Sampling Discrete Gaussians Efficiently and Obliviously

Shweta Agrawal [*] Craig Gentry [†] Shai Halevi [‡] Amit Sahai[§]

Abstract

In this work we construct an algorithm for sampling Discrete Gaussians efficiently and obliviously. Previously discrete Gaussian samplers have been constructed in [GPV08, Pei10], where the algorithms take as input a "high quality" basis and produce an output whose quality depends on the input basis quality. Our algorithm produces a discrete Gaussian of somewhat worse quality than [GPV08, Pei10] but with the advantage that it does not require access to an explicit description of the underlying lattice, for example it suffices for our purposes to have encryptions of lattice vectors under an additively homomorphic encryption scheme. At the heart of our work is the fundamental question *how do sums of discrete Gaussians behave?* Unlike their continuous counterparts, discrete Gaussians are not that well understood. We believe that our work fills in some important gaps of this understanding. Our results are already important in enabling the exciting new work on multilinear maps [GGH12], and since the questions we resolve arise naturally, we believe that our work will find application in other areas as well.

The second and third authors were supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D11PC20202. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

[*]UCLA. Email: shweta@cs.ucla.edu
[†]IBM Research. Email: craigbgentry@gmail.com
[‡]IBM Research. Email: shaih@alum.mit.edy
[§]UCLA. Email: sahai@cs.ucla.edu.

1 Introduction

Lattice based cryptography has recently acquired a great deal of attention and interest due to several attractive features – worst case guarantees, resistance to quantum attacks and suitability for constructing rich and exciting cryptographic primitives such as fully homomorphic encryption [Gen09, Gen10, vDGHV10, BV11a, BV11b, BGV12], functional encryption [GPV08, CHKP10, ABB10, AFV11, ABV$^+$12], digital signatures [GPV08, Boy10], PRFs [BPR12] and many more. At the heart of many of these schemes lies a *discrete Gaussian sampler* – an algorithm that samples points on a lattice, with probability proportional to a Gaussian distribution. Discrete Gaussian distributions show up frequently in the techniques of lattice based cryptography, most notably in the famous "Learning with Errors" assumption, but also often simplifying proofs and yielding better analysis [MR04, Reg09]. The first Gaussian sampler was proposed by Gentry et al, which immediately enabled numerous important applications [GPV08]. Subsequently, Peikert described a different sampler, which is more efficient than GPV's sampler and can be parallelized easily at the price of producing a slightly "lower-quality" Gaussian distribution. Both these samplers take as input an explicit description of a "high quality basis" of the relevant lattice, and the quality of their output distribution is related to the quality of the input basis.

In this work we describe yet another Gaussian sampler, one that can get by with having only an *implicit description* of the input lattice. For example, suppose we are given a set of lattice vectors, encrypted under an additively homomorphic scheme. Then, our sampler can be use to sample (encryptions of) well behaved discrete Gaussian on the underlying lattice.

1.1 Fundamental Questions

At the core of our new sampling algorithm lies the fundamental question: *how do sums of discrete Gaussians behave*? Formally, the discrete Gaussian distribution $\mathcal{D}_{L,s,\vec{c}}$ over a lattice L, with standard deviation s and center $\vec{c}$, assigns a point $\vec{x} \in L$ probability proportional to $e^{-\pi\|\vec{x}-\vec{c}\|^2/s^2}$, and for $\vec{x} \notin L$, assigns probability 0. Despite their utility and our excellent understanding of continuous Gaussians, there are curious gaps in our understanding of discrete Gaussians.

We know that, if x and y are two independent random variables that are normally distributed, then $x+y$ is also normally distributed, with its mean being the sum of the two means, and its variance being the sum of the two variances. However, things become more complicated when the distributions are discrete. It is natural to ask: if x and y are chosen according to discrete Gaussian distributions, is the distribution of $x+y$ statistically close to a discrete Gaussian? This is not true in general. Consider for example, the sum of two discrete Gaussians $\mathcal{D}_{\mathbb{Z},5}$ and $\mathcal{D}_{100\cdot\mathbb{Z},500}$. Clearly, in this distribution, the numbers 100 and 200 have a good chance of being chosen, while 150 does not. The "shape" of the sum of two discrete Gaussian distributions looks very non-Gaussian in general.

More generally, given an n dimensional lattice L, and vectors $X = [\vec{x}_1|\vec{x}_2|\dots|\vec{x}_m]^\top$ where $\vec{x}_i \in L$, we can ask the following fundamental questions: If $\vec{z} \sim \mathcal{D}_{\mathbb{Z}^m,s'}$, then under what conditions (on X and s') is $X^\top\vec{z}$ statistically close to a discrete Gaussian over L? Can we control the "shape" of the resultant discrete Gaussian, and make it spherical, or close to spherical? The first question has not been answered satisfactorily *even in the one-dimensional setting*: given fixed $\vec{y} \in \mathbb{Z}^m$, how large does s' need to be for $\langle\vec{y},\vec{z}\rangle$ to be statistically close to a (one-dimensional) Gaussian over $\mathbb{Z}$ when $\vec{z} \sim \mathcal{D}_{\mathbb{Z}^m,s'}$?

1.2 Sampling Discrete Gaussians Efficiently and Obliviously

The above questions are interesting not just from the theoretical perspective but also have great relevance from the application standpoint. One application is the construction of efficient spherical Gaussian samplers [GPV08, Pei10]. Peikert's sampler [Pei10] is elegant and its complexity is difficult to beat: the only online computation is to compute $\vec{c} - B_1\lfloor B_1^{-1}(\vec{c} - \vec{x}_2)\rceil$, where $\vec{c}$ is the center of the Gaussian, B_1 is the sampler's basis for its lattice L, and $\vec{x}_2$ is a vector that is generated in an offline phase (freshly for each sampling) in a way designed to "cancel" the covariance of B_1 so as to induce a purely spherical Gaussian. However, it is conceivable that a faster and even more natural sampler Gaussian exists, at least when $\vec{c} = 0$. Specifically, consider the following sampler. In an offline phase, for $m > n$, the sampler samples a set of short vectors $X = [\vec{x}_1|\vec{x}_2|\dots|\vec{x}_m]^\top$ from L – e.g., using GPV or Peikert's algorithm. Then, in the online phase, the sampler generates $\vec{z} \in \mathbb{Z}^m$ per the Gaussians $z_i \sim \mathcal{D}_{\mathbb{Z},s_i}$, and simply outputs $X^\top\vec{z}$. Conceivably, since this sampler just directly takes an integer linear combination of lattice vectors, and does not require extra precision for handling the inverse B_1^{-1}, it might outperform Peikert's in some situations. But does this simpler sampler work – i.e., can we say anything about its output distribution? Also, how small can we make the dimension m of $\vec{z}$ and how small can we make the entries of $\vec{z}$? Ideally m would be not much larger than the dimension of the lattice and the entries of $\vec{z}$ have small variance – e.g., $\tilde{O}(\sqrt{n})$.

As another natural cryptographic application, suppose you want to sample a lattice point according to a canonical near-spherical Gaussian, without explicit access to a good quality basis of the lattice? For example, suppose you are given lattice points encrypted under an additively homomorphic encryption scheme and wish to use them to generate an encrypted well behaved Gaussian on the underlying lattice. Previous samplers [GPV08, Pei10] are too complicated to use within an additively homomorphic encryption scheme. As noted by [Pei10], one can indeed generate an ellipsoidal Gaussian distribution over the lattice given a basis B by just outputting $\vec{y} \leftarrow B \cdot \vec{z}$ where $\vec{z} \sim \mathcal{D}_{\mathbb{Z}^n,s}$, but this ellipsoidal Gaussian distribution would typically be very skewed. So, it seems natural to ask: can we get a "more spherical" Gaussian by using a slightly larger set $X = [\vec{x}_1|\vec{x}_2|\dots|\vec{x}_m]^\top$ of lattice vectors?

The generation and randomization of ciphertexts in certain lattice-based homomorphic encryption schemes [Gen09, Gen10, vDGHV10, BV11a, BV11b, BGV12] also raise questions about the sum of discrete Gaussians. A common strategy in these schemes is to encrypt a plaintext m by taking the sum of m with a random linear combination of encryptions of 0 that are provided in the public key. Similarly, a common way to "randomize" a ciphertext so as to provide "function privacy" (to hide what function was evaluated on the plaintexts) is to add a random encryption of 0 that "drowns" the original ciphertext. In most of these schemes [Gen09, Gen10, BV11a, BV11b, BGV12], ciphertexts live in a finite ring, and arguing security involves a straightforward application of the leftover hash lemma [ILL89, HILL99]. However, in the integer-based fully homomorphic encryption scheme [vDGHV10], ciphertexts live in $\mathbb{Z}$. van Dijk et al. still manage to apply the left-over hash lemma in this setting, but awkwardly; they complicate the encryption procedure by reducing the linear combination of ciphertexts modulo a large ciphertext, so as to bring the scheme back in to the realm of finite rings where the leftover hash lemma is naturally applied. Rothblum [Rot11] proves a nice information theoretic lemma (incomparable, but in a similar spirit, to our results here) and uses this lemma to show that the complication in [vDGHV10] can be eliminated by restricting plaintexts to bits. However, this and other previous results do not answer whether the integer-based fully homomorphic encryption scheme can use the "natural" sum-of-ciphertexts encryption procedure without restricting its plaintext space to bits.[1] Can we use the sum-of-ciphertexts approach when the sum is a Gaussian linear combination? To answer this, we need to resolve basic questions about sums of discrete Gaussians in the one-dimensional setting.

[1] The security proof in [vDGHV10] relies on plaintext space being bits, but can be adapted easily to larger spaces.

