the vectors of the dual basis. From this it follows that $d(L^*) = 1/d(L)$.

We say that a lattice is *unique* if its shortest (nonzero) vector is strictly shorter than all other non-parallel vectors. Moreover, a lattice is $f(n)$-*unique* if its shortest vector is shorter than all other non-parallel vectors by a factor of more than $f(n)$. In the shortest vector problem we are interested in finding the shortest vector in a lattice. In this paper we will be interested in the $f(n)$-unique shortest vector problem ($f(n)$-uSVP) where in addition, we are promised that the lattice is $f(n)$-unique. Let $\lambda(L)$ denote the length of the shortest nonzero vector in the lattice L. We also denote the shortest vector (or one of the shortest vectors) by $\tau(L)$. Most of the lattices that appear is this paper are unique lattices and in these cases $\tau(L)$ is unique up to sign.

One particularly useful type of basis is an LLL-reduced basis. Such a basis can be found in polynomial time [19]. Hence, we will often assume without loss of generality that lattices are given by an LLL-reduced basis. The properties of LLL-reduced bases that we use are summarized in Lemma 3.2.

We define a negligible amount in n as an amount that is asymptotically smaller than n^{-c} for any constant $c > 0$. The parameter n will indicate the input size. Similarly, a non-negligible amount is one which is at least n^{-c} for some $c > 0$. Finally, exponentially small in n means an expression that is at most $2^{-\Omega(n)}$. We say that an algorithm $\mathcal{A}$ with oracle access is a distinguisher between two distributions if its acceptance probability when the oracle outputs samples of the first distribution and its acceptance probability when the oracle outputs samples of the second distribution differ by a non-negligible amount. In addition, an algorithm $\mathcal{A}$ is said to distinguish between the distribution T and the set of distributions $\mathcal{T}$ if for any distribution $T' \in \mathcal{T}$, $\mathcal{A}$ distinguishes between T and T'.

For two continuous random variables X and Y having values in $[0, 1)$ with density functions T_1 and T_2 respectively we define their statistical distance as

$$\Delta(X, Y) = \frac{1}{2} \int_0^1 |T_1(r) - T_2(r)| dr.$$

A similar definition holds for discrete random variables. One important fact that we use is that the statistical distance cannot increase by applying a (possibly randomized) function f, i.e.,

$$\Delta(f(X), f(Y)) \leq \Delta(X, Y), \tag{1}$$

see, e.g., [23]. In particular, this implies that the acceptance probability of any algorithm on inputs from X differs from its acceptance probability on inputs from Y by at most $\Delta(X, Y)$.

The set $\{1, 2, \ldots, n\}$ is denoted by $[n]$. All logarithms are of base 2 unless otherwise specified. We use $\tilde{c}$ to denote an unspecified constant. That is, whenever $\tilde{c}$ appears we can replace it with some universal constant. For example, the expression $\tilde{c} + 7 = \tilde{c}$ is true because we can substitute 1 and 8 for the constants. Other constants will be denoted by c with a letter as the subscript, e.g., c_{m}.

For two real numbers x and $y > 0$ we define $x \bmod y$ as $x - \lfloor x/y \rfloor y$. For $x \in \mathbb{R}$ we define $\lfloor x \rceil$ as the integer nearest to x or, in case two such integers exist, the smaller of the two. We also use the notation $\operatorname{frc}(x) := |x - \lfloor x \rceil|$, i.e., the distance of a real x to the nearest integer. Notice that for all $x, y \in \mathbb{R}$, $0 \leq \operatorname{frc}(x) \leq \frac{1}{2}$, $\operatorname{frc}(x) \leq |x|$ and $\operatorname{frc}(x + y) \leq \operatorname{frc}(x) + \operatorname{frc}(y)$.

Recall that the *normal distribution* with mean 0 and variance σ^2 is the distribution on $\mathbb{R}$ given by the density function $\frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{1}{2}(\frac{x}{\sigma})^2\right)$ where $\exp(y)$ denotes e^y. Also recall that the sum of two independent normal variables with mean 0 and variances σ_1^2 and σ_2^2 is a normal variable with mean 0 and variance $\sigma_1^2 + \sigma_2^2$. We define the *standard Gaussian* distribution as the distribution on $\mathbb{R}^n$ in which each coordinate is an independent normal random variables with mean 0 and standard deviation $1/\sqrt{2\pi}$. In other words, a standard Gaussian distribution is given by the density function $\exp\left(-\pi \|x\|^2\right)$ on $\mathbb{R}^n$.

For clarity, we present some of our reductions in a model that allows operations on real numbers. It is possible to modify them in a straightforward way so that they operate in a model that approximates real numbers up to an error of 2^{-n^c} for arbitrary large constant c in time polynomial in n. Therefore, if we say that two continuous distributions on $[0, 1)$ are indistinguishable (in the

real model) then for any $c > 0$ discretizing the distributions up to error 2^{-n^c} for any c yields two indistinguishable distributions.

2.1 Several Distributions

We define several useful distributions on the segment $[0, 1)$. The distribution U is simply the uniform distribution with the density function $U(r) = 1$. For $\beta \in \mathbb{R}^+$ the distribution Q_β is a normal distribution with mean 0 and variance $\frac{\beta}{2\pi}$ reduced modulo 1 (i.e., a periodization of the normal distribution):

$$Q_\beta(r) := \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{\beta}} \cdot \exp\left(-\frac{\pi}{\beta}(r-k)^2\right).$$

The 2π factor helps to simplify notation later in the paper. Note that the standard deviation of Q_β is proportional to $\sqrt{\beta}$. Clearly, one can efficiently sample from Q_β by sampling a normal variable and reducing the result modulo 1. Another distribution is $T_{h,\beta}$ where $h \in \mathbb{N}$ and $\beta \in \mathbb{R}^+$ (see Figure 1). Its density function is defined as

$$T_{h,\beta}(r) := Q_\beta(rh \bmod 1) = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{\beta}} \cdot \exp\left(-\frac{\pi}{\beta}(rh-k)^2\right).$$

We can efficiently sample a value $z \in [0, 1)$ according to $T_{h,\beta}$ as follows. First choose a value $x \in \{0, 1, \ldots, h-1\}$ uniformly at random and then choose a value y according to Q_β. The result is $(x + y)/h$.

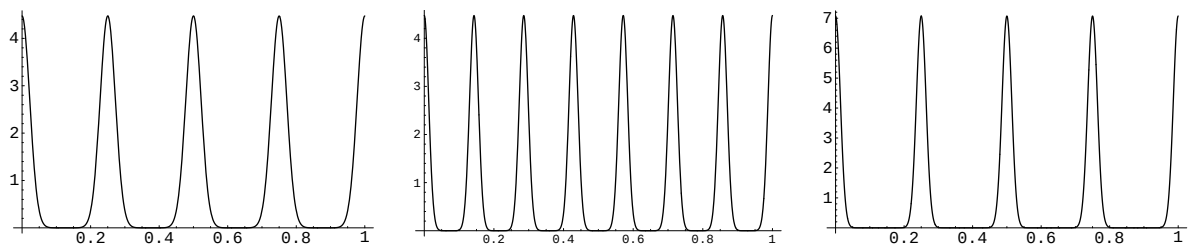

Figure 1: $T_{4,0.05}$, $T_{7,0.05}$ and $T_{4,0.02}$

We also define the following set of distributions:

$$\mathcal{T}_{n,g} := \left\{ T_{h,\beta} \;\middle|\; h \in \mathbb{N},\ h \le 2^{4n^2},\ \beta \in \left[\frac{n}{g^2}, \frac{4n}{g^2}\right) \right\}.$$

The reason for this choice of parameter is our main theorem, which we explain next.

3 Main Theorem

In this section we present a reduction from $g(n)$-uSVP to the problem of distinguishing between two types of distributions on $[0, 1)$. The proof is obtained by combining the four reductions shown later in this section.

Theorem 3.1 *Let $g(n)$ be any function such that $4\sqrt{n} \leq g(n) \leq \mathrm{poly}(n)$. If there exists a distinguisher between U and $\mathcal{T}_{n,g(n)}$ then there exists a solution to $g(n)$-*uSVP*.*

Proof: Let $p(n)$ be a prime larger than $g(n)$ and at most (say) $2g(n)$. We can now apply Lemmas 3.5, 3.7, 3.15 and 3.17 in order to obtain the theorem. ■

The following technical lemma provides some rough bounds on numbers arising from LLL-reduced bases: the coefficients of the shortest vector are not too big, the length of the shortest vector is known to be in a certain range, and the vectors in the dual basis are not too long. These properties will be used in the proof of Theorem 3.1.

Lemma 3.2 *Let $(v_1, \ldots, v_n)$ be an LLL-reduced basis of a lattice L and let $\sum_{i=1}^{n} a_i v_i$ be its shortest vector. Then $|a_i| \leq 2^{2n}$ for all $i \in [n]$ and $\lambda(L) \leq \|v_1\| \leq 2^n \lambda(L)$. Moreover, if $(v_1^*, \ldots, v_n^*)$ is the dual basis, then $\|v_i^*\| \leq \frac{\sqrt{n}}{\lambda(L)} 2^{2n}$ for all $i \in [n]$.*

Proof: Let $(v_1^\dagger, \ldots, v_n^\dagger)$ denote the Gram-Schmidt orthogonalization of $(v_1, \ldots, v_n)$, i.e., $v_i^\dagger$ is the component of v_i orthogonal to the subspace spanned by $v_1, \ldots, v_{i-1}$. Clearly, $\left\langle v_i^\dagger, v_j \right\rangle = 0$ for $i > j$. Recall that in an LLL-reduced basis $\|v_i^\dagger\| \leq \sqrt{2}\|v_{i+1}^\dagger\|$ and for $i < j$,

$$\left| \left\langle v_i^\dagger, v_j \right\rangle \right| \leq \frac{1}{2} \|v_i^\dagger\|^2.$$

In addition, recall that $\min_i \|v_i^\dagger\|$ is a lower bound on $\lambda(L)$. Then for any $i \in [n]$, $\|v_1^\dagger\| \leq 2^{(i-1)/2} \|v_i^\dagger\|$ and therefore $\|v_1^\dagger\| \leq 2^{(n-1)/2} \lambda(L)$. Using $v_1^\dagger = v_1$, we see that

$$\lambda(L) \leq \|v_1\| \leq 2^n \lambda(L).$$

Consider now the representation of $(v_1, \ldots, v_n)$ in the orthonormal basis

$$\left(\frac{v_1^\dagger}{\|v_1^\dagger\|}, \ldots, \frac{v_n^\dagger}{\|v_n^\dagger\|} \right).$$

It is given by the columns of the matrix $B = (b_{i,j})_{1 \leq i,j \leq n}$ where $b_{i,j} = \left\langle v_j, v_i^\dagger \right\rangle / \|v_i^\dagger\|$. Notice that this matrix is upper triangular and that its diagonal is $b_{i,i} = \|v_i^\dagger\|$. Also note that by the properties of an LLL-reduced basis, $|b_{i,j}| \leq \frac{1}{2}\|v_i^\dagger\|$ for $i < j$. The shortest vector is

$$\sum_{i=1}^{n} a_i v_i = \sum_{i=1}^{n} \left(\sum_{j=i}^{n} a_j b_{i,j} \right) \frac{v_i^\dagger}{\|v_i^\dagger\|}.$$

Since its length is at most $2^n \|v_i^\dagger\|$ the absolute value of each of its coordinates is at most $2^n \|v_i^\dagger\|$. Hence, $\left| \sum_{j=i}^{n} a_j b_{i,j} \right| \leq 2^n \|v_i^\dagger\|$ for every $i \in [n]$. By taking $i = n$ we get that $|a_n b_{n,n}| \leq 2^n \|v_n^\dagger\|$ and hence $|a_n|$ is at most 2^n. We continue inductively and show that $|a_k| \leq 2^{2n-k}$. Assume that the claim holds for $a_{k+1}, \ldots, a_n$. Then,

$$\left| \sum_{j=k+1}^{n} a_j b_{k,j} \right| \leq \frac{1}{2} \left| \sum_{j=k+1}^{n} a_j \right| \|v_k^\dagger\| \leq \frac{1}{2} \left(\sum_{j=k+1}^{n} 2^{2n-j} \right) \|v_k^\dagger\| \leq \frac{1}{2} \cdot 2^{2n-k} \|v_k^\dagger\|.$$

By the triangle inequality,

$$|a_k b_{k,k}| \leq \left| \sum_{j=k+1}^{n} a_j b_{k,j} \right| + \left| \sum_{j=k}^{n} a_j b_{k,j} \right| \leq \left(\frac{1}{2} \cdot 2^{2n-k} + 2^n \right) \|v_k^\dagger\| \leq 2^{2n-k} \|v_k^\dagger\|$$

and the proof of the first part is completed.

The basis of the dual lattice is given by the columns of $(B^T)^{-1}$. Since $\min_i |b_{i,i}| \geq \frac{\lambda(L)}{2^n}$ and $|b_{i,j}| \leq \frac{1}{2}|b_{i,i}|$, the following claim implies that the entries of $(B^T)^{-1}$ are at most $\frac{1}{\lambda(L)} 2^{2n}$ in absolute value. Therefore, the length of each column vector is at most $\frac{\sqrt{n}}{\lambda(L)} 2^{2n}$.

Claim 3.3 *Let $B = (b_{i,j})_{1 \leq i,j \leq n}$ be an $n \times n$ upper triangular matrix such that for all $i < j \leq n$, $|b_{i,j}| \leq |b_{i,i}|$. Then, the entries of $(B^T)^{-1}$ have an absolute value of at most $\frac{1}{\min_i |b_{i,i}|} 2^n$.*

Proof: First, let D denote the diagonal matrix with values $b_{i,i}$ on the diagonal. Then B can be written as DM where M is an upper triangular matrix with ones on the diagonal and all other entries have an absolute value of at most 1. Then, $(B^T)^{-1} = (M^T D^T)^{-1} = D^{-1}(M^T)^{-1}$. Therefore, it is enough to show that the entries of $L := (M^T)^{-1}$ are at most 2^n in absolute value. The diagonal of L is all ones and it is lower triangular. We can define it recursively by

$$l_i = e_i - \sum_{k<i} l_k m_{k,i}$$

where l_i denotes the ith row of L and e_i is the vector that has 1 in position i and 0 everywhere else. In other words, the entry $l_{i,j}$ for $i > j$ can be defined by $-\sum_{j \leq k < i} l_{k,j} m_{k,i}$. Therefore,

$$|l_{i,j}| = \left| \sum_{j \leq k < i} l_{k,j} m_{k,i} \right| \leq \sum_{j \leq k < i} |l_{k,j} m_{k,i}| \leq \sum_{j \leq k < i} |l_{k,j}|$$

from which we get the bound $|l_{i,j}| \leq 2^{i-j}$ for $i \geq j$. ■

■

3.1 Reduction to a Decision Problem

We reduce uSVP to the following decision problem.

Definition 3.4 (Divisibility SVP with parameter p (dSVP_p)) *For an integer $p \geq 2$, the input to the dSVP_p is an arbitrary basis $(v_1, \ldots, v_n)$ of a unique lattice L and a number α such that $\lambda(L) < \alpha \leq 2\lambda(L)$. Let $\tau(L) = \sum_{i=1}^{n} a_i v_i$ be the coefficients of the shortest vector. Then the goal is to output YES if p divides a_1 and NO otherwise.*

Lemma 3.5 *Let $p = p(n) > 2$ be a prime number that is at most polynomial in n.[2] There exists a reduction from finding the shortest vector in a unique lattice L to dSVP_p.[3] Moreover, if L is an $f(n)$-unique lattice then all the calls to the dSVP oracle are also with an $f(n)$-unique lattice.*

[2]The result holds for the case $p = 2$ as well with some technical differences.

[3]One can guarantee the uniqueness of the shortest vector in any lattice by adding tiny perturbations to the basis vectors. Therefore, the assumption that L is unique can be avoided.

Proof: It is convenient to have a bound on the coefficients of the shortest vector. So we assume, without loss of generality, that we are given an LLL-reduced basis $(v_1, \ldots, v_n)$ of L. Hence, by Lemma 3.2, we get that the coefficients of the shortest vector satisfy $|a_i| \leq 2^{2n}$ and $\frac{\|v_1\|}{2^n} \leq \lambda(L) \leq \|v_1\|$. These are the only properties that we need from the basis and in fact, other bases used throughout this proof will not necessarily be LLL-reduced. In the following we describe a procedure $\mathcal{B}(\alpha)$ that finds the shortest vector in L given access to a dSVP oracle and an α that satisfies $\lambda(L) < \alpha \leq 2\lambda(L)$. We apply the procedure n times with $\alpha = 2^{j-n} \cdot \|v_1\|$ for $j = 1, 2, \ldots, n+1$. Notice that when we call $\mathcal{B}$ with the wrong value of α it can error by either outputting a non-lattice vector or a lattice vector that is longer than the shortest vector. We can easily ignore these errors by checking that the returned vector is a lattice vector and then taking the shortest one. Therefore, it is sufficient to show that when α satisfies $\lambda(L) < \alpha \leq 2\lambda(L)$, $\mathcal{B}(\alpha)$ returns the shortest vector. Clearly, one can modify the dSVP oracle so that it finds whether $p \mid a_i$ for any $i \in [n]$ (and not just $i = 1$) by simply changing the order of the vectors in the basis given to it.

