Worst-case to Average-case Reductions based on Gaussian Measures*

Daniele Micciancio † Oded Regev ‡

December 14, 2005

Abstract

We show that finding small solutions to random modular linear equations is at least as hard as approximating several lattice problems in the worst case within a factor almost linear in the dimension of the lattice. The lattice problems we consider are the shortest vector problem, the shortest independent vectors problem, the covering radius problem, and the guaranteed distance decoding problem (a variant of the well known closest vector problem). The approximation factor we obtain is $n \log^{O(1)} n$ for all four problems. This greatly improves on all previous work on the subject starting from Ajtai's seminal paper (STOC, 1996), up to the strongest previously known results by Micciancio (SIAM J. on Computing, 2004). Our results also bring us closer to the limit where the problems are no longer known to be in NP intersect coNP.

Our main tools are Gaussian measures on lattices and the high-dimensional Fourier transform. We start by defining a new lattice parameter which determines the amount of Gaussian noise that one has to add to a lattice in order to get close to a uniform distribution. In addition to yielding quantitatively much stronger results, the use of this parameter allows us to simplify many of the complications in previous work.

Our technical contributions are two-fold. First, we show tight connections between this new parameter and existing lattice parameters. One such important connection is between this parameter and the length of the shortest set of linearly independent vectors. Second, we prove that the distribution that one obtains after adding Gaussian noise to the lattice has the following interesting property: the distribution of the noise vector when conditioning on the final value behaves in many respects like the original Gaussian noise vector. In particular, its moments remain essentially unchanged.

1 Introduction

Lattice problems have received considerable attention as a potential source of computational hardness to be used in cryptography, after a breakthrough result of Ajtai [2] showing that if certain lattice problems are computationally hard to solve in the *worst case*, then *average-case* one-way functions (a fundamental cryptographic primitive) exist. Ajtai's one-way function is essentially the generalized subset sum function over the additive group of n-dimensional vectors modulo q: functions are described by m group elements $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{Z}_q^n$, and the associated function maps bit-string $x_1, \ldots, x_m \in \{0,1\}$ to $f_{\mathbf{A}}(x_1, \ldots, x_m) = \sum_i \mathbf{a}_i x_i$. The main worst-case lattice problem considered by Ajtai is that of finding a set of n linearly independent lattice vectors in an arbitrary lattice of length within a polynomial (in n) factor from the

*A preliminary version of this paper appears in the Proceedings of the 45th Annual Symposium on Foundations of Computer Science - FOCS 2004. Rome, Italy. Oct. 2004. IEEE, pp. 372-381.

†UC San Diego, La Jolla, CA 92093. E-Mail: daniele@cs.ucsd.edu. Research supported in part by NSF Career Award CCR-0093029 and a Sloan Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

‡Department of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Work supported by the Israel Science Foundation, by the Binational Science Foundation, by an Alon Fellowship, by the Army Research Office grant DAAD19-03-1-0082, and by NSF grant CCR-9987845.

shortest such set.[1] This problem in turn, is related, using standard techniques, to various other lattice problems, like approximating the length of the shortest nonzero lattice vector in the worst case, within factors polynomial in n.

No polynomial time algorithm is known to solve any of these worst-case problems, so it is reasonable to conjecture that the problems are hard for any polynomial approximation factor. Still, since the problems get easier and easier as the factor increases, it is theoretically interesting and practically important to determine the smallest factors for which the hardness of approximating these lattice problems in the worst case implies that the function $f_{\mathbf{A}}$ is one-way on the average. The factors implicit in Ajtai's proof are rather large: [9] estimates all these factors to be larger than n^8. In subsequent developments the factors have been improved, leading to the currently best known results of Micciancio [21]: the subset-sum function $f_{\mathbf{A}}$ is hard to invert (in fact, even collision resistant) on the average, provided any of the following problems is hard in the worst case:

- Computing a set of n linearly independent lattice vectors in an n-dimensional lattice of length within a factor[2] $\tilde{O}(n^{2.5})$ from the shortest such set;
- Approximating the length of the shortest nonzero vector in an n-dimensional lattice within a factor $\tilde{O}(n^3)$;
- Approximating the covering radius of an n-dimensional lattice within a factor $\tilde{O}(n^{2.5})$;
- Finding a lattice vector within distance at most $\tilde{O}(n^{2.5})$ times the covering radius from any given target point.

Micciancio [21] also showed that the above factors can be further reduced by $\sqrt{n}$ if certain sequences of "almost perfect" easily decodable lattices exist, and conjectured a reduction achieving factors as low as $\tilde{O}(n^{1.5})$. In a recent work of Regev [24], a similar result was shown based on worst-case instances of a problem known as the $\tilde{O}(n^{1.5})$-*unique* shortest vector problem. This problem is a special case of the shortest vector problem in which the lattices have a special structure (namely, their shortest vector is *unique*, in the sense that the next shortest linearly independent vector is longer than the shortest nonzero vector by $\tilde{O}(n^{1.5})$). Although the connection factor $\tilde{O}(n^{1.5})$ is better than the factors of [21], a major drawback of the reduction in [24] is that the unique shortest vector problem is potentially easier to solve than the shortest vector problem; in fact, it is not even known to be NP-hard for small constant approximation factors.[3]

Our results: We substantially improve all of the above results and prove that the subset-sum function $f_{\mathbf{A}}$ is hard to invert (and collision resistant) on the average provided any of the following problems is hard in the worst case:

- Computing a set of n linearly independent lattice vectors in an n-dimensional lattice of length within a factor $\tilde{O}(n)$ from the shortest such set;
- Approximating the length of the shortest nonzero vector in an n-dimensional lattice within a factor $\tilde{O}(n)$;
- Approximating the covering radius of an n-dimensional lattice within a factor $\tilde{O}(n)$;
- Finding a lattice vector within distance at most $\tilde{O}(n)$ times the covering radius from any given target point.

[1] The length of a finite set of vectors is defined as the length of the longest vector in the set. The problem can be defined with respect to any norm, but the Euclidean norm is the most common.

[2] A function $g(n)$ is in $\tilde{O}(f(n))$ if there exist constants $a, c \geq 0$ such that $g(n) \leq af(n)\log^c f(n)$ for all sufficiently large n.

[3] The main result of [24] is a lattice based *encryption scheme*. This encryption scheme, as the one in the original work of Ajtai and Dwork [3], is also based on the unique shortest vector problem. Constructing an *encryption scheme* based on other lattice problems such as the shortest vector problem is a major open problem.

In other words, the connection factor is $\tilde{O}(n)$ for all four lattice problems. This proves Micciancio's conjecture [21], and in fact provides even better connection factors. Our results are significant for two reasons:

- On the technical side, we present a new approach to worst-case to average-case reductions for lattice problems, based on the use of Gaussian measures. The results in [21] were making an essentially optimal use of previous reduction techniques; the results presented in this paper require some new techniques that might be of independent interest. Another important technical contribution of this paper is the study of Gaussian distributions on lattices. These issues are discussed in Subsection 1.1.
- On the theoretical side, our improvements bring us closer to factors for which lattice problems are not known to be in NP $\cap$ coNP. This is discussed in Subsection 1.2.

1.1 Our techniques

The reduction: In this paper, as in previous work [2, 12, 9, 21, 24], we consider the problem of reducing worst-case instances of lattice approximation problems (e.g., finding short lattice vectors) to the problem of finding small solutions to random linear equations with coefficients in $\mathbb{Z}_q^n$. So, in order to perform such a reduction, one needs to sample (almost uniformly at random) the group $\mathbb{Z}_q^n$ in a way that is somehow related to an underlying lattice problem (for an arbitrary lattice) as we now explain. The core of the reduction is a (polynomial time) sampling procedure that allows to draw pairs consisting of a group element and a corresponding short "offset" vector (not necessarily in the lattice) having the following property: any (integer) solution to the homogeneous linear equation defined by the group elements maps the corresponding short offset vectors to a vector in the underlying lattice. The length of the resulting lattice vector depends on the size of the integer solution used to combine the short offset vectors. If we can find a *small* solution to the group equation (e.g., using the average-case oracle), then we can find a *short* lattice vector, essentially solving the underlying lattice problem. We remark that for the average-case oracle to work, the coefficients of the equation must be distributed almost uniformly at random in the group.

The high level approach outlined above to worst-case to average-case reduction is common to all works, including this paper. The difference is in the way group elements (and corresponding short offset vectors) are sampled. Essentially all previous works were based on the following approach: given an arbitrary lattice $\mathcal{L}(\mathbf{B})$, consider a sufficiently large region of space C which is approximately equal to a hypercube of size ℓ (with vertices in $\mathcal{L}(\mathbf{B})$). Then divide each side into q equal parts. This results in q^n subcubes of size ℓ/q, each corresponding to a group element in $\mathbb{Z}_q^n$. Next we sample lattice points from $\mathcal{L}(\mathbf{B}) \cap C$, and for each sample consider the corresponding subcube and offset within the subcube (e.g., with respect to the center of the subcube). If each subcube contains approximately the same number of lattice points, then the induced distribution on group elements is almost uniform over $\mathbb{Z}_q^n$. The correctness of the reduction is based on the following two important properties of the sampling procedure:

- Each subregion should be small enough, so that the offset vectors are short, and the final output of the reduction is a short lattice vector.
- Each subregion should be large enough, so that the number of lattice points in each region is about the same and the chosen group element is almost uniform in $\mathbb{Z}_q^n$.

These two contradicting requirements end up determining the connection factor obtained by the reduction.

In this paper we develop a new technique to generate random group elements that does not require starting from a large hypercube C. Instead of considering large regions of space and counting the number of lattice points in them, we simply start from a lattice point, and add some Gaussian noise to it. Our goal is to use an amount of noise sufficiently large so that the resulting point (which does not belong to the lattice in general) is distributed almost uniformly in space.

Technically, we pick a random noise vector with a Gaussian distribution, and reduce it modulo the basis of the lattice, to obtain a vector distributed almost uniformly at random over the fundamental parallelepiped

of the lattice. Next we divide the fundamental parallelepiped into q^n equal regions, and use each of them to represent a group element in $\mathbb{Z}_q^n$. Notice that none of the regions contains any lattice point. Notice also that using this approach, it is not important that the regions have a nice (approximately hypercubic) shape: since all regions have the same volume, a reduced noise vector distibuted almost uniformly over the fundamental parallelepiped will induce an almost uniform distribution over $\mathbb{Z}_q^n$.

As an additional remark, we point out that the previous best reductions produced group elements whose distribution is only moderately close to uniform. In order to get almost uniformly distributed group elements, they generated a small (super-logarithmic) number of group elements, and added them all up. Our technique avoids this complication since it directly gives group elements whose distribution is extremely close to uniform, and does not require to add up many samples. We believe that this fact, together with the fact that we do not need to start from a large cube, allows us to obtain a much cleaner and simpler reduction. The ideas and techniques presented in this paper have been recently used in [22] to obtain analogous improvements and simplifications for similar results about cyclic lattices.

Gaussian distributions: The use of Gaussian distributions in the study of lattices is standard in mathematics (see, for example, [5]). In computer science, they have been recently used in [8, 24, 1]. In [1], for example, Gaussian distributions are used to prove that certain lattice problems are in coNP.

We believe that a large part of our technical contribution is in the study of these Gaussian distributions. We start by defining the *smoothing parameter* of a lattice, a new lattice parameter with the following fundamental property:[4] if one picks a noise vector from a Gaussian distribution with radius at least as large as the smoothing parameter, and reduces the noise vector modulo the fundamental parallelepiped of the lattice, then the resulting distribution is very close to uniform. We then relate this parameter to standard lattice parameters such as the length of the shortest dual vector and the length of the shortest set of independent vectors. The proof of the former is based on a lemma by Banaszczyk [5] while the proof of the latter is, to the best of our knowledge, novel.

We then go on to consider the discrete Gaussian distribution on a lattice. Let $\mathbf{c}$ be any point in space. Let $\mathbf{y}$ be obtained by adding to $\mathbf{c}$ a vector chosen from a Gaussian distribution whose size is at least the smoothing parameter of the lattice. Then, consider the distribution of $\mathbf{y}$ conditioned on it being in the lattice (this will be made rigorous later). This distribution is illustrated in Figure 1. Essentially, it is a Gaussian distribution around $\mathbf{c}$ restricted to the lattice. Interestingly, we prove that this distribution behaves in many respects like the (continuous) Gaussian distribution around $\mathbf{c}$. For example, its center is very close to $\mathbf{c}$ and its average square distance from $\mathbf{c}$ is also very close to that of the continuous Gaussian distribution. From these two facts we can derive relatively easily all the properties needed for the worst-case to average-case reduction.

1.2 Complexity of lattice problems

Since many lattice problems are NP-hard to approximate within small factors, connections between the average-case and worst-case complexity of lattice problems can be regarded as progress toward the ambitious goal of constructing one-way functions based on the assumption that P $\neq$ NP. Unfortunately, there is still a big gap between factors for which lattice problems are known to be NP-hard and those known to imply the existence of one-way functions. The strongest known hardness result (for the problems considered in this paper) is the NP-hardness of approximating the length of the shortest linearly independent set within any constant and, under the stronger assumption NP $\not\subseteq$ DTIME($2^{\mathrm{polylog}(n)}$), within $2^{(\log n)^{1-\epsilon}}$ for any $\epsilon > 0$ [6]. For the shortest vector problem, hardness within any constant approximation factor or

[4] The actual definition of smoothing parameter involves the dual lattice, and it is rather technical. Here we only state a fundamental property of the smoothing parameter that conveys the intuition behind our definition. See Definition 3.1 for the actual definition.

factors of the form $2^{(\log n)^{1/2-\epsilon}}$ (for any $\epsilon > 0$) has been shown [16] under the assumption[5] that NP $\neq$ RP or NP $\not\subseteq$ BPTIME($2^{\text{polylog}(n)}$) respectively. No hardness result (under deterministic or probabilistic reductions) is currently known for the covering radius problem, although the problem is conceivably hard. (See [14] for further discussion of the complexity of the covering radius problem.)

Beside the fact that all known hardness results are only for subpolynomial approximation factors, all three problems have been shown to be in coAM for $O(\sqrt{n/\log n})$ approximation factors [11, 14] (see also [1, 14] where the problems are shown to be in coNP for $O(\sqrt{n})$ factors), giving evidence[6] that the problems are not NP-hard within such factors. Still, one might conjecture that some of these problems are NP-hard to approximate for factors close to $\sqrt{n/\log n}$, say, $n^{1/2-\epsilon}$ for any $\epsilon > 0$.

The results in this paper, showing that there exist hard on average problems based on the inapproximability of lattice problems within $\tilde{O}(n)$, bring us closer to factors $O(\sqrt{n})$, below which the lattice problems are not known to be in coNP, and therefore may be NP-hard. However, it is not clear how our techniques can be used to obtain factors below $\tilde{O}(n)$.

2 Preliminaries

General: For any real x, $\lfloor x \rfloor$ denotes the largest integer not greater than x. For a vector $\mathbf{x} = (x_1, \ldots, x_n)$ we define $\lfloor \mathbf{x} \rfloor$ as $(\lfloor x_1 \rfloor, \ldots, \lfloor x_n \rfloor)$. We write log for the logarithm to the base 2, and $\log_q$ when the base q is any number possibly different from 2. We use $\omega(f(n))$ to denote the set of functions growing faster than $c \cdot f(n)$ for any $c > 0$. A function $\epsilon(n)$ is *negligible* if $\epsilon(n) < 1/n^c$ for any $c > 0$ and all sufficiently large n.

The n-dimensional Euclidean space is denoted $\mathbb{R}^n$. We use bold lower case letters (e.g., $\mathbf{x}$) to denote vectors, and bold upper case letters (e.g., $\mathbf{M}$) to denote matrices. The ith coordinate of $\mathbf{x}$ is denoted x_i. For a set $S \subseteq \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$ and $a \in \mathbb{R}$, we let $S + \mathbf{x} = \{\mathbf{y} + \mathbf{x} \colon \mathbf{y} \in S\}$ denote the translate of S by $\mathbf{x}$, and $aS = \{a\mathbf{y} \colon \mathbf{y} \in S\}$ denote the scaling of S by a. The Euclidean norm (also known as the ℓ_2 norm) of a vector $\mathbf{x} \in \mathbb{R}^n$ is $\|\mathbf{x}\| = (\sum_i x_i^2)^{1/2}$, and the associated distance is $\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$. The distance function is extended to sets in the customary way: $\text{dist}(\mathbf{x}, S) = \text{dist}(S, \mathbf{x}) = \min_{\mathbf{y} \in S} \text{dist}(\mathbf{x}, \mathbf{y})$. We often use matrix notation to denote sets of vectors. For example, matrix $\mathbf{S} \in \mathbb{R}^{n \times m}$ represents the set of n-dimensional vectors $\{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$, where $\mathbf{s}_1, \ldots, \mathbf{s}_m$ are the columns of $\mathbf{S}$. We denote by $\|\mathbf{S}\|$ the maximum length of a vector in $\mathbf{S}$. The linear space spanned by a set of m vectors $\mathbf{S}$ is denoted $\text{span}(\mathbf{S}) = \{\sum_i x_i \mathbf{s}_i : \ x_i \in \mathbb{R} \text{ for } 1 \leq i \leq m\}$. For any set of n linearly independent vectors $\mathbf{S}$, we define the half-open parallelepiped $\mathcal{P}(\mathbf{S}) = \{\sum_i x_i \mathbf{s}_i : 0 \leq x_i < 1 \text{ for } 1 \leq i \leq n\}$. Finally, we denote by $\mathcal{B}$ the closed Euclidean ball of radius 1 around the origin, $\mathcal{B} = \{\mathbf{w} \in \mathbb{R}^n : \|\mathbf{w}\| \leq 1\}$.

Statistical Distance: Statistical distance is a measure of distance between two probability distributions and is a convenient tool in the analysis of randomized algorithms and reductions. Here we define it and state some simple facts that will be used in the rest of the paper. These facts are easily verified; for more details the reader is referred to [23, Chapter 8].

Definition 2.1 *We define the statistical distance between two discrete random variables X and Y over a (countable) set A as*

$$\Delta(X, Y) = \frac{1}{2} \sum_{a \in A} |\Pr\{X = a\} - \Pr\{Y = a\}|.$$

[5] No true NP-hardness result (i.e., under deterministic polynomial time reductions) is currently known for SVP even in its exact version. However, [19] showed that if a certain number theoretic conjecture on the distribution of square-free smooth numbers holds true, then SVP is NP-hard (under deterministic polynomial time Karp reductions) for any factor $\gamma < \sqrt{2}$.

[6] Specifically, since the first two problems are in NP even in their exact version, they cannot be NP-hard to approximate within $O(\sqrt{n/\log n})$ (resp. $O(\sqrt{n})$) unless NP $\subseteq$ coAM (resp. NP = coNP.) For the covering radius problem the situation is more complicated because the exact version of the problem is not known to be in NP. See [11, 1] for further discussion of the implications of these results.

Similarly, for two continuous random variables X and Y over $\mathbb{R}^n$ with probability density functions T_1 and T_2 respectively, the statistical distance is defined as

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

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 $\Delta(X,Y)$. Another useful property of the statistical distance is the following. Let $X_1,\ldots,X_k$ and $Y_1,\ldots,Y_k$ be two lists of totally independent random variables. Then

$$\Delta((X_1,\ldots,X_k),(Y_1,\ldots,Y_k)) \leq \sum_{i=1}^{k} \Delta(X_i,Y_i).$$

Lattices: We now describe some basic definitions related to lattices. For a more in-depth discussion, see [23]. An n-dimensional *lattice* is the set of all integer combinations

$$\Big\{ \sum_{i=1}^{n} x_i \mathbf{b}_i \colon x_i \in \mathbb{Z} \text{ for } 1 \leq i \leq n \Big\}$$

of n linearly independent vectors $\mathbf{b}_1,\ldots,\mathbf{b}_n$ in $\mathbb{R}^n$.[7] The set of vectors $\mathbf{b}_1,\ldots,\mathbf{b}_n$ is called a *basis* for the lattice. A basis can be represented by the matrix $\mathbf{B} = [\mathbf{b}_1,\ldots,\mathbf{b}_n] \in \mathbb{R}^{n\times n}$ having the basis vectors as columns. The lattice generated by $\mathbf{B}$ is denoted $\mathcal{L}(\mathbf{B})$. Notice that $\mathcal{L}(\mathbf{B}) = \{\mathbf{Bx} \colon \mathbf{x} \in \mathbb{Z}^n\}$, where $\mathbf{Bx}$ is the usual matrix-vector multiplication.