Finally, a recent construction of (lattice-based) cryptographic multilinear maps [GGH12] requires an efficient procedure for randomizing encodings. In their scheme, encodings are somewhat like encryptions, but are not semantically secure by design (the encodings permit an equality test). For security, they need to argue that the encodings generated by legitimate parties have nice distributions, even though the encodings are generated via a fast sum-of-encodings procedure that is oblivious to the underlying lattice structure. Our theorem allows them to argue this.

1.3 Our Results

Consider an n dimensional lattice L and vectors $X = [\vec{x}_1|\vec{x}_2|\dots|\vec{x}_m]^\top$ where $\vec{x}_i \leftarrow \mathcal{D}_{L,S}$ for some matrix S. Let $\vec{z} \leftarrow \mathcal{D}_{\mathbb{Z}^m,s'}$. We analyze the conditions under which the vector $X^\top\vec{z}$ is statistically close to a "near-spherical" discrete Gaussian. Formally, consider:

$$\mathcal{E}_{X,s'} \stackrel{\text{def}}{=} \{X^\top\vec{z} : \vec{z} \leftarrow \mathcal{D}_{\mathbb{Z}^m,s'}\}.$$

Then, we prove that $\mathcal{E}_{X,s'}$ is close to a discrete Gaussian over L of moderate "width". Specifically, we show that for large enough s', with overwhelming probability over the choice of X:

1. The distribution $\mathcal{E}_{X,s'}$ is statistically close to the ellipsoid Gaussian $\mathcal{D}_{L,s'X}$, over L.
2. The singular values of the matrix X are of size roughly $s\sqrt{m}$, hence the shape of $\mathcal{D}_{L,s'X}$ is "roughly spherical". Moreover, the "width" of $\mathcal{D}_{L,s'X}$ is roughly $s's\sqrt{m} = \text{poly}(n)$.

We emphasize that it is straightforward to show that the covariance matrix of $\mathcal{E}_{X,s'}$ is exactly $s'^2X^\top X$. However, the technical challenge lies in showing that $\mathcal{E}_{X,s'}$ is close to a discrete Gaussian for a non-square X. Also note that for a square X, the shape of the covariance matrix $X^\top X$ will typically be very "skewed" (i.e., the least singular value of X is typically much smaller than the largest singular value).

Another way to view our result is as an extension of the leftover hash lemma to infinite rings – that is, to a setting where we are not permitted to perform modular reduction on the final result.

1.4 Our Techniques

Our main result can be argued along the following broad outline. Our first theorem (Theorem 2) says that the distribution of $X^\top\vec{z} \leftarrow \mathcal{E}_{X,s'}$ is indeed statistically close to a discrete Gaussian over L, as long as s' exceeds the smoothing parameter of the "orthogonal lattice" of X (denoted A). Next, theorem 3 clarifies that A will have a small smoothing parameter as long as X is "regularly shaped" in a certain sense. Finally, we argue in Lemma 8 that when the columns of X are chosen from a discrete Gaussian, $\vec{x}_i \leftarrow \mathcal{D}_{L,S}$, then X is "regularly shaped," i.e. has singular values all close to $\sigma_n(S)\sqrt{m}$.

The analysis of the smoothing parameter of X's "orthogonal lattice" A is particularly challenging and requires careful analysis of a certain "dual lattice" related to A. Specifically, we proceed by first embedding A into a full rank lattice A_q and then move to study M_q – the (scaled) dual of A_q. Here we obtain a lower bound on $\lambda_{n+1}(M_q)$, i.e. the $n+1^{th}$ minima of M_q. Next, we use a theorem by Banasczcyk to convert the lower bound on $\lambda_{n+1}(M_q)$ to an upper bound on $\lambda_{m-n}(A_q)$, obtaining $m-n$ linearly independent, bounded vectors in A_q. We argue that these vectors belong to A, thus obtaining an upper bound on $\lambda_{m-n}(A)$. Relating $\lambda_{m-n}(A)$ to $\eta_\epsilon(A)$ using a lemma by Micciancio and Regev completes the analysis.

To argue that X is regularly shaped, we begin with the literature of random matrices which establishes that for a matrix $H \in \mathbb{R}^{m\times n}$, where each row of H is distributed as $\vec{h}_i \sim \mathcal{N}(0, s^2)$ and m is sufficiently

greater than n, then the singular values of H are all of size roughly $s\sqrt{m}$. We extend this result to discrete Gaussians – showing that as long as each row $\vec{x}_i \leftarrow \mathcal{D}_{L,S}$ where S is "not too small" and "not too skewed", then with high probability the singular values of X are all of size roughly $s\sqrt{m}$.

1.5 Related Work

Discrete Gaussian samplers have been studied by [GPV08] and [Pei10]. Both these works describe a discrete Gaussian sampling algorithm that takes as input a 'high quality' basis B for an n dimensional lattice L and output a sample from $\mathcal{D}_{L,s,\vec{c}}$. In [GPV08], $s \geq \|\tilde{B}\| \cdot \omega(\sqrt{\log n})$, and $\tilde{B} = \max_i \|\tilde{b}_i\|$ is the Gram Schmidt orthogonalization of B. In contrast, the algorithm of [Pei10] requires $s \geq \sigma_1(B)$, i.e. the largest singular value of B, but is fully parallelizable. In contrast, our sampler produces a Gaussian with worse width – roughly $ss'\sqrt{m}$ where $s \geq \eta_\epsilon(L)$, $s' \geq 4mn\ln(1/\epsilon)$, but has the advantage of being *linearly computable*, thus making it possible to sample from the required distribution given only encryptions (under an additively homomorphic encryption scheme) of lattice vectors.

More recently, Boneh and Freeman [BF11] observed that, under certain conditions, a sum of two discrete Gaussian distribution over different lattices may be a discrete Gaussian over the sum of the two lattices. However, the deviations of the Gaussians needed to achieve this are quite large.

2 Preliminaries

We begin by defining some notation that will be used throughout the paper. We say that a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is negligible if for all $d > d_0$ we have $f(\lambda) < 1/\lambda^d$ for sufficiently large λ. We write $f(\lambda) < \mathsf{negl}(\lambda)$. For two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ over some set Ω we define the statistical distance $\mathrm{SD}(\mathcal{D}_1, \mathcal{D}_2)$ as

$$\mathrm{SD}(\mathcal{D}_1, \mathcal{D}_2) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{x \in \Omega} \left| \Pr_{\mathcal{D}_1}[x] - \Pr_{\mathcal{D}_2}[x] \right|$$

We say that two distribution ensembles $\mathcal{D}_1(\lambda)$ and $\mathcal{D}_2(\lambda)$ are statistically close or statistically indistinguishable if $\mathrm{SD}\big(\mathcal{D}_1(\lambda), \mathcal{D}_2(\lambda)\big)$ is a negligible function of λ.

2.1 Gaussian Distributions

For any real $s > 0$ and vector $\vec{c} \in \mathbb{R}^n$, define the (spherical) Gaussian function on $\mathbb{R}^n$ centered at $\vec{c}$ with parameter s as $\rho_{s,\vec{c}}(\vec{x}) = \exp(-\pi\|\vec{x} - \vec{c}\|^2/s^2)$ for all $\vec{x} \in \mathbb{R}^n$. The *normal distribution* with mean μ and deviation σ, denoted $\mathcal{N}(\mu, \sigma^2)$, assigns to each real number $x \in \mathbb{R}$ the probability density $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \rho_{\sigma\sqrt{2\pi},\mu}(x)$. The n-dimensional (spherical) continuous Gaussian distribution with center $\vec{c}$ and uniform deviation σ^2, denoted $\mathcal{N}^n(\vec{c}, \sigma^2)$, just chooses each entry of a dimension-n vector independently from $\mathcal{N}(c_i, \sigma^2)$.

The n-dimensional spherical Gaussian function generalizes naturally to ellipsoid Gaussians, where the different coordinates are jointly Gaussian but are neither identical nor independent. In this case we replace the single variance parameter $s^2 \in \mathbb{R}$ by the covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ (which must be positive-definite and symmetric). To maintain consistency of notations between the spherical and ellipsoid cases, below we let S be a matrix such that $S^\top \times S = \Sigma$. Such a matrix S always exists for a symmetric Σ, but it is not unique. (In fact there exist such S'es that are not even n-by-n matrices, below we often work with such rectangular S'es.)

For a rank-n matrix $S \in \mathbb{R}^{m\times n}$ and a vector $\vec{c} \in \mathbb{R}^n$, the ellipsoid Gaussian function on $\mathbb{R}^n$ centered at $\vec{c}$ with parameter S is defined by

$$\rho_{S,\vec{c}}(\vec{x}) = \exp\left(-\pi(\vec{x}-\vec{c})^\top (S^\top S)^{-1}(\vec{x}-\vec{c})\right) \ \forall \vec{x} \in \mathbb{R}^n.$$

Obviously this function only depends on $\Sigma = S^\top S$ and not on the particular choice of S. It is also clear that the spherical case can be obtained by setting $S = sI_n$, with I_n the n-by-n identity matrix. Below we use the shorthand $\rho_s(\cdot)$ (or $\rho_S(\cdot)$) when the center of the distribution is $\vec{0}$.

2.2 Matrices and Singular Values

In this note we often use properties of rectangular (non-square) matrices. For $m \geq n$ and a rank-n matrix $X \in \mathbb{R}^{m\times n}$, the pseudoinverse of X is the (unique) m-by-n matrix Y such that $X^\top Y = Y^\top X = I_n$ and the columns of Y span the same linear space as the columns of X. It is easy to see that Y can be expressed as $Y = X(X^\top X)^{-1}$ (note that $X^\top X$ is invertible since X has rank n).

For a rank-n matrix $X \in \mathbb{R}^{m\times n}$, denote $U_X = \{\|X\vec{u}\| : \vec{u} \in \mathbb{R}^n, \|\vec{u}\| = 1\}$. The *least singular value* of X is then defined as $\sigma_n(X) = \inf(U_X)$ and similarly the *largest singular value* of X is $\sigma_1(X) = \sup(U_X)$. Some properties of singular values that we use later in the text are stated in Fact 1.