The procedure $\mathcal{B}$ is based on changes to the basis $(v_1, \ldots, v_n)$. Throughout the procedure we maintain the invariant that the lattice spanned by the current basis is a sublattice of the original lattice and that the shortest vector is unchanged. Notice that this implies that if the original lattice is an $f(n)$-unique lattice then all intermediate lattices are also $f(n)$-unique and hence all the calls to the dSVP oracle are with an $f(n)$-unique lattice, as required. In addition, since the shortest vector is unchanged, the estimate α can be used whenever we call the dSVP oracle with an intermediate lattice. The changes to the basis are meant to decrease the coefficients of the shortest vector. We let $a_1, \ldots, a_n$ denote the coefficients of the shortest vector according to the current basis. We will show that when the procedure ends all the coefficients of the shortest vector are zero except a_i for some $i \in [n]$. This implies that the shortest vector is v_i. In the following we describe a routine $\mathcal{C}$ that will later be used in $\mathcal{B}$.

The routine $\mathcal{C}(i, j)$ where $i, j \in [n]$ applies a sequence of changes to the basis. Only the vectors v_i and v_j in the basis are modified. When the routine finishes it returns the new basis and a bit. If the bit is zero then we are guaranteed that the coefficient a_i of the shortest vector in the new basis is zero and that a_j is unchanged. Otherwise, the bit is one and we are guaranteed that $|a_j| \leq \frac{1}{2}|a_i|$ and that a_i is nonzero. Moreover, the value of $|a_i|$ does not increase by $\mathcal{C}(i, j)$.

The routine is composed of the following two steps. In the first step we replace v_i with $p \cdot v_i$ as long as the dSVP oracle says that $p \mid a_i$ and not more than $2n$ times. By multiplying v_i by p when $p \mid a_i$, we obtain a sublattice that still contains the same shortest vector. The coefficient a_i decreases by a factor of p. Since we began with $|a_i| < 2^{2n}$, if this happens $2n$ times then $a_i = 0$ and therefore in this case we return the current basis and output a zero bit. Otherwise, we are guaranteed that in the current lattice $p \nmid a_i$.

In the second step we consider p different bases where v_i is replaced with one of $v_i - \frac{p-1}{2} v_j, \ldots, v_i - v_j, v_i, v_i + v_j, \ldots, v_i + \frac{p-1}{2} v_j$. Notice that all p bases span the same lattice. Also note that the coefficient a_j changes to $a_j + \frac{p-1}{2} a_i, \ldots, a_j + a_i, a_j, a_j - a_i, \ldots, a_j - \frac{p-1}{2} a_i$ respectively while all other coefficients remain the same. Since $p \nmid a_i$, one of the bases must satisfy that $p \mid a_j$ and we can find it by calling the dSVP_p oracle. We choose that basis and then multiply v_j by p. We repeat the above steps (of choosing one of the p bases and multiplying by p) $2n$ times and then output the

resulting lattice with the bit one. With each step, the new $|a_j|$ becomes at most

$$\frac{\frac{p-1}{2}|a_i| + |a_j|}{p} = \left(\frac{1}{2} - \frac{1}{2p}\right)|a_i| + \frac{|a_j|}{p}.$$

Hence, after $2n$ applications, the new $|a_j|$ is at most

$$\left(\frac{1}{2} - \frac{1}{2p}\right)\left(1 + \frac{1}{p} + \ldots + \frac{1}{p^{2n-1}}\right)|a_i| + \frac{|a_j|}{p^{2n}} < \frac{1}{2}|a_i| + \frac{|a_j|}{p^{2n}} < \frac{1}{2}|a_i| + \frac{1}{4}$$

and since a_j is integer this implies $|a_j| \leq \frac{1}{2}|a_i|$. This completes the description of $\mathcal{C}$.

We now check that $\mathcal{C}$ runs in polynomial time. Indeed, $\mathcal{C}$ can be seen as consisting of basic operations where in each basic operation we either multiply one of the basis vectors by p, or add a multiple of one basis vector to another. The number of basic operations is clearly at most polynomial in n. Consider the maximum over all i of the number of bits needed to represent v_i. Each basic operation increases this value by at most $O(n \log p)$ (since each v_i is an n-dimensional vector) and hence it is always at most polynomial in n. This implies that each basic operation can be performed in polynomial time.

The procedure $\mathcal{B}$ works by maintaining a set Z of possibly non-zero coefficients that is initially set to $[n]$. As long as $|Z| \geq 2$ we perform the following operations. Assume without loss of generality that $1, 2 \in Z$. We alternatively call $\mathcal{C}(1,2)$ and $\mathcal{C}(2,1)$ until the bit returned in one of the calls is zero. This indicates that one of the coefficients is zero (either a_1 or a_2 depending on which call returns the zero bit) and we remove it from the set Z. In order to show that the procedure runs in polynomial time, it is enough to show that an element is removed from Z after at most a polynomial number of steps. Notice that after each pair of calls to $\mathcal{C}$ that returned the bit one $|a_1|$ decreases by a factor of at least 4. Therefore, after at most $2n$ calls to $\mathcal{C}$, a_1 becomes zero and $\mathcal{C}(1,2)$ must return the bit zero. ■

3.2 Reduction to a Promise Problem

We continue our sequence of reductions by reducing dSVP to the following promise problem.

Definition 3.6 (Promise SVP with parameter $g(n)$ ($\mathsf{pSVP}_{g(n)}$)) *For $2 \leq g(n) \leq \mathrm{poly}(n)$, the input to the $\mathsf{pSVP}_{g(n)}$ is a lattice L given by an LLL-reduced basis. In YES instances, $\lambda(L) \in \left[\frac{\sqrt{n}}{g(n)}, \frac{2\sqrt{n}}{g(n)}\right)$ and all vectors not parallel to $\tau(L)$ are of length more than $\sqrt{n}$. In NO instances, $\lambda(L) > \sqrt{n}$. The goal is to distinguish between these two cases with the promise that L satisfies one of them.*

Lemma 3.7 *Let $g(n) < p(n)$ be such that $p(n)$ is a prime and both are at most polynomial in n. Then, there is a reduction from $\mathsf{dSVP}_{p(n)}$ on $g(n)$-unique lattices to $\mathsf{pSVP}_{g(n)}$.*

Proof: The input to the $\mathsf{dSVP}_{p(n)}$ problem is a basis $(v_1, \ldots, v_n)$ of a $g(n)$-unique lattice L and a number α such that $\lambda(L) < \alpha \leq 2\lambda(L)$. Let L' be the lattice L scaled by a factor $\frac{2\sqrt{n}}{\alpha \cdot g(n)}$, i.e., the lattice spanned by the basis

$$(v'_1, \ldots, v'_n) := \frac{2\sqrt{n}}{\alpha \cdot g(n)}(v_1, \ldots, v_n).$$

Notice that $\tau(L') = \sum_{i=1}^{n} a_i v'_i$ is of length in $\left[\frac{\sqrt{n}}{g(n)}, \frac{2\sqrt{n}}{g(n)}\right)$ and any vector not parallel to $\tau(L')$ is of length more than $g(n) \cdot \frac{\sqrt{n}}{g(n)} = \sqrt{n}$. Now, let M be the lattice spanned by the basis $(p(n)v'_1, v'_2, \ldots, v'_n)$. We output the answer obtained by applying the $\mathsf{pSVP}_{p(n)}$ oracle to an LLL-reduced basis of M.

If $p(n) \mid a_1$ then $\tau(M) = \tau(L')$ and therefore its length is in $\left[\frac{\sqrt{n}}{g(n)}, \frac{2\sqrt{n}}{g(n)}\right)$. Also, since $M \subseteq L'$, any vector in M not parallel to $\tau(M)$ is of length more than $\sqrt{n}$. If $p(n) \nmid a_1$ then the shortest multiple of $\tau(L')$ that is contained in M is $p(n) \cdot \tau(L')$ whose length is at least $p(n) \cdot \frac{\sqrt{n}}{g(n)} > \sqrt{n}$. Hence, in this case all non-zero vectors are of length more than $\sqrt{n}$. ■

3.3 Gaussian Distributions on Lattices

In this section, we describe how to reduce pSVP to a problem of distinguishing between two distributions. The reduction itself appears as Lemma 3.15. Before we get there, we will analyze what happens when one adds noise to a lattice. The main intuitive idea used in this section is that by adding noise to L^*, we can essentially 'erase' its fine structure. This is also the underlying idea in the work of Ajtai and Dwork [2]. We will demonstrate this idea for two cases: the first is the case where L^* is a very dense lattice. Here, adding noise erases the entire structure and effectively transforms the lattice into a uniform distribution. Intuitively, the lattice point are so close together that adding this noise transforms them into one uniform 'blur'. This intuition will be made precise in Lemma 3.11. In the second case, L^* is contained in well-separated $n-1$-dimensional hyperplanes. Inside each of these hyperplanes, L^* is very dense. In this case, adding noise effectively erases the structure inside the hyperplanes. The distribution that we obtain has the form of fuzzy hyperplanes (see Figure 3). This intuition will be made precise in Lemma 3.14.

Before getting to the technical part of this section, let us explain how we deal with distributions in n-dimensional space. Our distributions all have the property that they are periodic on the lattice L^*. For example, the most important of these distributions is the one obtained by adding Gaussian noise to the lattice L^*. How should we define it formally? We can say: "randomly choose a uniform lattice point and add some Gaussian noise to it". The problem with this definition is that there is no way to choose a random lattice point. One possible solution, which has been used in the past, is to choose the lattice point uniformly at random from all lattice points inside some very large cube. This is illustrated on the left side of Figure 2. In applying this solution, however, one has to deal with some annoying technical issues. For example, one has to show that the probability that the lattice point falls close to the edge of the cube is very small given that the cube is large enough.

In this paper, we will instead follow the approach taken in [20]. This approach yields a much cleaner analysis and avoids all the technical difficulties of the 'large cube' approach. The idea is simple: we only need to consider distributions on the basic parallelepiped of the lattice. For example, in order to represent the distribution mentioned above, we sample from a Gaussian centered around the origin and reduce the result modulo the basic parallelepiped of the lattice. See the right side of Figure 2. There is no need to 'choose a random lattice point' since this is already captured by the fact that we reduce modulo the basic parallelepiped. More precisely, for a given lattice L, we consider the distribution D_{L^*} on $\mathcal{P}(L^*)$ given by the density function

$$D_{L^*}(x) = \sum_{y \in L^*} \exp\left(-\pi \|y + x\|^2\right),$$

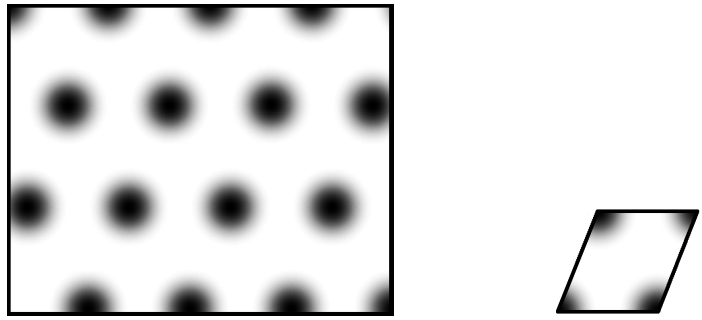

Figure 2: A Gaussian distribution around lattice points as a distribution on a large hypercube (left) and on the basic parallelepiped (right)

or, if for a countable set A we define

$$\rho(A) = \sum_{x \in A} \exp\left(-\pi \|x\|^2\right),$$

then the above becomes

$$D_{L^*}(x) = \rho(L^* + x).$$

A simple calculation shows that D_{L^*} is indeed a density function:

$$\begin{aligned} \int_{\mathcal{P}(L^*)} D_{L^*}(x)dx &= \int_{\mathcal{P}(L^*)} \sum_{y \in L^*} \exp\left(-\pi \|y+x\|^2\right) dx \\ &= \sum_{y \in L^*} \int_{\mathcal{P}(L^*)} \exp\left(-\pi \|y+x\|^2\right) dx \\ &= \int_{\mathbb{R}^n} \exp\left(-\pi \|x\|^2\right) dx = 1. \end{aligned}$$

It is important to realize that the parallelepiped is used only as a convenient technical tool. The reader might benefit from thinking about D_{L^*} (and other distributions in this section) as a function from $\mathbb{R}^n$ to $\mathbb{R}^+$. As such, D_{L^*} becomes a periodic function on L^*, i.e., $D_{L^*}(x) = D_{L^*}(x+y)$ for any $y \in L^*$ and any $x \in \mathbb{R}^n$. For example, if L^* is a very sparse lattice, then D_{L^*} looks like a Gaussian centered around each lattice point, as in Figure 2. Another point worth mentioning is the following. Consider two basic parallelepipeds of L^*, say $\mathcal{P}_1$ and $\mathcal{P}_2$. Then by restricting D_{L^*} to each of them, we obtain two seemingly different distributions: one on $\mathcal{P}_1$ and the other on $\mathcal{P}_2$. However, in many respects, these two distributions are *equivalent*. For example, one can show that sampling from the first distribution and reducing the result modulo $\mathcal{P}_2$ yields the second distribution. This, again, demonstrates that the parallelepiped serves only as a technical tool and is not an inherent part of the distribution.

Our main technical tool is the following lemma by Banaszczyk. It says that for any lattice L the contribution to the Gaussian weight $\rho(L)$ from points whose norm is more than $\sqrt{n}$ is negligible. We use B_n to denote the Euclidean unit ball.

Lemma 3.8 ([3], Lemma 1.5(i) with $c = 1$) *For any lattice L, $\rho(L \setminus \sqrt{n}B_n) < 2^{-\Omega(n)}\rho(L)$.*

The proof of this lemma is not straightforward; a somewhat easier proof can be found in Štefankovič's thesis [28]. Let us mention the following continuous variant of this lemma. Consider the Gaussian measure on $\mathbb{R}^n$ given by $\exp\left(-\pi\|x\|^2\right)$. Then the measure of $\mathbb{R}^n \setminus \sqrt{n}B_n$ is exponentially small (or, equivalently, the measure of $\sqrt{n}B_n$ is exponentially close to 1). The proof of this statement is actually easy and follows from the fact that in high dimensions, the Gaussian measure is highly concentrated around points of norm $\sqrt{n/(2\pi)}$ (so essentially nothing of the measure reaches out beyond norm $\sqrt{n}$). What Lemma 3.8 says is that the same property holds for Gaussian measures on discrete subgroups of $\mathbb{R}^n$ (i.e., lattices).

A simple corollary of this lemma is

Corollary 3.9 *For any lattice L, $\rho(L \setminus \sqrt{n}B_n) < 2^{-\Omega(n)}\rho(L \cap \sqrt{n}B_n)$.*

Proof: By Lemma 3.8,

$$\rho(L \setminus \sqrt{n}B_n) < 2^{-\Omega(n)}\rho(L) = 2^{-\Omega(n)}\left(\rho(L \setminus \sqrt{n}B_n) + \rho(L \cap \sqrt{n}B_n)\right).$$

Therefore,

$$\rho(L \setminus \sqrt{n}B_n) < \frac{2^{-\Omega(n)}}{1-2^{-\Omega(n)}}\rho(L \cap \sqrt{n}B_n) = 2^{-\Omega(n)}\rho(L \cap \sqrt{n}B_n).$$

■

We also need the following lemma, which is special case of Lemma 1.1(i) in [3] (specifically, choose $a = \pi$, $b = 1$, $y = 0$ in Lemma 1.1(i) in [3] and take the real part of both sides of the equation).

Lemma 3.10 *For any lattice L and any vector $z \in \mathbb{R}^n$, $\rho(L^*+z) = d(L)\cdot\sum_{x\in L}\cos(2\pi\langle x,z\rangle)\rho(\{x\})$.*

This lemma is in fact an easy corollary of the Poisson summation formula, a basic formula in Fourier analysis. Essentially, this formula says that for any function f on $\mathbb{R}^n$ and for any lattice L, the sum of f over L^* is equal to $d(L)$ times the sum of the Fourier transform of f over L. Hence, if we take f to be $\rho(\{x\})$, which is its own Fourier transform, we obtain Lemma 3.10 for the case $z = 0$, namely $\rho(L^*) = d(L)\rho(L)$. To get some intuition on this equality, try to consider the case where $L = \{kc \mid k \in \mathbb{Z}\}$ for some $c > 0$ is a one-dimensional lattice. Lemma 3.10 for arbitrary z follows similarly by taking f to be $\rho(\{x+z\})$. See Section 2.3 in [6] for a more complete treatment of the Poisson summation formula.