For any lattice basis $\mathbf{B}$ and point $\mathbf{x}$, there exists a unique vector $\mathbf{y} \in \mathcal{P}(\mathbf{B})$ such that $\mathbf{y} - \mathbf{x} \in \mathcal{L}(\mathbf{B})$. This vector is denoted $\mathbf{y} = \mathbf{x} \bmod \mathbf{B}$, and it can be computed in polynomial time given $\mathbf{B}$ and $\mathbf{x}$. The dual of a lattice Λ is the set

$$\Lambda^* = \{\mathbf{x} \colon \forall \mathbf{y} \in \Lambda \ \langle \mathbf{x}, \mathbf{y}\rangle \in \mathbb{Z}\}$$

of all vectors that have integer scalar product ($\langle \mathbf{x}, \mathbf{y}\rangle = \sum_i x_i y_i$) with all lattice vectors. The dual of a lattice is a lattice, and if $\Lambda = \mathcal{L}(\mathbf{B})$ is the lattice generated by basis $\mathbf{B}$, then $\mathbf{B}^* = (\mathbf{B}^T)^{-1}$ is a basis for the dual lattice, where $\mathbf{B}^T$ is the transpose of $\mathbf{B}$. A sub-lattice of $\mathcal{L}(\mathbf{B})$ is a lattice $\mathcal{L}(\mathbf{S})$ such that $\mathcal{L}(\mathbf{S}) \subseteq \mathcal{L}(\mathbf{B})$. The *determinant* of a lattice $\det(\mathcal{L}(\mathbf{B}))$ is the (n-dimensional) volume of the fundamental parallelepiped $\mathcal{P}(\mathbf{B})$ and is given by $|\det(\mathbf{B})|$.

The *minimum distance* of a lattice Λ, denoted $\lambda_1(\Lambda)$, is the minimum distance between any two distinct lattice points, and equals the length of the shortest nonzero lattice vector:

$$\begin{aligned} \lambda_1(\Lambda) &= \min\{\mathrm{dist}(\mathbf{x},\mathbf{y}) : \mathbf{x} \neq \mathbf{y} \in \Lambda\} \\ &= \min\{\|\mathbf{x}\| : \mathbf{x} \in \Lambda \setminus \{\mathbf{0}\}\}\ . \end{aligned}$$

This definition can be generalized to define the ith successive minimum as the smallest λ_i such that $\lambda_i \mathcal{B}$ contains i linearly independent lattice points:

$$\lambda_i(\Lambda) = \min\{r \colon \dim(\mathrm{span}(\Lambda \cap r\mathcal{B})) \geq i\}.$$

Another important constant associated to a lattice is the *covering radius* $\nu(\Lambda)$, defined as

$$\nu(\Lambda) = \max_{\mathbf{x}\in\mathbb{R}^n} \{\mathrm{dist}(\mathbf{x},\Lambda)\}.$$

We often abuse notation and write $\lambda_1(\mathbf{B})$ instead of $\lambda_1(\mathcal{L}(\mathbf{B}))$ and similarly for other lattice parameters.

[7] Strictly speaking, this is the definition of a *full-rank* lattice. Since only full-rank lattices are used in this paper, all definitions are restricted to the full-rank case.

Lattice problems: We consider the following lattice problems. For simplicity, we consider some of our problems in their promise version.[8] It is easy to see that a solution to any of the promise problems below implies a solution to the corresponding optimization problem (that is, the problem that asks for an approximation to the corresponding lattice parameter, e.g., λ_1). The reader is referred to [23] for further discussion of these lattice problems. The following definitions are parameterized by a positive (and typically monotone) real valued function $\gamma\colon \mathbb{Z}^+ \to \mathbb{R}^+$ of the lattice dimension.

Definition 2.2 (Shortest Vector Problem) *An input to* GAPSVP_γ *is a pair* $(\mathbf{B}, d)$ *where* $\mathbf{B}$ *is an* n*-dimensional lattice basis and* d *is a rational number. In* YES *inputs* $\lambda_1(\mathbf{B}) \leq d$ *and in* NO *inputs* $\lambda_1(\mathbf{B}) > \gamma(n) \cdot d$.

Definition 2.3 (Closest Vector Problem) *An input to* GAPCVP_γ *is a triple* $(\mathbf{B}, \mathbf{t}, d)$ *where* $\mathbf{B}$ *is an* n*-dimensional lattice basis,* $\mathbf{t}$ *is a target vector, and* d *is a rational number. In* YES *inputs* $\text{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) \leq d$ *and in* NO *inputs* $\text{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) > \gamma(n) \cdot d$.

Definition 2.4 (Covering Radius Problem) *An input to* GAPCRP_γ *is a pair* $(\mathbf{B}, d)$ *where* $\mathbf{B}$ *is an* n*-dimensional lattice basis and* d *is a rational number. In* YES *inputs* $\nu(\mathbf{B}) \leq d$ *and in* NO *inputs* $\nu(\mathbf{B}) > \gamma(n) \cdot d$.

The remaining lattice problems are given in their search version.

Definition 2.5 (Shortest Independent Vectors Problem) *An input to* SIVP_γ *is an* n*-dimensional lattice basis* $\mathbf{B}$*. The goal is to output a set of* n *linearly independent lattice vectors* $\mathbf{S} \subset \mathcal{L}(\mathbf{B})$ *such that* $\|\mathbf{S}\| \leq \gamma(n) \cdot \lambda_n(\mathbf{B})$ *where* $\|\mathbf{S}\|$ *is the maximum length of a vector in* $\mathbf{S}$.

A generalization of SIVP is the following somewhat less standard lattice problem.

Definition 2.6 (Generalized Independent Vectors Problem) *An input to* GIVP^ϕ_γ *is an* n*-dimensional lattice basis* $\mathbf{B}$*. The goal is to output a set of* n *linearly independent lattice vectors* $\mathbf{S} \subset \mathcal{L}(\mathbf{B})$ *such that* $\|\mathbf{S}\| \leq \gamma(n) \cdot \phi(\mathbf{B})$.

In the above, ϕ denotes any arbitrary function on lattices. Choosing $\phi = \lambda_n$ results in the SIVP. In this paper, we usually take ϕ to be the smoothing parameter, defined in the next section.

Definition 2.7 (Guaranteed Distance Decoding) *An input to* GDD^ϕ_γ *is an* n*-dimensional lattice basis* $\mathbf{B}$ *and a target point* $\mathbf{t}$*. The goal is to output a lattice point* $\mathbf{x} \in \mathcal{L}(\mathbf{B})$ *such that* $\text{dist}(\mathbf{t}, \mathbf{x}) \leq \gamma(n) \cdot \phi(\mathbf{B})$.

In this problem, we usually take $\phi = \nu$ to be the covering radius of the lattice. Notice that for any lattice basis $\mathbf{B}$ and target $\mathbf{t} \in \mathbb{R}^n$, there is always a lattice point within distance $\nu(\mathbf{B})$ of $\mathbf{t}$. The GDD^ν_γ problem can be seen as a variant of the CVP in which the quality of the solution is measured with respect to the worst possible distance $\max_{\mathbf{x} \in \mathbb{R}^n} \text{dist}(\mathbf{x}, \mathcal{L}(\mathbf{B}))$ instead of the distance of the given target $\text{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B}))$.

Gaussian measures: For any vectors $\mathbf{c}, \mathbf{x}$ and any $s > 0$, let

$$\rho_{s,\mathbf{c}}(\mathbf{x}) = e^{-\pi \|(\mathbf{x}-\mathbf{c})/s\|^2}$$

be a Gaussian function centered in $\mathbf{c}$ scaled by a factor of s. The total measure associated to $\rho_{s,\mathbf{c}}$ is $\int_{\mathbf{x} \in \mathbb{R}^n} \rho_{s,\mathbf{c}}(\mathbf{x}) d\mathbf{x} = s^n$. Therefore, we can define the (continuous) Gaussian distribution around $\mathbf{c}$ with parameter s by its probability density function

$$\forall \mathbf{x} \in \mathbb{R}^n,\ D_{s,\mathbf{c}}(\mathbf{x}) = \frac{\rho_{s,\mathbf{c}}(\mathbf{x})}{s^n}.$$

[8] Promise problems are a generalization of decision problems where one is asked whether a given input satisfies one of two mutually exclusive properties. Unlike decision problems, these two properties are not necessarily exhaustive. The problem is, under the promise that the given input satisfies one of the two conditions, tell which of the two properties is satisfied. If the input satisfies neither property, then any answer is acceptable.

It can be seen that the expected square distance from $\mathbf{c}$ of a vector chosen from this distribution is $ns^2/(2\pi)$. So, intuitively, one can think of $D_{s,\mathbf{c}}$ as a sphere of radius $s\sqrt{n/(2\pi)}$ centered around $\mathbf{c}$.

Notice that $D_{s,\mathbf{c}}$ can be expressed as the sum of n orthogonal 1-dimensional Gaussian distributions, and each of them can be efficiently approximated with arbitrary precision using standard techniques. So, the distribution $D_{s,\mathbf{c}}$ can be efficiently approximated. For simplicity, in this paper we work with real numbers and assume we can sample from $D_{s,\mathbf{c}}$ exactly. In practice, when only finite precision is available, $D_{s,\mathbf{c}}$ can be approximated by picking a fine grid, and choosing points from the grid with probability approximately proportional to $D_{s,\mathbf{c}}$. All our arguments can be made rigorous by selecting a sufficiently fine grid.

When $\mathbf{c}$ or s are not specified, we assume that they are the origin and 1 respectively. Functions are extended to sets in the usual way; e.g., $\rho_{s,\mathbf{c}}(A) = \sum_{\mathbf{x}\in A} \rho_{s,\mathbf{c}}(\mathbf{x})$ for any countable set A.

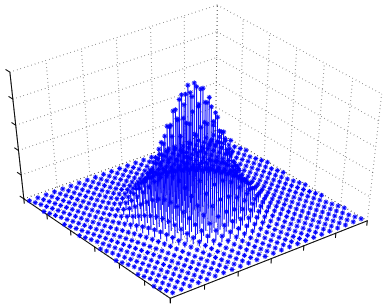

Figure 1: A discrete Gaussian distribution

For any vector $\mathbf{c}$, real $s > 0$, and lattice Λ, define the probability distribution $D_{\Lambda,s,\mathbf{c}}$ over Λ by

$$\forall \mathbf{x} \in \Lambda,\ D_{\Lambda,s,\mathbf{c}}(\mathbf{x}) = \frac{D_{s,\mathbf{c}}(\mathbf{x})}{D_{s,\mathbf{c}}(\Lambda)} = \frac{\rho_{s,\mathbf{c}}(\mathbf{x})}{\rho_{s,\mathbf{c}}(\Lambda)}.$$

We refer to $D_{\Lambda,s,\mathbf{c}}$ as a discrete Gaussian distribution (see Figure 1) and as before, we sometimes omit s or $\mathbf{c}$. We will later use the following connection between $D_{s,\mathbf{c}}$ and $D_{\Lambda,s,\mathbf{c}}$: if $\mathbf{x}$ is distributed according to $D_{s,\mathbf{c}}$ and we condition on $\mathbf{x} \in \Lambda$, the conditional distribution of $\mathbf{x}$ is $D_{\Lambda,s,\mathbf{c}}$. To see why this is true, recall that our vector $\mathbf{x}$ is in fact chosen from some very fine grid:[9] then, the probability of obtaining some grid point $\mathbf{x}$ in a sample from $D_{s,\mathbf{c}}$ is very close to $\alpha D_{s,\mathbf{c}}(\mathbf{x})$, where α is the volume of one cell in our grid, whereas the probability of $\mathbf{x} \in \Lambda$ is very close to $\alpha D_{s,\mathbf{c}}(\Lambda)$. All our arguments can be made rigorous by working with a fine enough grid.

We will show that for a large enough s, $D_{\Lambda,s,\mathbf{c}}$ behaves in many respects like the continuous Gaussian distribution $D_{s,\mathbf{c}}$. In particular, vectors distributed according to $D_{\Lambda,s,\mathbf{c}}$ have an average value very close to $\mathbf{c}$ and expected squared distance from $\mathbf{c}$ very close to $s^2 n/2\pi$ (for vectors chosen from $D_{s,\mathbf{c}}$, these quantities are exactly $\mathbf{c}$ and $s^2 n/2\pi$). In fact, we define a new lattice parameter that tells us how big s has to be in order for this to happen. We name this parameter the *smoothing parameter*. We then relate this parameter to other lattice parameters such as the length of the shortest vector in the dual lattice and the length of the shortest maximal set of independent vectors.

Fourier transform: We briefly review some of the important properties of the Fourier transform. For a more precise and in-depth treatment, see, e.g., [10]. The Fourier transform of a function $h : \mathbb{R}^n \mapsto \mathbb{R}$ is defined to be $\hat{h}(\mathbf{w}) = \int_{\mathbb{R}^n} h(\mathbf{x}) e^{-2\pi i \langle \mathbf{x},\mathbf{w}\rangle} d\mathbf{x}$. From the definition we can obtain several useful formulas; first, if h is defined by $h(\mathbf{x}) = g(\mathbf{x} + \mathbf{v})$ for some function g and vector $\mathbf{v}$ then

$$\hat{h}(\mathbf{w}) = e^{2\pi i \langle \mathbf{v},\mathbf{w}\rangle} \hat{g}(\mathbf{w}). \tag{2}$$

[9] Although not needed in this paper, one can also define the conditional probability on the continuous random variables directly. This requires some care as it involves conditioning on an event of probability zero.

Similarly, if h is defined by $h(\mathbf{x}) = e^{2\pi i\langle \mathbf{x},\mathbf{v}\rangle} g(\mathbf{x})$ for some function g and vector $\mathbf{v}$ then

$$\hat{h}(\mathbf{w}) = \hat{g}(\mathbf{w} - \mathbf{v}). \tag{3}$$

Also, if we denote by $h^{\mathbf{u}}$ the derivative of h in the direction of some unit vector $\mathbf{u}$ then its Fourier transform is

$$\widehat{h^{\mathbf{u}}}(\mathbf{w}) = 2\pi i\langle \mathbf{u}, \mathbf{w}\rangle \cdot \hat{h}(\mathbf{w}). \tag{4}$$

Another important fact is that the Gaussian is its own Fourier transform, i.e., $\hat{\rho} = \rho$. More generally, for any $s > 0$ it holds that $\widehat{\rho_s} = s^n \rho_{1/s}$. We use the following formulation of the Poisson summation formula.

Lemma 2.8 *For any lattice Λ and any*[10] *function $f : \mathbb{R}^n \to \mathbb{C}$, $f(\Lambda) = \det(\Lambda^*)\hat{f}(\Lambda^*)$ where $\hat{f}$ denotes the Fourier transform of f.*

An immediate application of the Poisson summation formula is the fact that the Gaussian measure $\rho_{s,\mathbf{c}}(\Lambda)$ is maximized when the center is a lattice point $\mathbf{c} \in \Lambda$.

Lemma 2.9 *For any lattice Λ, positive real $s > 0$ and vector $\mathbf{c}$, $\rho_{s,\mathbf{c}}(\Lambda) \leq \rho_s(\Lambda)$.*

Proof: Using Lemma 2.8 twice, and Equation (2) we get

$$\begin{aligned}
\rho_{s,\mathbf{c}}(\Lambda) &= \det(\Lambda^*)\widehat{\rho_{s,\mathbf{c}}}(\Lambda^*) \\
&= \det(\Lambda^*) \sum_{\mathbf{y}\in\Lambda^*} \widehat{\rho_{s,\mathbf{c}}}(\mathbf{y}) \\
&= \det(\Lambda^*) \sum_{\mathbf{y}\in\Lambda^*} e^{-2\pi i\langle \mathbf{c},\mathbf{y}\rangle} \widehat{\rho_s}(\mathbf{y}) \\
&\leq \det(\Lambda^*) \sum_{\mathbf{y}\in\Lambda^*} \widehat{\rho_s}(\mathbf{y}) = \rho_s(\Lambda)
\end{aligned}$$

where we used that $\widehat{\rho_s} = s^n \rho_{1/s}$ is a positive function. ■

We will also use the following lemma by Banaszczyk.

Lemma 2.10 ([5], Lemma 1.5) *For any $c > 1/\sqrt{2\pi}$, n-dimensional lattice Λ, and vector $\mathbf{v} \in \mathbb{R}^n$,*

$$\rho(\Lambda \setminus c\sqrt{n}\mathcal{B}) < C^n \cdot \rho(\Lambda) \tag{5}$$

$$\rho((\Lambda + \mathbf{v}) \setminus c\sqrt{n}\mathcal{B}) < 2C^n \cdot \rho(\Lambda) \tag{6}$$

where $C = c\sqrt{2\pi e} \cdot e^{-\pi c^2} < 1$.

Sum of independent vectors: We conclude this section with a simple lemma which will be used in Section 5 to bound the length of the sum of Gaussian random variables. The lemma essentially shows that when summing m independent random variables, the expected length of the sum grows with $\sqrt{m}$ and not m. (As an example to illustrate the use of the lemma, consider the case $\epsilon = 0$ and $\mathbf{z} = (1, \ldots, 1)$.)

Lemma 2.11 *Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be m vectors chosen independently from probability distributions $V_1, \ldots, V_m$ such that $\mathrm{Exp}[\|\mathbf{v}_i\|^2] \leq l$ and $\|\mathrm{Exp}[\mathbf{v}_i]\|^2 \leq \epsilon$ for every $i = 1, \ldots, m$. Then, for any $\mathbf{z} \in \mathbb{R}^m$, the expected squared norm of $\sum \mathbf{v}_i z_i$ is at most $\mathrm{Exp}[\|\sum_{i=1}^m \mathbf{v}_i z_i\|^2] \leq (l + \epsilon \cdot m)\|\mathbf{z}\|^2$.*

[10] For this formula to hold, f needs to satisfy certain niceness assumptions. These assumptions always hold in our applications. See [10] for more details.

Proof: By linearity of expectation and inequality $\sum_i |z_i| \leq \sqrt{m}\|\mathbf{z}\|$, we get

$$\begin{aligned}
\mathrm{Exp}\left[\left\|\sum_i \mathbf{v}_i z_i\right\|^2\right] &= \sum_{i,j} z_i z_j \,\mathrm{Exp}[\langle \mathbf{v}_i, \mathbf{v}_j\rangle] \\
&= \sum_i z_i^2 \,\mathrm{Exp}[\|\mathbf{v}_i\|^2] + \sum_{i\neq j} z_i z_j \langle \mathrm{Exp}[\mathbf{v}_i], \mathrm{Exp}[\mathbf{v}_j]\rangle \\
&\leq \|\mathbf{z}\|^2 l + \Big(\sum_i |z_i|\Big)^2 \epsilon \\
&\leq \|\mathbf{z}\|^2 (l + \epsilon m).
\end{aligned}$$

■

3 The Smoothing Parameter

In this section we define a new lattice parameter related to Gaussian measures on lattices. We name it the *smoothing parameter*:

Definition 3.1 *For an n-dimensional lattice Λ, and positive real $\epsilon > 0$, we define its smoothing parameter $\eta_\epsilon(\Lambda)$ to be the smallest s such that $\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) \leq \epsilon$.*

Notice that $\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})$ is a continuous and strictly decreasing function of s such that $\lim_{s\to 0} \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) = \infty$ and $\lim_{s\to\infty} \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) = 0$. So, the parameter $\eta_\epsilon(\Lambda)$ is well defined for any $\epsilon > 0$, and $\epsilon \mapsto \eta_\epsilon(\Lambda)$ is the inverse function of $s \mapsto \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})$. In particular, $\eta_\epsilon(\Lambda)$ is also a continuous and strictly decreasing function of ϵ.

In this paper we are mostly interested in sequences of lattices Λ_n (in increasing dimension n) and the corresponding smoothing parameters $\eta_{\epsilon(n)}(\Lambda_n)$, where $\epsilon(n)$ is some negligible function of n. So, $\eta_{\epsilon(n)}(\Lambda_n)$ is the smallest s such that a Gaussian measure on the dual lattice Λ_n^* with parameter $1/s$ gives all but a negligible amount of its weight to the origin, for some negligible function $\epsilon(n)$ of the lattice dimension.

The motivation for this definition (and the name 'smoothing parameter') is presented in Lemma 4.1. Intuitively, it says that if we start from a uniformly random lattice point in Λ and perturb it by a Gaussian of radius $\eta_\epsilon(\Lambda)$, then the resulting distribution is $\epsilon/2$ close to uniform on the entire space.[11][12] The next two lemmas relate the smoothing parameter to some standard lattice parameters.

Lemma 3.2 *For any n-dimensional lattice Λ, $\eta_\epsilon(\Lambda) \leq \sqrt{n}/\lambda_1(\Lambda^*)$ where $\epsilon = 2^{-n}$.*

Proof: We use Lemma 2.10 with $c = 1$ and $C = \sqrt{2\pi e}\cdot e^{-\pi} < 1/4$. By separating the right hand side of (5) as the sum over points in $\sqrt{n}\mathcal{B}$ and over points outside $\sqrt{n}\mathcal{B}$ and rearranging, we obtain that for any lattice Λ,

$$\rho(\Lambda \setminus \sqrt{n}\mathcal{B}) < \frac{C^n}{1-C^n}\rho(\Lambda \cap \sqrt{n}\mathcal{B}).$$

Now, let s be such that $s > \sqrt{n}/\lambda_1(\Lambda^*)$. We have,

$$\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) = \rho(s\Lambda^* \setminus \{\mathbf{0}\}) = \rho(s\Lambda^* \setminus \sqrt{n}\mathcal{B}) < \frac{C^n}{1-C^n}\rho(s\Lambda^* \cap \sqrt{n}\mathcal{B}) = \frac{C^n}{1-C^n} < 2^{-n}$$

where we used that the shortest vector in $s\Lambda^*$ is longer than $\sqrt{n}$, and therefore $s\Lambda^* \setminus \sqrt{n}\mathcal{B} = s\Lambda^* \setminus \{\mathbf{0}\}$ and $s\Lambda^* \cap \sqrt{n}\mathcal{B} = \{\mathbf{0}\}$. ■

[11] In fact, no uniform probability distribution can be defined over a lattice (or other countably infinite set) or over the entire space. Formally, in order to define this property we follow [18] and capture the intuition of "starting from a random lattice point" by working modulo the lattice. See Section 4 for details, and [18] for more motivations and explanations about working modulo the lattice.