Fact 1. *For rank-n matrices $X, Y \in \mathbb{R}^{m\times n}$ with $m \geq n$, the following holds:*

1. *If $X^\top X = Y^\top Y$ then X, Y have the same singular values.*
2. *If Y is the (pseudo)inverse of X then the singular values of X, Y are reciprocals.*
3. *If X is a square matrix (i.e., $m = n$) then X, X^T have the same singular values.*
4. *If $\sigma_1(Y) \leq \delta\sigma_n(X)$ for some constant $\delta < 1$, then $\sigma_1(X+Y) \in [1-\delta, 1+\delta]\sigma_1(X)$ and $\sigma_n(X+Y) \in [1-\delta, 1+\delta]\sigma_n(X)$.* □

It is well known that when m is sufficiently larger than n, then the singular values of a "random matrix" $X \in \mathbb{R}^{m\times n}$ are all of size roughly $\sqrt{m}$. For example, Lemma 1 below is a special case of [LPRTJ05, Thm 3.1], and Lemma 2 can be proven along the same lines of (but much simpler than) the proof of [Tao12, Corollary 2.3.5].

Lemma 1. *There exists a universal constant $C > 1$ such that for any $m > 2n$, if the entries of $X \in \mathbb{R}^{m\times n}$ are drawn independently from $\mathcal{N}(0,1)$ then $\Pr[\sigma_n(X) < \sqrt{m}/C] < \exp(-O(m))$.* □

Lemma 2. *There exists a universal constant $C > 1$ such that for any $m > 2n$, if the entries of $X \in \mathbb{R}^{m\times n}$ are drawn independently from $\mathcal{N}(0,1)$ then $\Pr[\sigma_1(X) > C\sqrt{m}] < \exp(-O(m))$.* □

Corollary 1. *There exists a universal constant $C > 1$ such that for any $m > 2n$ and $s > 0$, if the entries of $X \in \mathbb{R}^{m\times n}$ are drawn independently from $\mathcal{N}(0, s^2)$ then*

$$\Pr\left[s\sqrt{m}/C \ < \ \sigma_n(X) \ \leq \ \sigma_1(X) \ < \ sC\sqrt{m}\right] > 1 - \exp(-O(m)). \qquad \square$$

Remark. The literature on random matrices is mostly focused on analyzing the "hard cases" of more general distributions and m which is very close to n (e.g., $m = (1+o(1))n$ or even $m = n$). For our purposes, however, we only need the "easy case" where all the distributions are Gaussian and $m \gg n$ (e.g., $m = n^2$), in which case all the proofs are much easier (and the universal constant from Corollary 1 get closer to one).

2.3 Lattices and their Dual

A lattice $L \subset \mathbb{R}^n$ is an additive discrete sub-group of $\mathbb{R}^n$. We denote by $\text{span}(L)$ the linear subspace of $\mathbb{R}^n$, spanned by the points in L. The rank of $L \subset \mathbb{R}^n$ is the dimension of $\text{span}(L)$, and we say that L has full rank if its rank is n. In this work we often consider lattices of less than full rank.

Every (nontrivial) lattice has bases: a basis for a rank-k lattice L is a set of k linearly independent points $\vec{b}_1, \ldots, \vec{b}_k \in L$ such that $L = \{\sum_{i=1}^{k} z_i \vec{b}_i : z_i \in \mathbb{Z}\ \forall i\}$. If we arrange the vectors $\vec{b}_i$ as the columns of a matrix $B \in \mathbb{R}^{n \times k}$ then we can write $L = \{B\vec{z} : \vec{z} \in \mathbb{Z}^k\}$. If B is a basis for L then we say that B spans L.

Definition 1 (Dual of a Lattice)**.** *For a lattice $L \subset \mathbb{R}^n$, its* dual lattice *consists of all the points in* $\text{span}(L)$ *that are orthogonal to L modulo one, namely:*

$$L^* = \{\vec{y} \in \text{span}(L) : \forall \vec{x} \in L, \langle \vec{x}, \vec{y} \rangle \in \mathbb{Z}\}$$

Clearly, if L is spanned by the columns of some rank-k matrix $X \in \mathbb{R}^{n \times k}$ then L^* is spanned by the columns of the pseudoinverse of X. It follows from the definition that for two lattices $L \subseteq M$ we have $M^* \cap \text{span}(L) \subseteq L^*$.

Banasczcyk provided strong transference theorems that relate the size of short vectors in L to the size of short vectors in L^*. Recall that $\lambda_i(L)$ denotes the i-th minimum of L (i.e., the smallest s such that L contains i linearly independent vectors of size at most s).

Theorem 1 (Banasczcyk [Ban93])**.** *For any rank-n lattice $L \subset \mathbb{R}^m$, and for all $i \in [n]$,*

$$1 \leq \lambda_i(L) \cdot \lambda_{n-i+1}(L^*) \leq n.$$

2.4 Gaussian Distributions over Lattices

The *ellipsoid discrete Gaussian distribution* over lattice L with parameter S, centered around $\vec{c}$, is

$$\forall\ \vec{x} \in L, \mathcal{D}_{L,S,\vec{c}}(\vec{x}) = \frac{\rho_{S,\vec{c}}(\vec{x})}{\rho_{S,\vec{c}}(L)}\ ,$$

where $\rho_{S,\vec{c}}(A)$ for set A denotes $\sum_{\vec{x} \in A} \rho_{S,\vec{c}}(\vec{x})$. In other words, the probability $\mathcal{D}_{L,S,\vec{c}}(\vec{x})$ is simply proportional to $\rho_{S,\vec{c}}(\vec{x})$, the denominator being a normalization factor. The same definitions apply to the spherical case, which is denoted by $\mathcal{D}_{L,s,\vec{c}}(\cdot)$ (with lowercase s). As before, when $\vec{c} = \vec{0}$ we use the shorthand $\mathcal{D}_{L,S}$ (or $\mathcal{D}_{L,s}$). The following useful fact that follows directly from the definition, relates the ellipsoid Gaussian distributions over different lattices:

Fact 2. *Let $L \subset \mathbb{R}^n$ be a full-rank lattice, $\vec{c} \in R^n$ a vector, and $S \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times n}$ two rank-n matrices, and denote $L' = \{B^{-1}\vec{v} : \vec{v} \in L\}$, $\vec{c'} = B^{-1}\vec{c}$, and $S' = S \times (B^\top)^{-1}$. Then the distribution $\mathcal{D}_{L,S,\vec{c}}$ is identical to the distribution induced by drawing a vector $\vec{v} \leftarrow \mathcal{D}_{L',S',\vec{c'}}$ and outputting $\vec{u} = B\vec{v}$.* □

A useful special case of Fact 2 is when L' is the integer lattice, $L' = \mathbb{Z}^n$, in which case L is just the lattice spanned by the basis B. In other words, the ellipsoid Gaussian distribution on $L(B)$, $\vec{v} \leftarrow \mathcal{D}_{L(B),S,\vec{c}}$, is induced by drawing an integer vector according to $\vec{z} \leftarrow \mathcal{D}_{\mathbb{Z}^n,S',\vec{c'}}$ and outputting $\vec{v} = B\vec{z}$, where $S' = S(B^{-1})^\top$ and $\vec{c'} = B^{-1}\vec{c}$.

Another useful special case is where $S = sB^\top$, so S is a square matrix and $S' = sI_n$. In this case the ellipsoid Gaussian distribution $\vec{v} \leftarrow \mathcal{D}_{L,S,\vec{c}}$ is induced by drawing a vector according to the *spherical Gaussian* $\vec{u} \leftarrow \mathcal{D}_{L',s,\vec{c'}}$ and outputting $\vec{v} = \frac{1}{s}S^\top\vec{u}$, where $\vec{c'} = s(S^\top)^{-1}\vec{c}$ and $L' = \{s(S^\top)^{-1}\vec{v} : \vec{v} \in L\}$.

Smoothing parameter. As in [MR07], for lattice L and real $\epsilon > 0$, the *smoothing parameter* of L, denoted $\eta_\epsilon(L)$, is defined as the smallest s such that $\rho_{1/s}(L^* \setminus \{\vec{0}\}) \leq \epsilon$. Intuitively, for a small enough ϵ, the number $\eta_\epsilon(L)$ is sufficiently larger than L's fundamental parallelepiped so that sampling from the corresponding Gaussian "wipes out the internal structure" of L. Thus, the sparser the lattice, the larger its smoothing parameter.

It is well known that for a spherical Gaussian with parameter $s > \eta_\epsilon(L)$, the size of vectors drawn from $\mathcal{D}_{L,s}$ is bounded by $s\sqrt{n}$ whp (cf. [MR07, Lemma 4.4]). The following lemma (that follows easily from the spherical case and Fact 2) is a generalization to ellipsoid Gaussians.

Lemma 3. *For a rank-n lattice L, vector $\vec{c} \in \mathbb{R}^n$, constant $0 < \epsilon < 1$ and matrix S s.t. $\sigma_n(S) \geq \eta_\epsilon(L)$, we have that for $\vec{v} \leftarrow \mathcal{D}_{L,S,\vec{c}}$,*

$$\Pr_{\vec{v} \leftarrow \mathcal{D}_{L,S,\vec{c}}}\left(\|\vec{v} - \vec{c}\| \geq \sigma_1(S)\sqrt{n}\right) \leq \frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n}$$

Proof. We can assume w.l.o.g. that S is a square matrix (since $\mathcal{D}_{L,S,c}$ depends only on $S^\top S$, and all the matrices that agree on $S^\top S$ have the same singular values). Letting $s = \sigma_n(S)$, we apply Fact 2 with $B = \frac{1}{s}S^\top$, so we have $S' = sI_n$, $\vec{c'} = s(S^\top)^{-1}\vec{c}$, and $L' = \{s(S^\top)^{-1}\vec{v} : \vec{v} \in L\}$). Namely the ellipsoid Gaussian distribution $\vec{v} \leftarrow \mathcal{D}_{L,S,\vec{c}}$ is induced by drawing a vector according to the *spherical Gaussian* $\vec{u} \leftarrow \mathcal{D}_{L',s,\vec{c'}}$ and outputting $\vec{v} = \frac{1}{s}S^\top\vec{u}$.