We now get to the first lemma of this section. It shows that when L^* is dense enough, the distribution D_{L^*} is essentially uniform. Intuitively, this happens because adding the Gaussian noise makes the fine structure of the lattice L^* disappear. It turns out that a sufficient condition for this to happen is that the length of the shortest vector in L is more than $\sqrt{n}$. Interestingly, this characterization is quite tight: there are cases where the length of the shortest vector in L is $c\sqrt{n}$ for some constant c and D_{L^*} is far from uniform (see the references in [3]).

Lemma 3.11 *Let L be a lattice in which all non-zero vectors are of length more than $\sqrt{n}$ and let $U_{L^*}(x) = \frac{1}{d(L^*)} = d(L)$ be the uniform density function on $\mathcal{P}(L^*)$. Then, $\Delta(D_{L^*}, U_{L^*}) < 2^{-\Omega(n)}$.*

Proof: We first show that at any point $y \in \mathcal{P}(L^*)$, $D_{L^*}(y)$ and $U_{L^*}(y)$ are very close. By Lemma 3.10,

$$\begin{aligned} D_{L^*}(y) = \rho(L^* + y) &= d(L) \cdot \sum_{x \in L} \cos(2\pi\langle x, y\rangle)\rho(\{x\}) \\ &= d(L) \cdot \left(1 + \sum_{x \in L \setminus \{0\}} \cos(2\pi\langle x, y\rangle)\rho(\{x\})\right) \\ &= d(L) \cdot \left(1 + \sum_{x \in L \setminus \sqrt{n}B_n} \cos(2\pi\langle x, y\rangle)\rho(\{x\})\right) \end{aligned}$$

where we used $L \cap \sqrt{n}B_n = \{0\}$. Now,

$$\begin{aligned} |D_{L^*}(y) - U_{L^*}(y)| &= d(L) \cdot \left| \sum_{x \in L \setminus \sqrt{n}B_n} \cos(2\pi\langle x, y\rangle)\rho(\{x\}) \right| \\ &\leq d(L) \cdot \sum_{x \in L \setminus \sqrt{n}B_n} |\cos(2\pi\langle x, y\rangle)| \cdot \rho(\{x\}) \\ &\leq d(L) \cdot \sum_{x \in L \setminus \sqrt{n}B_n} \rho(\{x\}) \\ &= d(L) \cdot \rho(L \setminus \sqrt{n}B_n) \end{aligned}$$

Now, using Corollary 3.9,

$$\begin{aligned} d(L) \cdot \rho(L \setminus \sqrt{n}B_n) &< d(L) \cdot 2^{-\Omega(n)} \cdot \rho(L \cap \sqrt{n}B_n) \\ &= d(L) \cdot 2^{-\Omega(n)} \end{aligned}$$

where in the last inequality we used that $L \cap \sqrt{n}B_n = \{0\}$. We conclude the proof by integrating over $\mathcal{P}(L^*)$, whose volume is $d(L^*) = 1/d(L)$,

$$\Delta(D_{L^*}, U_{L^*}) < 2^{-\Omega(n)}.$$

■

We now turn to the second lemma of this section. Here, we consider the case where L has one short vector $\tau(L)$ and all other non-parallel vectors are of length more than $\sqrt{n}$. By the definition of the dual lattice, this implies that L^* is aligned on $n-1$-dimensional hyperplanes orthogonal to $\tau(L)$; the distance between two adjacent hyperplanes is $1/\lambda(L)$. Intuitively, the structure of the lattice on each of the hyperplanes is quite dense. This holds since all vectors not parallel to $\tau(L)$ in L are of length more than $\sqrt{n}$. After adding a Gaussian noise, the fine structure inside the hyperplanes disappears. We are left with a distribution that is essentially uniform on hyperplanes orthogonal to $\tau(L)$. In the direction of $\tau(L)$, the distribution is wavy. See Figure 3 for an illustration.

More formally, our current goal is to show that under the above conditions, D_{L^*} is very close to the distribution T_{L^*} on $\mathcal{P}(L^*)$ whose density function is given by

$$T_{L^*}(x) = \frac{d(L)}{\lambda(L)} \sum_{k \in \mathbb{Z}} \exp\left(-\pi \left(\frac{k + \langle \tau(L), x\rangle}{\lambda(L)}\right)^2\right).$$

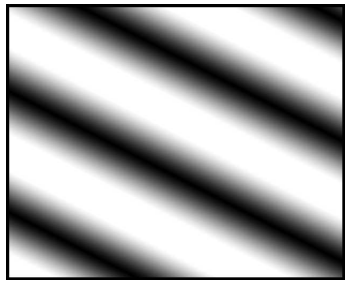

Figure 3: The distribution T_{L^*}

Notice that T_{L^*} depends only on $\langle \tau(L), x\rangle$. Actually, it depends only on $\langle \tau(L), x\rangle$ mod 1 because we sum over all $k \in \mathbb{Z}$. Its maximum is attained when $\langle \tau(L), x\rangle$ is an integer and its minimum is attained when $\langle \tau(L), x\rangle$ mod 1 is half. Hence, it corresponds to the hyperplane structure described above. Let us say again that it is helpful to think of T_{L^*} as a periodic function from $\mathbb{R}^n$ to $\mathbb{R}^+$, as shown in Figure 3.

We now describe an equivalent expression for $T_{L^*}(x)$. Consider the one-dimensional lattice M spanned by the number $\lambda(L)$, i.e., $M = \{k\lambda(L) \mid k \in \mathbb{Z}\}$. Clearly, the lattice M^* is spanned by the number $\frac{1}{\lambda(L)}$. According to Lemma 3.10, for any $a \in \mathbb{R}$,

$$\rho(M^* + a) = d(M) \sum_{b \in M} \cos(2\pi ab)\rho(\{b\}) = \lambda(L) \sum_{k \in \mathbb{Z}} \cos(2\pi ka\lambda(L))\rho(\{k\tau(L)\}).$$

Therefore, taking $a = \langle \tau(L), x\rangle / \lambda(L)$,

$$T_{L^*}(x) = \frac{d(L)}{\lambda(L)} \cdot \rho\left(M^* + \frac{\langle \tau(L), x\rangle}{\lambda(L)}\right) = d(L) \sum_{k \in \mathbb{Z}} \cos(2\pi k\langle \tau(L), x\rangle) \cdot \rho(\{k\tau(L)\}). \tag{2}$$

The following technical claim shows that T_{L^*} is indeed a density function.

Claim 3.12

$$\int_{\mathcal{P}(L^*)} T_{L^*}(x)dx = 1.$$

Proof: By the above,

$$\begin{aligned}
\int_{\mathcal{P}(L^*)} T_{L^*}(x)dx &= d(L) \int_{\mathcal{P}(L^*)} \sum_{k \in \mathbb{Z}} \cos(2\pi k\langle \tau(L), x\rangle) \cdot \rho(\{k\tau(L)\})dx \\
&= d(L) \sum_{k \in \mathbb{Z}} \int_{\mathcal{P}(L^*)} \cos(2\pi k\langle \tau(L), x\rangle) \cdot \rho(\{k\tau(L)\})dx \\
&= d(L) \cdot \left(d(L^*) + \sum_{k \in \mathbb{Z}\setminus\{0\}} \int_{\mathcal{P}(L^*)} \cos(2\pi k\langle \tau(L), x\rangle) \cdot \rho(\{k\tau(L)\})dx \right) \\
&= 1 + d(L) \cdot \sum_{k \in \mathbb{Z}\setminus\{0\}} \rho(\{k\tau(L)\}) \cdot \int_{\mathcal{P}(L^*)} \cos(2\pi k\langle \tau(L), x\rangle)dx.
\end{aligned}$$

Hence, it is enough to show that for any integer $k \neq 0$,

$$\int_{\mathcal{P}(L^*)} \cos(2\pi k\langle \tau(L), x\rangle)dx = 0.$$

Let $v_1^*, \ldots, v_n^*$ be the basis of L^* that forms $\mathcal{P}(L^*)$. Then since $\tau(L)$ is non-zero, there must exist an i_0 such that $\langle \tau(L), v_{i_0}^* \rangle$ is non-zero. Moreover, since $\tau(L) \in L$, $\langle \tau(L), v_{i_0}^* \rangle$ must be integer. Let $l = k\langle \tau(L), v_{i_0}^* \rangle$. Define

$$\mathcal{P}_1 = \left\{ \sum_{i=1}^{n} \alpha_i v_i^* \;\middle|\; \alpha_i \in [0,1) \text{ and } \lfloor 2l\alpha_{i_0} \rfloor \text{ is even} \right\},$$

$$\mathcal{P}_2 = \left\{ \sum_{i=1}^{n} \alpha_i v_i^* \;\middle|\; \alpha_i \in [0,1) \text{ and } \lfloor 2l\alpha_{i_0} \rfloor \text{ is odd} \right\}.$$

Clearly, the two sets are disjoint, $\mathcal{P}_1 \cup \mathcal{P}_2 = \mathcal{P}(L^*)$, and $\mathcal{P}_2 = \mathcal{P}_1 + v_{i_0}^*/(2l)$. Hence, the above integral can be written as

$$\begin{aligned}
&\int_{\mathcal{P}_1} \cos(2\pi k\langle \tau(L), x\rangle)dx + \int_{\mathcal{P}_2} \cos(2\pi k\langle \tau(L), x\rangle)dx \\
&= \int_{\mathcal{P}_1} \cos(2\pi k\langle \tau(L), x\rangle)dx + \int_{\mathcal{P}_1} \cos(2\pi k\langle \tau(L), x + v_{i_0}^*/(2l)\rangle)dx \\
&= \int_{\mathcal{P}_1} \cos(2\pi k\langle \tau(L), x\rangle)dx - \int_{\mathcal{P}_1} \cos(2\pi k\langle \tau(L), x\rangle)dx \\
&= 0
\end{aligned}$$

■

Remark: Notice that in the above proof we did not use any property of $\tau(L)$ except that it is a non-zero vector in L. In fact, one can define a distribution like T_{L^*} for any vector in L (and not just for $\tau(L)$).

We also need the following simple claim.

Claim 3.13

$$\forall x, r \in \mathbb{R}, \quad \sum_{k \in \mathbb{Z}} \exp\left(-\pi(kr + x)^2\right) \leq 1 + \frac{1}{r}$$

Proof: Let $k' \in \mathbb{Z}$ be such that $|k'r + x|$ is minimized. Then,

$$\begin{aligned}
\sum_{k \in \mathbb{Z}} \exp\left(-\pi(kr + x)^2\right) &\leq 1 + \sum_{k \in \mathbb{Z} \setminus \{k'\}} \exp\left(-\pi(kr + x)^2\right) \\
&= 1 + \frac{1}{r} \sum_{k \in \mathbb{Z} \setminus \{k'\}} r \cdot \exp\left(-\pi(kr + x)^2\right) \\
&\leq 1 + \frac{1}{r} \int_{-\infty}^{\infty} \exp\left(-\pi y^2\right) dy \\
&= 1 + \frac{1}{r}
\end{aligned}$$

where changing the sum to an integral is possible because the sum can be seen as the area under a function that lies completely below $\exp\left(-\pi y^2\right)$. ■

Lemma 3.14 *Let L be a lattice in which all vectors not parallel to $\tau(L)$ are of length more than $\sqrt{n}$. Then, $\Delta(D_{L^*}, T_{L^*}) < 2^{-\Omega(n)}(1 + \frac{1}{\lambda(L)})$. In particular, if $\lambda(L) \geq \frac{1}{n^c}$ for some $c > 0$ then $\Delta(D_{L^*}, T_{L^*}) < 2^{-\Omega(n)}$.*

Proof: As in the proof of Lemma 3.11, we first show that at any point $y \in \mathcal{P}(L^*)$, $D_{L^*}(y)$ and $T_{L^*}(y)$ are very close. By Lemma 3.10,

$$\begin{aligned} D_{L^*}(y) = \rho(L^* + y) &= d(L) \cdot \sum_{x \in L} \cos(2\pi\langle x, y\rangle)\rho(\{x\}) \\ &= d(L) \cdot \left(\sum_{x \in \{k\tau(L) | k \in \mathbb{Z}\}} \cos(2\pi\langle x, y\rangle)\rho(\{x\}) + \sum_{x \in L \setminus \{k\tau(L) | k \in \mathbb{Z}\}} \cos(2\pi\langle x, y\rangle)\rho(\{x\}) \right) \\ &= T_{L^*}(y) + d(L) \cdot \left(\sum_{x \in L \setminus \{k\tau(L) | k \in \mathbb{Z}\}} \cos(2\pi\langle x, y\rangle)\rho(\{x\}) \right) \end{aligned}$$

where we used (2). Now,

$$\begin{aligned} |D_{L^*}(y) - T_{L^*}(y)| &= d(L) \cdot \left| \sum_{x \in L \setminus \{k\tau(L) | k \in \mathbb{Z}\}} \cos(2\pi\langle x, y\rangle)\rho(\{x\}) \right| \\ &\leq d(L) \cdot \sum_{x \in L \setminus \{k\tau(L) | k \in \mathbb{Z}\}} |\cos(2\pi\langle x, y\rangle)| \cdot \rho(\{x\}) \\ &\leq d(L) \cdot \sum_{x \in L \setminus \{k\tau(L) | k \in \mathbb{Z}\}} \rho(\{x\}) \\ &= d(L) \cdot \rho(L \setminus \{k\tau(L) | k \in \mathbb{Z}\}) \\ &< d(L) \cdot \rho(L \setminus \sqrt{n}B_n) \end{aligned}$$

where we used $L \setminus \{k\tau(L) \mid k \in \mathbb{Z}\} \subseteq L \setminus \sqrt{n}B_n$. Now, using Corollary 3.9,

$$\begin{aligned} d(L) \cdot \rho(L \setminus \sqrt{n}B_n) &< d(L) \cdot 2^{-\Omega(n)} \cdot \rho(L \cap \sqrt{n}B_n) \\ &< d(L) \cdot 2^{-\Omega(n)} \cdot \rho(\{k\tau(L) \mid k \in \mathbb{Z}\}) \\ &\leq d(L) \cdot 2^{-\Omega(n)} \left(1 + \frac{1}{\lambda(L)}\right) \end{aligned}$$

where the last inequality follows from Claim 3.13 with $x = 0$. We conclude the proof by integrating over $\mathcal{P}(L^*)$.

$$\Delta(D_{L^*}, T_{L^*}) < 2^{-\Omega(n)} \left(1 + \frac{1}{\lambda(L)}\right).$$

■

Lemma 3.15 *Let $g(n)$ be at most polynomial in n. Then there exists a reduction from $\mathsf{pSVP}_{g(n)}$ to the following problem. Given a lattice L as an LLL-reduced basis and samples from some distribution, distinguish between the following two cases. Either the distribution is U_{L^*}, or $\lambda(L) \in \left[\frac{\sqrt{n}}{g(n)}, \frac{2\sqrt{n}}{g(n)}\right)$ and the distribution is T_{L^*}.*

Proof: The input to the $\mathsf{pSVP}_{g(n)}$ problem is a lattice L given as an LLL-reduced basis. Consider the distribution D_{L^*}. We can efficiently sample from it by sampling a standard Gaussian centered around the origin and reducing the result modulo $\mathcal{P}(L^*)$. According to Lemma 3.14, if L is a YES instance, then the distribution is exponentially close to T_{L^*}. On the other hand, if L is a NO instance, then Lemma 3.11 implies that the distribution is exponentially close to the uniform distribution U_{L^*}. We call the algorithm that distinguishes between T_{L^*} and U_{L^*} a polynomial number of times and take the majority. This makes the probability of error exponentially small. ■

3.4 One-dimensional Distributions

In this section we complete our sequence of reductions and the proof of the main theorem by reducing the n-dimensional problem of the last section to a one-dimensional problem. We begin with a technical claim.