[12] In fact, a stronger property holds: at any point, the density function of the resulting distribution is within $(1 \pm \epsilon)$ of that of the uniform distribution. Moreover, it can be shown that this stronger property is equivalent to the assumption $s \geq \eta_\epsilon(B)$.

Lemma 3.3 *For any n-dimensional lattice Λ and positive real $\epsilon > 0$,*

$$\eta_\epsilon(\Lambda) \leq \sqrt{\frac{\ln(2n(1+1/\epsilon))}{\pi}} \cdot \lambda_n(\Lambda).$$

In particular, for any superlogarithmic function $\omega(\log n)$, there exists a negligible function $\epsilon(n)$ such that $\eta_\epsilon(\Lambda) \leq \sqrt{\omega(\log n)} \cdot \lambda_n(\Lambda)$.

Proof: Let $s = \sqrt{\frac{\ln(2(1+1/\epsilon)n)}{\pi}} \cdot \lambda_n(\Lambda)$. Our goal is to show that $\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) \leq \epsilon$. The idea is to show that for any vector $\mathbf{v} \in \Lambda$ of length at most $\lambda_n(\Lambda)$, almost all the contribution to $\rho_{1/s}(\Lambda^*)$ comes from those points in Λ^* that lie on the hyperplane orthogonal to $\mathbf{v}$. Therefore, if we take n linearly independent vectors of length at most $\lambda_n(\Lambda)$, almost all the contribution to $\rho_{1/s}(\Lambda^*)$ must come from the intersection of the corresponding hyperplanes, which is simply the origin. Details follow.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a set of n linearly independent vectors in Λ each of length at most $\lambda_n(\Lambda)$. Define the set $S_{i,j} \subseteq \Lambda^*$ as the set of all points in Λ^* whose inner product with $\mathbf{v}_i$ is $j \in \mathbb{Z}$. Note that for any fixed i, the $S_{i,j}$'s form a partition of Λ^*. Moreover, since $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \Lambda$ are linearly independent, any nonzero vector in Λ^* must have a nonzero integer inner product with at least one of them, and hence $\Lambda^* \setminus \{\mathbf{0}\} = \bigcup_i (\Lambda^* \setminus S_{i,0})$.

For any index i let $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|^2$ be a vector of length $1/\|\mathbf{v}_i\| \geq 1/\lambda_n(\Lambda)$ in the same direction as $\mathbf{v}_i$. For all j,

$$\rho_{1/s}(S_{i,j}) = e^{-\pi \|js\mathbf{u}_i\|^2} \rho_{1/s}(S_{i,j} - j\mathbf{u}_i).$$

Now, $S_{i,j} - j\mathbf{u}_i$ is simply a shift of the set $S_{i,0}$. In other words, there exists some vector $\mathbf{w}$ (which is orthogonal to $\mathbf{u}_i$) such that $S_{i,j} - j\mathbf{u}_i = S_{i,0} - \mathbf{w}$. Therefore, by Lemma 2.9,

$$\rho_{1/s}(S_{i,j} - j\mathbf{u}_i) = \rho_{1/s}(S_{i,0} - \mathbf{w}) = \rho_{1/s,\mathbf{w}}(S_{i,0}) \leq \rho_{1/s}(S_{i,0}).$$

Using $\|\mathbf{u}_i\| \geq 1/\lambda_n(\Lambda)$, and the bound $\sum_{j \neq 0} x^{-j^2} \leq 2 \sum_{j>0} x^{-j} = 2/(x-1)$ (valid for all $x > 1$), we get

$$\begin{aligned}
\rho_{1/s}(\Lambda^* \setminus S_{i,0}) &= \sum_{j \neq 0} \rho_{1/s}(S_{i,j}) \\
&\leq \sum_{j \neq 0} e^{-\pi (s/\lambda_n)^2 j^2} \rho_{1/s}(S_{i,0}) \\
&\leq \frac{2}{e^{\pi (s/\lambda_n)^2} - 1} \rho_{1/s}(S_{i,0}) \\
&= \frac{2}{e^{\pi (s/\lambda_n)^2} - 1} (\rho_{1/s}(\Lambda^*) - \rho_{1/s}(\Lambda^* \setminus S_{i,0})).
\end{aligned}$$

Solving for $\rho_{1/s}(\Lambda^* \setminus S_{i,0})$, we get

$$\rho_{1/s}(\Lambda^* \setminus S_{i,0}) \leq \frac{2}{e^{\pi (s/\lambda_n)^2} + 1} \rho_{1/s}(\Lambda^*).$$

Since ρ is positive,

$$\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) \leq \sum_i \rho_{1/s}(\Lambda^* \setminus S_{i,0}) \leq \frac{2n}{e^{\pi (s/\lambda_n)^2} + 1} \rho_{1/s}(\Lambda^*).$$

Finally, using $\rho_{1/s}(\Lambda^*) = 1 + \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})$ and solving for $\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})$, we get

$$\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) \leq \frac{2n}{e^{\pi (s/\lambda_n)^2} + 1 - 2n} < \frac{2n}{e^{\pi (s/\lambda_n)^2} - 2n} = \epsilon$$

by our choice of s. ∎

4 Properties of Gaussian Distributions

In this section we prove several properties of Gaussian distributions related to lattices. Our first lemma below justifies the name given to the smoothing parameter.

Lemma 4.1 *For any $s > 0$, $\mathbf{c} \in \mathbb{R}^n$, and lattice $\mathcal{L}(\mathbf{B})$, the statistical distance between $D_{s,\mathbf{c}} \bmod \mathcal{P}(\mathbf{B})$ and the uniform distribution over $\mathcal{P}(\mathbf{B})$ is at most $\frac{1}{2}\rho_{1/s}(\mathcal{L}(\mathbf{B})^* \setminus \{\mathbf{0}\})$. In particular, for any $\epsilon > 0$ and any $s \geq \eta_\epsilon(\mathbf{B})$, the statistical distance is at most*

$$\Delta(D_{s,\mathbf{c}} \bmod \mathcal{P}(\mathbf{B}), U(\mathcal{P}(\mathbf{B}))) \leq \epsilon/2.$$

Proof: Let Y be the density function of the distribution over $\mathcal{P}(\mathbf{B})$ defined by $(D_{s,\mathbf{c}} \bmod \mathcal{P}(\mathbf{B}))$:

$$Y(\mathbf{x}) = \frac{1}{s^n} \sum_{\mathbf{y} \in \mathcal{L}(\mathbf{B})} \rho_{s,\mathbf{c}}(\mathbf{x}+\mathbf{y}) = \frac{1}{s^n}\rho_{s,\mathbf{c}-\mathbf{x}}(\mathcal{L}(\mathbf{B})).$$

By Equation (2), the Fourier transform of $\rho_{s,\mathbf{c}-\mathbf{x}}$ at point $\mathbf{w}$ is $e^{2\pi i\langle \mathbf{x}-\mathbf{c},\mathbf{w}\rangle} s^n \rho_{1/s}(\mathbf{w})$. Hence, using Lemma 2.8,

$$\begin{aligned} Y(\mathbf{x}) &= \det(\mathcal{L}(\mathbf{B})^*) \sum_{\mathbf{w} \in \mathcal{L}(\mathbf{B})^*} e^{2\pi i\langle \mathbf{x}-\mathbf{c},\mathbf{w}\rangle} \rho_{1/s}(\mathbf{w}) \\ &= \det(\mathcal{L}(\mathbf{B})^*) \left(1 + \sum_{\mathbf{w} \in \mathcal{L}(\mathbf{B})^* \setminus \{\mathbf{0}\}} e^{2\pi i\langle \mathbf{x}-\mathbf{c},\mathbf{w}\rangle} \rho_{1/s}(\mathbf{w})\right). \end{aligned}$$

The density function of the uniform distribution over $\mathcal{P}(\mathbf{B})$ is $U(\mathbf{x}) = 1/\mathrm{vol}(\mathcal{P}(\mathbf{B})) = \det(\mathcal{L}(\mathbf{B})^*)$. Therefore the statistical distance between Y and U is

$$\begin{aligned} \Delta(Y,U) &= \frac{1}{2}\int_{\mathbf{x} \in \mathcal{P}(\mathbf{B})} |Y(\mathbf{x}) - U(\mathbf{x})| d\mathbf{x} \\ &\leq \frac{1}{2}\mathrm{vol}(\mathcal{P}(\mathbf{B})) \cdot \max_{\mathbf{x} \in \mathcal{P}(\mathbf{B})} |Y(\mathbf{x}) - \det(\mathcal{L}(\mathbf{B})^*)| \\ &= \frac{1}{2}\mathrm{vol}(\mathcal{P}(\mathbf{B})) \cdot \det(\mathcal{L}(\mathbf{B})^*) \cdot \max_{\mathbf{x} \in \mathcal{P}(\mathbf{B})} \left| \sum_{\mathbf{w} \in \mathcal{L}(\mathbf{B})^* \setminus \{\mathbf{0}\}} e^{2\pi i\langle \mathbf{x}-\mathbf{c},\mathbf{w}\rangle} \rho_{1/s}(\mathbf{w}) \right| \\ &\leq \frac{1}{2} \cdot \rho_{1/s}(\mathcal{L}(\mathbf{B})^* \setminus \{\mathbf{0}\}) \end{aligned}$$

where the last inequality follows by the triangle inequality (and is in fact an equality). ■

Our second lemma shows that when s is large enough, some statistical properties of the discrete Gaussian distribution $D_{\Lambda,s,\mathbf{c}}$ are very close to those of the continuous Gaussian distribution $D_{s,\mathbf{c}}$.

Lemma 4.2 *For any n-dimensional lattice Λ, point $\mathbf{c} \in \mathbb{R}^n$, unit vector $\mathbf{u}$, and reals $0 < \epsilon < 1$, $s \geq 2\eta_\epsilon(\Lambda)$,*

$$\begin{aligned} \left| \operatorname*{Exp}_{\mathbf{x} \sim D_{\Lambda,s,\mathbf{c}}} [\langle \mathbf{x}-\mathbf{c}, \mathbf{u}\rangle] \right| &\leq \frac{\epsilon s}{1-\epsilon} \\ \left| \operatorname*{Exp}_{\mathbf{x} \sim D_{\Lambda,s,\mathbf{c}}} [\langle \mathbf{x}-\mathbf{c}, \mathbf{u}\rangle^2] - \frac{s^2}{2\pi} \right| &\leq \frac{\epsilon s^2}{1-\epsilon} \end{aligned}$$

Proof: For any positive real $s > 0$, define $\Lambda' = \Lambda/s$, $\mathbf{c}' = \mathbf{c}/s$. Notice that, for any $\mathbf{x}$,

$$\Pr\{D_{\Lambda,s,\mathbf{c}} = s\mathbf{x}\} = \frac{\rho_{s,\mathbf{c}}(s\mathbf{x})}{\rho_{s,\mathbf{c}}(\Lambda)} = \frac{\rho_{\mathbf{c}'}(\mathbf{x})}{\rho_{\mathbf{c}'}(\Lambda')} = \Pr\{D_{\Lambda',\mathbf{c}'} = \mathbf{x}\},$$

i.e., the distribution $D_{\Lambda,s,\mathbf{c}}$ is equal to $D_{\Lambda',\mathbf{c}'}$ scaled by a factor of s. Therefore, it is enough to prove the lemma for $s = 1$. The general case follows by scaling the lattice by a factor s.

In the rest of the proof, we assume $s = 1$. We want to estimate the quantity $\mathrm{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[\langle \mathbf{x}-\mathbf{c}, \mathbf{u}\rangle^j]$, for $j = 1, 2$. Without loss of generality, assume that $\mathbf{u}$ is the vector $(1, 0, \ldots, 0)$, and define the functions

$$g_j(\mathbf{x}) \quad = \quad (x_1 - c_1)^j \cdot \rho_{\mathbf{c}}(\mathbf{x}),$$

where x_1 and c_1 denote the first coordinate of $\mathbf{x}$ and $\mathbf{c}$ respectively. Notice that

$$\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[\langle \mathbf{x}-\mathbf{c}, \mathbf{u}\rangle^j] = \operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[(x_1 - c_1)^j] = \frac{g_j(\Lambda)}{\rho_{\mathbf{c}}(\Lambda)}.$$

Applying Poisson's summation formula (Lemma 2.8) to the numerator and denominator, the above fraction can be rewritten as

$$\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[\langle \mathbf{x}-\mathbf{c}, \mathbf{u}\rangle^j] = \frac{\det(\Lambda^*) \cdot \widehat{g_j}(\Lambda^*)}{\det(\Lambda^*) \cdot \widehat{\rho_{\mathbf{c}}}(\Lambda^*)} = \frac{\widehat{g_j}(\Lambda^*)}{\widehat{\rho_{\mathbf{c}}}(\Lambda^*)}. \tag{7}$$

The Fourier transform $\widehat{\rho_{\mathbf{c}}}$ is easily computed using Equation 2: $\widehat{\rho_{\mathbf{c}}}(\mathbf{y}) = \rho(\mathbf{y})e^{-2\pi i\langle \mathbf{y},\mathbf{c}\rangle}$. In particular, $\widehat{\rho_{\mathbf{c}}}(\mathbf{0}) = 1$, $|\widehat{\rho_{\mathbf{c}}}(\mathbf{y})| = \rho(\mathbf{y})$, and

$$|\widehat{\rho_{\mathbf{c}}}(\Lambda^*)| = \left|1 + \sum_{\mathbf{y}\in\Lambda^*\setminus\{\mathbf{0}\}} \widehat{\rho_{\mathbf{c}}}(\mathbf{y})\right| \geq 1 - \rho(\Lambda^* \setminus \{\mathbf{0}\}) \geq 1 - \epsilon \tag{8}$$

where the last inequality uses $\eta_\epsilon(\Lambda) \leq \frac{1}{2} \leq 1$.

We evaluate the Fourier transform $\widehat{g_j}$ (for $j = 1, 2$) as follows. For any $j \geq 0$, let

$$\rho_{\mathbf{c}}^{(j)}(\mathbf{x}) = \left(\frac{\partial}{\partial x_1}\right)^j \rho_{\mathbf{c}}(\mathbf{x})$$

be the jth partial derivative of $\rho_{\mathbf{c}}(\mathbf{x})$ with respect to x_1. It is easy to see that

$$\begin{aligned} \rho_{\mathbf{c}}^{(1)}(\mathbf{x}) &= -2\pi(x_1 - c_1)\rho_{\mathbf{c}}(\mathbf{x}) \\ \rho_{\mathbf{c}}^{(2)}(\mathbf{x}) &= (4\pi^2(x_1 - c_1)^2 - 2\pi)\rho_{\mathbf{c}}(\mathbf{x}). \end{aligned}$$

Taking linear combinations of the previous equations, we can express the g_j functions as:

$$\begin{aligned} g_1 &= -\frac{1}{2\pi}\rho_{\mathbf{c}}^{(1)} \\ g_2 &= \frac{1}{4\pi^2}\rho_{\mathbf{c}}^{(2)} + \frac{1}{2\pi}\rho_{\mathbf{c}}. \end{aligned}$$

Using $\widehat{\rho_{\mathbf{c}}^{(j)}}(\mathbf{y}) = (2\pi i y_1)^j \widehat{\rho_{\mathbf{c}}}(\mathbf{y})$ (see Equation 4) and the linearity of the Fourier transform, we get

$$\widehat{g_1}(\mathbf{y}) = -iy_1\widehat{\rho_{\mathbf{c}}}(\mathbf{y}) \tag{9}$$

$$\widehat{g_2}(\mathbf{y}) = \left(\frac{1}{2\pi} - y_1^2\right)\widehat{\rho_{\mathbf{c}}}(\mathbf{y}) \tag{10}$$

We are now ready to evaluate expression (7). For $j = 1$, using (9) and (8), we get

$$\left|\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[\langle \mathbf{x}-\mathbf{c}, \mathbf{u}\rangle]\right| \leq \frac{\sum_{\mathbf{y}\in\Lambda^*} |y_1| \cdot |\widehat{\rho_{\mathbf{c}}}(\mathbf{y})|}{1-\epsilon}.$$

We use $|y_1| \leq \sqrt{\|\mathbf{y}\|^2} \leq e^{\|\mathbf{y}\|^2/2}$ and $|\widehat{\rho_{\mathbf{c}}}(\mathbf{y})| = \rho(\mathbf{y})$ to bound the numerator:

$$\begin{aligned}\sum_{\mathbf{y}\in\Lambda^*} |y_1||\widehat{\rho_{\mathbf{c}}}(\mathbf{y})| &= \sum_{\mathbf{y}\in\Lambda^*\setminus\{\mathbf{0}\}} |y_1|\cdot\rho(\mathbf{y}) \\ &\leq \sum_{\mathbf{y}\in\Lambda^*\setminus\{\mathbf{0}\}} e^{\|\mathbf{y}\|^2/2}\cdot e^{-\pi\|\mathbf{y}\|^2} \\ &\leq \sum_{\mathbf{y}\in\Lambda^*\setminus\{\mathbf{0}\}} e^{-\pi\|\mathbf{y}/2\|^2} = \rho_2\left(\Lambda^*\setminus\{\mathbf{0}\}\right) \leq \epsilon\end{aligned}$$

where the last inequality uses $\eta_\epsilon(\Lambda) \leq \frac{1}{2}$. This completes the proof for $j = 1$.

For $j = 2$, combining (7), (8), (10), and $|\widehat{\rho_{\mathbf{c}}}(\mathbf{y})| = \rho(\mathbf{y})$, we get

$$\begin{aligned}\left|\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[\langle\mathbf{x}-\mathbf{c},\mathbf{u}\rangle^2] - \frac{1}{2\pi}\right| &= \frac{|\sum_{\mathbf{y}\in\Lambda^*} y_1^2\cdot\widehat{\rho_{\mathbf{c}}}(\mathbf{y})|}{\widehat{\rho_{\mathbf{c}}}(\Lambda^*)} \\ &\leq \frac{\sum_{\mathbf{y}\in\Lambda^*} y_1^2\cdot\rho(\mathbf{y})}{1-\rho(\Lambda^*\setminus\{\mathbf{0}\})}.\end{aligned}$$

This time we use $y_1^2 \leq \|\mathbf{y}\|^2 \leq e^{\|\mathbf{y}\|^2}$ to bound the numerator:

$$\sum_{\mathbf{y}\in\Lambda^*} y_1^2\cdot\rho(\mathbf{y}) \leq \sum_{\mathbf{y}\in\Lambda^*\setminus\{\mathbf{0}\}} e^{\|\mathbf{y}\|^2}\cdot e^{-\pi\|\mathbf{y}\|^2} \leq \sum_{\mathbf{y}\in\Lambda^*\setminus\{\mathbf{0}\}} e^{-\pi\|\mathbf{y}/2\|^2} = \rho_2\left(\Lambda^*\setminus\{\mathbf{0}\}\right) \leq \epsilon.$$

■

As a corollary, we obtain the following lemma.

Lemma 4.3 *For any n-dimensional lattice Λ, vector $\mathbf{c}\in\mathbb{R}^n$, and reals $0<\epsilon<1$, $s\geq 2\eta_\epsilon(\Lambda)$, we have*

$$\begin{aligned}\left\|\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,s,\mathbf{c}}}[\mathbf{x}-\mathbf{c}]\right\|^2 &\leq \left(\frac{\epsilon}{1-\epsilon}\right)^2 s^2 n \\ \operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,s,\mathbf{c}}}[\|\mathbf{x}-\mathbf{c}\|^2] &\leq \left(\frac{1}{2\pi}+\frac{\epsilon}{1-\epsilon}\right) s^2 n.\end{aligned}$$

Proof: Take any orthonormal basis $\mathbf{u}_1,\ldots,\mathbf{u}_n$. Using Lemma 4.2, we get

$$\left\|\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,s,\mathbf{c}}}[\mathbf{x}-\mathbf{c}]\right\|^2 = \sum_{i=1}^n \left(\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,s,\mathbf{c}}}[\langle\mathbf{x}-\mathbf{c},\mathbf{u}_i\rangle]\right)^2 \leq ns^2\cdot\left(\frac{\epsilon}{1-\epsilon}\right)^2$$

and

$$\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,s,\mathbf{c}}}[\|\mathbf{x}-\mathbf{c}\|^2] = \sum_{i=1}^n \operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,s,\mathbf{c}}}[\langle\mathbf{x}-\mathbf{c},\mathbf{u}_i\rangle^2] \leq ns^2\cdot\left(\frac{1}{2\pi}+\frac{\epsilon}{1-\epsilon}\right).$$

■

The remaining lemmas describe some additional properties of the discrete Gaussian distribution. These lemmas are only used in our GapSVP result of Subsection 5.4.