We recall that the largest singular value of $(S^\top)^{-1}$ is the reciprocal of the least singular value of $S^\top$ (which is $\sigma_n(S^\top) = \sigma_n(S) = s$), namely $\sigma_1((S^\top)^{-1}) = 1/s$. Hence the singular values of the matrix $s(S^\top)^{-1}$ are all at most one, which means that multiplying by $s(S^\top)^{-1}$ is "shrinking", $\|s(S^\top)^{-1}\vec{v}\| \leq \|\vec{v}\|$ for all $\vec{v}$. Since the lattice L' is obtained from L by "shrinking" all the vectors $\vec{v} \in L$ as above, it follows that the smoothing parameter of L' is no larger than that of L, so $s = \sigma_n(S) \geq \eta_\epsilon(L) \geq \eta_\epsilon(L')$.

Applying now [MR07, Lemma 4.4] for the spherical case, when drawing a vector $\vec{u} \leftarrow \mathcal{D}_{L',s,\vec{c'}}$ we get $\|\vec{u}\| \leq s\sqrt{n}$ except with probability at most $\frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n}$. Hence we can bound whp the norm of $\vec{v}$ by $\|\vec{v}\| = \|\frac{1}{s}S^\top\vec{u}\| \leq \frac{1}{s} \cdot \sigma_1(S^\top) \cdot \|\vec{u}\| = \frac{1}{s} \cdot \sigma_1(S) \cdot s\sqrt{n} = \sigma_1(S)\sqrt{n}$. □

The next lemma says that the Gaussian distribution with parameter $s \geq \eta_\epsilon(L)$ is so smooth and "spread out" that it covers the approximately the same number of L-points regardless of where the Gaussian is centered. This is again well known for spherical distributions (cf. [GPV08, Lemma 2.7]) and the generalization to ellipsoid distributions is immediate using Fact 2.

Lemma 4. *For any rank-n lattice L, real $\epsilon \in (0,1)$, vector $c \in \mathbb{R}^n$, and rank-n matrix $S \in \mathbb{R}^{m \times n}$ such that $\sigma_n(S) \geq \eta_\epsilon(L)$, we have $\rho_{S,\vec{c}}(L) \in [\frac{1-\epsilon}{1+\epsilon}, 1] \cdot \rho_S(L)$.* □

Regev also proved that drawing a point from L according to a spherical discrete Gaussian and adding to it a spherical continuous Gaussian, yields a probability distribution close to a continuous Gaussian (independent of the lattice), provided that both distributions have parameters sufficiently larger than the smoothing parameter of L.

Lemma 5 (Claim 3.9 of [Reg09])**.** *Fix any n-dimensional lattice $L \subset \mathbb{R}^n$, real $\epsilon \in (0, 1/2)$, and two reals s, r such that $\frac{rs}{\sqrt{r^2+s^2}} \geq \eta_\epsilon(L)$, and denote $t = \sqrt{r^2+s^2}$.*

Let $\mathcal{R}_{L,r,s}$ be a distribution induced by choosing $\vec{x} \leftarrow \mathcal{D}_{L,s}$ from the spherical discrete Gaussian on L and $\vec{y} \leftarrow \mathcal{N}^n(0, r^2/2\pi)$ from a continuous Gaussian, and outputting $\vec{z} = \vec{x} + \vec{y}$. Then for any point $\vec{u} \in \mathbb{R}^n$,

the probability density $\mathcal{R}_{L,r,s}(\vec{u})$ is close to the probability density under the spherical continuous Gaussian $\mathcal{N}^n(0, t^2/2\pi)$ upto a factor of $\frac{1-\epsilon}{1+\epsilon}$:

$$\tfrac{1-\epsilon}{1+\epsilon}\mathcal{N}^n(0,t^2/2\pi)(\vec{u}) \;\le\; \mathcal{R}_{L,r,s}(\vec{u}) \;\le\; \tfrac{1+\epsilon}{1-\epsilon}\mathcal{N}^n(0,t^2/2\pi)(\vec{u})$$

In particular, the statistical distance between $\mathcal{R}_{L,r,s}$ and $\mathcal{N}^n(0, t^2/2\pi)$ is at most 4ϵ.

More broadly, Lemma 5 implies that for any event $E(\vec{u})$, we have

$$\Pr_{\vec{u}\leftarrow\mathcal{N}(0,t^2/2\pi)}[E(\vec{u})]\cdot\tfrac{1-\epsilon}{1+\epsilon} \;\le\; \Pr_{\vec{u}\leftarrow\mathcal{R}_{L,r,s}}[E(\vec{u})] \;\le\; \Pr_{\vec{u}\leftarrow\mathcal{N}(0,t^2/2\pi)}[E(\vec{u})]\cdot\tfrac{1+\epsilon}{1-\epsilon}$$

Another useful property of "wide" discrete Gaussian distributions is that they do not change much by short shifts. Specifically, if we have an arbitrary subset of the lattice, $T \subseteq L$, and an arbitrary *short vector* $\vec{v} \in L$, then the probability mass of T is not very different than the probability mass of $T - \vec{v} = \{\vec{u} - \vec{v} : \vec{u} \in T\}$. Below let $\mathsf{erf}(\cdot)$ denote the Gauss error function.

Lemma 6. *Fix a lattice $L \subset \mathbb{R}^n$, a positive real $\epsilon > 0$, and two parameters s, c such that $c > 2$ and $s \ge (1+c)\eta_\epsilon(L)$. Then for any subset $T \subset L$ and any additional vector $\vec{v} \in L$, it holds that $\mathcal{D}_{L,s}(T) - \mathcal{D}_{L,s}(T-\vec{v}) \;\le\; \frac{\mathsf{erf}(q(1+4/c)/2)}{\mathsf{erf}(2q)}\cdot\frac{1+\epsilon}{1-\epsilon}$, where $q = \|v\|\sqrt{\pi}/s$.*

Proof. Clearly for any fixed $\vec{v}$, the set that maximizes $\mathcal{D}_{L,s}(T) - \mathcal{D}_{L,s}(T-\vec{v})$ is the set of all vectors $\vec{u} \in L$ for which $\mathcal{D}_{L,s}(\vec{u}) > \mathcal{D}_{L,s}(\vec{u}-\vec{v})$, which we denote by $T_{\vec{v}} \stackrel{\text{def}}{=} \{\vec{u} \in L : \mathcal{D}_{L,s}(\vec{u}) > \mathcal{D}_{L,s}(\vec{u}-\vec{v})\}$. Observe that for any $\vec{u} \in L$ we have $\mathcal{D}_{L,s}(\vec{u}) > \mathcal{D}_{L,s}(\vec{u}-\vec{v})$ iff $\rho_s(\vec{u}) > \rho_s(\vec{u}-\vec{v})$, which is equivalent to $\|\vec{u}\| < \|\vec{u}-\vec{v}\|$. That is, $\vec{u}$ must lie in the half-space whose projection on $\vec{v}$ is less than half of $\vec{v}$, namely $\langle\vec{u},\vec{v}\rangle < \|\vec{v}\|^2/2$. In other words we have

$$T_{\vec{v}} = \{\vec{u} \in L : \langle\vec{u},\vec{v}\rangle < \|\vec{v}\|^2/2\},$$

which also means that $T_{\vec{v}} - \vec{v} = \{\vec{u} \in L : \langle\vec{u},\vec{v}\rangle < -\|\vec{v}\|^2/2\} \subseteq T_{\vec{v}}$. We can therefore express the difference in probability mass as $\mathcal{D}_{L,s}(T_{\vec{v}}) - \mathcal{D}_{L,s}(T_{\vec{v}}-\vec{v}) = \mathcal{D}_{L,s}(T_{\vec{v}} \setminus (T_{\vec{v}}-\vec{v}))$. Below we denote this set-difference by

$$H_{\vec{v}} \;\stackrel{\text{def}}{=}\; T_{\vec{v}} \setminus (T_{\vec{v}} - \vec{v}) \;=\; \left\{\vec{u} \in L : \langle\vec{u},\vec{v}\rangle \in (-\tfrac{\|\vec{v}\|^2}{2}, \tfrac{\|\vec{v}\|^2}{2}]\right\}.$$

That is, $H_{\vec{v}}$ is the "slice" in space of width $\|\vec{v}\|$ in the direction of $\vec{v}$, which is symmetric around the origin. The arguments above imply that for any set T we have $\mathcal{D}_{L,s}(T) - \mathcal{D}_{L,s}(T-\vec{v}) \le \mathcal{D}_{L,s}(H_{\vec{v}})$. The rest of the proof is devoted to upper-bounding the probability mass of that slice, i.e., $\mathcal{D}_{L,s}(H_{\vec{v}}) = \Pr_{\vec{u}\leftarrow\mathcal{D}_{L,s}}[\vec{u} \in H_{\vec{v}}]$.