Claim 3.16 *For any $a, x, y \in \mathbb{R}, a \geq 0$ and any $b > \frac{1}{\sqrt{2\pi}} + 1$,*

$$\left| \frac{d}{dx} \sum_{k \in \mathbb{Z}} \exp\left(-\pi(bk + ax + y)^2\right) \right| \leq \tilde{c}a.$$

Proof: Let z denote $ax + y$. Then,

$$\begin{aligned}
\left| \frac{d}{dx} \sum_{k \in \mathbb{Z}} \exp\left(-\pi(bk + ax + y)^2\right) \right| &= a \left| \frac{d}{dz} \sum_{k \in \mathbb{Z}} \exp\left(-\pi(bk + z)^2\right) \right| \\
&= a \left| \sum_{k \in \mathbb{Z}} -2\pi(bk + z) \exp\left(-\pi(bk + z)^2\right) \right| \\
&\leq a \sum_{k \in \mathbb{Z}} \left| 2\pi(bk + z) \exp\left(-\pi(bk + z)^2\right) \right| \\
&\leq a \sum_{k \in \mathbb{Z}} \eta(bk + z)
\end{aligned}$$

where $\eta(r)$ denotes $|2\pi r \cdot \exp\left(-\pi r^2\right)|$. In the following we will upper bound

$$\sum_{k \in \{0,1,\ldots\}} \eta(bk + z)$$

by a constant for any $z \geq 0$. It can be seen that the original expression is at most $2a$ times this value. Notice that η is increasing from 0 to $\frac{1}{\sqrt{2\pi}}$ where it attains the maximum value of $\sqrt{2\pi e}$. After that point it is monotonically decreasing. Hence,

$$\begin{aligned}
\sum_{k \in \{0,1,\ldots\}} \eta(bk + z) &= \eta(z) + \sum_{k \in \{1,2,\ldots\}} \eta(bk + z) \\
&\leq \eta(z) + \sum_{k \in \{1,2,\ldots\}} \eta(b + k - 1 + z) \\
&\leq \sqrt{2\pi e} + \sum_{k \in \{1,2,\ldots\}} \eta(b + k - 1 + z)
\end{aligned}$$

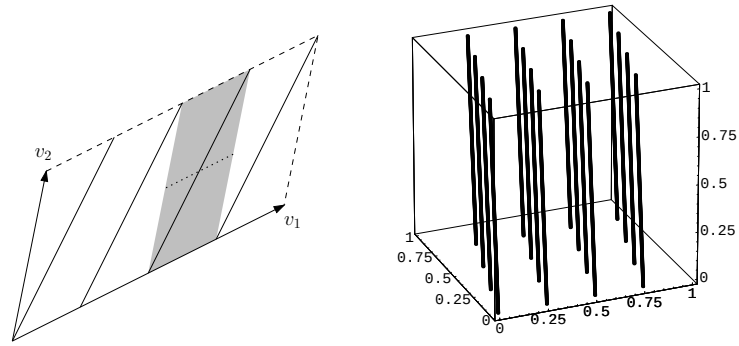

Figure 4: The line connecting the origin with w with $K = 4$ in two dimensions with $\mathcal{P}(v_1, v_2)$ and in three dimensions with the unit cube. The gray area on the left is mapped by f to the segment $\left[\frac{1}{2}, \frac{3}{4}\right)$. The dotted line in its center is the set $S(\frac{5}{8})$.

where the first inequality holds since $b+k-1 \leq bk$ and η is monotonically decreasing on $(b+z, \infty)$. The sum can be upper bounded by

$$\sum_{k \in \{1,2,\ldots\}} \eta(b+k-1+z) \leq \int_{b+z-1}^{\infty} \eta(z)dz \leq \int_{0}^{\infty} \eta(z)dz = 1$$

where the first inequality holds since η is monotonically decreasing on $(b+z-1, \infty)$. ■

Lemma 3.17 *Let $g(n) \geq 4\sqrt{n}$ be at most polynomial in n. Then, if there exists a distinguisher between U and $\mathcal{T}_{n,g(n)}$ then there also exists a distinguisher that, given an LLL-reduced basis for a lattice L and samples from some distribution, distinguishes between the following two cases. Either the distribution is U_{L^*}, or $\lambda(L) \in \left[\frac{\sqrt{n}}{g(n)}, \frac{2\sqrt{n}}{g(n)}\right)$ and the distribution is T_{L^*}.*

Proof: The proof is based on a mapping f from $\mathcal{P}(L^*)$ to $[0,1)$. The mapping has the property that it maps the uniform distribution U_{L^*} on $\mathcal{P}(L^*)$ to the uniform distribution U on $[0,1)$ and that it maps the distribution T_{L^*} on $\mathcal{P}(L^*)$ to a distribution on $[0,1)$ that is very close to one of the distributions in $\mathcal{T}_{n,g(n)}$. Hence, a distinguisher between U and $\mathcal{T}_{n,g(n)}$ implies a distinguisher between U_{L^*} and T_{L^*}.

We start by describing the mapping f. Let $v_1, \ldots, v_n$ denote the LLL-reduced basis of L and let $v_1^*, \ldots, v_n^*$ be the dual basis of L^*, i.e., a basis of L^* such that $\langle v_i, v_j^* \rangle = \delta_{ij}$. For a very large K, we partition the fundamental parallelepiped into K^{n-1} 'narrow and long' cells, by partitioning the coefficients of $v_1^*, \ldots, v_{n-1}^*$ into small intervals of size $1/K$, and letting the coefficient of v_n^* range over the entire interval $[0,1)$. That is, the cell with index $\langle r_1, r_2, \ldots, r_{n-1} \rangle$ is

$$\left\{ \sum_{i=1}^{n} a_i v_i^* \;\middle|\; \forall i \in [n-1]\ a_i \in \left[\frac{r_i}{K}, \frac{r_i+1}{K}\right) \text{ and } a_n \in [0,1) \right\}$$

where $r_1, \ldots, r_{n-1}$ are all in $\{0, \ldots, K-1\}$. See the gray area in Figure 4. These cells are ordered lexicographically, starting from $\langle 0, \ldots, 0, 0 \rangle$, $\langle 0, \ldots, 0, 1 \rangle$ and so on. For $j = 1, \ldots, K^{n-1}$, the jth cell is mapped by f to the interval $[(j-1)/K^{n-1}, j/K^{n-1})$. More precisely, if $x \in \mathcal{P}(L^*)$ is in

the jth cell and its v_n^* coefficient is a_n, then $f(x)$ is the number $(j + a_n - 1)/K^{n-1} \in [0, 1)$. An alternative, more succinct way to describe f is as the function that maps the vector $v = \sum_{i=1}^n a_i v_i^*$ in $\mathcal{P}(L^*)$ to

$$\frac{\lfloor Ka_1 \rfloor}{K} + \frac{\lfloor Ka_2 \rfloor}{K^2} + \ldots + \frac{\lfloor Ka_{n-1} \rfloor}{K^{n-1}} + \frac{Ka_n}{K^n} \in [0, 1).$$

The set of points mapped to each $r \in [0, 1)$, which we denote by $S(r)$, is an $n-1$-dimensional parallelepiped obtained by taking a slice of one of the cells. See the dotted line in Figure 4. More precisely, we define

$$S(r) := \left\{ \sum_{i=1}^n a_i v_i^* \;\middle|\; \forall i \in [n-1]\ a_i \in \left[\frac{r_i}{K}, \frac{r_i + 1}{K}\right) \text{ and } a_n = \frac{r_n}{K} \right\}.$$

where $r_1, \ldots, r_{n-1} \in \{0, 1, \ldots, K-1\}$ and $r_n \in [0, K)$ are the unique numbers such that

$$r = \frac{r_1}{K} + \frac{r_2}{K^2} + \ldots + \frac{r_{n-1}}{K^{n-1}} + \frac{r_n}{K^n}.$$

Notice that $S(r)$ is an $n-1$-dimensional parallelepiped whose diameter is at most $\frac{1}{K}\sum_{i=1}^{n-1} \|v_i^*\|$. This fact is crucial in our reduction.

The reduction works by sampling a point from the given distribution on $\mathcal{P}(L^*)$ and applying f, thereby obtaining a distribution on $[0, 1)$. Notice that f can be computed efficiently. By starting from a uniform distribution on $\mathcal{P}(L^*)$ we obtain the uniform distribution on $[0, 1)$ (this holds since the volume of $S(r)$ is independent of r). Hence, it is enough to consider the case where the given distribution is T_{L^*}. Here, we show that when K is large enough, the resulting distribution is close to one of the distributions in $\mathcal{T}_{n,g(n)}$. The density of the resulting distribution at any $r \in [0, 1)$ is given by averaging (i.e., integrating) the density function T_{L^*} over $S(r)$. We first note that when K is large enough then all the points in $S(r)$ have almost the same density under T_{L^*}. This requires K to be large enough so that the diameter of $S(r)$ is small compared with the derivative of T_{L^*}. Hence, for such K, the density of the resulting distribution at any $r \in [0, 1)$ is closely approximated by $T_{L^*}(x)$ for any $x \in S(r)$. We then choose a specific point in each $S(r)$, namely, $rw \bmod \mathcal{P}(L^*)$ for some w to be defined later, and note that

$$T_{L^*}(rw \bmod P)$$

is in fact a distribution in $\mathcal{T}_{n,g(n)}$.

Let us explain the above in more detail. Choose $K = 2^{3n}$. By averaging over $S(r)$ and multiplying by a normalization factor, we see that the distribution that we obtain on $[0, 1)$ is given by

$$Z(r) := \frac{d(L^*)}{\mathrm{vol}(S(r))} \int_{S(r)} T_{L^*}(x)dx.$$

We now use the crucial fact that the diameter of $S(r)$ is very small: $Z(r)$, which is $d(L^*)$ times the average of T_{L^*} over $S(r)$, can be closely approximated by $d(L^*)$ times the value of T_{L^*} at any point in $S(r)$. More precisely, in Claim 3.18 we will show that $Z(r)$ is exponentially close to $d(L^*)T_{L^*}(z)$ for any $z \in S(r)$.

In the next step, we choose one point from each $S(r)$. Let $w \in L^*$ denote the vector $v_1^* + Kv_2^* + \ldots + K^{n-1}v_n^*$. Since v_n^* has the largest coefficient, w is almost parallel to v_n^*. Consider the line

connecting the origin and w, $\{rw \mid r \in [0,1)\}$. We reduce this line modulo $\mathcal{P}(L^*)$ and obtain

$$\{rw \bmod \mathcal{P}(L^*) \mid r \in [0,1)\}.$$

This set has the form of K^{n-1} segments running through $\mathcal{P}(L^*)$; each segment is contained in a different cell. Figure 4 illustrates this with $K = 4$. We now claim that for any $r \in [0,1)$, $rw \bmod \mathcal{P}(L^*)$ is in $S(r)$. Let $r_1, \ldots, r_{n-1} \in \{0, 1, \ldots, K-1\}$ and $r_n \in [0, K)$ be the unique numbers such that

$$r = \frac{r_1}{K} + \frac{r_2}{K^2} + \ldots + \frac{r_{n-1}}{K^{n-1}} + \frac{r_n}{K^n}.$$

Then,

$$rw \bmod \mathcal{P}(L^*) = (r \bmod 1) \cdot v_1^* + (Kr \bmod 1) \cdot v_2^* + \ldots + (K^{n-1} r \bmod 1) \cdot v_n^*.$$

It is now easy to check that for each $i \in [n-1]$ the coefficient of v_i^* is in $[r_i/K, (r_i+1)/K)$ and that the coefficient of v_n^* is r_n/K. Hence, $rw \bmod \mathcal{P}(L^*) \in S(r)$.

The last crucial observation we make is that $d(L^*)T_{L^*}(rw \bmod \mathcal{P}(L^*))$ is in fact a density function in $\mathcal{T}_{n,g(n)}$. Indeed,

$$\begin{aligned} d(L^*) \cdot T_{L^*}(rw \bmod \mathcal{P}(L^*)) &= \frac{1}{\lambda(L)} \sum_{k \in \mathbb{Z}} \exp\left(-\pi \left(\frac{k + r\langle \tau(L), w\rangle}{\lambda(L)}\right)^2\right) \\ &= \frac{1}{\lambda(L)} \sum_{k \in \mathbb{Z}} \exp\left(-\pi \left(\frac{r\,|\langle \tau(L), w\rangle| - k}{\lambda(L)}\right)^2\right) \\ &= T_{|\langle \tau(L), w\rangle|, \lambda(L)^2}(r). \end{aligned}$$

The distribution $T_{|\langle \tau(L), w\rangle|, \lambda(L)^2}$ is in $\mathcal{T}_{n,g(n)}$ for the following reasons. First, $\lambda(L)^2$ is in $\left[\frac{n}{g^2}, 4\frac{n}{g^2}\right)$. Moreover, recall that $w = \sum_{i=1}^n K^{i-1} v_i^*$ and $\tau(L) = \sum_{i=1}^n a_i v_i$ where all $|a_i| \le 2^{2n}$ by Lemma 3.2. Since $\left\langle v_i, v_j^* \right\rangle = \delta_{ij}$, the inner product $\langle \tau(L), w\rangle$ is integer and its absolute value is at most $n \cdot 2^{2n} \cdot K^n \le 2^{4n^2}$.

It remains to prove the following claim.

Claim 3.18 *For all $r \in [0,1)$ and all $z \in S(r)$,*

$$|Z(r) - d(L^*)T_{L^*}(z)| \le 2^{-\Omega(n)}.$$

Proof: It is enough to show that

$$\left| \frac{1}{\mathrm{vol}(S(r))} \int_{S(r)} T_{L^*}(x)dx - T_{L^*}(z) \right| \le d(L) \cdot 2^{-\Omega(n)}.$$

The left expression inside the absolute value is the average of T_{L^*} in $S(r)$. The right expression is the value of T_{L^*} at a point in $S(r)$. Hence, in the following, it will be enough to prove that for any two points $x, y \in S(r)$,

$$|T_{L^*}(x) - T_{L^*}(y)| \le d(L) \cdot 2^{-\Omega(n)}.$$

Consider the derivative of T_{L^*} in some direction $u \in \mathbb{R}^n$ (i.e., u is a unit vector). In order to calculate this derivative, write

$$\begin{aligned} J_{u,x}(t) = T_{L^*}(x+tu) &= \frac{d(L)}{\lambda(L)} \sum_{k\in\mathbb{Z}} \exp\left(-\pi\left(\frac{k+\langle\tau(L), x+tu\rangle}{\lambda(L)}\right)^2\right) \\ &= \frac{d(L)}{\lambda(L)} \sum_{k\in\mathbb{Z}} \exp\left(-\pi\left(\frac{1}{\lambda(L)}\cdot k + \frac{\langle\tau(L),u\rangle}{\lambda(L)}\cdot t + \frac{\langle\tau(L),x\rangle}{\lambda(L)}\right)^2\right). \end{aligned}$$

The derivative of T_{L^*} in the direction u at point x is given by $J'_{u,x}(0)$. Using Claim 3.16 and $1/\lambda(L) \geq \frac{g(n)}{2\sqrt{n}} \geq 2$, the absolute value of this derivative can be upper bounded by

$$\frac{d(L)}{\lambda(L)} \cdot \tilde{c} \cdot \frac{\langle\tau(L),u\rangle}{\lambda(L)} \leq \tilde{c} \cdot \frac{d(L)}{\lambda(L)}$$

since both $\tau(L)/\lambda(L)$ and u are unit vectors. Having bounded the derivative of T_{L^*} in any direction and at any point, we now apply the mean value theorem and obtain that for any $x, y \in S(r)$,

$$\begin{aligned} |T_{L^*}(x) - T_{L^*}(y)| &\leq \tilde{c} \cdot \frac{d(L)}{\lambda(L)} \cdot \mathrm{diam}(S(r)) \\ &\leq \tilde{c} \cdot \frac{d(L)}{\lambda(L)} \cdot \frac{1}{K} \sum_{i=1}^{n-1} \|v_i^*\| \\ &\leq \tilde{c} \cdot \frac{d(L)}{\lambda(L)} \cdot \frac{1}{K} \cdot n \cdot \frac{\sqrt{n}}{\lambda(L)} \cdot 2^{2n} \\ &\leq \tilde{c} \cdot d(L) \cdot \frac{1}{K} \cdot 2^{2n} \cdot \mathrm{poly}(n) \\ &\leq d(L) \cdot 2^{-\Omega(n)} \end{aligned}$$

where we used Lemma 3.2. ■

■

4 From Worst-case to Average-case

We start with a few technical claims.

Claim 4.1 *For any $h \in \mathbb{N}, \beta \in \mathbb{R}^+$, let X, Y be two independent random variables; X is distributed uniformly over $\{0, \frac{1}{h}, \ldots, \frac{h-1}{h}\}$ and Y is normal with mean 0 and variance $\frac{\beta}{2\pi h^2}$. Then $T_{h,\beta}$ is equivalent to the distribution of the sum of X and Y reduced modulo 1.*

Proof:

$$\begin{aligned} T_{h,\beta}(r) = Q_\beta(hr \bmod 1) &= \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{\beta}} \cdot \exp\left(-\frac{\pi}{\beta}(hr-k)^2\right) \\ &= \sum_{l=0}^{h-1} \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{\beta}} \cdot \exp\left(-\frac{\pi}{\beta}(hr-hk-l)^2\right) \\ &= \sum_{l=0}^{h-1} \frac{1}{h} \sum_{k=-\infty}^{\infty} \frac{h}{\sqrt{\beta}} \cdot \exp\left(-\frac{\pi h^2}{\beta}\left(r-k-\frac{l}{h}\right)^2\right) \end{aligned}$$

■

Claim 4.2 *For any $h \in \mathbb{N}$ and any $\beta, \mu \in \mathbb{R}^+$, $T_{h,\beta} + Q_\mu \bmod 1 = T_{h,\beta+\mu h^2}$.*

Proof: According to Claim 4.1, $T_{h,\beta}$ can be viewed as the sum of two random variables X and Y reduced modulo 1. Therefore, $T_{h,\beta} + Q_\mu \bmod 1 = X + Y + Q_\mu \bmod 1$. But since both Y and Q_μ are normal, their sum modulo 1 is exactly $Q_{\frac{\beta}{h^2}+\mu}$ and we conclude the proof by using Claim 4.1 again. ■

Before describing the main theorem of this section, we need to extend the definition of $T_{h,\beta}$ to non-integer h. This is done by adding a normalization factor, i.e., for a real $h > 0$ and $r \in [0, 1)$ we define

$$T_{h,\beta}(r) = \frac{1}{\int_0^1 Q_\beta(xh \bmod 1)dx} Q_\beta(rh \bmod 1).$$

It is easy to see that $T_{h,\beta}$ is still efficiently samplable. Namely, for a real $h > 0$, first choose a value $x \in \{0, 1, \ldots, \lceil h \rceil - 1\}$ and then choose a value y according to Q_β. If $\frac{x+y}{h} < 1$ then return it as the result. Otherwise, repeat the process again. It is easy to see that the distribution obtained is indeed $T_{h,\beta}$ and that the process is efficient for (say) $h \geq 1$.