Lemma 4.4 *For any n-dimensional lattice Λ, vector $\mathbf{c}\in\mathbb{R}^n$, and reals $0<\epsilon<1$, $s\geq\eta_\epsilon(\Lambda)$, we have*

$$\Pr_{\mathbf{x}\sim D_{\Lambda,s,\mathbf{c}}}\{\|\mathbf{x}-\mathbf{c}\| > s\sqrt{n}\} \leq \frac{1+\epsilon}{1-\epsilon}\cdot 2^{-n}.$$

Proof: As in the proof of Lemma 4.2, it is enough to prove the lemma for $s = 1$. We can write

$$\Pr_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}\{\|\mathbf{x}-\mathbf{c}\| > \sqrt{n}\} = \frac{\rho((\Lambda-\mathbf{c})\setminus\sqrt{n}\mathcal{B})}{\rho_{\mathbf{c}}(\Lambda)}.$$

By Lemma 2.10 with $c = 1$, the numerator is at most $2^{-n}\rho(\Lambda)$. By the Poisson summation formula (Lemma 2.8),

$$\begin{aligned}\rho_{\mathbf{c}}(\Lambda) &= \det(\Lambda^*)\widehat{\rho_{\mathbf{c}}}(\Lambda^*)\\ &= \det(\Lambda^*)\sum_{\mathbf{y}\in\Lambda^*}\widehat{\rho_{\mathbf{c}}}(\mathbf{y})\\ &= \det(\Lambda^*)\sum_{\mathbf{y}\in\Lambda^*}e^{-2\pi i\langle\mathbf{c},\mathbf{y}\rangle}\widehat{\rho}(\mathbf{y})\\ &= \det(\Lambda^*)(1+\delta)\end{aligned}$$

where $|\delta| \le |\rho(\Lambda^*\setminus\{\mathbf{0}\})| \le \epsilon$. Therefore $\rho_{\mathbf{c}}(\Lambda) \ge \det(\Lambda^*)(1-\epsilon)$, $\rho(\Lambda) \le \det(\Lambda^*)(1+\epsilon)$, and $2^{-n}\rho(\Lambda)/\rho_{\mathbf{c}}(\Lambda) \le 2^{-n}\frac{1+\epsilon}{1-\epsilon}$. ■

Lemma 4.5 *Let Λ be an n-dimensional lattice, $\mathbf{c}, \mathbf{v}$ be two points in $\mathbb{R}^n$, $0 < \epsilon < 1$ and $s \ge \eta_\epsilon(\Lambda)$ such that $\mathrm{dist}(\mathbf{v}, \Lambda^*) \ge \sqrt{n}/s$. Then,*

$$\left|\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,s,\mathbf{c}}}[e^{2\pi i\langle\mathbf{x},\mathbf{v}\rangle}]\right| \quad\le\quad \frac{1+\epsilon}{1-\epsilon}\cdot 2^{-n}.$$

Proof: Define $\Lambda' = \Lambda/s$, $\mathbf{c}' = \mathbf{c}/s$, and $\mathbf{v}' = s\mathbf{v}$. As in the proof of Lemma 4.2, the distribution $D_{\Lambda,s,\mathbf{c}}$ is equal to $D_{\Lambda',\mathbf{c}'}$ scaled by a factor of s. Therefore, it is enough to prove the lemma for the case $s = 1$.

Define the function

$$g(\mathbf{x}) \quad = \quad e^{2\pi i\langle\mathbf{x},\mathbf{v}\rangle}\cdot\rho_{\mathbf{c}}(\mathbf{x})$$

and notice that

$$\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[e^{2\pi i\langle\mathbf{x},\mathbf{v}\rangle}] = \frac{g(\Lambda)}{\rho_{\mathbf{c}}(\Lambda)}.$$

Applying Poisson's summation formula (Lemma 2.8) to the numerator and denominator, the above fraction can be rewritten as

$$\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[e^{2\pi i\langle\mathbf{x},\mathbf{v}\rangle}] = \frac{\det(\Lambda^*)\cdot\widehat{g}(\Lambda^*)}{\det(\Lambda^*)\cdot\widehat{\rho_{\mathbf{c}}}(\Lambda^*)} = \frac{\widehat{g}(\Lambda^*)}{\widehat{\rho_{\mathbf{c}}}(\Lambda^*)}. \tag{11}$$

As in the proof of Lemma 4.2, we have $\widehat{\rho_{\mathbf{c}}}(\mathbf{y}) = \rho(\mathbf{y})e^{-2\pi i\langle\mathbf{y},\mathbf{c}\rangle}$ and $|\widehat{\rho_{\mathbf{c}}}(\Lambda^*)| \ge 1-\rho(\Lambda^*\setminus\{\mathbf{0}\})$. By Equation 3, the Fourier transform of g is given by

$$\widehat{g}(\mathbf{y}) = \widehat{\rho_{\mathbf{c}}}(\mathbf{y}-\mathbf{v}) = \rho(\mathbf{y}-\mathbf{v})e^{-2\pi i\langle\mathbf{y}-\mathbf{v},\mathbf{c}\rangle}.$$

Combined with (11), we obtain

$$\left|\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[e^{2\pi i\langle\mathbf{x},\mathbf{v}\rangle}]\right| \le \frac{\rho(\Lambda^*-\mathbf{v})}{1-\rho(\Lambda^*\setminus\{\mathbf{0}\})}.$$

Since $\mathrm{dist}(\mathbf{v}, \Lambda^*) \ge \sqrt{n}$, Lemma 2.10 with $c = 1$ implies that

$$\rho(\Lambda^*-\mathbf{v}) \le 2^{-n}\rho(\Lambda^*) = 2^{-n}(1+\rho(\Lambda^*\setminus\{\mathbf{0}\})$$

so we have

$$\left|\operatorname*{Exp}_{\mathbf{x}\sim D_{\Lambda,\mathbf{c}}}[e^{2\pi i\langle\mathbf{x},\mathbf{v}\rangle}]\right| \le 2^{-n}\frac{1+\rho(\Lambda^*\setminus\{\mathbf{0}\})}{1-\rho(\Lambda^*\setminus\{\mathbf{0}\})}.$$

■

Using the lemma, we obtain the following easy corollary.

Corollary 4.6 *Let Λ be an n-dimensional lattice, $\mathbf{w}, \mathbf{c}, \mathbf{v} \in \mathbb{R}^n$, $0 < \epsilon < 1$ and $s \geq \eta_\epsilon(\Lambda)$ such that $\mathrm{dist}(\mathbf{v}, \Lambda^*) \geq \sqrt{n}/s$. Then,*

$$\Big| \operatorname*{Exp}_{\mathbf{x} \sim D_{\Lambda,s,\mathbf{c}}} [\cos(2\pi\langle \mathbf{x}+\mathbf{w}, \mathbf{v}\rangle)] \Big| \quad \leq \quad \frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n}$$

Proof:

$$\begin{aligned} \Big| \operatorname*{Exp}_{\mathbf{x} \sim D_{\Lambda,s,\mathbf{c}}} [\cos(2\pi\langle \mathbf{x}+\mathbf{w}, \mathbf{v}\rangle)] \Big| &= \Big| \Re\Big(\operatorname*{Exp}_{\mathbf{x} \sim D_{\Lambda,s,\mathbf{c}}} [e^{2\pi i \langle \mathbf{x}+\mathbf{w}, \mathbf{v}\rangle}] \Big) \Big| \\ &\leq \Big| \operatorname*{Exp}_{\mathbf{x} \sim D_{\Lambda,s,\mathbf{c}}} [e^{2\pi i \langle \mathbf{x}+\mathbf{w}, \mathbf{v}\rangle}] \Big| \\ &= \Big| \operatorname*{Exp}_{\mathbf{x} \sim D_{\Lambda,s,\mathbf{c}}} [e^{2\pi i \langle \mathbf{x}, \mathbf{v}\rangle}] \Big| \end{aligned}$$

■

5 Worst-case to Average-case Connection

In this section we show that if various lattice problems are hard to solve in the worst case, then a certain computational problem is hard to solve on the average. We start in Subsection 5.1 with a description of the average-case problem. We then describe our reductions. Following [21, 22], the reductions are performed in two steps. First, in Subsection 5.2, we present a reduction from an intermediate worst-case lattice problem to the average-case problem. This is the core of our proof. Then, in Subsection 5.3, we show that the intermediate worst-case lattice problem is at least as hard as various other computational problems on lattices, such as SIVP and GapCRP. We remark that the intermediate worst-case problem is introduced to present the worst-case to average-case reduction in a simpler setting where the worst-case algorithm makes a single call to the average-case oracle. This allows for a cleaner and simpler probabilistic analysis, and it is well worth the effort of introducing one additional and perhaps artificial problem. Work prior to [21, 22] reduced standard worst-case lattice problems (like SIVP) directly to the average-case problem by making (polynomially) many random calls to the average-case oracle, resulting in an overall more complex probabilistic argument.

In Subsection 5.4 we present our reduction from $\mathrm{GapSVP}_{\tilde{O}(n)}$ to the average-case problem. The proof of this result requires some additional machinery, and relies on the results proved in Subsections 5.2 and 5.3 as well as techniques from [1]. We remark that a weaker result can be derived directly from the results in Subsection 5.3. Namely, using standard reductions between lattice problems (see [23, Theorem 7.12]), our $\tilde{O}(n)$ approximation to SIVP immediately implies a $\tilde{O}(n^2)$ approximation to GapSVP. Hence, Subsection 5.4 is only needed in order to reduce the approximation factor to $\tilde{O}(n)$.

5.1 The average-case problem

Our average-case problem is the problem of finding small nonzero solutions to random linear systems of modular equations.

Definition 5.1 *The* small integer solution *problem* SIS *(in the ℓ_2 norm) is: given an integer q, a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and a real β, find a nonzero integer vector $\mathbf{z} \in \mathbb{Z}^m \setminus \{\mathbf{0}\}$ such that $\mathbf{A}\mathbf{z} = \mathbf{0} \bmod q$ and $\|\mathbf{z}\| \leq \beta$.*

Equivalently, the SIS problem asks to find a vector $\mathbf{z} \in \Lambda_q(\mathbf{A}) \setminus \{\mathbf{0}\}$ with $\|\mathbf{z}\| \leq \beta$ where

$$\Lambda_q(\mathbf{A}) = \{\mathbf{z} \in \mathbb{Z}^m \colon \mathbf{A}\mathbf{z} = \mathbf{0} \bmod q\}$$

is the set of all integer solutions to the system of linear equations modulo q defined by matrix $\mathbf{A}$.

In the definition of SIS it is implicitly assumed that a solution of length $\|\mathbf{z}\| \leq \beta$ exists (for otherwise the problem is trivially hard). The following lemma gives sufficient conditions under which SIS instances are guaranteed to have a solution.

Lemma 5.2 *For any q, $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and $\beta \geq \sqrt{m} q^{n/m}$, the* SIS *instance $(q, \mathbf{A}, \beta)$ admits a solution, i.e., there exists a vector $\mathbf{z} \in \mathbb{Z}^m \setminus \{\mathbf{0}\}$ such that $\mathbf{A}\mathbf{z} = \mathbf{0} \bmod q$ and $\|\mathbf{z}\| \leq \beta$.*

Proof: The proof is by the pigeon-hole principle. Consider all vectors $\mathbf{z} \in \mathbb{Z}^m$ with coordinates in $\{0, \ldots, q^{n/m}\}$. There are more than q^n such vectors and hence there must exist two such vectors $\mathbf{z}_1 \neq \mathbf{z}_2$ for which $\mathbf{A}\mathbf{z}_1 = \mathbf{A}\mathbf{z}_2 \bmod q$. Then, $\mathbf{z}_1 - \mathbf{z}_2 \neq \mathbf{0}$ satisfies $\mathbf{A}(\mathbf{z}_1 - \mathbf{z}_2) = \mathbf{0} \bmod q$ and moreover, $\|\mathbf{z}_1 - \mathbf{z}_2\| \leq \sqrt{m} q^{n/m}$ since all its coordinates are between $-q^{n/m}$ and $q^{n/m}$. ■

We want to study the average-case complexity of the SIS problem when $\beta \geq \sqrt{m} q^{n/m}$ satisfies the condition in Lemma 5.2, and SIS instances $(q, \mathbf{A}, \beta)$ are guaranteed to have a solution. In order to define probability ensembles over SIS instances, it is convenient to use the number of equations n as a security parameter, and consider families of SIS instances indexed by functions $q(n)$, $m(n)$ and $\beta(n)$ that express the other parameters in terms of n.

Definition 5.3 *For any functions $q(n)$, $m(n)$ and $\beta(n)$, let*

$$\mathrm{SIS}_{q,m,\beta} = \{(q(n), U(\mathbb{Z}_{q(n)}^{n \times m(n)}), \beta(n))\}_n$$

be the probability ensemble over SIS *instances $(q(n), \mathbf{A}, \beta(n))$ where $\mathbf{A}$ is chosen uniformly at random among all $n \times m(n)$ integer matrices modulo $q(n)$. When $\beta(n) = \sqrt{m(n)} q(n)^{n/m(n)}$ is the bound specified in Lemma 5.2, the parameter $\beta(n)$ is often omitted, and we simply write* $\mathrm{SIS}_{q,m}$.

Notice that for the instances of $\mathrm{SIS}_{q,m,\beta}$ to be of size polynomial in n, the number of variables must be a polynomially bounded function $m(n) = n^{O(1)}$, but $q(n)$ and $\beta(n)$ can be exponentially large. However, we will be mostly interested in instances where $q(n)$ and $\beta(n)$ are also polynomially bounded functions of the security parameter n. Moreover, we typically choose values of q and m satisfying $q(n)^{n/m(n)} = O(1)$, so that $\beta(n) = \sqrt{m(n)} q(n)^{n/m(n)} = O(\sqrt{m(n)})$. In the next two subsections we show that for an appropriate choice of parameters q, m and β, solving $\mathrm{SIS}_{q,m,\beta}$ on the average is as hard as solving worst-case instances of several standard lattice problems such as SIVP and GapCRP. The reduction from GapSVP is shown in Subsection 5.4. For technical reasons, in that reduction we need to consider a variant of the SIS problem, defined below, which extends SIS with the additional requirement that the solution vector must contain at least one odd coordinate.

Definition 5.4 *The* SIS′ *problem (in the ℓ_2 norm) is: given an integer q, a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and a real β, find an integer vector $\mathbf{z} \in \mathbb{Z}^m \setminus 2\mathbb{Z}^m$ such that $\mathbf{A}\mathbf{z} = \mathbf{0} \bmod q$ and $\|\mathbf{z}\| \leq \beta$.*

The distribution ensemble $\mathrm{SIS}'_{q,m,\beta} = \{(q(n), \mathbf{A}, \beta(n))\}_n$ is defined analogously to $\mathrm{SIS}_{q,m,\beta}$ by choosing matrix $\mathbf{A} \in \mathbb{Z}_{q(n)}^{n \times m(n)}$ uniformly at random. Similarly, when $\beta(n) = \sqrt{m(n)} q(n)^{n/m(n)}$, we omit the parameter β, and simply write $\mathrm{SIS}'_{q,m}$. Clearly, any solution to $\mathrm{SIS}'_{q,m,\beta}$ is also a solution to $\mathrm{SIS}_{q,m,\beta}$ because $\mathbf{0} \notin \mathbb{Z}^m \setminus 2\mathbb{Z}^m$. The next lemma shows that when the modulus $q(n)$ is odd, $\mathrm{SIS}'_{q,m,\beta}$ is not any harder than $\mathrm{SIS}_{q,m,\beta}$.

Lemma 5.5 *For any odd integer $q \in 2\mathbb{Z} + 1$, and* SIS′ *instance $I = (q, \mathbf{A}, \beta)$, if I has a solution as an instance of* SIS*, then it also has a solution as an instance of* SIS′*. Moreover, there is a polynomial time algorithm that on input a solution to a* SIS *instance I, outputs a solution to the same* SIS′ *instance I.*

Proof: Assume q is odd, and let $\mathbf{z}$ be a solution to SIS instance $(q, \mathbf{A}, \beta)$, i.e., assume $\mathbf{z} \in \mathbb{Z}^m \setminus \{\mathbf{0}\}$, $\mathbf{Az} = \mathbf{0} \bmod q$ and $\|\mathbf{z}\| \leq \beta$. Compute the largest power i such that 2^i divides all the coordinates of $\mathbf{z}$, and output $\mathbf{z}/2^i$. Since $\mathbf{z}$ is nonzero, i is well defined and can be easily computed. Moreover, $\mathbf{z}/2^i$ has at least one odd coordinate, and since the modulus $q(n)$ is odd, $\mathbf{z}/2^i$ satisfies $\mathbf{A}(\mathbf{z}/2^i) = \mathbf{0} \bmod q$. ■

We end this subsection with two simple observations on the average-case hardness of SIS. These observations are not used in the following subsections and can be safely skipped at first reading.

First, observe that for any $\mathbf{A}$, $\Lambda_q(\mathbf{A})$ forms a lattice. Therefore, the SIS problem is closely related to the shortest vector problem (SVP) on lattices of the form $\Lambda_q(\mathbf{A})$. More specifically, finding shortest nonzero vectors in a random lattice $\Lambda_q(\mathbf{A})$ is at least as hard as solving $\mathrm{SIS}_{q,m}$ on the average. So, all our results can be formulated as reductions from solving various lattice problems (including GAPSVP_γ for factors $\gamma(n) = \tilde{O}(n)$) in the worst case to solving SVP on the average (for random lattices of the form $\Lambda_q(\mathbf{A})$).

Next, we observe that SIS can be reduced to the problem of finding collisions for an appropriately defined family of hash functions. For any $q(n), m(n)$ and $d(n)$, define the family of functions

$$\mathcal{H}_{q,m,d} = \{f_{\mathbf{A}} \colon \{0, \ldots, d(n)-1\}^{m(n)} \to \mathbb{Z}_{q(n)}^n \mid \mathbf{A} \in \mathbb{Z}_{q(n)}^{n \times m(n)}\}$$

where n is a security parameter and $f_{\mathbf{A}}(\mathbf{x}) = \mathbf{Ax} \bmod q(n)$. A typical choice of parameters is $q(n) = n^{O(1)}$, $d(n) = 2$ and $m(n) > n \log_2 q(n) = \Theta(n \log n)$. A collision is a pair of distinct inputs $\mathbf{x} \neq \mathbf{y}$ (both in the domain $\{0, \ldots, d(n)-1\}^{m(n)}$ of $f_{\mathbf{A}}$) that are mapped to the same output $f_{\mathbf{A}}(\mathbf{x}) = f_{\mathbf{A}}(\mathbf{y})$. Notice that if $m(n) > n \log_{d(n)} q(n)$, then the domain $\{0, \ldots, d(n)-1\}^{m(n)}$ is larger than the range $\mathbb{Z}_{q(n)}^n$, and, by the pigeon hole principle, the functions $f_{\mathbf{A}}$ are guaranteed to have collisions $(\mathbf{x}, \mathbf{y})$. We argue that these collisions are computationally hard to find when $\mathbf{A}$ is chosen at random. Observe that if $(\mathbf{x}, \mathbf{y})$ is a collision for $f_{\mathbf{A}}$, then $\mathbf{z} = \mathbf{x} - \mathbf{y} \in \Lambda_q(\mathbf{A}) \setminus \{\mathbf{0}\}$ is a nonzero lattice vector of length at most $\beta(n) = (d(n)-1)\sqrt{m(n)}$. So, finding collisions on the average when $\mathbf{A}$ is chosen uniformly at random is at least as hard as solving random instances of $\mathrm{SIS}_{q,m,\beta}$ for the same value of $q(n)$ and $m(n)$, and $\beta(n) = (d(n)-1)\sqrt{m(n)}$. This gives collision resistant hash functions that are provably secure based on the worst-case intractability assumption of lattice approximation problems (e.g., SIVP_γ, GAPCRP_γ, GAPSVP_γ) for approximation factors $\gamma(n) = \tilde{O}(n)$ almost linear in the dimension of the lattice.

5.2 Incremental guaranteed distance decoding

In this section we show that solving SIS on the average with non-negligible probability is at least as hard as solving worst-case instances of the following INCGDD problem (originally introduced in [22] in a slightly different form). We remind the reader that we introduce INCGDD for the sole purpose of simplifying the worst-case to average-case reduction. In particular, we will show that INCGDD can be solved (in the worst-case) by making a single call to the average-case SIS oracle, resulting in a simpler probabilistic analysis compared to reductions that make several oracle calls. In the next subsection we show that several other more interesting lattice problems (like SIVP and GAPCRP) can be solved in the worst-case by making many calls to an INCGDD oracle. Although these reductions require the solution of several INCGDD instances, they are conceptually easier to analyze because they are standard worst-case to worst-case reductions.