To this end we consider the slightly thicker slice, say $H'_{\vec{v}} = (1+\frac{4}{c})H_{\vec{v}}$, and the random variable $\vec{w}$, which is obtained by drawing $\vec{u} \leftarrow \mathcal{D}_{L,s}$ and adding to it a continuous Gaussian variable of "width" s/c. We argue that $\vec{w}$ is somewhat likely to fall outside of the thick slice $H'_{\vec{v}}$, but coditioning on $\vec{u} \in H_{\vec{v}}$ we have that $\vec{w}$ is very unlikely to fall outside of $H'_{\vec{v}}$. Putting these two arguments together, we get that $\vec{u}$ must have significant probability of falling outside $H_{\vec{v}}$, thereby getting our upper bound.

In more detail, denoting $r = s/c$ we consider drawing $\vec{u} \leftarrow \mathcal{D}_{L,s}$ and $\vec{z} \leftarrow \mathcal{N}^n(0, r^2/2\pi)$, and setting $\vec{w} = \vec{u} + \vec{z}$. Denoting $t = \sqrt{r^2+s^2}$, we have that $s \le t \le s(1+\frac{1}{c})$ and $rs/t \ge s/(c+1) \ge \eta_\epsilon(L)$. Thus the conditions of Lemma 5 are met, and we get that $\vec{w}$ is distributed close to a normal random variable $\mathcal{N}^n(0, t^2/2\pi)$, upto a factor of at most $\frac{1+\epsilon}{1-\epsilon}$.

Since the continuous Gaussian distribution is spherical, we can consider expressing it in an orthonormal basis with one vector in the direction of $\vec{v}$. When expressed in this basis, we get the event $\vec{z} \in H'_{\vec{v}}$

exactly when the coefficient in the direction of $\vec{v}$ (which is distributed close to the 1-diemsnional Gaussian $\mathcal{N}(0, t^2/2\pi)$) exceeds $\|\vec{v}(1+\frac{4}{c})/2\|$ in magnitude. Hence we have

$$\begin{aligned}\Pr[\vec{w} \in H'_{\vec{v}}] &\leq \Pr_{\alpha \leftarrow \mathcal{N}(0,t^2/2\pi)}[|\alpha| \leq \|\vec{v}\|] \cdot \frac{1+\epsilon}{1-\epsilon} \\ &= \mathsf{erf}\left(\frac{\|\vec{v}\|\sqrt{\pi}(1+\frac{4}{c})}{2t}\right) \cdot \frac{1+\epsilon}{1-\epsilon} \leq \mathsf{erf}\left(\frac{\|\vec{v}\|\sqrt{\pi}(1+\frac{4}{c})}{2s}\right) \cdot \frac{1+\epsilon}{1-\epsilon}\end{aligned}$$

On the other hand, consider the conditional probability $\Pr[\vec{w} \in H'_{\vec{v}} | \vec{u} \in H_{\vec{v}}]$: Let $H''_{\vec{v}} = \frac{4}{c} H_{\vec{v}}$, then if $\vec{u} \in H_{\vec{v}}$ and $\vec{z} \in H''_{\vec{v}}$, then it must be the case that $\vec{w} = \vec{u} + \vec{z} \in H'_{\vec{v}}$. As before, we can consider the continuous Gaussian on $\vec{z}$ in an orthonormal basis with one vector in the direction of $\vec{v}$, and we get

$$\begin{aligned}\Pr[\vec{w} \in H'_{\vec{v}} | \vec{u} \in H_{\vec{v}}] &\geq \Pr[\vec{z} \in H''_{\vec{v}} | \vec{u} \in H_{\vec{v}}] = \Pr[\vec{z} \in H''_{\vec{v}}] \\ &= \Pr_{\beta \leftarrow \mathcal{N}(0,r^2/2\pi)}[|\beta| \leq 2\|\vec{v}\|/c] = \mathsf{erf}(\|\vec{v}\|2\sqrt{\pi}/cr) = \mathsf{erf}(2\|\vec{v}\|\sqrt{\pi}/s)\end{aligned}$$

Putting the last two bounds together, we get

$$\begin{aligned}\mathsf{erf}\left(\frac{\|\vec{v}\|\sqrt{\pi}(1+\frac{4}{c})}{2s}\right) \cdot \frac{1+\epsilon}{1-\epsilon} &\geq \Pr[\vec{w} \in H'_{\vec{v}}] \geq \Pr[\vec{u} \in H_{\vec{v}}] \cdot \Pr[\vec{w} \notin H'_{\vec{v}} | \vec{u} \in H_{\vec{v}}] \\ &\geq \Pr[\vec{u} \in H_{\vec{v}}] \cdot \mathsf{erf}\left(\frac{\|\vec{v}\|2\sqrt{\pi}}{s}\right)\end{aligned}$$

from which we conclude that $\Pr[\vec{u} \in H_{\vec{v}}] \leq \frac{\mathsf{erf}\left(\|\vec{v}\|\sqrt{\pi}(1+4/c)/2s\right)}{\mathsf{erf}\left(\|\vec{v}\|2\sqrt{\pi}/s\right)} \cdot \frac{1+\epsilon}{1-\epsilon}$, as needed. □

One useful special case of Lemma 6 is when $c = 100$ (say) and $\|\vec{v}\| \approx s$, where we get a bound $\mathcal{D}_{L,s}(T) - \mathcal{D}_{L,s}(T-\vec{v}) \leq \frac{\mathsf{erf}(0.52\sqrt{\pi})}{\mathsf{erf}(2\sqrt{\pi})} \cdot \frac{1+\epsilon}{1-\epsilon} \approx 0.81$. We note that when $\frac{\|\vec{v}\|}{s} \to 0$, the bound from Lemma 6 tends to (just over) $1/4$, but we note that we can make it tend to zero with a different choice of parameters in the proof (namely making $H'_{\vec{v}}$ and $H''_{\vec{v}}$ thicker, e.g. $H''_{\vec{v}} = H_{\vec{v}}$ and $H'_{\vec{v}} = 2H_{\vec{v}}$). Lemma 6 extends easily also to the ellipsoid Gaussian case, using Fact 2:

Corollary 2. *Fix a lattice $L \subset \mathbb{R}^n$, a positive real $\epsilon > 0$, a parameter $c > 2$ and a rank-n matrix S such that $s \stackrel{\text{def}}{=} \sigma_n(S) \geq (1+c)\eta_\epsilon(L)$. Then for any subset $T \subset L$ and any additional vector $\vec{v} \in L$, it holds that $\mathcal{D}_{L,S}(T) - \mathcal{D}_{L,S}(T-\vec{v}) \leq \frac{\mathsf{erf}(q(1+4/c)/2)}{\mathsf{erf}(2q)} \cdot \frac{1+\epsilon}{1-\epsilon}$, where $q = \|v\|\sqrt{\pi}/s$.*

Micciancio and Regev give the following bound on the smoothing parameter in terms of the primal lattice.

Lemma 7. *[Lemma 3.3 of [MR07]] For any n-dimensional lattice L and positive real $\epsilon > 0$,*

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

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

3 Sums of Discrete Gaussians

Consider a full rank lattice $L \subseteq \mathbb{Z}^n$, some negligible $\epsilon = \epsilon(n)$, the corresponding smoothing parameter $\eta = \eta_\epsilon(L)$ and parameters $s > \Omega(\eta)$, $m > \Omega(n \log n)$, and $s' > \Omega(\mathrm{poly}(n)\log(1/\epsilon))$. The process that we analyze begins by choosing "once and for all" m points in L, drawn independently from a discrete Gaussian with parameter s, $\vec{x}_i \leftarrow \mathcal{D}_{L,s}$.[2]

Once the $\vec{x}_i$'s are fixed, we arrange them as the rows of an m-by-n matrix $X = (\vec{x}_1|\vec{x}_2|\ldots|\vec{x}_m)^\top$, and consider the distribution $\mathcal{E}_{X,s'}$, induced by choosing an integer vector $\vec{v}$ from a discrete spherical Gaussian with parameter s' and outputting $\vec{y} = X^\top \vec{v}$:

$$\mathcal{E}_{X,s'} \stackrel{\mathrm{def}}{=} \{X^\top \vec{v} : \vec{v} \leftarrow \mathcal{D}_{\mathbb{Z}^m,s'}\}. \tag{1}$$

Our goal is to prove that $\mathcal{E}_{X,s'}$ is close to the ellipsoid Gaussian $\mathcal{D}_{L,s'X}$, over L. We begin by proving that the singular values of X are all roughly of the size $s\sqrt{m}$.

Lemma 8. *There exists a universal constant $K > 1$ such that for all $m \geq 2n$, $\epsilon > 0$ and every n-dimensional real lattice $L \subset \mathbb{R}^n$, the following holds: Choosing the rows of an m-by-n matrix X independently at random from a spherical discrete Gaussian on L with parameter $s > 2K\eta_\epsilon(L)$, $X \leftarrow (\mathcal{D}_{L,s})^m$, we have*

$$\Pr\left[s\sqrt{2\pi m}/K < \sigma_n(X) \leq \sigma_1(X) < sK\sqrt{2\pi m}\right] > 1 - (4m\epsilon + O(\exp(-m/K))).$$

Proof. Let C be the universal constant from Corollary 1, and we set $K = \max(3C, 2C^2)$. Denote $r = s/K$, and consider the process of first choosing X as in the lemma statement, then choosing the rows of an m-by-n matrix Y independently from the continuous n-dimensional Normal distribution $\mathcal{N}(0, r^2/2\pi)$, then setting $Z = X + Y$. Note that for these parameters r, s we have

$$\frac{rs}{\sqrt{r^2+s^2}} = \frac{s(s/K)}{\sqrt{s^2+(s/K)^2}} = \frac{s}{\sqrt{1+K^2}} > s/2K > \eta_\epsilon(L).$$