Definition 4.3 *Given a density function X on $[0, 1)$ and a real $\eta \geq 1$, we define its* compression *by η as the distribution on $[0, 1)$ given by*

$$\frac{1}{\int_0^1 X(\eta x \bmod 1)dx} X(\eta r \bmod 1).$$

We denote the result by $C_\eta(X)$.

Using the above definition, for any real $h > 0$, $T_{h,\beta}$ is a compression of Q_β by a factor of h. Notice that if we can sample efficiently from X then we can also sample efficiently from its compression. This is done in a way similar to that used to sample from $T_{h,\beta}$.

Claim 4.4 *For any $h \in \mathbb{N}$ and any real $\eta \geq 1$, the compression of $T_{h,\beta}$ by a factor η is $T_{\eta h,\beta}$.*

Proof: The proof follows directly from the definition of $T_{h,\beta}$. ■

We now prove the main theorem of this section.

Theorem 4.5 *Let $g(n)$ be any function such that $4\sqrt{n} \leq g(n) \leq \mathrm{poly}(n)$. Let h be chosen uniformly from $[2^{4n^2}, 2 \cdot 2^{4n^2})$ and β be chosen uniformly from $[4n/g(n)^2, 8n/g(n)^2)$. Assume there exists a distinguisher $\mathcal{A}$ that with probability at least $1/\mathrm{poly}(n)$ over the choice of h and β distinguishes between U and $T_{h,\beta}$. Then, there exists a solution to $g(n)$-uSVP.*

Proof: Let $p_{\mathcal{A}}(D)$ denote the probability that $\mathcal{A}$ accepts given samples from some distribution D on $[0, 1)$. Then, our assumption above says that there exists some $c > 0$ such that with probability at least $1/\mathrm{poly}(n)$ over our choice of h and β, $|p_{\mathcal{A}}(U) - p_{\mathcal{A}}(T_{h,\beta})| \geq n^{-c}$. In the following we construct a distinguisher $\mathcal{B}$ that distinguishes between U and *any* $T_{h,\beta} \in \mathcal{T}_{n,g(n)}$. In other words, our goal is to construct $\mathcal{B}$ such that the acceptance probability with U and the acceptance probability with $T_{h,\beta}$

differ by a non-negligible amount. Using the main theorem, this implies a solution to $g(n)$-uSVP. Recall that neither h nor β are given to $\mathcal{B}$. The idea is to perform a random modification to the given distribution. The modification is such that the uniform distribution remains uniform while $T_{h,\beta}$ transforms to $T_{h',\beta'}$ for some h', β' that are in the range in which $\mathcal{A}$ works. Modifying h is done by compressing the input distribution; modifying β is done by adding some noise.

Let R denote the unknown distribution given to $\mathcal{B}$. We start by choosing $\tilde{h}$ uniformly from the set $\{1, 2, 4, \ldots, 2^{4n^2}\}$. In addition, we choose δ uniformly from $[1, 4)$ and s uniformly from $\left[0, 32n/g(n)^2\right)$. Then, consider the distribution

$$R' = C_{\delta 2^{4n^2}/\tilde{h}}(R + Q_{s/\tilde{h}^2} \bmod 1),$$

i.e., we first add a normal variable to R and then compress the result by a factor of $\delta 2^{4n^2}/\tilde{h}$. We call $\mathcal{A}$ a polynomial number of times with samples taken from this distribution (each time with as many samples as required by $\mathcal{A}$). This allows us to obtain, with probability exponentially close to 1, an estimate on $p_{\mathcal{A}}(R')$ that is accurate up to an additive error of $\frac{1}{8n^c}$. We then do a similar process with samples taken from U and obtain an estimate on $p_{\mathcal{A}}(U)$ with the same additive error. If the two estimates differ by more than $\frac{1}{2n^c}$, $\mathcal{B}$ accepts. Otherwise, $\mathcal{B}$ rejects.

We first claim that when R is the uniform distribution, $\mathcal{B}$ rejects with high probability. The distribution $R + Q_{s/\tilde{h}^2} \bmod 1$ is still a uniform distribution on $[0, 1)$ and so is R' as can be easily seen from the definition of the compression. Hence, $p_{\mathcal{A}}(U) = p_{\mathcal{A}}(R')$ and the probability that our two estimates differ by more than $\frac{1}{2n^c}$ is exponentially small.

Now assume that R is the distribution $T_{h,\beta}$ for some fixed integer $h \leq 2^{4n^2}$ and $\beta \in [n/g(n)^2, 4n/g(n)^2)$ and we claim that $\mathcal{B}$ accepts with non-negligible probability. According to Claim 4.2, $R + Q_{s/\tilde{h}^2} \bmod 1$ is $T_{h,\beta+s(h/\tilde{h})^2}$. Hence, according to Claim 4.4, R' is $T_{\delta 2^{4n^2} h/\tilde{h}, \beta+s(h/\tilde{h})^2}$. Let X denote the event that

$$h \leq \tilde{h} < 2h, \ \delta h/\tilde{h} \in [1, 2), \ \text{and } \beta + s(h/\tilde{h})^2 \in \left[\frac{4n}{g(n)^2}, \frac{8n}{g(n)^2}\right).$$

We now show that X happens with probability $1/\text{poly}(n)$ over our choice of $\tilde{h}, \delta, s$. First, with probability $\frac{1}{4n^2}$, $h \leq \tilde{h} < 2h$. From now on, condition on this event happening. Then, $\delta h/\tilde{h}$, which is uniformly distributed in $[h/\tilde{h}, 4h/\tilde{h})$, satisfies that the probability that $\delta h/\tilde{h} \in [1, 2)$ is at least $\frac{1}{3}$. Moreover, $\beta + s(h/\tilde{h})^2$ is distributed uniformly in

$$\left[\beta, \quad \beta + 32n/g(n)^2 \cdot (h/\tilde{h})^2\right).$$

Since $h/\tilde{h} \in (\frac{1}{2}, 1]$, the length of this segment is at most $32n/g(n)^2$ and it always contains $\left[4n/g(n)^2, 8n/g(n)^2)\right)$ (recall that $\beta \in \left[n/g(n)^2, 4n/g(n)^2\right)$). Therefore, the probability on the choice of s that

$$\beta + s(h/\tilde{h})^2 \in \left[\frac{4n}{g(n)^2}, \frac{8n}{g(n)^2}\right)$$

is at least $\frac{4}{32} = \frac{1}{8}$. To sum up, the probability of X is at least

$$\frac{1}{4n^2} \cdot \frac{1}{3} \cdot \frac{1}{8} = \frac{1}{\text{poly}(n)}.$$

Finally, notice that conditioned on X, the distribution of $\delta 2^{4n^2} h/\tilde{h}$ and $\beta + s(h/\tilde{h})^2$ is the same as the distribution of h and β in our assumption on $\mathcal{A}$. Therefore, with probability at least $1/\text{poly}(n)$,

$p_{\mathcal{A}}(U)$ and $p_{\mathcal{A}}(R')$ differ by at least n^{-c}. Then, except with exponentially small probability, our estimates are good enough and $\mathcal{B}$ accepts.

■

5 A Public Key Cryptosystem

For a security parameter n, let N be 2^{8n^2} and let m be $c_{\mathsf{m}} n^2$ where c_{m} is a constant to be specified later. Let $\gamma(n) = \omega(n\sqrt{\log n})$, i.e., any function that satisfies $\frac{\gamma(n)}{n\sqrt{\log n}} \to \infty$ as n goes to infinity. On one hand, choosing a smaller function $\gamma(n)$ yields a stronger security guarantee. On the other hand, it also makes decryption errors more likely. Our choice of $\omega(n\sqrt{\log n})$ is the smallest possible $\gamma(n)$ that still leaves the probability of decryption error negligible. For concreteness, one can choose $\gamma(n) = n \log n$. Let us now describe the cryptosystem.

- **Private Key:** Let $H = \{h \in [\sqrt{N}, 2\sqrt{N}) \mid \mathrm{frc}\,(h) < \frac{1}{16m}\}$. Choose $h \in H$ uniformly at random. Let d denote $\frac{N}{h}$. The private key is the number h.
- **Public Key:** Choose $\beta \in \left[4/(\gamma(n))^2, 8/(\gamma(n))^2\right)$ uniformly at random. We choose m values $z_1, \ldots, z_m$ from $T_{h,\beta}$ by choosing $x_1, \ldots, x_m$ and $y_1, \ldots, y_m$ as described above Definition 4.3. Let i_0 be an index such that x_{i_0} is odd (such an i_0 exists with probability exponentially close to 1). For $i \in [m]$, let a_i denote $\lfloor N \cdot z_i \rfloor$. The public key is $(a_1, \ldots, a_m, i_0)$.
- **Encryption:** In order to encrypt a bit we choose a random subset S of $[m]$. The encryption is $\sum_{i \in S} a_i \bmod N$ if the bit is 0 and $\sum_{i \in S} a_i + \lfloor \frac{a_{i_0}}{2} \rfloor \bmod N$ if the bit is 1.
- **Decryption:** On receiving $w \in \{0, \ldots, N-1\}$ we decrypt 0 if $\mathrm{frc}\left(\frac{w}{d}\right) < \frac{1}{4}$ and 1 otherwise.

5.1 Analysis

We start with a simple tail bound on the normal distribution.

Claim 5.1 *The probability that the distance of a normal variable with variance σ^2 from its mean is more than t is at most $\sqrt{\frac{2}{\pi}} \cdot \frac{\sigma}{t} \cdot \exp\left(-\frac{t^2}{2\sigma^2}\right)$.*

Proof:

$$\begin{aligned}
\int_t^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx &\leq \int_t^\infty \left(1 + \frac{\sigma^2}{x^2}\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\
&= -\frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{\sigma^2}{x} \exp\left(-\frac{x^2}{2\sigma^2}\right)\Bigg|_{x=t}^{\infty} \\
&= \frac{\sigma}{\sqrt{2\pi}t} \exp\left(-\frac{t^2}{2\sigma^2}\right).
\end{aligned}$$

■

In the following lemma we prove the correctness of the encryption scheme. The idea is the following. We prove that an encryption of 0 is close to a multiple of d and that an encryption of 1 is far from a multiple of d. Let us explain how we prove this for an encryption of 0, the other case

being similar. Each number a_i is chosen to be close to some multiple of d with standard deviation roughly $d/\gamma(n)$. Hence $\sum_{i\in S} a_i$, which is a sum of at most m of them, is also distributed around some multiple of d and has standard deviation $d\sqrt{m}/\gamma(n) = d/\omega(\sqrt{\log n})$. Hence, the probability that its distance from a multiple of d is more than, say, $d/8$ is at most $2^{-\omega(\log n)}$, which is a negligible amount. We also have to deal with the fact that the encryption is in fact $\sum_{i\in S} a_i \bmod N$ and not $\sum_{i\in S} a_i$. Notice that for an integer h, the distance to the nearest multiple of d is the same for both expressions. We show that since h is close to an integer, the distance to the nearest multiple of d is almost the same for both expressions.

Lemma 5.2 (Correctness) *The probability of a decryption error is at most* $2^{-\Omega(\frac{(\gamma(n))^2}{m})}$ *plus some exponentially small terms.*

Note that the above probability is negligible since $\gamma(n) = \omega(n\sqrt{\log n})$.

Proof: First consider an encryption of the bit 0. Probabilities are taken over the choices of the private and public keys and the randomization in the encryption process. Let S denote the subset of indices that are included in the sum and let $w := \sum_{i\in S} a_i \bmod N$. Since $\sum_{i\in S} a_i \le m \cdot N$,

$$\left| w - \left(\sum_{i\in S} a_i \bmod d\lfloor h \rceil \right) \right| \le m \cdot |N - d\lfloor h \rceil| = m \cdot d \cdot \mathrm{frc}\,(h) < \frac{1}{16} d$$

and by the triangle inequality,

$$\mathrm{frc}\left(\frac{w}{d}\right) < \frac{1}{16} + \mathrm{frc}\left(\frac{\sum_{i\in S} a_i \bmod d\lfloor h \rceil}{d}\right) = \frac{1}{16} + \mathrm{frc}\left(\frac{\sum_{i\in S} a_i}{d}\right) < \frac{1}{16} + \frac{m}{d} + \mathrm{frc}\left(\frac{N}{d}\sum_{i\in S} z_i\right)$$

where the last inequality uses $|N \cdot z_i - a_i| < 1$. Notice that

$$\mathrm{frc}\left(\frac{N}{d}\sum_{i\in S} z_i\right) = \mathrm{frc}\left(\sum_{i\in S}(x_i + y_i)\right) = \mathrm{frc}\left(\sum_{i\in S} y_i\right).$$

Hence,

$$\mathrm{frc}\left(\frac{w}{d}\right) < \frac{1}{16} + \frac{m}{d} + \mathrm{frc}\left(\sum_{i\in S} y_i\right) < \frac{1}{8} + \mathrm{frc}\left(\sum_{i\in S} y_i\right)$$

where we used the fact that d is much larger than m. With probability exponentially close to 1, all x_i's are strictly less than $\lceil h \rceil - 1$. Conditioned on that, the distribution of y_i is Q_β and the distribution of $\sum_{i\in S} y_i \bmod 1$ is $Q_{|S|\beta}$ where $|S|\beta \le m \cdot \beta = O(\frac{m}{(\gamma(n))^2})$. Therefore, according to Claim 5.1, the probability of $\mathrm{frc}\left(\sum_{i\in S} y_i\right) > \frac{1}{16}$ is at most $2^{-\Omega(\frac{(\gamma(n))^2}{m})}$ and hence

$$\mathrm{frc}\left(\frac{w}{d}\right) < \frac{1}{8} + \frac{1}{16}, \tag{3}$$

which is less than $\frac{1}{4}$, as required.

The proof for the case of an encryption of 1 is similar. By using the fact that x_{i_0} is odd and that with probability exponentially close to 1, $\mathrm{frc}\,(y_{i_0}) < \frac{1}{16}$ we get $\mathrm{frc}\left(\frac{\lfloor a_{i_0}/2 \rfloor}{d}\right) > \frac{1}{2} - \frac{1}{32} - \frac{1}{d}$. This, combined with (3) gives

$$\mathrm{frc}\left(\frac{w}{d}\right) > \mathrm{frc}\left(\frac{\lfloor a_{i_0}/2 \rfloor}{d}\right) - \frac{1}{8} - \frac{1}{16} > \frac{1}{4}$$

and the proof is completed. ■

The following claim is a special case of Lemma 1 in the appendix of [1]. We include a proof for completeness.

Claim 5.3 *For large enough c, when choosing $c \cdot l$ numbers $a_1, \ldots, a_{c \cdot l}$ uniformly from 0 to $2^l - 1$ the probability that the statistical distance between the uniform distribution on $\{0, \ldots, 2^l - 1\}$ and the distribution given by sums modulo 2^l of random subsets of $\{a_1, \ldots, a_{c \cdot l}\}$ is more than 2^{-l} is at most 2^{-l}.*

Proof: Let $X_{t,b}$ for $t \in \{0, \ldots, 2^l - 1\}, b \in \{0,1\}^{c \cdot l} \setminus \{0^{c \cdot l}\}$ denote the event that $\sum_{i=1}^{c \cdot l} b_i a_i \equiv t \pmod{2^l}$ where the probability is taken over the choice of $\{a_1, \ldots, a_{c \cdot l}\}$. Then, $E[X_{t,b}] = 2^{-l}$ and $V[X_{t,b}] < 2^{-l}$. Hence,

$$E[Y_t] = \frac{2^{c \cdot l} - 1}{2^l} = 2^{(c-1) \cdot l} - 2^{-l}$$

where Y_t denotes $\sum_{b \in \{0,1\}^{c \cdot l} \setminus 0^{c \cdot l}} X_{t,b}$. Moreover, for any $b \neq b'$, the events $X_{t,b}$ and $X_{t,b'}$ are independent. Therefore,

$$V[Y_t] < \frac{2^{c \cdot l} - 1}{2^l} < 2^{(c-1) \cdot l}.$$

Using the Chebyshev inequality,

$$\Pr\left(\left|Y_t - (2^{(c-1) \cdot l} - 2^{-l})\right| \geq 2^{(\frac{c-1}{2}+1) \cdot l}\right) \leq 2^{-2l}$$

and hence,

$$\Pr\left(\left|Y_t - 2^{(c-1) \cdot l}\right| \geq 2^{(\frac{c-1}{2}+1) \cdot l} + 2^{-l}\right) \leq 2^{-2l}.$$

Using the union bound,

$$\Pr\left(\exists t,\ \left|Y_t - 2^{(c-1) \cdot l}\right| \geq 2^{(\frac{c-1}{2}+1) \cdot l} + 2^{-l}\right) \leq 2^{-l}.$$

Therefore, with probability at least $1 - 2^{-l}$ on the choice of $\{a_1, \ldots, a_{c \cdot l}\}$, the number of subsets (including the empty subset) mapped to each number t is at most

$$2^{(\frac{c-1}{2}+1) \cdot l} + 2^{-l} + 1 \leq 2^{(\frac{c-1}{2}+2) \cdot l}$$

away from $2^{(c-1) \cdot l}$. This translates to a statistical distance of at most

$$2^{(\frac{c-1}{2}+2) \cdot l} \cdot 2^{-(c-1) \cdot l} < 2^{-l}$$

for large enough c. ■

Based on Theorem 4.5, we can now prove the security of the encryption scheme.