Definition 5.6 (Incremental Guaranteed Distance Decoding) *An input to* $\mathrm{INCGDD}_{\gamma,g}^{\phi}$ *is an* n*-dimensional lattice basis* $\mathbf{B}$*, a set of* n *linearly independent vectors* $\mathbf{S} \subset \mathcal{L}(\mathbf{B})$*, a target point* $\mathbf{t}$*, and a real* $r > \gamma(n) \cdot \phi(\mathbf{B})$*. The goal is to output a lattice vector* $\mathbf{s} \in \mathcal{L}(\mathbf{B})$ *such that* $\|\mathbf{s} - \mathbf{t}\| \leq (\|\mathbf{S}\|/g) + r$*.*

In other words, the INCGDD problem asks to find a lattice vector within distance $(\|\mathbf{S}\|/g) + r$ from the given target. One possible choice of parameters is, for example, $g = 4$, $\gamma(n) = \sqrt{n}/2$ and $\phi = \lambda_n$. Often, $\|\mathbf{S}\|$ is much larger than r, so the dominant part in the distance bound is $\|\mathbf{S}\|/g$, or $\|\mathbf{S}\|/4$ for our choice

of parameters. Notice that using the nearest plane algorithm [4] one can always find (in polynomial time) a lattice point within distance $(\sqrt{n}/2)\|\mathbf{S}\|$ from any target. Here we are trying to do much better than that. However, it is not always possible to find a lattice vector within distance $\|\mathbf{S}\|/4$ of a given target vector: for example, consider the integer lattice $\mathbb{Z}^n$ generated by the identity matrix $\mathbf{B} = \mathbf{I}$. If we choose the set $\mathbf{S} = \mathbf{I}$ and the target point $\mathbf{t} = (1/2, \ldots, 1/2)$ then there is no lattice point at distance strictly less than $\sqrt{n}/2 = (\sqrt{n}/2)\|\mathbf{S}\|$ from the target. The r term in the distance bound of the IncGDD problem is introduced to guarantee the existence of a solution. For example, using the above choice of parameters, we get $r > \gamma(n)\phi(\mathbf{B}) = (\sqrt{n}/2)\lambda_n(\mathbf{B})$, and a lattice point within this distance always exists by the nearest plane algorithm. To summarize, one can think of IncGDD as asking to find a lattice point within distance roughly $\|\mathbf{S}\|/g$ from the target, provided $\|\mathbf{S}\|$ is not too small.

We now give a high-level overview of the reduction. Our goal is to reduce worst-case instances of IncGDD to random instances of SIS. In other words, we want to solve an IncGDD instance $(\mathbf{B}, \mathbf{S}, \mathbf{t}, r)$ with the help of an oracle $\mathcal{F}$ that on input a random matrix $\mathbf{A}$, returns with non-negligible probability a short nonzero integer vector $\mathbf{z}$ such that $\mathbf{Az} = \mathbf{0}$. To fix some parameters, assume that we want to reduce IncGDD with, say, $g = 4$ (we ignore γ and ϕ in this discussion) to SIS with $q(n) = n^4$, $m(n) = n \log n$, and $\beta(n) = n$ (it is easy to check that for large enough n, this choice satisfies the conditions in Lemma 5.2). For now, let us make two simplifying assumption: the target vector $\mathbf{t}$ is the origin $\mathbf{0}$, and $\mathbf{S} = \mathbf{B}$. Although the former assumption makes the IncGDD instance trivial ($\mathbf{0} \in \mathcal{L}(\mathbf{B})$ is always a solution), it helps in explaining the main ideas in the reduction. We will later indicate how to avoid these assumptions.

With these assumptions in place, we can describe a simplified form of the reduction. At the core of the reduction is a sampling procedure $\mathcal{S}$. This procedure generates a pair $(\mathbf{c}, \mathbf{y})$ where $\mathbf{c}$ is distributed uniformly in $\mathcal{P}(\mathbf{B})$ and $\mathbf{y} \in \mathcal{L}(\mathbf{B})$ is a lattice vector close to $\mathbf{c}$. The reduction starts by applying the sampling procedure m times to obtain m pairs $(\mathbf{c}_1, \mathbf{y}_1), \ldots, (\mathbf{c}_m, \mathbf{y}_m)$. We then partition the parallelepiped $\mathcal{P}(\mathbf{B})$ into q^n smaller parallelepipeds, naturally corresponding to elements of $\mathbb{Z}_q^n$. For each $\mathbf{c}_i$, let $\tilde{\mathbf{c}}_i$ be the 'lower-left' corner of the parallelepiped of $\mathbf{c}_i$, that is, $\tilde{\mathbf{c}}_i = \mathbf{B}\lfloor q \cdot \mathbf{B}^{-1}\mathbf{c}_i \rfloor / q$. Notice that the distance between $\mathbf{c}_i$ and $\tilde{\mathbf{c}}_i$ is at most $n\|\mathbf{B}\|/q = \|\mathbf{B}\|/n^3$. Next, let $\mathbf{a}_i \in \mathbb{Z}_q^n$ be the group element corresponding to the parallelepiped that contains $\mathbf{c}_i$. More precisely, we define $\mathbf{a}_i = \lfloor q \cdot \mathbf{B}^{-1}\mathbf{c}_i \rfloor \bmod q$. Since each $\mathbf{c}_i$ is uniformly distributed in $\mathcal{P}(\mathbf{B})$, each $\mathbf{a}_i$ is uniformly distributed in $\mathbb{Z}_q^n$. We can therefore apply the oracle $\mathcal{F}$ to the matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$ to find a small combination of the $\mathbf{a}_i$ that sums to zero in $\mathbb{Z}_q^n$. That is, we find a vector $\mathbf{z}$ such that $\mathbf{Az} = \mathbf{0}$ and $\|\mathbf{z}\|_1 \leq \sqrt{m}\|\mathbf{z}\|_2 \leq \sqrt{m}\beta \leq n^2$. Crucially, the same combination applied to $\tilde{\mathbf{c}}_i$ yields a lattice vector: if we denote by $\tilde{\mathbf{C}}$ the matrix $[\tilde{\mathbf{c}}_1, \ldots, \tilde{\mathbf{c}}_m]$, we see that $\tilde{\mathbf{C}}\mathbf{z} \in \mathcal{L}(\mathbf{B})$. We complete the argument by noting that the vector $\mathbf{Cz}$ is close to both $\tilde{\mathbf{C}}\mathbf{z}$ and $\mathbf{Yz}$ (where $\mathbf{C}$ and $\mathbf{Y}$ are defined similarly to $\tilde{\mathbf{C}}$). Since the latter two vectors are lattice vectors, we obtain that $(\tilde{\mathbf{C}} - \mathbf{Y})\mathbf{z}$ is a lattice vector close to $\mathbf{0}$, as required. In slightly more detail, it turns out that the dominant part in the distance between $\tilde{\mathbf{C}}\mathbf{z}$ and $\mathbf{Yz}$ is typically that between $\mathbf{Cz}$ and $\tilde{\mathbf{C}}\mathbf{z}$. By a triangle inequality, this distance is at most $\|\mathbf{z}\|_1\|\mathbf{B}\|/n^3 \leq \|\mathbf{B}\|/n \ll \|\mathbf{B}\|/4$, hence we obtain a solution to IncGDD.

Let us indicate how to avoid the two simplifying assumptions we have made. First, IncGDD asks not for a lattice vector close to the origin, but for a lattice vector close to a given target $\mathbf{t}$. This is taken care of by modifying the sampling procedure so that it outputs a pair $(\mathbf{c}, \mathbf{y})$ where $\mathbf{y}$ is close to $\mathbf{c} + \mathbf{t}'$ (instead of $\mathbf{c}$) where $\mathbf{t}'$ is now an input to the sampling procedure. By carefully choosing the vectors $\mathbf{t}'$ used in each of the m applications of the sampling procedure, we can guarantee that with some reasonable probability, the output of the reduction will be a vector close to $\mathbf{t}$. The second issue to consider is that in general, $\mathbf{S}$ is not equal to $\mathbf{B}$ and typically, $\|\mathbf{S}\| \ll \|\mathbf{B}\|$. This is taken care of by first mapping the vectors $\mathbf{c}_i$ to the parallelepiped $\mathcal{P}(\mathbf{S})$ and then partitioning $\mathcal{P}(\mathbf{S})$ into q^n smaller parallelepipeds, as we did before with $\mathcal{P}(\mathbf{B})$. This makes the dominant distance roughly $\|\mathbf{S}\|/q$, as required. The mapping requires some care, as we want to map the uniform distribution over $\mathcal{P}(\mathbf{B})$ to the uniform distribution over $\mathcal{P}(\mathbf{S})$. Finally, let us mention that although we have ignored so far the distance between $\mathbf{Cz}$ and $\mathbf{Yz}$, this distance ends up determining the approximation factor achieved by the reduction. Because of this, in the reduction below we will make an effort to give a good bound on this distance.

We can now describe the reduction in more detail. We start with the sampling procedure $\mathcal{S}$. This procedure takes as input a lattice $\mathbf{B}$ and two additional parameters $\mathbf{t}$ and s. Provided s is not too small, the output of the procedure is a pair of vectors $(\mathbf{c}, \mathbf{y})$ with the following properties. The distribution of $\mathbf{c}$ is very close to uniform on $\mathcal{P}(\mathbf{B})$. The vector $\mathbf{y}$ is a lattice vector distributed according to a discrete Gaussian distribution with parameter s around $\mathbf{t} + \mathbf{c}$. Since s is typically small, we can think of the procedure as outputting a uniform vector $\mathbf{c} \in \mathcal{P}(\mathbf{B})$ and a lattice vector $\mathbf{y}$ close to $\mathbf{c} + \mathbf{t}$.

Lemma 5.7 (Sampling Lemma) *There is a probabilistic polynomial time algorithm $\mathcal{S}(\mathbf{B}, \mathbf{t}, s)$ that on input an n-dimensional lattice $\mathbf{B} \in \mathbb{R}^{n \times n}$, a vector $\mathbf{t} \in \mathbb{R}^n$, and a real $s \geq \eta_\epsilon(\mathbf{B})$ (for some $\epsilon > 0$), outputs a pair of vectors $(\mathbf{c}, \mathbf{y}) \in \mathcal{P}(\mathbf{B}) \times \mathcal{L}(\mathbf{B})$ such that*

- *the distribution of vector $\mathbf{c}$ is within statistical distance $\Delta(\mathbf{c}, U(\mathcal{P}(\mathbf{B}))) \leq \epsilon/2$ from the uniform distribution over $\mathcal{P}(\mathbf{B})$;*
- *for any $\hat{\mathbf{c}} \in \mathcal{P}(\mathbf{B})$, the conditional distribution of $\mathbf{y}$ given $\mathbf{c} = \hat{\mathbf{c}}$ is $D_{\mathcal{L}(\mathbf{B}), s, (\mathbf{t}+\hat{\mathbf{c}})}$.*

Proof: The sampling procedure $\mathcal{S}(\mathbf{B}, \mathbf{t}, s)$ is the following:

1. Generate a noise vector $\mathbf{r}$ with probability density $D_{s,\mathbf{t}}$.
2. Output $\mathbf{c} = -\mathbf{r} \bmod \mathcal{P}(\mathbf{B})$ and $\mathbf{y} = \mathbf{r} + \mathbf{c}$.

For the first property, notice that by Lemma 4.1 and $s \geq \eta_\epsilon(\mathbf{B})$, the statistical distance between the distribution of $\mathbf{c}$ and the uniform distribution is at most

$$\begin{aligned} \Delta(\mathbf{c}, U(\mathcal{P}(\mathbf{B}))) &= \Delta(-D_{s,\mathbf{t}} \bmod \mathcal{P}(\mathbf{B}), U(\mathcal{P}(\mathbf{B}))) \\ &= \Delta(D_{s,-\mathbf{t}} \bmod \mathcal{P}(\mathbf{B}), U(\mathcal{P}(\mathbf{B}))) \\ &\leq \epsilon/2. \end{aligned}$$

For the second property, fix any $\hat{\mathbf{c}} \in \mathcal{P}(\mathbf{B})$. Then, by definition, the distribution of $\mathbf{r} + \hat{\mathbf{c}}$ is $D_{s,\mathbf{t}+\hat{\mathbf{c}}}$. Conditioning on $\mathbf{c} = \hat{\mathbf{c}}$ is the same as conditioning on $\mathbf{r} + \hat{\mathbf{c}} \in \mathcal{L}(\mathbf{B})$. As discussed in Section 2, the distribution of $\mathbf{r} + \hat{\mathbf{c}}$ conditioned on $\mathbf{r} + \hat{\mathbf{c}} \in \mathcal{L}(\mathbf{B})$ is $D_{\mathcal{L}(\mathbf{B}), s, \mathbf{t}+\hat{\mathbf{c}}}$, as required. ■

Next, we describe a procedure which we call the combining procedure $\mathcal{A}$. This procedure is the heart of the worst-case to average-case reduction. It maps the vectors $\mathbf{c}_i$ to vectors in the parallelepiped $\mathcal{P}(\mathbf{S})$ and group elements $\mathbf{a}_i$, and then applies the oracle $\mathcal{F}$. At first reading, we suggest to skip the proof of the lemma, and jump directly to Theorem 5.9.

Lemma 5.8 (Combining Procedure) *There is a probabilistic polynomial time oracle algorithm $\mathcal{A}^{\mathcal{F}}(\mathbf{B}, \mathbf{S}, \mathbf{C}, q)$ that on input an n-dimensional lattice $\mathbf{B} \in \mathbb{R}^{n \times n}$, a full-rank sublattice $\mathbf{S} \subset \mathcal{L}(\mathbf{B})$, m vectors $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_m] \in \mathcal{P}(\mathbf{B})^m$, and a positive integer q, makes a single oracle call $\mathcal{F}(\mathbf{A}) = \mathbf{z}$ (with $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$) and outputs a vector $\mathbf{x} \in \mathbb{R}^n$ such that*

- *if the input matrix $\mathbf{C} \in \mathcal{P}(\mathbf{B})^m$ is distributed uniformly at random, then the query matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ is also uniformly distributed;*
- *if the oracle's answer $\mathbf{z} = \mathcal{F}(\mathbf{A})$ is in $\Lambda_q(\mathbf{A})$, then the output vector $\mathbf{x}$ belongs to the lattice $\mathcal{L}(\mathbf{B})$;*
- *the distance between the output vector $\mathbf{x}$ and $\mathbf{Cz}$ is at most $\sqrt{m}n\|\mathbf{S}\| \cdot \|\mathbf{z}\|/q$.*

Proof: The procedure $\mathcal{A}^{\mathcal{F}}(\mathbf{B}, \mathbf{S}, \mathbf{C}, q)$ is the following (see also the box labelled $\mathcal{A}^{\mathcal{F}}$ in Figure 2).

1. Generate m uniformly random lattice vectors $\mathbf{v}_i \in \mathcal{L}(\mathbf{B}) \bmod \mathcal{P}(\mathbf{S})$ (this can be done using standard techniques, see for example [21, Proposition 2.9]).

2. Define the matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ where $\mathbf{w}_i = \mathbf{v}_i + \mathbf{c}_i \bmod \mathcal{P}(\mathbf{S})$ for all $i = 1, \ldots, m$.
3. Define the query $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$ where $\mathbf{a}_i = \lfloor q \cdot \mathbf{S}^{-1}\mathbf{w}_i \rfloor \in \mathbb{Z}_q^n$ for all $i = 1, \ldots, m$.
4. Invoke the oracle $\mathcal{F}$ on input $\mathbf{A}$ to obtain an integer vector $\mathbf{z} = \mathcal{F}(\mathbf{A})$.
5. Output the vector $\mathbf{x} = (\mathbf{C} - \mathbf{W} + \mathbf{SA}/q)\mathbf{z}$.

We now prove the first property. We start by noting that if $\mathbf{c}$ is uniformly distributed in $\mathcal{P}(\mathbf{B})$ and $\mathbf{v}$ is chosen uniformly from the vectors in $\mathcal{L}(\mathbf{B}) \bmod \mathcal{P}(\mathbf{S})$, then $\mathbf{c} + \mathbf{v} \bmod \mathcal{P}(\mathbf{S})$ is distributed uniformly in $\mathcal{P}(\mathbf{S})$. This holds since the sets $(\mathbf{v} + \mathcal{P}(\mathbf{B})) \bmod \mathcal{P}(\mathbf{S})$ for all $\mathbf{v} \in \mathcal{L}(\mathbf{B}) \bmod \mathcal{P}(\mathbf{S})$ form a partition of $\mathcal{P}(\mathbf{S})$ into sets of equal volume. Thus, we see that if $\mathbf{C} \in \mathcal{P}(\mathbf{B})^m$ is distributed uniformly then $\mathbf{W}$ is distributed uniformly in $\mathcal{P}(\mathbf{S})^m$. From this, it easily follows that $\mathbf{A}$ is distributed uniformly in $\mathbb{Z}_q^{n \times m}$, as required.

We now prove the second property. Assume $\mathbf{z} \in \Lambda_q(\mathbf{A})$, and consider the output vector

$$\mathbf{x} = (\mathbf{C} - \mathbf{W} + \mathbf{SA}/q)\mathbf{z} = \sum_{i=1}^{m} (\mathbf{c}_i - \mathbf{w}_i) z_i + \mathbf{S}(\mathbf{Az}/q).$$

Notice that for any $i = 1, \ldots, m$ the vector

$$\mathbf{c}_i - \mathbf{w}_i = ((\mathbf{c}_i + \mathbf{v}_i) - \mathbf{w}_i) - \mathbf{v}_i$$

belongs to the lattice $\mathcal{L}(\mathbf{B})$ because $\mathbf{c}_i + \mathbf{v}_i \equiv \mathbf{w}_i$ modulo $\mathcal{L}(\mathbf{S}) \subseteq \mathcal{L}(\mathbf{B})$ and $\mathbf{v}_i \in \mathcal{L}(\mathbf{B})$. Also, $\mathbf{Az}/q$ is an integer vector because $\mathbf{Az} = \mathbf{0} \bmod q$. This proves that $\mathbf{x}$ belongs to the lattice $\mathcal{L}(\mathbf{B})$ because it is an integer linear combination of lattice vectors.

For the third property, we bound the distance between $\mathbf{x}$ and $\mathbf{Cz}$ as follows:

$$\begin{aligned} \|\mathbf{x} - \mathbf{Cz}\| &= \Big\| \sum_{i=1}^{m} (\mathbf{w}_i - (\mathbf{S}/q)\mathbf{a}_i) z_i \Big\| \\ &= \frac{1}{q} \Big\| \mathbf{S} \sum_{i=1}^{m} (\mathbf{u}_i - \lfloor \mathbf{u}_i \rfloor) z_i \Big\|, \end{aligned}$$

where $\mathbf{u}_i = q\mathbf{S}^{-1}\mathbf{w}_i$. Since for each i, all entries of $\mathbf{u}_i - \lfloor \mathbf{u}_i \rfloor$ are bounded by 1, the vector $\sum_{i=1}^{m} (\mathbf{u}_i - \lfloor \mathbf{u}_i \rfloor) z_i$ has all entries bounded by $\sum_i |z_i| \leq \sqrt{m}\|\mathbf{z}\|$. It follows by triangle inequality that

$$\Big\| \mathbf{S} \sum_{i=1}^{m} (\mathbf{u}_i - \lfloor \mathbf{u}_i \rfloor) z_i \Big\| \leq n\sqrt{m}\|\mathbf{z}\|\|\mathbf{S}\|$$

and $\|\mathbf{x} - \mathbf{Cz}\| \leq n\sqrt{m}\|\mathbf{z}\|\|\mathbf{S}\|/q$. ■

We are now ready to reduce IncGDD to SIS using the procedures $\mathcal{S}$ and $\mathcal{A}$ from the previous lemmas. In the theorem below, all parameters are implicitly assumed to have bit-size polynomial in the security parameter, i.e., $g(n), q(n) = 2^{n^{O(1)}}$. In fact, in most of the applications of this theorem considered in this paper, the parameters will be smaller, typically polynomial in n. For example, one can take, say, $g = 8$ to be a constant, $q(n) = n^3$ (or any other sufficiently large polynomial), $m(n) = n \log n$, and $\beta(n) = \sqrt{m(n)} q(n)^{n/m(n)} = 8\sqrt{n \log n}$ the bound from Lemma 5.2 so that $\mathrm{SIS}_{q,m,\beta}$ is guaranteed to admit a solution. It is easy to check that $q(n)$ satisfies the condition in the theorem below, yielding approximation factor $\gamma(n) = \beta(n)\sqrt{n} = 8n\sqrt{\log n} = O(n\sqrt{\log n})$.

Theorem 5.9 *For any function $g(n) > 0$, polynomially bounded functions $m(n), \beta(n) = n^{O(1)}$, negligible function $\epsilon(n) = n^{-\omega(1)}$, and $q(n) \geq g(n) n \sqrt{m(n)} \beta(n)$, there is a probabilistic polynomial time reduction from solving $\mathrm{IncGDD}^{\eta_\epsilon}_{\gamma,g}$ for $\gamma(n) = \beta(n)\sqrt{n}$ on n-dimensional instances in the worst case to solving $\mathrm{SIS}_{q,m,\beta}$ on the average with non-negligible probability.*

Proof: Many of our parameters depend on n; for notational convenience, we often omit this dependency and write m instead of $m(n)$, and similarly for $\gamma, \beta, g, q, \epsilon$, and δ. Let $\mathcal{F}$ be an oracle that solves $\mathrm{SIS}_{q,m,\beta}$ on the average. In other words, we assume that on input a uniformly random matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, the oracle call $\mathcal{F}(\mathbf{A})$ returns a nonzero vector $\mathbf{z} \in \Lambda_q(\mathbf{A})$ of length at most $\|\mathbf{z}\| \leq \beta$ with some non-negligible probability $\delta(n) = n^{-O(1)}$.