Thus the conditions of Lemma 5 are met, hence setting $t = \sqrt{s^2+r^2}$ we conclude that the statistical distance between the rows of Z and a continues n-dimensional Gaussian $\mathcal{N}^n(0, t^2/2\pi)$ is at most 4ϵ. Namely we can bound by $4m\epsilon$ the statistical distance between Z and a matrix whose entries are all chosen independently from $\mathcal{N}(0, t^2/2\pi)$. Therefore, by Corollary 1 we have that

$$\Pr\left[t\sqrt{2\pi m}/C < \sigma_n(Z) \leq \sigma_1(Z) < tC\sqrt{2\pi m}\right] > 1 - (4m\epsilon + O(\exp(-m/C))),$$

and since $s < t < 2s$ then with at least the same probability we have $s\sqrt{2\pi m}/C < \sigma_n(Z) \leq \sigma_1(Z) < 2sC\sqrt{2\pi m}$. At the same time, again by Corollary 1 we have that $\Pr[\sigma_n(Y) > Cr\sqrt{2\pi m}] < O(\exp(-m/C))$, and our parameters choice imply that

$$Cr\sqrt{2\pi m} = s/K \cdot C\sqrt{2\pi m} \leq \frac{Cs\sqrt{2\pi m}}{2C^2} = s\sqrt{2\pi m}/2C.$$

We conclude that except with probability $4m\epsilon + O(\exp(-m/C))$, we have both $\sigma_n(Z) \geq s\sqrt{2\pi m}/C$ and $\sigma_1(-Y) = \sigma_1(Y) \leq s\sqrt{2\pi m}/2C$. In this case, since $X = Z - Y$, we can apply Fact 1 (with $\delta = 1/2$) to

[2] More generally, we can consider drawing the vectors $\vec{x}_i$ from an ellipsoid discrete Gaussian, $\vec{x}_i \leftarrow \mathcal{D}_{L,S}$, so long as the least singular value of S is at least s.

conclude that $\sigma_n(X) \geq (1-\frac{1}{2})s\sqrt{2\pi m}/C > s\sqrt{2\pi m}/K$ and $\sigma_n(X) \leq (1+\frac{1}{2})2sC\sqrt{2\pi m} \leq sK\sqrt{2\pi m}$. In summary, we have

$$\begin{aligned}
&\Pr\left[s\sqrt{2\pi m}/K < \sigma_n(X) \leq \sigma_1(X) < sK\sqrt{2\pi m}\right] \\
&\geq \Pr\left[2\sigma_1(Y) < s\sqrt{2\pi m}/C < \sigma_n(Z) \leq \sigma_1(Z) < sC\sqrt{2\pi m}\right] \\
&\geq 1-(4m\epsilon + O(\exp(-m/C))) \geq 1-(4m\epsilon + O(\exp(-m/K))),
\end{aligned}$$

as needed. □

3.1 The Distribution $\mathcal{E}_{X,s'}$ Over $\mathbb{Z}^n$

We next move to show that with high probability over the choice of X, the distribution $\mathcal{E}_{X,s'}$ is statistically close to the ellipsoid discrete Gaussian $\mathcal{D}_{L,s'X}$. We first prove this for the special case of the integer lattice, $L = \mathbb{Z}^n$, and then use that special case to prove the same statement for general lattices. In either case, we analyze the setting where the rows of X are chosen from an ellipsoid Gaussian which is "not too small" and "not too skewed."

Parameters. Below n is the security parameters and $\epsilon = \text{negligible}(n)$. Let S be an n-by-n matrix such that $\sigma_n(S) \geq 2K\eta_\epsilon(\mathbb{Z}^n)$, and denote $s_1 = \sigma_1(S)$, $s_n = \sigma_n(S)$, and $w = s_1/s_n$. (We consider w to be a measure for the "skewness" of S.) Also let m, q, s' be parameters satisfying $m \geq 10n\log q$, $q > 8(mn)^{1.5}s_1 w$, and $s' \geq 4mnw\ln(1/\epsilon)$. An example setting of parameters to keep in mind is $m = n^2$, $s_n = \sqrt{n}$ (which implies $\epsilon \approx 2^{-\sqrt{n}}$), $s_1 = n$ (so $w = \sqrt{n}$), $q = 8n^6$, and $s' = 4n^4$.

Theorem 2. *For ϵ negligible in n, let $S \in \mathbb{R}^{n\times n}$ be a matrix such that $s_n = \sigma_n(S) \geq 18K\eta_\epsilon(\mathbb{Z}^n)$, and denote $s_1 = \sigma_1(S)$ and $w = s_1/s_n$. Also let m, s' be parameters such that $m \geq 10n\log(8(mn)^{1.5}s_1 w)$ and $s' \geq 4mnw\ln(1/\epsilon)$.*

Then, when choosing the rows of an m-by-n matrix X from the ellipsoid Gaussian over $\mathbb{Z}^n$, $X \leftarrow (\mathcal{D}_{\mathbb{Z}^n,S})^m$, we have with all but probability $2^{-O(m)}$ over the choice of X, that the statistical distance between $\mathcal{E}_{X,s'}$ and the ellipsoid Gaussian $\mathcal{D}_{\mathbb{Z}^n,s'X}$ is bounded by 2ϵ.

The rest of this subsection is devoted to proving Theorem 2. We begin by showing that with overwhelming probability, the rows of X span all of $\mathbb{Z}^n$, which means also that the support of $\mathcal{E}_{X,s'}$ includes all of $\mathbb{Z}^n$.

Lemma 9. *With parameters as above, when drawing the rows of an m-by-n matrix X independently at random from $\mathcal{D}_{\mathbb{Z}^n,S}$ we get $X^\top\mathbb{Z}^m = \mathbb{Z}^n$ with all but probability $2^{-O(m)}$.*

Proof. Consider choosing the rows one by one, and we show that (a) as long as the current rows only $\mathbb{R}$-span a subspace of $\mathbb{R}^n$ then it is likely that the next row falls outside that subspace, and (b) once the current matrix has full rank, as long as the current rows only $\mathbb{Z}$-span a sub-lattice of $\mathbb{Z}^n$, it is likely that the next one falls outside that sub-lattice. Combining these two arguments, the lemma follows.

For $i = 1, 2, \ldots, m$, consider the binary random variable χ_i, which is defined as follows over the choice of the rows $\vec{x}_i$ of X: At any step i we consider only the "short vectors" among the previous $\vec{x}_i$'s, namely $X_{i-1} \stackrel{\text{def}}{=} \{\vec{x}_j : j < i, \|\vec{x}_j\| \leq s\sqrt{n}\}$.

1. If the vectors in X_{i-1} only $\mathbb{R}$-span a proper linear subspace of $\mathbb{R}^n$, then we define $\chi_i = 1$ if $\|\vec{x}_i\| \leq s\sqrt{n}$ and $\vec{x}_i$ falls outside that linear subspace, and $\chi_i = 0$ otherwise;

2. If the vectors in X_{i-1} only $\mathbb{Z}$-span a sub-lattice of $\mathbb{Z}^n$ but $\mathbb{R}$-span the entire $\mathbb{R}^n$, then we define $\chi_i = 1$ if $\|\vec{x}_i\| \le s\sqrt{n}$ and $\vec{x}_i$ falls outside that sub-lattice, and $\chi_i = 0$ otherwise;

3. Else (if $\vec{x}_1, \ldots, \vec{x}_{i-1}$ $\mathbb{Z}$-span the entire $\mathbb{Z}^n$), we defined $\chi_i = 1$.

It is clear from the definition of the χ_i's that $\sum_{i=1}^m \chi_i \ge n$ implies that the $\vec{x}_i$'s $\mathbb{R}$-span all of of $\mathbb{R}^n$. Moreover we claim that if $\sum_{i=1}^m \chi_i \ge n(\log(s\sqrt{n}) + 1)$ then the $\vec{x}_i$'s must $\mathbb{Z}$-span the entire lattice $\mathbb{Z}^n$. To see this, consider the first n vectors $\vec{x}_i$ for which $\chi_i = 1$: they must be linearly independent and they are all shorter than $s\sqrt{n}$, hence they $\mathbb{Z}$-span a full-rank sub-lattice of $\mathbb{Z}^n$ of determinant less than $(s\sqrt{n})^n$. As long as the $\vec{x}_i$ do not yet $\mathbb{Z}$-span the entire integer lattice, any subsequent $\vec{x}_i$ for which $\chi_i = 1$ corresponds to a refinement of the current sub-lattice, which must reduce the determinant by at least a factor of 2. Hence after at most $\log((s\sqrt{n})^n) = n\log(s\sqrt{n})$ such vectors the determinant is reduced to 1, which means that the $\vec{x}_i$'s must $\mathbb{Z}$-span the entire integer lattice. We therefore have

$$\Pr[X^\top \mathbb{Z}^m = \mathbb{Z}^n] \;\ge\; \Pr\left[\sum_i \chi_i \ge n(\log(s\sqrt{n}) + 1)\right].$$

It is left to lower-bound the last expression. We claim that regardless of the previous $\vec{x}_{i'}$'s for $i' < i$, we always have $\Pr[\chi_i = 1] \ge 1/4$. This is obvious if χ_i is assigned according to the third rule above, so we only need to prove it for the first two rules. To see why this is true for the first rule, note that as long as the vectors in X_{i-1} only $\mathbb{R}$-span a proper sub-space of $\mathbb{R}^n$, there must exists at least one standard unit vector $\vec{e}_j$ outside that sub-space. Letting $T_{i-1} \subset \mathbb{Z}^n$ be the sub-lattice of $\mathbb{Z}^n$ that lies in the sub-space of X_{i-1}, we have that $T_{i-1} - \vec{e}_j$ is disjoint from T_{i-1}. Since $\|\vec{e}_j\| = 1$ and $s > \eta_\epsilon(\mathbb{Z}^n) \ge \sqrt{n}$, then Corollary 2 (with $c = 9$) says that

$$\Pr[\vec{x}_i \in T_{i-1}] - \Pr[\vec{x}_i \in T_{i-1} - \vec{e}_j] \le \underbrace{\frac{\mathsf{erf}(0.75\sqrt{\pi/n})}{\mathsf{erf}(2\sqrt{\pi/n})}}_{\approx 0.75/2 = 0.375} \cdot \frac{1+\epsilon}{1-\epsilon} < 0.4,$$

which means that $\Pr[\vec{x}_i \in T_{i-1}] < \frac{1+0.4}{2} = 0.7$. Hence

$$\begin{aligned}\Pr[\chi_i = 1] \;&\ge\; \Pr[\vec{x}_i \notin T_{i-1} \text{ and } \|\vec{x}_i\| \le \sqrt{n}] \;\ge\; \Pr[\vec{x}_i \notin T_{i-1}] - \frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n} \\ &\ge\; 0.3 - \frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n} \;>\; 0.25\end{aligned}$$

The argument for the second rule is nearly identical, using the fact that for any proper sub-lattice of $\mathbb{Z}^n$ there must be at least one standard unit vector $\vec{e}_j$ outside that sub-lattice.