Lemma 5.4 (Security) *For a large enough c_{m}, if there exists a polynomial time algorithm $\mathcal{A}$ that distinguishes between encryptions of 0 and 1 then there exists an algorithm $\mathcal{B}$ that with probability at least $1/\mathrm{poly}(n)$ over the choice of h uniformly from $[2^{4n^2}, 2 \cdot 2^{4n^2})$ and β uniformly from $\left[4/(\gamma(n))^2, 8/(\gamma(n))^2\right)$, distinguishes between the distributions U and $T_{h,\beta}$.*

Proof: Let p_0 be the acceptance probability of $\mathcal{A}$ on input $((a_1, \ldots, a_m, i_0), w)$ where w is an encryption of 0 with the public key $(a_1, \ldots, a_m, i_0)$ and the probability is taken over the choice of private and public keys and the encryption algorithm. We define p_1 similarly for encryptions of 1 and let p_u be the acceptance probability of $\mathcal{A}$ on inputs $((a_1, \ldots, a_m, i_0), w)$ where $a_1, \ldots, a_m, i_0$ are again chosen according to the private and public keys distribution but w is chosen uniformly from $\{0, \ldots, N-1\}$. We would like to construct an $\mathcal{A}'$ that distinguishes between the case where w is an encryption of 0 and the case where w is random. According to our hypothesis, $|p_0 - p_1| \geq \frac{1}{n^c}$ for some $c > 0$. Therefore, either $|p_0 - p_u| \geq \frac{1}{2n^c}$ or $|p_1 - p_u| \geq \frac{1}{2n^c}$. In the former case $\mathcal{A}$ is itself the required distinguisher. In the latter case $\mathcal{A}$ distinguishes between the case where w is an encryption of 1 and the case where w is random. We construct $\mathcal{A}'$ as follows. On input $((a_1, \ldots, a_n, i_0), w)$, $\mathcal{A}'$ calls $\mathcal{A}$ with $((a_1, \ldots, a_n, i_0), w + \lfloor \frac{a_{i_0}}{2} \rfloor \bmod N)$. Notice that this maps the distribution on encryptions of 0 to the distribution on encryptions of 1 and the uniform distribution to itself. Therefore, $\mathcal{A}'$ is the required distinguisher.

Let $p_0(h, \beta)$ be the probability that $\mathcal{A}'$ accepts on input $((a_1, \ldots, a_m, i_0), w)$ where $(a_1, \ldots, a_m, i_0)$ is chosen as a public key with some fixed choice of h and β, and w is an encryption of 0 with the public key $(a_1, \ldots, a_m, i_0)$. Similarly, define $p_u(h, \beta)$ to be the acceptance probability of $\mathcal{A}'$ where $(a_1, \ldots, a_m, i_0)$ is chosen as a public key with some fixed choice of h and β, and w is now chosen uniformly at random from $\{0, \ldots, N-1\}$. Define

$$Y = \left\{ (h, \beta) \;\middle|\; |p_0(h, \beta) - p_u(h, \beta)| \geq \frac{1}{4n^c} \right\}.$$

By an averaging argument we get that with probability at least $\frac{1}{4n^c}$ on the choice of (h, β) in the encryption scheme, $(h, \beta) \in Y$ for otherwise $\mathcal{A}'$ would have a gap of less than $\frac{1}{2n^c}$. Notice that if instead of choosing h as in the encryption scheme, we choose it uniformly from $[\sqrt{N}, 2\sqrt{N})$, we get that the probability that $(h, \beta) \in Y$ is at least $\frac{1}{8m} \cdot \frac{1}{4n^c} = 1/\mathrm{poly}(n)$ since with probability $\frac{1}{8m}$, $\mathrm{frc}\,(h) < \frac{1}{16m}$. Hence, it is enough to show a distinguisher $\mathcal{B}$ that distinguishes between U and $T_{h,\beta}$ for any $(h, \beta) \in Y$.

In the following we describe the distinguisher $\mathcal{B}$. We are given a distribution R that is either U or $T_{h,\beta}$ for some $(h, \beta) \in Y$. We take m samples $a_1, \ldots, a_m$ from $\lfloor N \cdot R \rfloor$ and let i_0 be chosen uniformly at random from $[m]$. Let $p_0(a_1, \ldots, a_m, i_0)$ be the probability that $\mathcal{A}'$ accepts on input $((a_1, \ldots, a_m, i_0), w)$ where the probability is taken only on the choice of w as an encryption of the bit 0. Similarly, let $p_u(a_1, \ldots, a_m, i_0)$ be the probability that $\mathcal{A}'$ accepts on input $((a_1, \ldots, a_m, i_0), w)$ where the probability is taken over the choice of w as a random element of $\{0, \ldots, N-1\}$. We estimate both $p_0(a_1, \ldots, a_m, i_0)$ and $p_u(a_1, \ldots, a_m, i_0)$ up to an additive error of $\frac{1}{64n^c}$. If the two estimates differ by more than $\frac{1}{16n^c}$, $\mathcal{B}$ accepts. Otherwise, $\mathcal{B}$ rejects.

We first claim that when R is the uniform distribution, $\mathcal{B}$ rejects with high probability. In this case, $a_1, \ldots, a_m$ are chosen uniformly from $\{0, \ldots, N-1\}$ and according to Claim 5.3, if c_{m} is a large enough constant, then with probability exponentially close to 1, the distribution on w obtained by encryptions of 0 is exponentially close to the uniform distribution on $\{0, \ldots, N-1\}$. Therefore, except with exponentially small probability,

$$|p_0(a_1, \ldots, a_m, i_0) - p_u(a_1, \ldots, a_m, i_0)|$$

is exponentially small. Hence, our two estimates differ by at most $\frac{1}{32n^c}$, and $\mathcal{B}$ rejects.

Next, we show that if R is $T_{h,\beta}$ for some $(h,\beta) \in Y$ then $\mathcal{B}$ accepts with probability $1/\text{poly}(n)$. Notice that $p_0(h,\beta)$ (respectively, $p_u(h,\beta)$) is the average of $p_0(a_1,\ldots,a_m,i_0)$ (respectively, $p_u(a_1,\ldots,a_m,i_0)$) taken over the choice of $a_1,\ldots,a_m,i_0$ in the encryption scheme. From $|p_0(h,\beta) - p_u(h,\beta)| \geq \frac{1}{4n^c}$ we obtain by an averaging argument that

$$|p_u(a_1,\ldots,a_m,i_0) - p_0(a_1,\ldots,a_m,i_0)| \geq \frac{1}{8n^c}$$

with probability at least $\frac{1}{8n^c}$ on the choice of $a_1,\ldots,a_m,i_0$ in the encryption scheme. Now $\mathcal{B}$ chooses $a_1,\ldots,a_m$ in the same way they are chosen by the encryption scheme, i.e., from $\lfloor N \cdot T_{h,\beta} \rfloor$. The index i_0, however, is chosen randomly. This implies that with probability at least $\frac{1}{8n^c} \cdot \frac{1}{m} = 1/\text{poly}(n)$, $\mathcal{B}$ chooses a tuple $(a_1,\ldots,a_m,i_0)$ such that

$$|p_u(a_1,\ldots,a_m,i_0) - p_0(a_1,\ldots,a_m,i_0)| \geq \frac{1}{8n^c}.$$

Since our estimates are accurate to within $\frac{1}{64n^c}$, the difference between them is more than $\frac{1}{16n^c}$ and $\mathcal{B}$ accepts. ■

By combining the two lemmas above and using Theorem 4.5 we get,

Theorem 5.5 *For a large enough c_{m}, our public key cryptosystem makes decryption errors with negligible probability and its security is based on the worst-case hardness of $\sqrt{n} \cdot \gamma(n)$-*uSVP*.*

6 A Family of Collision Resistant Hash Functions

For a security parameter n, let N be 2^{8n^2} and let m be $c_{\mathsf{m}} n^2$ where $c_{\mathsf{m}} > 0$ is any constant. Choose m numbers $a_1,\ldots,a_m$ uniformly in $\{0,1,\ldots,N-1\}$. The function $f : \{0,1\}^m \to \{0,1,\ldots,N-1\}$ is defined as

$$f(b) = \sum_{i=1}^{m} b_i a_i \bmod N.$$

Notice that if $c_{\mathsf{m}} > 8$ then f indeed compresses the size of the input and collisions are guaranteed to exist.

6.1 Analysis

We start with a simple bound on the statistical distance between joint distributions.

Claim 6.1 *Let $X_1,\ldots,X_m,Y_1,\ldots,Y_m$ be mutually independent random variables. Then the statistical distance between the joint distributions satisfies*

$$\Delta((X_1,\ldots,X_m),(Y_1,\ldots,Y_m)) \leq \sum_{i=1}^{m} \Delta(X_i,Y_i).$$

Proof: We consider the case $m = 2$. The claim follows for $m > 2$ by induction. According to the triangle inequality,

$$\Delta((X_1,X_2),(Y_1,Y_2)) \leq \Delta((X_1,X_2),(X_1,Y_2)) + \Delta((X_1,Y_2),(Y_1,Y_2)).$$

Since X_1 is independent of X_2 and Y_2,

$$\Delta((X_1, X_2), (X_1, Y_2)) = \Delta(X_2, Y_2)$$

and similarly

$$\Delta((X_1, Y_2), (Y_1, Y_2)) = \Delta(X_1, Y_1).$$

■

Claim 6.2 *Let $X_1, \ldots, X_m$ be m independent normal random variables with mean 0 and standard deviation σ. For any vector $b \in \mathbb{R}^m$, the random variable $\sum_{i=1}^m b_i X_i$ has a normal distribution with mean 0 and standard deviation $\|b\| \cdot \sigma$.*

Proof: The joint distribution $(X_1, \ldots, X_m)$ is a Gaussian distribution in $\mathbb{R}^m$ that is invariant under rotations. Hence we can equivalently consider the inner product of $(\|b\|, 0, \ldots, 0)$ and a Gaussian distribution. We complete the proof by noting that the first coordinate of the Gaussian has a normal distribution with mean 0 and standard deviation σ. ■

We now define two distributions on the segment $[0, 1)$. These distributions are obtained by restricting $T_{h,\beta}$ to an interval and scaling appropriately. The first is a restriction to the interval $[a, a + 1/\tilde{h})$ and the second is a restriction to the interval $[a, a + 1/h)$. In the technical claim that follows, we show that these distributions are close given that $\tilde{h}$ is close to h.

Definition 6.3 *For any $h \in \mathbb{N}$, $\tilde{h}, \beta \in \mathbb{R}$ and any $a \in [0, 1)$ we define the following two density functions on $[0, 1)$:*

$$\begin{aligned} S_{\tilde{h},h,\beta,a}(r) &:= \frac{1}{\tilde{h}\int_a^{a+1/\tilde{h}} T_{h,\beta}(x)dx} T_{h,\beta}\left(a + \frac{r}{\tilde{h}}\right), \\ S'_{h,\beta,a}(r) &:= T_{h,\beta}\left(a + \frac{r}{h}\right) = Q_\beta(a \cdot h + r \bmod 1). \end{aligned}$$

Claim 6.4 *If $h \leq \tilde{h} < (1+\delta)h$ where $h \in \mathbb{N}$, $\tilde{h} \in \mathbb{R}$, $\delta > 0$ and $\beta \leq \frac{1}{4}$ then $\Delta(S_{\tilde{h},h,\beta,a}, S'_{h,\beta,a}) \leq \frac{\tilde{c}}{\beta}\delta$.*

Proof: According to Claim 3.13,

$$T_{h,\beta}(x) = Q_\beta(hx \bmod 1) \leq (1 + \sqrt{\beta})/\sqrt{\beta} \leq 2/\sqrt{\beta}$$

for any $x \in \mathbb{R}$. Therefore,

$$\int_a^{a+1/h} - \int_a^{a+1/\tilde{h}} T_{h,\beta}(x)dx \leq \frac{2}{\sqrt{\beta}}\left(\frac{1}{h} - \frac{1}{\tilde{h}}\right) = \frac{2}{\sqrt{\beta} \cdot \tilde{h}}\left(\frac{\tilde{h}}{h} - 1\right) \leq \frac{2\delta}{\sqrt{\beta} \cdot \tilde{h}}.$$

But $\int_a^{a+1/h} T_{h,\beta}(x)dx = \frac{1}{h}$ and therefore we see that

$$\frac{\tilde{h}}{h} - \tilde{h}\int_a^{a+1/\tilde{h}} T_{h,\beta}(x)dx \leq \frac{2\delta}{\sqrt{\beta}}.$$

Let $S''_{\tilde{h},h,\beta,a}(r) := T_{h,\beta}(a + r/\tilde{h})$. Then,

$$\begin{aligned}
\int_0^1 \left| S_{\tilde{h},h,\beta,a}(r) - S''_{\tilde{h},h,\beta,a}(r) \right| dr &= \left| 1 - \tilde{h} \int_a^{a+1/\tilde{h}} T_{h,\beta}(x)dx \right| \cdot \int_0^1 S_{\tilde{h},h,\beta,a}(r)dr \\
&= \left| 1 - \tilde{h} \int_a^{a+1/\tilde{h}} T_{h,\beta}(x)dx \right| \\
&\leq \left| 1 - \frac{\tilde{h}}{h} \right| + \left| \frac{\tilde{h}}{h} - \tilde{h} \int_a^{a+1/\tilde{h}} T_{h,\beta}(x)dx \right| \leq \left(1 + \frac{2}{\sqrt{\beta}}\right) \delta.
\end{aligned}$$

Now, using the mean value theorem, for any $r \in [0, 1)$

$$\begin{aligned}
\left| S'_{h,\beta,a}(r) - S''_{\tilde{h},h,\beta,a}(r) \right| &\leq \left(\frac{1}{h} - \frac{1}{\tilde{h}}\right) \max_x \left| \frac{d}{dx} T_{h,\beta}(x) \right| \\
&= \left(\frac{1}{h} - \frac{1}{\tilde{h}}\right) \max_x \left| \frac{d}{dx} \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{\beta}} \exp\left(-\pi(\frac{h}{\sqrt{\beta}}x - \frac{1}{\sqrt{\beta}}k)^2\right) \right|
\end{aligned}$$

which, according to Claim 3.16 using $\frac{1}{\sqrt{\beta}} \geq 2 > \frac{1}{\sqrt{2\pi}} + 1$, is at most

$$\left(\frac{1}{h} - \frac{1}{\tilde{h}}\right) \cdot \frac{\tilde{c}}{\sqrt{\beta}} \cdot \frac{h}{\sqrt{\beta}} = \frac{\tilde{c}}{\beta}\left(1 - \frac{h}{\tilde{h}}\right) \leq \frac{\tilde{c}}{\beta} \cdot \delta.$$

To sum up,

$$\begin{aligned}
2\Delta(S_{\tilde{h},h,\beta,a}, S'_{h,\beta,a}) &\leq \int_0^1 \left| S_{\tilde{h},h,\beta,a}(r) - S''_{\tilde{h},h,\beta,a}(r) \right| dr + \int_0^1 \left| S'_{h,\beta,a}(r) - S''_{\tilde{h},h,\beta,a}(r) \right| dr \\
&\leq \left(\frac{\tilde{c}}{\beta} + 1 + \frac{2}{\sqrt{\beta}}\right) \delta \leq \frac{\tilde{c}}{\beta} \cdot \delta.
\end{aligned}$$

■

Let $\gamma(n) = \omega(n\sqrt{\log n})$ be any function growing faster than $n\sqrt{\log n}$. For concreteness, we can choose $\gamma(n) = n \log n$. For technical reasons, we also need some polynomial upper bound on $\gamma(n)$ so assume that $\gamma(n) \leq n^2$.

Theorem 6.5 *If there exists an algorithm $\mathcal{A}$ that given a list $a_1, \ldots, a_m \in \{0, 1, \ldots, N-1\}$ chosen uniformly at random finds a nonzero vector $b \in \mathbb{Z}^m$ such that $\|b\| \leq \sqrt{m}$ and $\sum_{i=1}^m b_i a_i \equiv 0 (\mathrm{mod}\ N)$ with probability at least $n^{-c_{\mathsf{a}}}$ where $c_{\mathsf{a}} > 0$ is some constant then there exists a solution to $\sqrt{n} \cdot \gamma(n)$-uSVP.*

Note that in particular, if $b \in \{-1, 0, 1\}^m$ then $\|b\| \leq \sqrt{m}$ and hence this theorem includes collision finding algorithms.