On input an $\textsc{IncGDD}^{\eta_\epsilon}_{\gamma,g}$ instance $(\mathbf{B}, \mathbf{S}, \mathbf{t}, r)$, the reduction performs the following operations (see Figure 2 for a high-level overview). The goal of the first step is to 'guess' how to choose the vectors $\mathbf{t}_i$ given to the sampling procedure. As we shall see later, with reasonable probability, this guess causes the output of the reduction to be a lattice vector close to $\mathbf{t}$.

1. Pick an index $j \in M = \{1, \ldots, m\}$ and integer $\alpha \in B = \{-\beta, \ldots, -1, 1, \ldots, \beta\}$ uniformly at random. For each $i \in M$, define the vector
$$\mathbf{t}_i = \begin{cases} -\mathbf{t}/\alpha & \text{if } i = j \\ \mathbf{0} & \text{otherwise} \end{cases}$$
2. For each $i = 1, \ldots, m$, compute the pair
$$(\mathbf{c}_i, \mathbf{y}_i) = \mathcal{S}(\mathbf{B}, \mathbf{t}_i, 2r/\gamma)$$
using the sampling procedure of Lemma 5.7, each time with independent randomness.
3. Define the matrices $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_m]$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m]$.
4. Finally, call the combining algorithm $\mathcal{A}^{\mathcal{F}}(\mathbf{B}, \mathbf{S}, \mathbf{C}, q) = \mathbf{x}$ using $\mathcal{F}$ as an oracle, and output the vector $\mathbf{s} = \mathbf{x} - \mathbf{Y}\mathbf{z}$, where $\mathbf{z} = \mathcal{F}(\mathbf{A})$ is the answer returned by $\mathcal{F}$ to $\mathcal{A}$'s query $\mathbf{A}$.

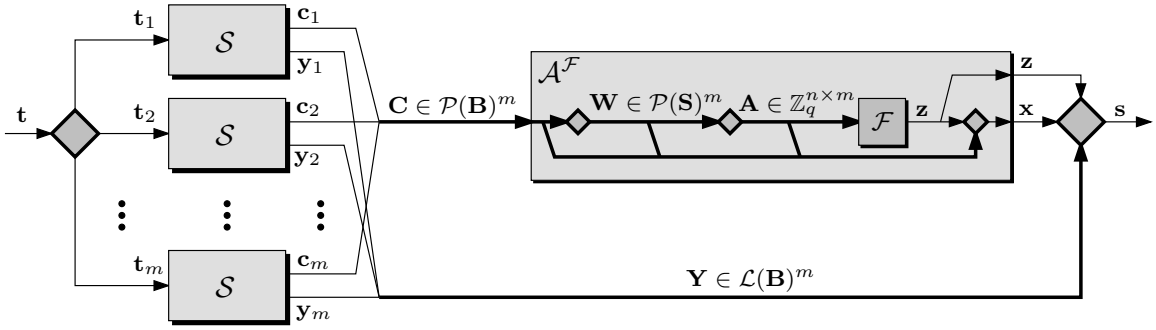

Figure 2: A diagram of the worst-case to average-case reduction

We want to bound from below the success probability of the reduction, i.e., the probability that the output vector $\mathbf{s}$ satisfies $\mathbf{s} \in \mathcal{L}(\mathbf{B})$ and $\|\mathbf{s} - \mathbf{t}\| \leq (\|\mathbf{S}\|/g) + r$. We start by finding some 'good' values for j and α. To this end, consider the output $\mathbf{z}' = \mathcal{F}(\mathbf{A}')$ of the oracle on a *uniformly random* input $\mathbf{A}' \in \mathbb{Z}_q^{n \times m}$. For each $j' \in M, \alpha' \in B$ denote by $\delta_{j',\alpha'}$ the probability that this output $\mathbf{z}'$ satisfies
$$(z'_{j'} = \alpha') \wedge (\mathbf{z}' \in \Lambda_q(\mathbf{A}') \setminus \{\mathbf{0}\}) \wedge (\|\mathbf{z}'\| \leq \beta).$$
Since any nonzero vector $\mathbf{z} \in \mathbb{Z}^m$ with $\|\mathbf{z}\| \leq \beta$ must have at least one coordinate in the set B,
$$\sum_{j' \in M, \alpha' \in B} \delta_{j',\alpha'} \geq \delta.$$
Hence, there must exist some $j' \in M, \alpha' \in B$ for which $\delta_{j',\alpha'} \geq \delta/2\beta m$ and is thus non-negligible. In the rest of the proof we consider the execution of the reduction for any fixed values of j and α and show that its

success probability is at least $\delta_{j,\alpha}/3$, up to a negligible term. In particular, for $j = j', \alpha = \alpha'$, the reduction is successful with a non-negligible probability. This would complete the proof since the event $j = j', \alpha = \alpha'$ happens with probability $1/2\beta m$, which is non-negligible.

So in the rest of the proof, fix some values of j and α. Let H be the event

$$H \stackrel{def}{=} (z_j = \alpha) \wedge (\mathbf{z} \in \Lambda_q(\mathbf{A}) \setminus \{\mathbf{0}\}) \wedge (\|\mathbf{z}\| \leq \beta) \tag{12}$$

where $\mathbf{A}$ and $\mathbf{z}$ are the random variables that appear in the reduction. In other words, H is the event that oracle $\mathcal{F}$ is successful and the values j and α satisfy the desired condition $z_j = \alpha$. We now show that the input $\mathbf{C}$ to the combining procedure is very close to uniform, and that this implies that H happens with probability very close to $\delta_{j,\alpha}$. We will later see that conditioned on H, the reduction succeeds with probability $1/3$.

Each column $\mathbf{c}_i$ of $\mathbf{C}$ is chosen by running the sampling algorithm $\mathcal{S}(\mathbf{B}, \mathbf{t}_i, s) = (\mathbf{c}_i, \mathbf{y}_i)$ for some vector $\mathbf{t}_i \in \mathbb{R}^n$ and $s = 2r/\gamma > 2\eta_\epsilon(\mathbf{B})$. It follows from Lemma 5.7 that for each i, the statistical distance between $\mathbf{c}_i$ and the uniform distribution over $\mathcal{P}(\mathbf{B})$ is at most $\epsilon/2$. Since the vectors $\mathbf{c}_i$ are independent, we have

$$\Delta(\mathbf{C}, U(\mathcal{P}(\mathbf{B})^m)) \leq \sum_{i=1}^{m} \Delta(\mathbf{c}_i, U(\mathcal{P}(\mathbf{B}))) \leq \epsilon m/2. \tag{13}$$

By Lemma 5.8, on input a uniformly random matrix $\mathbf{C}' \in \mathcal{P}(\mathbf{B})^m$, the distribution of the query $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ asked by $\mathcal{A}^{\mathcal{F}}(\mathbf{B}, \mathbf{S}, \mathbf{C}', q)$ is also uniform. Therefore, on a uniform input $\mathbf{C}'$, the vector $\mathbf{z} = \mathcal{F}(\mathbf{A})$ obtained by the combining procedure satisfies the conditions in (12) with probability $\delta_{j,\alpha}$. By (13) and the properties of the statistical distance, it follows that the event H holds with probability at least $\delta_{j,\alpha} - \epsilon m/2$, which is $\delta_{j,\alpha}$ up to a negligible term.

To complete the proof, we show that the success probability of the reduction conditioned on H is at least $1/3$. We in fact show the stronger fact that the success probability of the reduction is at least $1/3$ conditioned on any fixed values of $\mathbf{C}$, the oracle query $\mathbf{A}$, and the answer $\mathbf{z} = \mathcal{F}(\mathbf{A})$ for which H is satisfied. So in the following, fix some j, α, $\mathbf{C}$, $\mathbf{A}$, and $\mathbf{z}$ for which $z_j = \alpha$, $\mathbf{z} \in \Lambda_q(\mathbf{A}) \setminus \{\mathbf{0}\}$ and $\|\mathbf{z}\| \leq \beta$. In particular, if we define $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_m]$, we get $\mathbf{Tz} = -\mathbf{t}$. We know from Lemma 5.8 that the vector $\mathbf{x} = \mathcal{A}^{\mathcal{F}}(\mathbf{B}, \mathbf{S}, \mathbf{C}, q)$ belongs to the lattice $\mathcal{L}(\mathbf{B})$ and is within distance $n\sqrt{m}\|\mathbf{z}\|\|\mathbf{S}\|/q \leq \|\mathbf{S}\|/g$ from $\mathbf{Cz}$. We also know from Lemma 5.7 that the vectors $\mathbf{y}_i$ are distributed independently according to $D_{\mathcal{L}(\mathbf{B}), s, (\mathbf{c}_i + \mathbf{t}_i)}$, where $s = 2r/\gamma > 2\eta_\epsilon$. Since $\mathbf{x}$ and $\mathbf{y}_1, \ldots, \mathbf{y}_m$ are all lattice vectors, and $\mathbf{z} \in \mathbb{Z}^m$, also $\mathbf{s} = \mathbf{x} - \mathbf{Yz}$ belongs to the lattice $\mathcal{L}(\mathbf{B})$. We need to compute the probability that $\|\mathbf{s} - \mathbf{t}\| \leq (\|\mathbf{S}\|/g) + r$. By the triangle inequality,

$$\|\mathbf{s} - \mathbf{t}\| \leq \|\mathbf{x} - \mathbf{Cz}\| + \|(\mathbf{C} - \mathbf{Y})\mathbf{z} - \mathbf{t}\| \leq \frac{\|\mathbf{S}\|}{g} + \|(\mathbf{Y} - (\mathbf{C} + \mathbf{T}))\mathbf{z}\|.$$

So, all we have to do is to bound the probability that $\|(\mathbf{Y} - (\mathbf{C} + \mathbf{T}))\mathbf{z}\| \leq r$. By Lemma 4.3, since each vector $\mathbf{y}_i$ is distributed according to $D_{\mathcal{L}(\mathbf{B}), s, (\mathbf{c}_i + \mathbf{t}_i)}$, we have

$$\mathrm{Exp}[\|\mathbf{y}_i - (\mathbf{c}_i + \mathbf{t}_i)\|^2] \leq \left(\frac{1}{2\pi} + \frac{\epsilon}{1 - \epsilon}\right) s^2 n$$

and

$$\|\,\mathrm{Exp}[\mathbf{y}_i - (\mathbf{c}_i + \mathbf{t}_i)]\|^2 \leq \left(\frac{\epsilon}{1 - \epsilon}\right)^2 s^2 n.$$

Since the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_m$ are chosen independently, we can apply Lemma 2.11 and get

$$\begin{aligned} \mathrm{Exp}\left[\left\|\sum_{i=1}^{m} (\mathbf{y}_i - (\mathbf{c}_i + \mathbf{t}_i)) z_i\right\|^2\right] &\leq \left(\frac{1}{2\pi} + \frac{\epsilon}{1 - \epsilon} + \left(\frac{\epsilon}{1 - \epsilon}\right)^2 m\right) \|\mathbf{z}\|^2 s^2 n \\ &\leq \frac{\|\mathbf{z}\|^2 s^2 n}{6} \end{aligned}$$

for all sufficiently large n. Finally, using $\|\mathbf{z}\| \leq \beta$, $s = 2r/\gamma$ and $\gamma = \beta\sqrt{n}$, we get

$$\mathrm{Exp}[\|(\mathbf{Y} - (\mathbf{C} + \mathbf{T}))\mathbf{z}\|^2] \leq \frac{\|\mathbf{z}\|^2 s^2 n}{6} \leq \frac{2}{3} r^2$$

and, by Markov inequality, we get

$$\Pr\{\|(\mathbf{Y} - (\mathbf{C} + \mathbf{T}))\mathbf{z}\| > r\} = \Pr\{\|(\mathbf{Y} - (\mathbf{C} + \mathbf{T}))\mathbf{z}\|^2 > r^2\} \leq \frac{2}{3}.$$

This proves that the conditional probability of $\|\mathbf{s} - \mathbf{t}\| \leq (\|\mathbf{S}\|/g) + r$ is at least $1/3$. ■

5.3 Other worst-case problems

In Section 5.2 we have shown that solving SIS on the average is at least as hard as solving IncGDD in the worst case. In this section we prove that solving SIS on the average is at least as hard as solving many other standard lattice problems, like SIVP and GapCRP. These results are obtained as corollaries to Theorem 5.9 using straightforward worst-case to worst-case reductions among lattice problems. We now describe three such reductions. The first two are taken from [22] and for completeness, we include a sketch of their proof.

Lemma 5.10 *For any $\gamma(n) \geq 1$ and any ϕ, there exists a reduction from $\mathrm{GIVP}^{\phi}_{8\gamma}$ to $\mathrm{IncGDD}^{\phi}_{\gamma,8}$.*

Proof: Given a basis $\mathbf{B}$, our goal is to construct a set of n linearly independent vectors $\mathbf{S}$ of length $\|\mathbf{S}\| \leq 8\gamma(n)\phi(\mathbf{B})$. We do this by an iterative process. Initially, we set $\mathbf{S} = \mathbf{B}$. At each step, we identify the longest vector in $\mathbf{S}$, say $\mathbf{s}_i$. We then take $\mathbf{t}$ to be a vector orthogonal to $\mathbf{s}_1, \ldots, \mathbf{s}_{i-1}, \mathbf{s}_{i+1}, \ldots, \mathbf{s}_n$ of length $\|\mathbf{S}\|/2$. We apply the IncGDD oracle with the instance $(\mathbf{B}, \mathbf{S}, \mathbf{t}, \|\mathbf{S}\|/8)$. If it fails, we abort and output $\mathbf{S}$. Otherwise, we obtain a lattice vector $\mathbf{u}$ within distance at most $(\|\mathbf{S}\|/8) + \|\mathbf{S}\|/8 = \|\mathbf{S}\|/4$ from $\mathbf{t}$. Notice that $\|\mathbf{u}\| \leq 3\|\mathbf{S}\|/4$ and that it is linearly independent from the vectors in $\mathbf{s}_1, \ldots, \mathbf{s}_{i-1}, \mathbf{s}_{i+1}, \ldots, \mathbf{s}_n$. We then replace $\mathbf{s}_i$ with $\mathbf{u}$ and repeat the process.

Notice that when the oracle call fails, it must be the case that $\|\mathbf{S}\|/8 \leq \gamma(n)\phi(\mathbf{B})$, and hence $\|\mathbf{S}\| \leq 8\gamma(n)\phi(\mathbf{B})$, as required. Moreover, it is not difficult to argue that this procedure terminates after a polynomial number of steps. For instance, one can note that $\log \Pi_i \|\mathbf{s}_i\|$ decreases by a constant at each step, and that its initial value is polynomial in the input size. ■

Lemma 5.11 *For any $\gamma(n) \geq 1$ and any ϕ, there exists a reduction from $\mathrm{GDD}^{\phi}_{3\gamma}$ to $\mathrm{IncGDD}^{\phi}_{\gamma,8}$.*

Proof: Given a basis $\mathbf{B}$ and a vector $\mathbf{t}$, our goal is to find a lattice vector within distance $3\gamma(n)\phi(\mathbf{B})$ of $\mathbf{t}$. First, we apply the reduction in Lemma 5.10 to obtain a set $\mathbf{S}$ of n linearly independent vectors of length at most $\|\mathbf{S}\| \leq 8\gamma(n)\phi(\mathbf{B})$. We then search for a value r for which an oracle call with $(\mathbf{B}, \mathbf{S}, \mathbf{t}, r/2)$ fails but an oracle call with $(\mathbf{B}, \mathbf{S}, \mathbf{t}, r)$ succeeds. Since the former oracle call fails, it must be the case that $r \leq 2\gamma(n)\phi(\mathbf{B})$. The latter oracle call yields a lattice vector within distance $\|\mathbf{S}\|/8 + r \leq \gamma(n)\phi(\mathbf{B}) + 2\gamma(n)\phi(\mathbf{B}) = 3\gamma(n)\phi(\mathbf{B})$, as required. ■

Lemma 5.12 *For any $\gamma(n)$, there exists a randomized reduction from GapCRP_{γ} to $\mathrm{GDD}^{\lambda_n}_{\gamma/4}$.*

Proof: Let $(\mathbf{B}, d)$ be an instance of GapCRP_{γ}. The reduction picks a point $\mathbf{t} \in \mathcal{P}(\mathbf{B})$ uniformly at random and then calls the $\mathrm{GDD}^{\lambda_n}_{\gamma/4}$ oracle with the instance $(\mathbf{B}, \mathbf{t})$ to obtain a lattice vector $\mathbf{x} \in \mathcal{L}(\mathbf{B})$ within distance $(\gamma/4)\lambda_n(\mathbf{B})$ from the target $\mathbf{t}$. If $\|\mathbf{t} - \mathbf{x}\| \leq \gamma d/2$ then we accept, otherwise we reject. If $\nu(\mathbf{B}) \leq d$, then

$$\|\mathbf{t} - \mathbf{x}\| \leq \gamma\lambda_n(\mathbf{B})/4 \leq \gamma\nu(\mathbf{B})/2 \leq \gamma d/2$$

where we used that for any lattice Λ, $\nu(\Lambda) \geq \lambda_n(\Lambda)/2$ [23, Theorem 7.9]. So, Yes instances are always accepted. On the other hand, assume that $\nu(\mathbf{B}) > \gamma d$. In [14] it is shown that a random $\mathbf{t}$ chosen as above satisfies $\mathrm{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) \geq \nu(\mathbf{B})/2$ with probability at least $1/2$. Hence, $\mathrm{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) > \gamma d/2$ with probability $1/2$, and No instances are rejected with probability $1/2$. ■

By combining these reductions with Theorem 5.9, we obtain several useful corollaries. The first relates GIVP to SIS and we give it here in its most general form, as it will later be used in the GapSVP reduction.

Corollary 5.13 *For any polynomially bounded functions $\beta(n), m(n) = n^{O(1)}$, any negligible function $\epsilon(n)$, and any $q(n) \geq 8n\sqrt{m(n)}\beta(n)$, there is a probabilistic polynomial time reduction from solving $\mathrm{GIVP}_{\gamma}^{\eta_\epsilon}$ in the worst case with $\gamma(n) = 8\beta(n)\sqrt{n}$ to solving $\mathrm{SIS}_{q,m,\beta}$ on the average with non-negligible probability.*

For the remaining corollaries, it is helpful to specialize Theorem 5.9 to the $\mathrm{SIS}_{q,m}$ problem where solutions are guaranteed to exist. This is done by choosing $\beta(n) = \sqrt{m(n)}q(n)^{n/m(n)}$. Observe that for this value of β, the condition $q(n) \geq g(n)n\sqrt{m(n)}\beta(n)$ is equivalent to $q(n) \geq (g(n)nm(n))^{1+n/(m(n)-n)}$.

Corollary 5.14 *For any function $g(n) > 0$, polynomially bounded function $m(n) = n^{O(1)}$, negligible function $\epsilon(n) = n^{-\omega(1)}$, and $q(n) \geq (g(n)nm(n))^{1+n/(m(n)-n)}$, there is a probabilistic polynomial time reduction from solving $\mathrm{IncGDD}_{\gamma,g}^{\eta_\epsilon}$ for $\gamma(n) = \sqrt{nm(n)} \cdot q(n)^{n/m(n)}$ on n-dimensional instances in the worst case to solving $\mathrm{SIS}_{q,m}$ on the average with non-negligible probability.*

We continue with some other connections. For simplicity, from now on we consider a specific choice of parameters. Other choices can be handled similarly.

Theorem 5.15 *For any $m(n) = \Theta(n \log n)$, there exists some $q(n) = O(n^2 \log n)$ and $\gamma(n) = O(n\sqrt{\log n})$ such that for any negligible function $\epsilon(n)$, solving $\mathrm{SIS}_{q,m}$ on the average with non-negligible probability is at least as hard as solving any of the following worst-case problems:*

- $\mathrm{GIVP}_{\gamma}^{\eta_\epsilon}$
- $\mathrm{GDD}_{\gamma}^{\eta_\epsilon}$

Proof: Notice that for any $m(n) = \Theta(n \log n)$ there exists a $q(n) = O(n^2 \log n)$ that satisfies the conditions in Corollary 5.14 with $g(n) = 8$ a constant. This yields a solution to $\mathrm{IncGDD}_{\gamma,8}^{\eta_\epsilon}$ for some $\gamma(n) = O(n\sqrt{\log n})$. It remains to apply Lemmas 5.10 and 5.11. ■

Theorem 5.16 *For any $m(n) = \Theta(n \log n)$ there exists a $q(n) = O(n^2 \log n)$ such that for any function $\gamma(n) = \omega(n \log n)$, solving $\mathrm{SIS}_{q,m}$ on the average with non-negligible probability is at least as hard as solving any of the following worst-case problems:*

- SIVP_{γ} *(or equivalently,* $\mathrm{GIVP}_{\gamma}^{\lambda_n}$*),*
- $\mathrm{GDD}_{\gamma}^{\lambda_n}$*,*
- GapCRP_{γ}.