It follows that $\Pr[\sum_i \chi_i < n(\log(s\sqrt{n}) + 1)]$ is upper-bounded by the same probability expression applied to m Bernoulli-$\frac{1}{4}$ variables, which is $2^{-O(m/4 - n(\log(s\sqrt{n})+1))} = 2^{-O(m)}$. □

From now on we assume that the rows of X indeed span all of $\mathbb{Z}^n$. Now let $A = A(X)$ be the $(m-n)$-dimensional lattice in $\mathbb{Z}^m$ orthogonal to all the columns of X, and for any $\vec{z} \in \mathbb{Z}^n$ we denote by $A_{\vec{z}} = A_{\vec{z}}(X)$ the $\vec{z}$ coset of A:

$$A = A(X) \stackrel{\text{def}}{=} \{\vec{v} \in \mathbb{Z}^m : X^\top \vec{v} = 0\} \text{ and } A_{\vec{z}} = A_{\vec{z}}(X) \stackrel{\text{def}}{=} \{\vec{v} \in \mathbb{Z}^m : X^\top \vec{v} = \vec{z}\}.$$

Since the rows of X span all of $\mathbb{Z}^n$ then $A_{\vec{z}}$ is nonempty for every $\vec{z} \in \mathbb{Z}^n$, and we have $A_{\vec{z}} = \vec{v}_{\vec{z}} + A$ for any arbitrary point $\vec{v}_{\vec{z}} \in A_{\vec{z}}$.

Below we prove that the smoothing parameter of A is small (whp), and use that to bound the distance between $\mathcal{E}_{X,s'}$ and $\mathcal{D}_{\mathbb{Z}^n,s'X}$. First we show that if the smoothing parameter of A is indeed small (i.e., smaller than the parameter s' used to sample the coefficient vector $\vec{v}$), then $\mathcal{E}_{X,s'}$ and $\mathcal{D}_{\mathbb{Z}^n,s'X}$ must be close.

Lemma 10. *Fix X and $A = A(X)$ as above. If $s' \geq \eta_\epsilon(A)$, then for any point $\vec{z} \in \mathbb{Z}^n$, the probability mass assigned to $\vec{z}$ by $\mathcal{E}_{X,s'}$ differs from that assigned by $\mathcal{D}_{\mathbb{Z}^n,s'X}$ by at most a factor of $(1-\epsilon)/(1+\epsilon)$, namely*

$$\mathcal{E}_{X,s'}(\vec{z}) \in \left[\tfrac{1-\epsilon}{1+\epsilon}, 1\right] \cdot \mathcal{D}_{\mathbb{Z}^n,s'X}(\vec{z}).$$

In particular, if $\epsilon < 1/3$ then the statistical distance between $\mathcal{E}_{X,s'}$ and $\mathcal{D}_{\mathbb{Z}^n,s'X}$ is at most 2ϵ.

Proof. Fix some $\vec{z} \in \mathbb{Z}^n$. The probability mass assigned to $\vec{z}$ by $\mathcal{E}_{X,s'}$ is the probability of drawing a random vector according to the discrete Gaussian $\mathcal{D}_{\mathbb{Z}^m,s'}$ and hitting some $\vec{v} \in \mathbb{Z}^m$ for which $X^\top \vec{v} = \vec{z}$. In other words, this is exactly the probability mass assigned by $\mathcal{D}_{\mathbb{Z}^m,s'}$ to the coset $A_{\vec{z}}$. Below let $T = T(X) \subseteq \mathbb{R}^m$ be the linear subspace containing the lattice A, and $T_{\vec{z}} = T_{\vec{z}}(X) \subseteq \mathbb{R}^m$ be the affine subspace containing the coset $A_{\vec{z}}$:

$$T = T(X) = \{\vec{v} \in \mathbb{R}^m : X^\top \vec{v} = 0\}, \text{ and } T_{\vec{z}} = T_{\vec{z}}(X) = \{\vec{v} \in \mathbb{R}^m : X^\top \vec{v} = \vec{z}\}.$$

Let Y be the pseudoinverse of X (i.e. $Y^\top X = I_n$ and the columns of Y span the same linear sub-space as the columns of X). Let $\vec{u}_{\vec{z}} = Y\vec{z}$, and we note that $\vec{u}_{\vec{z}}$ is the point in the affine space $T_{\vec{z}}$ closest to the origin: To see this, note that $\vec{u}_{\vec{z}} \in T_{\vec{z}}$ since $X^\top \vec{u}_{\vec{z}} = X^\top \times Y\vec{z} = \vec{z}$. In addition, $\vec{u}_{\vec{z}}$ belongs to the column space of Y, so also to the column space of X, and hence it is orthogonal to T.

Since $\vec{u}_{\vec{z}}$ is the point in the affine space $T_{\vec{z}}$ closest to the origin, it follows that for every point in the coset $\vec{v} \in A_{\vec{z}}$ we have $\|\vec{v}\|^2 = \|\vec{u}_{\vec{z}}\|^2 + \|\vec{v} - \vec{u}_{\vec{z}}\|^2$, and therefore

$$\rho_{s'}(\vec{v}) \;=\; e^{-\pi(\|\vec{v}\|/s')^2} \;=\; e^{-\pi(\|\vec{u}_{\vec{z}}\|/s')^2} \cdot e^{-\pi(\|\vec{v}-\vec{u}_{\vec{z}}\|/s')^2} \;=\; \rho_{s'}(\vec{u}_{\vec{z}}) \cdot \rho_{s'}(\vec{v} - \vec{u}_{\vec{z}}).$$

This, in turn, implies that the total mass assigned to $A_{\vec{z}}$ by $\rho_{s'}$ is

$$\rho_{s'}\big(A_{\vec{z}}\big) \;=\; \sum_{\vec{v} \in A_{\vec{z}}} \rho_{s'}(\vec{v}) \;=\; \rho_{s'}(\vec{u}_{\vec{z}}) \cdot \sum_{\vec{v} \in A_{\vec{z}}} \rho_{s'}(\vec{v} - \vec{u}_{\vec{z}}) \;=\; \rho_{s'}(\vec{u}_{\vec{z}}) \cdot \rho_{s'}\big(A_{\vec{z}} - \vec{u}_{\vec{z}}\big). \tag{2}$$

Fix one arbitrary point $\vec{w}_{\vec{z}} \in A_{\vec{z}}$, and let $\vec{\delta}_{\vec{z}}$ be the distance from $\vec{u}_{\vec{z}}$ to that point, $\vec{\delta}_{\vec{z}} = \vec{u}_{\vec{z}} - \vec{w}_{\vec{z}}$. Since $A_{\vec{z}} = \vec{w}_{\vec{z}} + A$, we get $A_{\vec{z}} - \vec{u}_{\vec{z}} = A - \vec{\delta}_{\vec{z}}$, and together with the equation above we have:

$$\begin{aligned} \rho_{s'}\big(A_{\vec{z}}\big) \;&=\; \rho_{s'}(\vec{u}_{\vec{z}}) \cdot \rho_{s'}\big(A_{\vec{z}} - \vec{u}_{\vec{z}}\big) \;=\; \rho_{s'}(\vec{u}_{\vec{z}}) \cdot \rho_{s'}\big(A - \vec{\delta}_{\vec{z}}\big) \\ &=\; \rho_{s'}(\vec{u}_{\vec{z}}) \cdot \rho_{s',\vec{\delta}_{\vec{z}}}(A) \;\overset{\text{Lemma 4}}{=}\; \rho_{s'}(\vec{u}_{\vec{z}}) \cdot \rho_{s'}\big(A\big) \cdot \big[\tfrac{1-\epsilon}{1+\epsilon}, 1\big]. \end{aligned} \tag{3}$$

As a last step, recall that $\vec{u}_{\vec{z}} = Y\vec{z}$ where Y is the pseudoinverse of X (which implies that $Y^\top Y = (X^\top X)^{-1}$). Thus we have

$$\rho_{s'}(\vec{u}_{\vec{z}}) \;=\; \rho_{s'}(Y\vec{z}) \;=\; \exp(-\pi|\vec{z}^\top Y^\top Y \vec{z}|/s'^2) = \exp\left(-\pi\big|\vec{z}^\top \big((s'X)^\top (s'X)\big)^{-1}\vec{z}\big|\right) \;=\; \rho_{(s'X)}(\vec{z})$$

Putting everything together we get

$$\mathcal{E}_{X,s'}(\vec{z}) \;=\; \mathcal{D}_{\mathbb{Z}^m,s'}\big(A_{\vec{z}}\big) \;=\; \frac{\rho_{s'}\big(A_{\vec{z}}\big)}{\rho_{s'}(\mathbb{Z}^m)} \;\in\; \rho_{(s'X)}(\vec{z}) \cdot \frac{\rho_{s'}\big(A\big)}{\rho_{s'}(\mathbb{Z}^m)} \cdot \big[\tfrac{1-\epsilon}{1+\epsilon}, 1\big]$$

The term $\frac{\rho_{s'}(A)}{\rho_{s'}(\mathbb{Z}^m)}$ is a normalization factor independent of $\vec{z}$, hence the probability mass $\mathcal{E}_{X,s'}(\vec{z})$ is proportional to $\rho_{(s'X)}(\vec{z})$, upto some "deviation factor" in $[\frac{1-\epsilon}{1+\epsilon}, 1]$. □

3.1.1 The smoothing parameter of A

We now turn our attention to proving that A is "smooth enough". Specifically, for the parameters above we prove that with high probability over the choice of X, the smoothing parameter $\eta_\epsilon(A)$ is bounded below $s' = 4mnw\ln(1/\epsilon)$.