Proof: The existence of $\mathcal{A}$ implies the existence of an algorithm $\mathcal{A}'$ that, given a list $z_1, \ldots, z_m \in [0, 1)$ chosen uniformly at random finds a nonzero vector $b \in \mathbb{Z}^m$ such that $\|b\| \leq \sqrt{m}$ and

$$\mathrm{frc}\left(\sum_{i=1}^m b_i z_i\right) \leq \frac{m}{N}$$

with probability at least $n^{-c_{\mathfrak{a}}}$. Notice that $\frac{m}{N}$ is extremely small, essentially on the order of $1/N$. Indeed, given a list $z_1, \ldots, z_m \in [0,1)$ we can define $a_i = \lfloor N \cdot z_i \rfloor$ and call $\mathcal{A}$. The returned vector b satisfies

$$\mathrm{frc}\left(\sum_{i=1}^{m} b_i z_i\right) = \mathrm{frc}\left(\sum_{i=1}^{m} \frac{1}{N} \cdot b_i \cdot N z_i\right) \leq \mathrm{frc}\left(\sum_{i=1}^{m} \frac{1}{N} \cdot b_i a_i\right) + \sum_{i=1}^{m} \frac{1}{N} \cdot b_i = \sum_{i=1}^{m} \frac{1}{N} \cdot b_i \leq \frac{m}{N}.$$

From now on, we will use $\mathcal{A}'$ instead of $\mathcal{A}$.

According to Theorem 3.1 it is enough to construct a distinguisher $\mathcal{B}$ between U and $\mathcal{T}_{n,\sqrt{n}\cdot\gamma(n)}$. The distinguisher $\mathcal{B}$ works by calling the routine $\mathcal{C}$ described below n times with each value $\tilde{h} = (1+n^{-c_{\tilde{\mathfrak{h}}}})^i$, $i \in [\log_{1+n^{-c_{\tilde{\mathfrak{h}}}}} N]$. The constant $c_{\tilde{\mathfrak{h}}}$ will be specified later. If there exists an $\tilde{h}$ for which all n calls to $\mathcal{C}$ accept, $\mathcal{B}$ accepts. Otherwise, for any $\tilde{h}$ there exists one call where $\mathcal{C}$ rejects and $\mathcal{B}$ rejects.

The routine $\mathcal{C}(\tilde{h})$ samples m values $x_1, \ldots, x_m$ from the given distribution, which we denote by R. It also chooses m values $y_1, \ldots, y_m$ uniformly in $[0, 1/\tilde{h})$. Let $z_i = x_i - y_i \bmod 1$. We call $\mathcal{A}'$ with $z_1, \ldots, z_m$. If $\mathcal{A}'$ fails we repeat the process again (choosing x_i, y_i and calling $\mathcal{A}'$). If after $n^{c_{\mathfrak{a}}+1}$ calls $\mathcal{A}'$ still fails, $\mathcal{C}$ accepts. Otherwise, we have a vector $b \in \mathbb{Z}^m$ such that $\|b\| \leq \sqrt{m}$ and $\mathrm{frc}\left(\sum_{i=1}^{m} b_i z_i\right) \leq \frac{m}{N}$. The routine $\mathcal{C}(\tilde{h})$ accepts if $\mathrm{frc}\left(\sum_{i=1}^{m} b_i \tilde{h} y_i\right) < \frac{1}{4}$ and rejects otherwise. We summarize the routine $\mathcal{C}(\tilde{h})$ in the following.

1. For each $i \in [m]$, choose x_i according to R and choose y_i uniformly from $[0, 1/\tilde{h})$
2. For each $i \in [m]$, set $z_i = x_i - y_i \bmod 1$
3. Call $\mathcal{A}'$ with $z_1, \ldots, z_m$
4. If $\mathcal{A}'$ fails, go back to step (1) (and after $n^{c_{\mathfrak{a}}+1}$ times accept). Otherwise, let $b_1, \ldots, b_m$ be its answer.
5. Accept iff $\mathrm{frc}\left(\sum_{i=1}^{m} b_i \tilde{h} y_i\right) < \frac{1}{4}$

We now make an important observation. For each $i \in [n]$, the first two steps of $\mathcal{C}(\tilde{h})$ essentially sample the triple (x_i, y_i, z_i) from some distribution D on (x, y, z). In this distribution, x and y are independent: x is chosen from R and y is chosen uniformly from $[0, 1/\tilde{h})$. Moreover, z is a function of x and y. Let us describe an equivalent way to obtain a sample (x_i, y_i, z_i) from D. First choose z_i from D_z, the marginal distribution of z. Then choose x_i and y_i from $D|z = z_i$, the conditional distribution on x and y given that z is z_i. Clearly, a triple (x_i, y_i, z_i) chosen this way has exactly the same distribution D. Hence, if we replace the first two steps in $\mathcal{C}(\tilde{h})$ with the following, we obtain a procedure that has the same acceptance probability:

1. For each $i \in [m]$, choose z_i from D_z
2. For each $i \in [m]$, choose x_i and y_i according to $D|z = z_i$

This modified procedure is no longer implementable as we do not know how to sample from $D|z = z_i$ (the distribution R is unknown). This, however, raises no difficulties since we only use the modified procedure for the analysis. Notice now that step (2) is independent of steps (3) and (4). Hence, the above can be equivalently written as

1. For each $i \in [m]$, choose z_i from D_z
2. Call $\mathcal{A}'$ with $z_1, \ldots, z_m$
3. If $\mathcal{A}'$ fails, go back to step (1) (and after $n^{c_{\mathsf{a}}+1}$ times accept). Otherwise, let $b_1, \ldots, b_m$ be its answer.
4. For each $i \in [m]$, choose x_i and y_i according to $D|z = z_i$
5. Accept iff $\mathrm{frc}\left(\sum_{i=1}^m b_i \tilde{h} y_i\right) < \frac{1}{4}$

We use this equivalent procedure in our analysis of $\mathcal{C}(\tilde{h})$.

We first show that if R is the uniform distribution then for any $\tilde{h}$, $\mathcal{C}(\tilde{h})$ accepts with probability roughly $\frac{1}{2}$. From this it will follow that the probability that n calls to $\mathcal{C}(\tilde{h})$ accept is exponentially small, i.e., $\mathcal{B}$ rejects with probability exponentially close to 1. The distribution D_z is clearly uniform. Hence, each z_i is uniform in $[0, 1)$ and according to our assumption, $\mathcal{A}'$ succeeds with probability at least $n^{-c_{\mathsf{a}}}$. The probability that $n^{c_{\mathsf{a}}+1}$ calls fail is at most $(1 - n^{-c_{\mathsf{a}}})^{n^{c_{\mathsf{a}}+1}} < \exp(-n)$, which is exponentially small. So now assume that $\mathcal{A}'$ accepts in one of the calls and fix the values of $z_1, \ldots, z_m$ and $b_1, \ldots, b_m$. Then, in step (4), we choose each y_i from the conditional distribution which in this case is uniform in $[0, 1/\tilde{h})$. The distribution of $\tilde{h} y_i$ is uniform in $[0, 1)$. Since $b_1, \ldots, b_m$ are not all zero, $\mathrm{frc}\left(\sum_{i=1}^m b_i \tilde{h} y_i\right)$ is distributed uniformly in $[0, \frac{1}{2})$. The probability that $\mathrm{frc}\left(\sum_{i=1}^m b_i \tilde{h} y_i\right) < \frac{1}{4}$ is therefore $\frac{1}{2}$, as required.

Now consider the case that R is $T_{h,\beta}$ where $\beta \leq \frac{4}{(\gamma(n))^2}$. We claim that when $\tilde{h}$ is the smallest such that $\tilde{h} \geq h$, $\mathcal{C}(\tilde{h})$ rejects with probability at most $\tilde{c} m n^{4-c_{\tilde{\mathsf{h}}}}$. Therefore, the probability that $\mathcal{B}$ sees a rejection after n calls is at most $\tilde{c} m n^{4-c_{\tilde{\mathsf{h}}}+1}$ and it therefore accepts with probability close to 1 if we choose a large enough $c_{\tilde{\mathsf{h}}}$. Notice that such an $\tilde{h}$ satisfies $h \leq \tilde{h} < (1 + n^{-c_{\tilde{\mathsf{h}}}})h$.

In order for $\mathcal{C}(\tilde{h})$ to reject, we must get to step (4). Hence, it is enough to show that for any $z_1, \ldots, z_m$ and any $b_1, \ldots, b_m$ such that $\mathrm{frc}\left(\sum_{i=1}^m b_i z_i\right) \leq \frac{m}{N}$, steps (4) and (5) reject with probability at most $\tilde{c} m n^{4-c_{\tilde{\mathsf{h}}}}$. The conditional distribution from which we choose y_i is given by:

$$\frac{1}{\int_{z_i}^{z_i+1/\tilde{h}} T_{h,\beta}(x)dx} T_{h,\beta}(z_i + r) \quad \forall r \in \left[0, 1/\tilde{h}\right).$$

Hence the density function of the distribution of $\tilde{h} \cdot y_i$ is exactly $S_{\tilde{h},h,\beta,z_i}$. According to Claim 6.4 the statistical distance between $S_{\tilde{h},h,\beta,z_i}$ and S'_{h,β,z_i} is at most $\frac{\tilde{c}}{\beta} n^{-c_{\tilde{\mathsf{h}}}} \leq \tilde{c} n^{4-c_{\tilde{\mathsf{h}}}}$. Let $\xi_1, \ldots, \xi_m$ be m random variables chosen independently according to Q_β. Notice that the distribution of the random variable $\xi_i - h \cdot z_i \bmod 1$ is exactly S'_{h,β,z_i}. Hence, according to Claim 6.1, the statistical distance between the joint distributions $(\tilde{h} \cdot y_1, \ldots, \tilde{h} \cdot y_m)$ and $(\xi_1 - h \cdot z_1 \bmod 1, \ldots, \xi_m - h \cdot z_m \bmod 1)$ is at most $\tilde{c} m \cdot n^{4-c_{\tilde{\mathsf{h}}}}$. Now,

$$\sum_{i=1}^m b_i(\xi_i - h \cdot z_i) \bmod 1 = \sum_{i=1}^m b_i \xi_i - \sum_{i=1}^m b_i \cdot h \cdot z_i \bmod 1.$$

According to Claim 6.2, $\sum_{i=1}^m b_i \xi_i$ has a normal distribution with mean 0 and standard deviation

$$\|b\| \cdot \sqrt{\frac{\beta}{2\pi}} \leq \sqrt{\frac{m\beta}{2\pi}} \leq \sqrt{\frac{2m}{\pi(\gamma(n))^2}} = o\left(\frac{1}{\sqrt{\log n}}\right).$$

Therefore, according to Claim 5.1, the probability that $\mathrm{frc}\left(\sum_{i=1}^{m} b_i \xi_i\right) > \frac{1}{8}$ is negligible. Now,

$$\mathrm{frc}\left(\sum_{i=1}^{m} b_i \cdot h \cdot z_i\right) \leq h \cdot \mathrm{frc}\left(\sum_{i=1}^{m} b_i \cdot z_i\right) \leq \frac{hm}{N}.$$

Therefore, except with negligible probability,

$$\mathrm{frc}\left(\sum_{i=1}^{m} b_i(\xi_i - h \cdot z_i)\right) \leq \frac{1}{8} + \frac{hm}{N} < \frac{1}{4}$$

where we used the fact that $h \leq 2^{4n^2} = \sqrt{N}$. This implies that the probability that $\mathrm{frc}\left(\sum_{i=1}^{m} b_i \tilde{h} y_i\right) < \frac{1}{4}$ is at most $\tilde{c}m \cdot n^{4-c_{\hbar}}$ plus some negligible amount. ■

7 Quantum Computation

In this section we show a result related to a problem in quantum computation known as the dihedral hidden subgroup problem. Let us start by describing the hidden subgroup problem (HSP), a central problem in quantum computation. Here, we are given a black box that computes a function on elements of a group G. The function is known to be constant on left cosets of a subgroup $H \leqslant G$ and distinct on each coset. Our goal is to find H. Interestingly, almost all known quantum algorithms that run super-polynomially faster than classical (i.e., non-quantum) algorithms solve special cases of the HSP on *Abelian* groups (e.g., [27]). Also, it is known that solving the HSP on the *symmetric* group leads to a quantum solution to graph isomorphism [17]. This motivated research into possible extensions of the HSP to noncommutative groups (see, e.g., [13, 14, 26, 8]).

In this section we consider the HSP on the dihedral group. The dihedral group of order $2N$ is the group of symmetries of an N-sided regular polygon. It is isomorphic to the abstract group generated by the element ρ of order n and the element τ of order 2 subject to the relation $\rho\tau = \tau\rho^{-1}$. No efficient solution to the HSP on the dihedral group is known. The best known algorithm is due to Kuperberg [18] and runs in subexponential time $2^{O(\sqrt{\log N})}$ (the size of the input is $O(\log N)$).

A different approach was taken by Ettinger and Høyer [7]. They reduced the dihedral HSP to the classical problem of finding an integer k given access to the distribution Z_k on $\{0, 1, \ldots, N-1\}$ defined by

$$\Pr(Z_k = z) = 2/N \cdot \cos^2(\pi k z/N), \quad z = 0, 1, \ldots, N-1.$$

They presented an exponential time classical algorithm that solves this problem using only a polynomial number of samples from Z_k. Hence, a polynomial number of samples contains enough information to find k. The question of whether there exists an *efficient* algorithm remained open. In this section we show that such an efficient algorithm is unlikely to exist: its existence implies a (classical) solution to n^c-uSVP for some c.

Another related result is that of Regev [25]. He showed that under certain conditions (namely, that of coset sampling), an efficient solution to the dihedral HSP implies an efficient *quantum* algorithm for uSVP. Finding such an algorithm is a very important open question in quantum computation. Hence, another way to interpret the result of this section is the following: a solution to the classical problem of Ettinger and Høyer would not only lead to a *quantum* algorithm for uSVP but also to a *classical* algorithm and should therefore be considered unlikely.

We start by extending Theorem 3.1 to more general periodic distributions. Let D be an efficiently samplable distribution on $[0,1)$ such that its density function satisfies $D(r) \leq c_{\mathsf{D}}$ and $|D(r) - D(r+\epsilon \bmod 1)| \leq c_{\mathsf{D}}\epsilon$ for all $r, \epsilon \in [0,1)$ for some constant c_{D}. Essentially, this means that D is smooth enough. For $h \in \mathbb{N}$, define

$$T_h^D(r) = D(rh \bmod 1)$$

to be the distribution on $[0,1)$ given by h periods of D. Moreover, define

$$\mathcal{T}_n^D = \left\{ T_h^D \;\middle|\; h \in \mathbb{N},\ h \leq 2^{4n^2} \right\}$$

where n is the size parameter of the problem.

Lemma 7.1 *For D as above, if there exists a distinguisher between U and $\mathcal{T}_n^D$ then there exists a solution to n^c-*uSVP *for some $c > 0$.*

Proof: The proof is based on a reduction from distinguishing between U and $\mathcal{T}_{n,n^c}$ to distinguishing between U and $\mathcal{T}_h^D$. The idea is the following. Assume we are given either U or $T_{h,\beta}$ for some unknown h and small enough β. Assume we have a good estimate $\tilde{h}$ on h (we can obtain it by trying polynomially many possibilities for $\tilde{h}$). The reduction works by sampling a value from the unknown distribution and then adding to it a sample from $D/\tilde{h}$. Then, on one hand, the distribution $U + D/\tilde{h}$ is exactly the uniform distribution U. On the other hand, $T_{h,\beta} + D/\tilde{h}$ is shown to be close to T_h^D. Therefore, if one can distinguish between U and T_h^D then one could also distinguish between U and $T_{h,\beta}$.

The proof that $T_{h,\beta} + D/\tilde{h}$ is close to T_h^D is rather technical. First, we show that $T_{h,\beta} + D/\tilde{h}$ is close to $T_{h,\beta} + D/h$. We actually show the stronger fact that $D/\tilde{h}$ is close to D/h. This holds since D is smooth and $\tilde{h}$ is close to h. Then we show that $T_{h,\beta} + D/h$ is close to T_h^D. Intuitively, the limit of $T_{h,\beta} + D/h$ as β goes to 0 is exactly T_h^D. Our proof here shows that since D is smooth, the noise added by a non-zero β does not change the distribution much.

Let us now describe the proof in more detail. Assume $\mathcal{A}$ is a distinguisher between U and $\mathcal{T}_n^D$ and assume that it uses $n^{c_{\mathsf{A}}}$ samples of the given distribution for some $c_{\mathsf{A}} > 0$. Let p_u denote the acceptance probability of $\mathcal{A}$ on inputs from distribution U and for $h \in \mathbb{N}$ let p_h denote its acceptance probability on inputs from T_h^D. According to our hypothesis $|p_u - p_h| \geq n^{-c_{\mathsf{d}}}$ for all $h \in [2^{4n^2}]$ for some constant $c_{\mathsf{d}} > 0$.