Proof: Let $\alpha(n)$ be any function (e.g., $\alpha(n) = \sqrt{\gamma(n)/n \log n}$) such that $\alpha(n) = \omega(1)$ and $\gamma(n) = \omega(\alpha(n)n \log n)$. By Lemma 3.3, there exists a negligible $\epsilon(n)$ for which $\eta_\epsilon(\Lambda) \leq \alpha(n)\sqrt{\log n}\lambda_n(\Lambda)$ holds for any lattice. Hence, the first two claims follow from Theorem 5.15 for some approximation factor $O(\alpha(n)n \log n) < \gamma(n)$. The third claim follows from the second claim together with Lemma 5.12. ■

We complete this section with a discussion of *non-adaptive* reductions. These are reductions in which the oracle queries do not depend on the answers to previous queries and hence can be performed all at once. It is known that unless the polynomial hierarchy collapses, no average-case problem can be shown to be NP-hard under non-adaptive reductions. See [7] and references therein for a more accurate description of these results. Here, we observe that our reductions can be made non-adaptive with only a slight worsening of the approximation factors obtained.

Lemma 5.17 *For any functions $g(n), \gamma(n)$ such that $\gamma(n) < n^c$ for some $c > 0$, there exists a non-adaptive reduction from* $\mathrm{GDD}^{\lambda_n}_{\gamma'}$ *to* $\mathrm{IncGDD}^{\lambda_n}_{\gamma,g}$ *where $\gamma'(n) = (2^n/g(n)) + 2\gamma(n)$.*

Proof: Given a lattice $\mathbf{B}$ and a target $\mathbf{t}$, we want to find a lattice vector close to $\mathbf{t}$. Using the LLL lattice reduction algorithm [17], we can efficiently compute a basis $\mathbf{S}$ of $\mathcal{L}(\mathbf{B})$ such that $\|\mathbf{S}\| \leq 2^n \lambda_n(\mathbf{B})$. Let $\tilde{\lambda}_n = \|\mathbf{S}\|/2^n$ and notice that $\tilde{\lambda}_n \leq \lambda_n(\mathbf{B}) \leq 2^n \tilde{\lambda}_n$. The reduction then calls the IncGDD oracle on input $(\mathbf{B}, \mathbf{S}, \mathbf{t}, 2^i \cdot \tilde{\lambda}_n)$ for $i = 0, 1, \ldots, \lceil n + c \log n \rceil$ and outputs the lattice vector closest to $\mathbf{t}$ among the vectors returned.

Let i be the smallest index such that $2^i \tilde{\lambda}_n > \gamma(n)\lambda_n(\mathbf{B})$. Such an i exists since $2^{n+c\log n}\tilde{\lambda}_n = n^c \cdot 2^n \tilde{\lambda}_n > \gamma(n)\lambda_n(\mathbf{B})$. Notice that $2^i \tilde{\lambda}_n \leq 2\gamma(n)\lambda_n(\mathbf{B})$. It follows that the lattice vector returned by the IncGDD oracle on input $(\mathbf{B}, \mathbf{S}, \mathbf{t}, 2^i \tilde{\lambda}_n)$ is within distance

$$\frac{\|\mathbf{S}\|}{g(n)} + 2^i \tilde{\lambda}_n \leq \frac{2^n \lambda_n(\mathbf{B})}{g(n)} + 2\gamma(n)\lambda_n(\mathbf{B}) = \Big(\frac{2^n}{g(n)} + 2\gamma(n)\Big)\lambda_n(\mathbf{B})$$

from the target $\mathbf{t}$, as required. ■

Lemma 5.18 *For any $\gamma(n)$, there exists a non-adaptive reduction from* SIVP_γ *to* $\mathrm{GDD}^{\lambda_n}_{\gamma/4\sqrt{n}}$.

Proof: Let $\mathbf{B}$ be some instance of SIVP_γ. Using the LLL lattice reduction algorithm [17], we can efficiently compute a basis $\mathbf{S}$ of the same lattice such that $\|\mathbf{S}\| \leq 2^n \lambda_n(\mathbf{B})$. Notice that if $\gamma \geq 2^n$, we can simply output $\mathbf{S}$ so in the following assume $\gamma < 2^n$. Let $\tilde{\lambda}_n = 2^{-n-1}\|\mathbf{S}\|$ and notice that $2\tilde{\lambda}_n \leq \lambda_n(\mathbf{B}) \leq 2^{n+1}\tilde{\lambda}_n$. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be some orthonormal set of vectors. The reduction calls the $\mathrm{GDD}^{\lambda_n}_{\gamma/4\sqrt{n}}$ oracle on input $(\mathbf{B}, 2^i \tilde{\lambda}_n \mathbf{e}_j)$ for $i = 0, \ldots, 2n-1$ and $j = 1, \ldots, n$. For $i = 0, \ldots, 2n-1$, let $\mathbf{S}_i$ denote the set of n vectors returned by the oracle on queries corresponding to i. Among the sets $\mathbf{S}_i$ that contain n linearly independent vectors, the reduction outputs the one that minimizes $\|\mathbf{S}_i\|$. We need to prove that there exists an index i such that the vectors $\mathbf{S}_i$ are linearly independent and $\|\mathbf{S}_i\| \leq \gamma\lambda_n(\mathbf{B})$.

Let $i \in \{0, \ldots, 2n-1\}$ be the smallest index such that $2^i \tilde{\lambda}_n > \gamma\lambda_n(\mathbf{B})/4$. Notice that such an i exists and that $2^i \tilde{\lambda}_n \leq \gamma\lambda_n(\mathbf{B})/2$. Each column of $\mathbf{S}_i$ is within distance $\gamma\lambda_n(\mathbf{B})/4\sqrt{n}$ from the corresponding vector $2^i \tilde{\lambda}_n \mathbf{e}_j$. Since the length of the latter is strictly greater than $\gamma\lambda_n(\mathbf{B})/4$, it follows that the columns of $\mathbf{S}_i$ are linearly independent (see, e.g., [20]). Finally, by the triangle inequality, each vector in $\mathbf{S}_i$ has length at most

$$2^i \tilde{\lambda}_n + \gamma\lambda_n(\mathbf{B})/4\sqrt{n} \leq \gamma\lambda_n(\mathbf{B})/2 + \gamma\lambda_n(\mathbf{B})/4\sqrt{n} \leq \gamma\lambda_n(\mathbf{B}).$$

■

Theorem 5.19 *There exist functions $q(n) = 2^{O(n)}$ and $m(n) = n^{O(1)}$ such that for any function $\alpha(n) = \omega(\sqrt{\log n})$, solving* $\mathrm{SIS}_{q,m}$ *on the average with non-negligible probability is at least as hard (via non-adaptive reductions) as solving any of the following worst-case problems:*

- $\mathrm{GDD}^{\lambda_n}_{\gamma}$ *for some $\gamma(n) = O(n^{1.5}\alpha(n))$,*
- GapCRP_γ *for some $\gamma(n) = O(n^{1.5}\alpha(n))$,*
- SIVP_γ *for some $\gamma(n) = O(n^2\alpha(n))$.*

Proof: By Lemma 3.3, we can choose a negligible function $\epsilon(n)$ such that for any lattice Λ, $\eta_\epsilon(\Lambda) \leq \alpha(n)\lambda_n(\Lambda)$. Let $q(n) = n^3 2^n$, $m(n) = n^2$ and $g(n) = 2^n/4$. Notice that this choice satisfies the hypothesis in Corollary 5.14. Moreover, notice that the reduction in Theorem 5.9 is non-adaptive since it makes only one oracle query. Therefore, by Corollary 5.14, there is a non-adaptive reduction from solving worst-case instances of $\mathrm{IncGDD}^{\eta_\epsilon}_{\gamma,g}$ with $\gamma(n) \leq 4n^{1.5}$ to solving $\mathrm{SIS}_{q,m}$ on the average with non-negligible probability. By our choice of ϵ, this is also a reduction from $\mathrm{IncGDD}^{\lambda_n}_{\gamma',g}$ where $\gamma'(n) = \gamma(n)\alpha(n)$.

The first claim follows from Lemma 5.17. The second claim follows directly from the first together with Lemma 5.12. The only thing to notice is that the reduction in that lemma is non-adaptive since it makes only one oracle call. The third follows similarly with the use of Lemma 5.18.

■

5.4 Shortest vector problem

In this subsection we reduce GapSVP to SIS′. Let us first recall Hoeffding's inequality [15], which states the following. Let $X_1, \ldots, X_N$ be N independent random variables, such that for all i, $X_i \in [a, b]$. Then $S_N = \sum_i X_i$ satisfies

$$\Pr\{S_N \geq \mathrm{Exp}[S_N] + N\epsilon\} \leq e^{-N\epsilon^2/(b-a)^2}. \tag{14}$$

We will also need the following lemma from [1]. For completeness, we include its proof in the appendix.

Lemma 5.20 ([1], Lemma 6.2) *Let σ, K, ℓ be some positive numbers and let D be a distribution on $\mathbb{R}^n$ such that for any fixed unit vector $\mathbf{u}$,*

$$\mathop{\mathrm{Exp}}_{\mathbf{w}\sim D}[\langle \mathbf{u}, \mathbf{w}\rangle^2] \leq \ell^2$$

and, moreover,

$$\Pr_{\mathbf{w}\sim D}\{\|\mathbf{w}\| \geq K\ell\} \leq \sigma.$$

Let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ be a matrix obtained by picking each column independently at random according to distribution $\mathbf{w}_i \sim D$. Then, with probability at least $1 - e^{-N/K^4}(4\sqrt{n}K^2)^n - N\sigma$ (over the choice of matrix $\mathbf{W}$) the maximum eigenvalue of the $n \times n$ matrix $\mathbf{W}\mathbf{W}^T$ is at most $3N\ell^2$.

We now define a variant of the closest vector problem that will be used as an intermediate step in our reduction from GapSVP to SIS′.

Definition 5.21 *An input to GapCVP'_γ is a triple $(\mathbf{B}, \mathbf{t}, d)$ where $\mathbf{B}$ is an n-dimensional lattice basis, $\mathbf{t}$ is a target vector, and d is a rational number. In Yes inputs $\mathrm{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) \leq d$. In No inputs $\lambda_1(\mathbf{B}) > \gamma(n) \cdot d$ and for any odd $k \in \mathbb{Z}$, $\mathrm{dist}(k\mathbf{t}, \mathcal{L}(\mathbf{B})) > \gamma(n) \cdot d$.*

The difference between GapCVP′ and the standard problem GapCVP, is that when the target is far from the lattice, also any odd multiple of the target is far and the minimum distance of the lattice is large. In [13] it is shown that there is a polynomial time reduction from GapSVP_γ to GapCVP_γ. We observe that the reduction given in [13] is also a reduction from GapSVP_γ to GapCVP'_γ, as shown in the following lemma.

Lemma 5.22 *For any approximation factor $\gamma(n)$, there is a polynomial time reduction from GapSVP_γ to GapCVP'_γ.*

Proof: In [13] it is shown that for any γ, there is a deterministic Cook reduction from GapSVP_γ to GapCVP_γ (see also [23]). Here we observe that the same reduction can be used as a reduction from GapSVP_γ to GapCVP'_γ. To see this, it suffices to know that on input GapSVP_γ instance $(\mathbf{B}, d)$, all the GapCVP_γ calls made by the reduction have the form $(\mathbf{B}_i, \mathbf{b}_i, d)$, where $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$ and $\mathbf{B}_i = [\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, 2\mathbf{b}_i, \mathbf{b}_{i+1}, \ldots, \mathbf{b}_n]$. Moreover, the reduction outputs Yes if and only if any of the calls is answered Yes. Since GapCVP_γ and GapCVP'_γ have the same set of Yes instances, if the reduction is guaranteed to output Yes given a GapCVP_γ oracle (e.g., when the input is a Yes instance), then it outputs Yes also when given access to a GapCVP'_γ oracle. Now, let us consider the case when the input $(\mathbf{B}, d)$ is a No instance, and therefore all calls made by the reduction would receive a No answer from a GapCVP_γ oracle. Notice that for any odd integer k, $\mathrm{dist}(k\mathbf{b}_i, \mathcal{L}(\mathbf{B}_i)) = \mathrm{dist}(\mathbf{b}_i, \mathcal{L}(\mathbf{B}_i))$ because $2\mathbf{b}_i \in \mathcal{L}(\mathbf{B}_i)$. Moreover, $\lambda_1(\mathbf{B}_i) \geq \lambda_1(\mathbf{B}) > \gamma d$. So, if $(\mathbf{B}_i, \mathbf{b}_i, d)$ is a No instance of GapCVP_γ, then it is also a No instance of GapCVP'_γ. Therefore all calls made by the reduction receive a No answer by the GapCVP'_γ oracle as well, and the final output of the reduction is No. ■

We now show how to solve GapCVP'_γ in the worst case given access to an oracle that solves SIS′ on the average. By Lemma 5.22 this also implies a reduction from GapSVP_γ to SIS′ (or SIS when the modulus q is odd).

Theorem 5.23 *For any polynomially bounded functions $\beta(n), m(n), q(n) = n^{O(1)}$, with $q(n) \geq 4\sqrt{m(n)}n^{1.5}\beta(n)$, and $\gamma(n) = 14\pi\sqrt{n}\beta(n)$, there is a probabilistic polynomial time reduction from solving GAPCVP'_γ in the worst case to solving $\text{SIS}'_{q,m,\beta}$ on the average with non-negligible probability.*

In particular, for any $m(n) = \Theta(n \log n)$, there exist $q(n) = O(n^{2.5} \log n)$, and $\gamma(n) = O(n\sqrt{\log n})$, such that solving $\text{SIS}'_{q,m}$ on the average is at least as hard as solving GAPSVP_γ in the worst case.

Proof: We adopt the notation of Theorem 5.9 and omit the dependence on n for the parameters $m, \gamma, \beta, q, \epsilon$, and δ. Let $\mathcal{F}$ be an oracle solving $\text{SIS}'_{q,m,\beta}$ on the average with non-negligible probability δ. Namely, $\mathcal{F}$ is an oracle that on input a random matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ returns a vector $\mathbf{z} = \mathcal{F}(\mathbf{A}) \in \Lambda_q(\mathbf{A}) \setminus 2\mathbb{Z}^m$ of length at most β with probability δ. Notice that since $\mathbf{0} \notin \Lambda_q(\mathbf{A}) \setminus 2\mathbb{Z}^m$, oracle $\mathcal{F}$ also solves $\text{SIS}_{q,m,\beta}$ with probability at least δ. We want to use $\mathcal{F}$ to solve GAPCVP'_γ. The main idea is to use the NP verifier for (the complement of) GAPCVP presented in [1] as a routine for solving GAPCVP'. To be able to do this, we need to be able to generate a good witness to that verifier. Such a witness is given by a set of short vectors sampled from the discrete Gaussian distribution in the dual lattice. Luckily, we can generate such a witness by using the sampling procedure and the combining procedure given in Subsection 5.2 (together with $\mathcal{F}$). In fact, to be able to use these procedures, we need a reasonably short set of linearly independent vectors $\mathbf{S}$. We obtain such a set by using Corollary 5.13.[13]

We start by describing the NP verifier of [1]. For our purposes, it is best to think of this NP verifier as an algorithm, call it $\mathcal{V}$. The input to $\mathcal{V}$ consists of a lattice $\mathbf{B}$, a vector $\mathbf{t}$, a number $d > 0$, and a sequence of vectors $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ in $\mathcal{L}(\mathbf{B})^*$ where $N = n^3m^3$. The algorithm $\mathcal{V}(\mathbf{B}, \mathbf{t}, d, \mathbf{W})$ performs three tests:

(a) Check that for all $i = 1, \ldots, N$, $\mathbf{w}_i \in \mathcal{L}(\mathbf{B})^*$,

(b) Check that $f_{\mathbf{W}}(\mathbf{t}) < 1/2$, where $f_{\mathbf{W}}$ is the function $f_{\mathbf{W}}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \cos(2\pi\langle \mathbf{x}, \mathbf{w}_i \rangle)$.

(c) Check that the largest eigenvalue of the $n \times n$ positive semidefinite matrix $\mathbf{W}\mathbf{W}^T$ is at most $N/(2\pi d)^2$.

If all three tests are satisfied, then $\mathcal{V}$ outputs YES, otherwise it outputs NO. It is shown in [1] that if $\text{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) \leq d$ then $\mathcal{V}(\mathbf{B}, \mathbf{t}, d, \mathbf{W})$ is guaranteed to output NO (for any matrix $\mathbf{W}$), while if $\text{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) > c\sqrt{n}d$ (for some absolute constant c) then there exist a matrix $\mathbf{W}$ that makes $\mathcal{V}$ output YES.

We now describe our GAPCVP' reduction. From now on, fix $\epsilon(n) = 2^{-n}$. First, using Corollary 5.13 (with $\mathcal{F}$ as an oracle), we obtain a set of n linearly independent vectors $\mathbf{S}$ in $\mathcal{L}(\mathbf{B})^*$ such that $\|\mathbf{S}\| \leq 8\beta\sqrt{n}\eta_\epsilon(\mathbf{B}^*)$. Define

$$s = \frac{2\sqrt{n}}{\gamma d}.$$

Consider the following procedure $\mathcal{W}(\mathbf{B}, \mathbf{S})$:

1. Run the sampling procedure $\mathcal{S}(\mathbf{B}^*, \mathbf{0}, s)$ of Lemma 5.7 on input a basis $\mathbf{B}^*$ of the dual lattice $\mathcal{L}(\mathbf{B})^*$. This procedure is run m times to generate m pairs of vectors $(\mathbf{c}_i, \mathbf{y}_i)$, and define the matrices $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_m]$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m]$.

2. Run the combining procedure $\mathcal{A}^{\mathcal{F}}(\mathbf{B}^*, \mathbf{S}, \mathbf{C}, q)$ of Lemma 5.8 with the oracle $\mathcal{F}$. Let $\mathbf{A}$ be the query asked by $\mathcal{A}$, and $\mathbf{z} = \mathcal{F}(\mathbf{A})$ the answer returned by the oracle.

3. If $\mathbf{z}$ is not a valid solution to SIS' instance $(q, \mathbf{A}, \beta)$, then $\mathcal{W}$ aborts the computation with no output. Otherwise, let $\mathbf{x}$ be the vector returned by $\mathcal{A}$, and output the vector $\mathbf{w} = \mathbf{x} - \mathbf{Y}\mathbf{z}$.

We apply $\mathcal{W}(\mathbf{B}, \mathbf{S})$ nN/δ times, each time with independent randomness. If the number of non-aborting runs of $\mathcal{W}$ is less than N, then the reduction terminates immediately with output YES. Otherwise, let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ be the vectors returned by the first N non-aborting runs of $\mathcal{W}$, and call $\mathcal{V}(\mathbf{B}, \mathbf{t}, d, \mathbf{W})$. If

[13] We remark that we could also use any polynomially longer set $\mathbf{S}$ with only a minor effect on our results.

$\mathcal{V}$ says YES, the reduction outputs NO; otherwise, the reduction outputs YES. This completes the description of the reduction.