We construct a distinguisher $\mathcal{B}$ between U and $\mathcal{T}_{n,n^c}$ for some large enough $c > 0$. The lemma then follows from Theorem 3.1. Let R denote the given distribution. First, $\mathcal{B}$ chooses a value $\tilde{h}$ uniformly from the set $\{1, 1+\mu, (1+\mu)^2, \ldots, 2^{4n^2}\}$ where $\mu = n^{-c_\mu}$ for some constant $c_\mu > 0$ to be chosen later. Then, define the distribution R' as

$$R' = R + \frac{D}{\tilde{h}} \bmod 1,$$

i.e., a sample from R' is given by $x + r/\tilde{h} \bmod 1$ where x is chosen from R and r is chosen from D. It then estimates the acceptance probability of $\mathcal{A}$ using sequences of samples from R' each of length $n^{c_{\mathsf{A}}}$. According to the Chernoff bound, using a polynomial number of sequences, we can obtain an estimate that with probability exponentially close to 1 is within $\frac{1}{4n^{c_{\mathsf{d}}}}$ of the actual acceptance

probability. If the estimate differs from p_u by more than $\frac{1}{2n^{c_\mathsf{d}}}$, $\mathcal{B}$ accepts; otherwise, it rejects. This completes the description of $\mathcal{B}$.

When R is the uniform distribution then R' is also uniform. Therefore, with probability exponentially close to 1, $\mathcal{B}$'s estimate is within $\frac{1}{4n^{c_\mathsf{d}}}$ of p_u and $\mathcal{B}$ rejects. Hence, it is remains to show that $\mathcal{B}$ accepts with some non-negligible probability when R is $T_{h,\beta}$ where $h \leq 2^{4n^2}$ and $\beta \leq n^{-c_\beta}$ for some large enough c_β.

Consider the event in which $h \leq \tilde{h} < (1+\mu)h$. Notice that it happens with non-negligible probability since $\tilde{h}$ is chosen from a set of size polynomial in n. The following technical claim will complete the proof by showing that the statistical distance between R' and T_h^D is smaller than $n^{-c_\mathsf{A}-c_\mathsf{d}}/4$. Using Claim 6.1, it follows that the statistical distance between a sequence of n^{c_A} elements of R' and a sequence of n^{c_A} elements of T_h^D is at most $n^{-c_\mathsf{d}}/4$. Finally, using Equation 1 in Section 2, this implies that $\mathcal{A}$'s success probability on sequences from R' is within $n^{-c_\mathsf{d}}/4$ from p_h and since $|p_u - p_h| \geq n^{-c_\mathsf{d}}$, $\mathcal{B}$ accepts.

Claim 7.2 *For $\tilde{h}$ as above and for large enough c_β and c_μ, the statistical distance $\Delta(R', T_h^D) \leq n^{-c_\mathsf{A}-c_\mathsf{d}}/4$.*

Proof: Consider the distribution R'' given by

$$R'' = T_{h,\beta} + \frac{D}{h}.$$

The distribution R'' can be seen as a random function of the distribution D: given a value $r \in D$ sample a value x from $T_{h,\beta}$ and output $x + r/h$. Notice that R' is given by applying the same function to the distribution $(h/\tilde{h})D$. Hence, using Equation 1,

$$\begin{aligned}
\Delta(R', R'') \leq \Delta\left(D, \frac{h}{\tilde{h}}D\right) &= \int_0^{h/\tilde{h}} \left|D(r) - D\left(\tilde{h}r/h\right)\right| dr + \int_{h/\tilde{h}}^1 D(r)dr \\
&\leq c_\mathsf{D}\left(1 - \frac{h}{\tilde{h}}\right) + \left(1 - \frac{h}{\tilde{h}}\right) c_\mathsf{D} \\
&\leq 2c_\mathsf{D}\mu = 2c_\mathsf{D} n^{-c_\mu}. \qquad (4)
\end{aligned}$$

We next bound the statistical distance between T_h^D and R''. Let X be a random variable distributed uniformly over $\{0, \frac{1}{h}, \ldots, \frac{h-1}{h}\}$. Then, it can be seen that

$$T_h^D = X + \frac{D}{h} \bmod 1.$$

Now, let Y be another random variable distributed normally with mean 0 and variance $\frac{\beta}{2\pi}$. Then, as in Claim 4.1, $T_{h,\beta} = X + Y/h \bmod 1$ and hence,

$$R'' = X + \frac{Y}{h} + \frac{D}{h} \bmod 1.$$

Therefore, T_h^D can be seen as a random function applied to a sample from $\frac{D}{h}$ while R'' can be seen as the same function applied to a sample from $\frac{Y}{h} + \frac{D}{h}$. From Equation 1 it follows that

$$\Delta(T_h^D, R'') \leq \Delta\left(\frac{1}{h}D,\ \frac{1}{h}(D+Y)\right) = \Delta(D, D+Y). \qquad (5)$$

Let $\widehat{Y}$ be the restriction of a normal distribution with mean 0 and variance $\frac{\beta}{2\pi}$ to the interval $[-n\sqrt{\beta}, n\sqrt{\beta}]$. More formally,

$$\widehat{Y}(r) = \frac{Y(r)}{\int_{-n\sqrt{\beta}}^{n\sqrt{\beta}} Y(r)dr}$$

for $r \in [-n\sqrt{\beta}, n\sqrt{\beta}]$ and $\widehat{Y}(r) = 0$ elsewhere. From Claim 5.1 it follows that the distribution of Y is very close to that of $\widehat{Y}$:

$$\Delta(Y, \widehat{Y}) \leq \sqrt{\frac{2}{\pi}} \cdot \frac{1}{n\sqrt{2\pi}} \cdot \exp\left(-\pi n^2\right) = 2^{-\Omega(n^2)}. \tag{6}$$

Now, using the fact that $\widehat{Y}$ always gets values of small absolute value,

$$\begin{aligned}
\left|D(r) - (D + \widehat{Y})(r)\right| &= \left|D(r) - \int_{-n\sqrt{\beta}}^{n\sqrt{\beta}} D(r-x)\widehat{Y}(x)dx\right| \\
&= \left|\int_{-n\sqrt{\beta}}^{n\sqrt{\beta}} (D(r) - D(r-x))\widehat{Y}(x)dx\right| \\
&\leq \int_{-n\sqrt{\beta}}^{n\sqrt{\beta}} |D(r) - D(r-x)|\, \widehat{Y}(x)dx \\
&\leq c_{\mathsf{D}} n\sqrt{\beta} \int_{-n\sqrt{\beta}}^{n\sqrt{\beta}} \widehat{Y}(x)dx \\
&= c_{\mathsf{D}} n\sqrt{\beta}.
\end{aligned}$$

where we used the triangle inequality and the fact that $\int_{-n\sqrt{\beta}}^{n\sqrt{\beta}} \widehat{Y}(x)dx = 1$. Since both $D(r)$ and $(D + \widehat{Y})(r)$ are zero for $r < -n\sqrt{\beta}$ and for $r > 1 + n\sqrt{\beta}$,

$$\begin{aligned}
\Delta(D, D + \widehat{Y}) &= \int_{-n\sqrt{\beta}}^{1+n\sqrt{\beta}} \left|D(r) - (D + \widehat{Y})(r)\right| dr \\
&\leq (1 + 2n\sqrt{\beta}) \cdot c_{\mathsf{D}} n\sqrt{\beta} \\
&\leq (1 + 2n^{1-c_\beta/2}) \cdot c_{\mathsf{D}} n^{1-c_\beta/2} \leq 2c_{\mathsf{D}} n^{1-c_\beta/2}
\end{aligned} \tag{7}$$

for large enough c_β. Finally, combining Equations 4, 5, 6, 7 and using the triangle inequality, we obtain

$$\Delta(R', T_h^D) \leq 2c_{\mathsf{D}} n^{-c_\mu} + 2^{-\Omega(n^2)} + 2c_{\mathsf{D}} n^{1-c_\beta/2} \leq n^{-c_{\mathsf{A}} - c_{\mathsf{d}}}/4$$

for large enough c_β and c_μ. ■

This completes the proof of Lemma 7.1. ■

We can now prove the main theorem of this section.

Theorem 7.3 *For $k \in \mathbb{N}$, $k < N$, define the distribution Z_k on $\{0, 1, \ldots, N-1\}$ by*

$$\Pr(Z_k = z) = 2/N \cdot \cos^2(\pi k z/N), \quad z = 0, 1, \ldots, N-1.$$

Assume there exists an algorithm $\mathcal{A}$ that given a polynomial (in $\log N$) number of samples from Z_k, returns k with probability exponentially close to 1. Then, there exists a solution to n^c-uSVP for some c.

We remark that it is possible to relax the assumptions of the theorem. It is enough if the algorithm returns k with non-negligible probability. Also, by using Theorem 4.5 instead of Theorem 3.1, one can show that it is enough if the algorithm finds k only for some non-negligible fraction of all possible k's.

Proof: Let D be the distribution on $[0,1)$ given by $D(r) = 2\cos^2(\pi r)$. An easy calculation shows that the absolute value of its derivative is at most 4π. Therefore, it satisfies the conditions stated before Lemma 7.1 with $c_{\mathsf{D}} = 4\pi$. Using Lemma 7.1, it is enough to show how to distinguish between U and $\mathcal{T}_n^D$. The idea is to notice that Z_h is essentially a discretization of T_h^D. Therefore, algorithm $\mathcal{A}$ can find the value h given T_h^D. From this, we construct a distinguisher between U and $\mathcal{T}_n^D$ by simply checking whether a sample from the unknown distribution is close to $1/h$. For T_h^D it is more likely to be close to $1/h$ than for U.

Given an unknown distribution R, let R' be the distribution given by $\lfloor N \cdot R \rfloor$ where N is chosen to be large enough, say, 2^{8n^2}. We call $\mathcal{A}$ with enough samples from R' and obtain a value k. Finally, we take one sample r from R and accept if $\mathrm{frc}\,(rk) < 1/4$ and reject otherwise.

First, consider the case where R is the uniform distribution. Then no matter which value of k we obtain, the probability that $\mathrm{frc}\,(rk) < 1/4$ is exactly $1/2$. Now consider the case where R is T_h^D for some $h \leq 2^{4n^2}$. For any $r = 0, \ldots, N-1$, the probability that $R' = r$ is given by

$$\int_{r/N}^{(r+1)/N} D(hx \bmod 1)dx = \int_{r/N}^{(r+1)/N} 2\cos^2(\pi hx)dx.$$

From the bound on the derivative of D mentioned above, we obtain that the distance of this integral from $2/N \cdot \cos^2(\pi hr/N)$ is at most $4\pi^2 h/N^2$. Therefore, the statistical distance between R' and Z_h is

$$\Delta(Z_h, R') \leq \frac{N}{2} \cdot 4\pi^2 h/N^2 = 2^{-\Omega(n^2)}.$$

Since the number of samples given to $\mathcal{A}$ is only polynomial in n, its input is still within statistical distance $2^{-\Omega(n^2)}$ of Z_h and it therefore outputs h with probability exponentially close to 1. Then, the probability that $\mathrm{frc}\,(rk) < 1/4$ is given by

$$\int_{-1/4}^{1/4} 2\cos^2(\pi r)dr = \frac{1}{2} + \frac{1}{\pi}.$$

■

8 Acknowledgements

I thank Irit Dinur for suggesting that I look at cryptographic constructions and Daniele Micciancio for many helpful comments on an earlier draft of this paper. I also thank the anonymous referee for many excellent suggestions.

References

[1] M. Ajtai. Generating hard instances of lattice problems. In *ECCCTR: Electronic Colloquium on Computational Complexity, technical reports*, 1996.

[2] M. Ajtai and C. Dwork. A public-key cryptosystem with worst-case/average-case equivalence. In *Proc. 29th ACM Symp. on Theory of Computing*, pages 284–293, 1997.

[3] W. Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(4):625–635, 1993.

[4] J.-Y. Cai. Applications of a new transference theorem to Ajtai's connection factor. In *14th IEEE Conference on Computational Complexity*, pages 205–214. IEEE Computer Soc., Los Alamitos, CA, 1999.

[5] J.-Y. Cai and A. P. Nerurkar. An improved worst-case to average-case connection for lattice problems (extended abstract). In *Proc. 38th IEEE Symp. on Foundations of Computer Science*, pages 468–477, 1997.

[6] W. Ebeling. *Lattices and codes*. Advanced Lectures in Mathematics. Friedr. Vieweg & Sohn, Braunschweig, revised edition, 2002. A course partially based on lectures by F. Hirzebruch.

[7] M. Ettinger and P. Høyer. On quantum algorithms for noncommutative hidden subgroups. *Adv. in Appl. Math.*, 25(3):239–251, 2000.

[8] K. Friedl, G. Ivanyos, F. Magniez, M. Santha, and P. Sen. Hidden translation and orbit coset in quantum computing. In *Proc. 35th ACM Symp. on Theory of Computing*, pages 1–9, 2003.

[9] O. Goldreich, S. Goldwasser, and S. Halevi. Collision-free hashing from lattice problems. In *ECCCTR: Electronic Colloquium on Computational Complexity, technical reports*, 1996.

[10] O. Goldreich, S. Goldwasser, and S. Halevi. Eliminating decryption errors in the Ajtai-Dwork cryptosystem. In *Advances in cryptology*, volume 1294 of *Lecture Notes in Comput. Sci.*, pages 105–111. Springer, 1997.

[11] O. Goldreich, S. Goldwasser, and S. Halevi. Public-key cryptosystems from lattice reduction problems. In *Advances in cryptology*, volume 1294 of *Lecture Notes in Comput. Sci.*, pages 112–131. Springer, 1997.

[12] O. Goldreich, D. Micciancio, S. Safra, and J.-P. Seifert. Approximating shortest lattice vectors is not harder than approximating closest lattice vectors. *Inform. Process. Lett.*, 71(2):55–61, 1999.

[13] M. Grigni, L. J. Schulman, M. Vazirani, and U. V. Vazirani. Quantum mechanical algorithms for the nonabelian hidden subgroup problem. In *Proc. 33rd ACM Symp. on Theory of Computing*, pages 68–74, 2001.

[14] S. Hallgren, A. Russell, and A. Ta-Shma. Normal subgroup reconstruction and quantum computation using group representations. In *Proc. 32nd ACM Symp. on Theory of Computing*, pages 627–635, 2000.

[15] J. Hoffstein, J. Pipher, and J. H. Silverman. NTRU: a ring-based public key cryptosystem. In *Algorithmic number theory*, volume 1423 of *Lecture Notes in Comput. Sci.*, pages 267–288. Springer, 1998.

[16] R. Impagliazzo and M. Naor. Efficient cryptographic schemes provably as secure as subset sum. *J. Cryptology*, 9(4):199–216, 1996.

[17] K. Johannes, S. Uwe, and T. Jacobo. *The graph isomorphism problem: its structural complexity*. Birkhäuser Boston Inc., 1993.

[18] G. Kuperberg. A subexponential-time quantum algorithm for the dihedral hidden subgroup problem. In *quant-ph/0302112, http://xxx.lanl.gov*, 2003.

[19] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász. Factoring polynomials with rational coefficients. *Math. Ann.*, 261(4):515–534, 1982.

[20] D. Micciancio. Improving lattice based cryptosystems using the hermite normal form. In *Cryptography and Lattices Conference (CaLC)*, volume 2146 of *Lecture Notes in Computer Science*, pages 126–145, Providence, Rhode Island, Mar. 2001. Springer-Verlag.

[21] D. Micciancio. Generalized compact knapsacks, cyclic lattices, and efficient one-way functions from worst-case complexity assumptions. In *Proc. 43rd IEEE Symp. on Foundations of Computer Science*, Vancouver, Canada, Nov. 2002.

[22] D. Micciancio. Improved cryptographic hash functions with worst-case/average-case connection. In *Proc. 34th ACM Symp. on Theory of Computing*, pages 609–618, Montreal, Canada, 2002.

[23] D. Micciancio and S. Goldwasser. *Complexity of Lattice Problems: a cryptographic perspective*, volume 671 of *The Kluwer International Series in Engineering and Computer Science*. Kluwer Academic Publishers, Boston, Massachusetts, Mar. 2002.

[24] D. Micciancio and O. Regev. Worst-case to average-case reductions based on Gaussian measures. Manuscript, 2004.

[25] O. Regev. Quantum computation and lattice problems. In *Proc. 43rd IEEE Symp. on Foundations of Computer Science*, Vancouver, Canada, Nov. 2002.

[26] M. Rötteler and T. Beth. Polynomial-time solution to the hidden subgroup problem for a class of non-abelian groups. In *quant-ph/9812070, http://xxx.lanl.gov*, 1998.

[27] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[28] D. Štefankovič. Fourier transforms in computer science. Master's Thesis, University of Chicago, Department of Computer Science, TR-2002-03.