By the properties of $\mathcal{V}$, it is clear that whenever $\mathrm{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) \leq d$, the reduction correctly outputs YES (either because the number of non-aborting runs of $\mathcal{W}$ is less than N, or because $\mathcal{V}(\mathbf{B}, \mathbf{t}, d, \mathbf{W})$ outputs NO). For completeness, let us sketch the proof that $\mathcal{V}$ outputs NO whenever $\mathrm{dist}(\mathbf{t}, \mathcal{L}(\mathbf{B})) \leq d$. Assume that the distance of $\mathbf{t}$ from $\mathcal{L}(\mathbf{B})$ is at most d and assume that tests (a) and (c) are satisfied. We show that test (b) must fail, and therefore $\mathcal{V}$ outputs NO. First, by the definition of $f_{\mathbf{W}}$ and the assumption that test (a) accepts, we have that $f_{\mathbf{W}}$ is periodic modulo $\mathcal{L}(\mathbf{B})$. Moreover, since the largest eigenvalue of $\mathbf{W}\mathbf{W}^T$ is bounded by $N/(2\pi d)^2$, we have that $\|\mathbf{W}^T\mathbf{x}\|^2 \leq N\|\mathbf{x}\|^2/(2\pi d)^2$ for any vector $\mathbf{x}$. Let $\tau(\mathbf{t})$ denote the lattice vector closest to $\mathbf{t}$. Notice that $\|\mathbf{t}-\tau(\mathbf{t})\| \leq d$. Since $f_{\mathbf{W}}$ is periodic modulo the lattice, $f_{\mathbf{W}}(\mathbf{t}) = f_{\mathbf{W}}(\mathbf{t}-\tau(\mathbf{t}))$. It thus suffices to prove that $f_{\mathbf{W}}(\mathbf{t}-\tau(\mathbf{t})) \geq 1/2$. Using the inequality $\cos x \geq 1 - x^2/2$ (valid for any $x \in \mathbb{R}$) we get:

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{t}-\tau(\mathbf{t})) &= \frac{1}{N}\sum_{i=1}^{N} \cos(2\pi\langle \mathbf{t}-\tau(\mathbf{t}), \mathbf{w}_i\rangle) \\ &\geq 1 - \frac{4\pi^2}{2N}\sum_{i=1}^{N} \langle \mathbf{t}-\tau(\mathbf{t}), \mathbf{w}_i\rangle^2 \\ &= 1 - \frac{2\pi^2}{N}\|\mathbf{W}^T(\mathbf{t}-\tau(\mathbf{t}))\|^2 \\ &\geq 1 - \frac{\|\mathbf{t}-\tau(\mathbf{t})\|^2}{2d^2} \geq \frac{1}{2}. \end{aligned}$$

It remains to show that the reduction outputs the correct answer when the input is a NO instance, i.e., when $\lambda_1(\mathbf{B}) > \gamma d$ and $\mathrm{dist}(k\mathbf{t}, \mathcal{L}(\mathbf{B})) > \gamma d$ for any odd $k \in \mathbb{Z}$. In order for our reduction to output NO, two conditions must be satisfied: at least N calls to $\mathcal{W}$ succeed, and $\mathcal{V}$ outputs YES. Let us first show that after $n \cdot N/\delta$ calls to $\mathcal{W}$, we obtain at least N vectors with high probability. By Lemma 3.2, we have that

$$\eta_\epsilon(\mathbf{B}^*) \leq \frac{\sqrt{n}}{\lambda_1(\mathbf{B})} < \frac{\sqrt{n}}{\gamma d} = \frac{s}{2} \tag{15}$$

and hence s satisfies $s > 2\eta_\epsilon(\mathbf{B}^*)$. Therefore, by Lemma 5.7, the pairs $(\mathbf{c}_i, \mathbf{y}_i)$ computed by the sampling procedure $\mathcal{S}(\mathbf{B}^*, \mathbf{0}, s)$ satisfy that $\mathbf{c}_i$ is within distance $\epsilon/2$ from the uniform distribution. It follows that $\mathbf{C}$ is within distance $m\epsilon/2$ from the uniform distribution over $\mathcal{P}(\mathbf{B}^*)^m$. Hence, by Lemma 5.8, the query $\mathbf{A}$ given to the oracle by the combining procedure $\mathcal{A}^{\mathcal{F}}(\mathbf{B}^*, \mathbf{S}, \mathbf{C}, q)$ is within negligible distance $\epsilon m/2$ from the uniform distribution, and the oracle returns a vector $\mathbf{z}$ such that $\mathbf{z} \in \Lambda_q(\mathbf{A}) \setminus 2\mathbb{Z}^m$ and $\|\mathbf{z}\| \leq \beta$ with probability at least $\delta - \epsilon m/2 > \delta/2$ for all sufficiently large n. So, the probability that out of $n \cdot N/\delta$ calls to $\mathcal{W}$ less than N are successful is at most $N(1-\delta/2)^{n/\delta} \leq Ne^{-n/2} < 2^{-n/2}$.

It remains to show that $\mathcal{V}$ outputs YES with high probability. The proof of this is based on [1]. However, in [1], it is only shown that there *exists* a good matrix $\mathbf{W}$ that makes $\mathcal{V}$ output YES. Here, we have to argue that the $\mathbf{W}$ given by $\mathcal{W}$ is good with high probability.

First, we observe that test (a) is always satisfied since $\mathcal{W}(\mathbf{B}, \mathbf{S})$ is guaranteed to output vectors in $\mathcal{L}(\mathbf{B})^*$. Indeed, $\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_m \in \mathcal{L}(\mathbf{B})^*$ and hence $\mathbf{x} - \mathbf{Y}\mathbf{z}$ also belongs to the lattice $\mathcal{L}(\mathbf{B})^*$. In the rest of the proof, we show that tests (b) and (c) are satisfied with high probability. To this end, notice that each vector $\mathbf{w}$ in $\mathbf{W}$ is distributed independently according to the distribution D defined as the output of $\mathcal{W}(\mathbf{B}, \mathbf{S})$ conditioned on a non-aborting run.

Consider test (b). Our goal is to show that $\Pr\{f_{\mathbf{W}}(\mathbf{t}) \geq 1/2\}$ is small. Below, we will show that

$$\left| \operatorname*{Exp}_{\mathbf{w}}[\cos(2\pi\langle \mathbf{t}, \mathbf{w}\rangle)] \right| \leq 2^{-n+1} \tag{16}$$

where $\mathbf{w}$ is distributed according to D. By Hoeffding's bound (14), this would imply that $f_{\mathbf{W}}(\mathbf{t}) \geq 1/2$ with probability at most $e^{-N(\frac{1}{2}-\mathrm{Exp}[f_{\mathbf{W}}(\mathbf{t})])^2/4} \leq e^{-N(\frac{1}{2}-2^{-n+1})^2/4} = 2^{-\Omega(N)}$. We now prove (16). We in fact show that it is true even when we condition on any fixed values of $\mathbf{C}$, $\mathbf{A}$, and $\mathbf{z} \in \Lambda_q(\mathbf{A}) \setminus 2\mathbb{Z}^m$. Furthermore, we condition on any fixed values of $\mathbf{y}_1, \ldots, \mathbf{y}_{j-1}, \mathbf{y}_{j+1}, \ldots, \mathbf{y}_m$ where j is some index for which z_j is odd. Notice that the only randomness left is in $\mathbf{y}_j$, which, by Lemma 5.7, is distributed according to $D_{\mathcal{L}(\mathbf{B})^*,s,\mathbf{c}_j}$. Hence, a sample $\mathbf{w} = \mathbf{x} - \mathbf{Y}\mathbf{z}$ can be written as $-z_j(\hat{\mathbf{w}} + \mathbf{y}_j)$ for some fixed vector $\hat{\mathbf{w}}$. Notice that

$$\cos(2\pi\langle \mathbf{t}, \mathbf{w}\rangle) = \cos(2\pi\langle \mathbf{t}, -z_j(\hat{\mathbf{w}} + \mathbf{y}_j)\rangle) = \cos(2\pi\langle -z_j\mathbf{t}, \hat{\mathbf{w}} + \mathbf{y}_j\rangle).$$

By Corollary 4.6, Equation (15), and $\mathrm{dist}(-z_j\mathbf{t}, \mathcal{L}(\mathbf{B})) > \gamma d = 2\sqrt{n}/s$, we obtain

$$\left|\operatorname*{Exp}_{\mathbf{y}_j}[\cos(2\pi\langle \mathbf{t}, \mathbf{w}\rangle)]\right| = \left|\operatorname*{Exp}_{\mathbf{y}_j}[\cos(2\pi\langle \hat{\mathbf{w}} + \mathbf{y}_j, -z_j\mathbf{t}\rangle)]\right| \leq \frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n} \leq 2^{-n+1}.$$

In the rest of the proof we show that test (c) is satisfied with high probability. We do this by applying Lemma 5.20 with $\ell = 2s\beta$, $N = n^3m^3$, $\sigma = N2^{-n}(1+\epsilon)/(1-\epsilon)$, $K = \sqrt{n \cdot m}$, and the distribution D. Assuming the hypothesis in that lemma holds, we get that the maximum eigenvalue of matrix $\mathbf{W}\mathbf{W}^T$ is bounded by

$$3N\ell^2 = 12Ns^2\beta^2 = \frac{48Nn\beta^2}{\gamma^2 d^2} < \frac{N}{(2\pi d)^2}$$

except possibly with probability

$$e^{-N/K^4}(4\sqrt{n}K^2)^n + N\sigma \leq (4e^{-m}n^{1.5}m)^n + N^2 2^{-n+1} \leq 2^{-n/2}.$$

Therefore, the probability that test (c) fails is exponentially small. It remains to check that the hypothesis of Lemma 5.20 is satisfied, i.e.,

$$\Pr_{\mathbf{w}}\{\|\mathbf{w}\| \geq 2\sqrt{n \cdot m}s\beta\} \leq \sigma \tag{17}$$

and for any unit vector $\mathbf{u}$,

$$\operatorname*{Exp}_{\mathbf{w}}[\langle \mathbf{u}, \mathbf{w}\rangle^2] \leq 4s^2\beta^2. \tag{18}$$

In the following, we show that (17) and (18) are true even when we condition on any fixed values of $\mathbf{C}$, $\mathbf{A}$, and $\mathbf{z}$. The only randomness left is in $\mathbf{y}_1, \ldots, \mathbf{y}_m$ where each $\mathbf{y}_i$ is distributed according to $D_{\mathcal{L}(\mathbf{B})^*,s,\mathbf{c}_i}$.

We first prove (17). We can write a vector $\mathbf{w}$ produced by $\mathcal{W}$ as

$$\mathbf{w} = \mathbf{x} - \mathbf{Y}\mathbf{z} = (\mathbf{x} - \mathbf{C}\mathbf{z}) - (\mathbf{Y} - \mathbf{C})\mathbf{z}.$$

By Lemma 5.8 and (15), the norm of the first term is at most

$$\begin{aligned} \|\mathbf{x} - \mathbf{C}\mathbf{z}\| &\leq \frac{\sqrt{m}n\|\mathbf{S}\| \cdot \|\mathbf{z}\|}{q} \\ &\leq \frac{8\sqrt{m} \cdot n^{1.5}\eta_\epsilon(\mathbf{B}^*) \cdot \beta^2}{q} \\ &\leq 2\beta\eta_\epsilon(\mathbf{B}^*) < s\beta \end{aligned}$$

with probability 1. By Lemma 4.4, for every i, the probability that $\|\mathbf{y}_i - \mathbf{c}_i\| > s\sqrt{n}$ is at most $2^{-n}(1+\epsilon)/(1-\epsilon) = \sigma/N$. Hence, by union bound and triangle inequality, the norm of the second term is bounded by

$$\|(\mathbf{Y} - \mathbf{C})\mathbf{z}\| \leq s\sqrt{n}\sum_{i=1}^{m}|z_i| \leq s\sqrt{n}\sqrt{m}\beta$$

with probability at least $1 - \sigma$. It follows that with probability at least $1 - \sigma$ the norm of $\mathbf{w}$ is bounded by $s\beta + \sqrt{nm}s\beta < 2\sqrt{nm}s\beta$ for all sufficiently large n, proving (17).

Next, we prove (18). Fix some unit vector $\mathbf{u}$ and let us bound the expected value of $\langle \mathbf{u}, \mathbf{w} \rangle^2$. Using the inequality $(a-b)^2 \leq 2a^2 + 2b^2$ (valid for all $a, b \in \mathbb{R}$), we can write

$$\begin{aligned} \langle \mathbf{u}, \mathbf{w} \rangle^2 &= (\langle \mathbf{u}, \mathbf{x} - \mathbf{C}\mathbf{z} \rangle - \langle \mathbf{u}, (\mathbf{Y} - \mathbf{C})\mathbf{z} \rangle)^2 \\ &\leq 2\langle \mathbf{u}, \mathbf{x} - \mathbf{C}\mathbf{z} \rangle^2 + 2\langle \mathbf{u}, (\mathbf{Y} - \mathbf{C})\mathbf{z} \rangle^2 \\ &\leq 2\|\mathbf{x} - \mathbf{C}\mathbf{z}\|^2 + 2\langle \mathbf{u}, (\mathbf{Y} - \mathbf{C})\mathbf{z} \rangle^2. \end{aligned}$$

Using Lemma 4.2, we obtain that for all $i = 1, \ldots, m$,

$$|\operatorname{Exp}[\langle \mathbf{u}, \mathbf{y}_i - \mathbf{c}_i \rangle]| \leq \frac{\epsilon s}{1-\epsilon}, \tag{19}$$

$$\operatorname{Exp}[\langle \mathbf{u}, \mathbf{y}_i - \mathbf{c}_i \rangle^2] \leq \Big(\frac{1}{2\pi} + \frac{\epsilon}{1-\epsilon}\Big) s^2. \tag{20}$$

Using Equations (19) and (20), and Lemma 2.11 with $\mathbf{v}_i = \langle \mathbf{u}, \mathbf{y}_i - \mathbf{c}_i \rangle$ as one-dimensional vectors, we obtain

$$\begin{aligned} \operatorname{Exp}[\langle \mathbf{u}, (\mathbf{Y} - \mathbf{C})\mathbf{z} \rangle^2] &= \operatorname{Exp}\Big[\Big(\sum_{i=1}^{m} \langle \mathbf{u}, \mathbf{y}_i - \mathbf{c}_i \rangle z_i\Big)^2\Big] \\ &\leq \Big(\Big(\frac{1}{2\pi} + \frac{\epsilon}{1-\epsilon}\Big) s^2 + \Big(\frac{\epsilon}{1-\epsilon}\Big)^2 s^2 m\Big) \|\mathbf{z}\|^2 \\ &\leq \Big(\frac{1}{2\pi} + \frac{\epsilon}{1-\epsilon} + \Big(\frac{\epsilon}{1-\epsilon}\Big)^2 m\Big) s^2 \beta^2 \\ &\leq s^2 \beta^2. \end{aligned}$$

Using this bound in the expression for $\operatorname{Exp}[\langle \mathbf{u}, \mathbf{w} \rangle^2]$ we get that

$$\operatorname{Exp}[\langle \mathbf{u}, \mathbf{w} \rangle^2] < 2(s\beta)^2 + 2(s\beta)^2 = 4s^2\beta^2.$$

■

6 Acknowledgments

Part of this work was done while both authors were visiting the Institute for Advanced Study, Princeton. We thank the anonymous referees for their helpful comments.

References

[1] D. Aharonov and O. Regev. Lattice problems in NP intersect coNP. *Journal of the ACM*, 52(5):749–765, 2005. Preliminary version in FOCS 2004.

[2] M. Ajtai. Generating hard instances of lattice problems. In *Proc. 28th ACM Symp. on Theory of Computing*, pages 99–108, 1996. Available from ECCC at `http://www.uni-trier.de/eccc/`.

[3] M. Ajtai and C. Dwork. A public-key cryptosystem with worst-case/average-case equivalence. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing - STOC '97*, pages 284–293, El Paso, TX, USA, May 1997. ACM.

[4] L. Babai. On Lovasz' lattice reduction and the nearest lattice point problem. *Combinatorica*, 6(1):1–13, 1986. Preliminary version in STACS 1985.

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

[6] J. Blömer and J.-P. Seifert. On the complexity of computing short linearly independent vectors and short bases in a lattice. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing - STOC '99*, pages 711–720, Atlanta, GA, USA, May 1999. ACM.

[7] A. Bogdanov and L. Trevisan. On worst-case to average-case reductions for NP problems. In *Proc. 44th Annual IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 308–317, 2003.

[8] J.-Y. Cai. A new transference theorem in the geometry of numbers and new bounds for Ajtai's connection factor. *Discrete Applied Mathematics*, 126(1):9–31, Mar. 2003. Preliminary version in CCC 1999.

[9] J.-Y. Cai and A. Nerurkar. An improved worst-case to average-case connection for lattice problems. In *Proc. 38th IEEE Symp. on Found. of Comp. Science*, pages 468–477, 1997.

[10] 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.

[11] O. Goldreich and S. Goldwasser. On the limits of nonapproximability of lattice problems. *Journal of Computer and System Sciences*, 60(3):540–563, 2000. Preliminary version in STOC 1998.

[12] O. Goldreich, S. Goldwasser, and S. Halevi. Collision-free hashing from lattice problems. Technical Report TR96-056, Electronic Colloquium on Computational Complexity (ECCC), 1996.

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

[14] V. Guruswami, D. Micciancio, and O. Regev. The complexity of the covering radius problem. *Computational Complexity*, 14:90–121, 2005. Preliminary version in CCC 2004.

[15] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[16] S. Khot. Hardness of approximating the shortest vector problem in lattices. In *Proc. 45th Annual IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 126–135, 2004.

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

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

[19] D. Micciancio. The shortest vector problem is NP-hard to approximate to within some constant. *SIAM Journal on Computing*, 30(6):2008–2035, Mar. 2001. Preliminary version in FOCS 1998.

[20] D. Micciancio. A note on the minimal volume of almost cubic parallelepiped. *Discrete and Computational Geometry*, 29(1):133–138, Dec. 2002.

[21] D. Micciancio. Almost perfect lattices, the covering radius problem, and applications to Ajtai's connection factor. *SIAM Journal on Computing*, 34(1):118–169, 2004. Preliminary version in STOC 2002.

[22] D. Micciancio. Generalized compact knapsacks, cyclic lattices, and efficient one-way functions from worst-case complexity assumptions. Technical Report TR04-095, ECCC Electronic Colloquium on Computational Complexity, 2004. Preliminary version in FOCS 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] O. Regev. New lattice-based cryptographic constructions. *Journal of the ACM*, 51(6):899–942, 2004. Preliminary version in STOC 2003.

A Proof of Lemma 5.20

The largest eigenvalue of $\mathbf{W} \cdot \mathbf{W}^T$ is at most $3N\ell^2$ if and only if

$$\frac{1}{N}\sum_{i=1}^{N} \langle \mathbf{u}, \mathbf{w}_i \rangle^2 \leq 3\ell^2$$

for all unit vectors $\mathbf{u} \in \mathbb{R}^n$. In the following, we show that this condition is satisfied with the desired probability. Let $\xi : \mathbb{R}^n \to \mathbb{R}^n$ be the function defined by $\xi(\mathbf{x}) = \mathbf{x}$ if $\|x\| \leq K\ell$ and $\xi(\mathbf{x}) = 0$ otherwise. Clearly, for any unit vector $\mathbf{u}$,

$$\operatorname*{Exp}_{\mathbf{w}\sim D}[\langle \mathbf{u}, \xi(\mathbf{w}) \rangle^2] \leq \operatorname*{Exp}_{\mathbf{w}\sim D}[\langle \mathbf{u}, \mathbf{w} \rangle^2] \leq \ell^2.$$

Moreover, the random variable $\langle \mathbf{u}, \xi(\mathbf{w}) \rangle^2$ takes values in the interval $[0, (K\ell)^2]$. Hence, Hoeffding's inequality (14) implies that for any unit vector $\mathbf{u}$, a sequence of samples $\mathbf{w}_1, \ldots, \mathbf{w}_N$ from D satisfies

$$\frac{1}{N}\sum_{i=1}^{N} \langle \mathbf{u}, \xi(\mathbf{w}_i) \rangle^2 \leq 2\ell^2 \tag{21}$$

with probability at least $1 - e^{-N/K^4}$.

Consider an ϵ-net A on the unit sphere with parameter $\epsilon = \frac{1}{2}K^{-2}$, i.e., a set of points A such that any point on the unit sphere is within distance ϵ from some point in A. It is possible to construct such nets of size at most $(2\sqrt{n}/\epsilon)^n$. For instance, let C be $[-1, 1]^n$, i.e., the n-dimensional cube of edge length 2. Notice that C contains the unit sphere. Partition C into $(2\sqrt{n}/\epsilon)^n$ small cubes of edge length $\epsilon/\sqrt{n}$. For each small cube that intersects the n-dimensional sphere, choose any point in the intersection and include it in A. It is easy to see that the collection of these points constitutes an ϵ-net on the sphere, because any point in the sphere belongs to one of the small cubes, and the diameter of each small cube is exactly ϵ.

We now apply the union bound on the set of all unit vectors $\mathbf{u}$ in A. It follows that (21) holds with probability at least $1 - e^{-N/K^4}(4\sqrt{n}K^2)^n$ for all $\mathbf{u}$ in the net A *simultaneously*.

Next, we show that if (21) holds for all $\mathbf{u} \in A$, then a slightly weaker version of it holds for *all* unit vectors. Consider an arbitrary unit vector $\mathbf{u}'$. Let $\mathbf{u} \in A$ be the closest point to $\mathbf{u}'$ in A. Notice that $\|\mathbf{u} - \mathbf{u}'\| \leq \epsilon$. Thus,

$$\begin{aligned}\Big|\frac{1}{N}\sum_{i=1}^{N} \langle \mathbf{u}', \xi(\mathbf{w}_i) \rangle^2 - \frac{1}{N}\sum_{i=1}^{N} \langle \mathbf{u}, \xi(\mathbf{w}_i) \rangle^2\Big| &\leq \frac{1}{N}\sum_{i=1}^{N} |\langle \mathbf{u}' - \mathbf{u}, \xi(\mathbf{w}_i) \rangle \langle \mathbf{u}' + \mathbf{u}, \xi(\mathbf{w}_i) \rangle| \\ &\leq 2\epsilon \max_i \|\xi(\mathbf{w}_i)\|^2 \leq 2\epsilon(K\ell)^2 = \ell^2.\end{aligned}$$

This yields that with probability at least $1 - e^{-N/K^4}(4\sqrt{n}K^2)^n$ over the choice of the $\mathbf{w}_i$'s it holds that

$$\frac{1}{N}\sum_{i=1}^{N} \langle \mathbf{u}, \xi(\mathbf{w}_i) \rangle^2 \leq 2\ell^2 + \ell^2 = 3\ell^2$$

for all unit vectors $\mathbf{u}$. It remains to notice that with probability at least $1 - N\sigma$, $\xi(\mathbf{w}_i) = \mathbf{w}_i$ for all i.