Recall again that $A = A(X)$ is the rank-$(m-n)$ lattice containing all the integer vectors in $\mathbb{Z}^m$ orthogonal to the columns of X. We extend A to a full-rank lattice as follows: First we extend the columns space of X, by throwing in also the scaled standard unit vectors $q\vec{e}_i$ for the integer parameter q mentioned above ($q \geq 8(mn)^{1.5}s_1 w$). That is, we let $M_q = M_q(X)$ be the full-rank m-dimensional lattice spanned by the columns of X and the vectors $q\vec{e}_i$,

$$M_q \;=\; \{X\vec{z} + q\vec{y} : \vec{z} \in \mathbb{Z}^n, \vec{y} \in \mathbb{Z}^m\} \;=\; \{\vec{u} \in \mathbb{Z}^m : \exists \vec{z} \in \mathbb{Z}_q^n \text{ s.t. } \vec{u} \equiv X^\top \vec{z} \pmod q\}$$

(where we identify $\mathbb{Z}_q$ above with the set $[-q/2, q/2) \cap \mathbb{Z}$). Next, let A_q be the dual of M_q, scaled up by a factor of q, i.e.,

$$\begin{aligned} A_q \;=\; qM_q^* &= \{\vec{v} \in \mathbb{R}^m : \forall \vec{u} \in M_q, \langle \vec{v}, \vec{u}\rangle \in q\mathbb{Z}\} \\ &= \{\vec{v} \in \mathbb{R}^m : \forall \vec{z} \in \mathbb{Z}_q^n, \vec{y} \in \mathbb{Z}^m,\;\; \vec{z}^\top X^\top \vec{v} + q\,\langle \vec{v}, \vec{y}\rangle \in q\mathbb{Z}\} \end{aligned}$$

It is easy to see that $A \subset A_q$, since any $\vec{v} \in A$ is an integer vector (so $q\,\langle \vec{v}, \vec{y}\rangle \in q\mathbb{Z}$ for all $\vec{y} \in \mathbb{Z}^m$) and orthogonal to the columns of X (so $\vec{z}^\top X^\top \vec{v} = 0$ for all $\vec{z} \in \mathbb{Z}_q^n$).

Obviously all the columns of X belong to M_q, and whp they are linearly independent and relatively short (i.e., of size roughly $s_1\sqrt{m}$). In Lemma 11 below we show, however, that whp over the choice of X's, these are essentially the *only* short vectors in M_q.

Lemma 11. *Recall that we choose X as $X \leftarrow (\mathcal{D}_{\mathbb{Z}^n,S})^m$, and let $w = \sigma_1(S)/\sigma_n(S)$ be a measure of the "skewness" of S. The $n+1$'st minima of the lattice $M_q = M_q(X)$ is at least $q/4nw$, except with negligible probability over the choice of X. Namely, $\Pr_{X \leftarrow (\mathcal{D}_{\mathbb{Z}^n,S})^m}[\lambda_{n+1}(M_q) < q/4nw] < 2^{-O(m)}$.*

Proof. We prove that with high probability over the choice of X, every vector in M_q which is *not* in the linear span of the columns of X is of size at least $q/4nw$.

Recall that every vector in M_q is of the form $X\vec{z} + q\vec{y}$ for some $\vec{z} \in \mathbb{Z}_q^n$ and $\vec{y} \in \mathbb{Z}^m$. Let us denote by $[\vec{v}]_q$ the modular reduction of all the entries in $\vec{v}$ into the interval $[-q/2, q/2)$, then clearly for every $\vec{z} \in \mathbb{Z}_q^n$

$$\|[X\vec{z}]_q\| \;=\; \inf\{\|X\vec{z} + q\vec{y}\| : \vec{y} \in \mathbb{Z}^m\}.$$

Moreover, for every $\vec{z} \in \mathbb{Z}_q^n, \vec{y} \in \mathbb{Z}^m$, if $X\vec{z} + q\vec{y} \neq [X\vec{z}]_q$ then $\|X\vec{z} + q\vec{y}\| \geq q/2$. Thus it suffices to show that every vector of the form $[X\vec{z}]_q$ which is not in the linear span of the columns of X has size at least $q/4nw$ (whp over the choice of X).

Fix a particular vector $\vec{z} \in \mathbb{Z}_q^n$ (i.e. an integer vector with entries in $[-q/2, q/2)$). For this fixed vector $\vec{z}$, let $i_{\max}$ be the index of the largest entry in $\vec{z}$ (in absolute value), and let $z_{\max}$ be the value of that entry. Considering the vector $\vec{v} = [X\vec{z}]_q$ for a random matrix X whose rows are drawn independently from the distribution $\mathcal{D}_{\mathbb{Z}^n,S}$, each entry of $\vec{v}$ is the inner product of the fixed vector $\vec{z}$ with a random vector $\vec{x}_i \leftarrow \mathcal{D}_{\mathbb{Z}^n,S}$, reduced modulo q into the interval $[-q/2, +q/2)$.

We now have two cases, either $\vec{z}$ is "small", i.e., $|z_{\max}| < q/2ns_1$ or it is "large", $|z_{\max}| \geq q/2ns_1$. Recall that by Lemma 3 for each $\vec{x}_i$ we have $\|\vec{x}_i\| \leq s_1\sqrt{n}$ except with probability 2^{-m}. If $\vec{z}$ is "small" then we get

$$|\,\langle \vec{z}, \vec{x}_i\rangle\,| \;\leq\; \|\vec{z}\| \cdot \|\vec{x}_i\| \;\leq\; |z_{\max}|\sqrt{n} \cdot s_1\sqrt{n} \;<\; q/2.$$

Hence except with probability $m2^{-m}$ all the entries of $X\vec{z}$ are smaller than $q/2$ in magnitude, which means that $[X\vec{z}]_q = X\vec{z}$, and so $[X\vec{z}]_q$ belongs to the column space of X. Using the union bound again, we get that with all but probability $q^n \cdot m2^{-m} < m2^{-9m/10}$, the vectors $[X\vec{z}]_q$ for all the "small" $\vec{z}$'s belong to the column space of X.

We next turn to analyzing "large" $\vec{z}$'s. Fix one "large" vector $\vec{z}$, and for that vector define the set of "bad" vectors $\vec{x} \in \mathbb{Z}^n$, i.e. the ones for which $|[\langle \vec{z}, \vec{x} \rangle]_q| < q/4nw$ (and the other vectors $\vec{x} \in \mathbb{Z}^n$ are "good"). Observe that if $\vec{x}$ is "bad", then we can get a "good" vector by adding to it the $i_{\max}$'th standard unit vector, scaled up by a factor of $\mu = \min\left(\lceil s_n \rceil\,,\ \lfloor q/|2z_{\max}| \rfloor\right)$, since

$$|[\langle \vec{z}, \vec{x} + \mu\vec{e}_{i_{\max}} \rangle]_q| \ = \ |[\langle \vec{z}, \vec{x} \rangle + \mu z_{\max}]_q| \ \geq \ \mu|z_{\max}| - |[\langle \vec{z}, \vec{x} \rangle]_q| \ \geq \ q/4nw.$$

(The last two inequalities follow since $q/2nw < \mu|z_{\max}| \leq q/2$ and $|[\langle \vec{z}, \vec{x} \rangle]_q| < q/4nw$.) Hence the injunction $\vec{x} \mapsto \vec{x} + \mu\vec{e}_{i_{\max}}$ maps "bad" $\vec{x}$'es to "good" $\vec{x}$'es. Moreover, since the $\vec{x}$'es are chosen according to the wide ellipsoid Gaussian $\mathcal{D}_{\mathbb{Z}^n,S}$ with $\sigma_n(S) = s_n \geq \eta_\epsilon(\mathbb{Z}^n)$, and since the scaled standard unit vectors are short, $\mu < s_n + 1$, then by Lemma 6 the total probability mass of the "bad" vectors $\vec{x}$ differs from the total mass of the "good" vectors $\vec{x} + \mu\vec{e}_{i_{\max}}$ by at most 0.81. It follows that when choosing $\vec{x} \leftarrow \mathcal{D}_{\mathbb{Z}^n,S}$, we have $\Pr_{\vec{x}}\left[|[\langle \vec{z}, \vec{x} \rangle]_q| < q/4nw\right] \leq (1 + 0.81)/2 < 0.91$. Thus the probability that all the entries of $[X\vec{z}]_q$ are smaller than $q/4nw$ in magnitude is bounded by $(0.91)^m = 2^{-0.14m}$. Since $m > 10n \log q$, we can use the union bound to conclude that the probability that there exists some "large" vector for which $\|[X\vec{z}]_q\| < q/4nw$ is no more than $q^n \cdot 2^{-0.14m} < 2^{-O(m)}$.

Summing up the two cases, with all but probability $2^{-O(m)})$ over the choice of X, there does not exist any vector $\vec{z} \in \mathbb{Z}_q^n$ for which $[X\vec{z}]_q$ is linearly independent of the columns of X and yet $|[X\vec{z}]_q| < q/4nw$. □

Corollary 3. *With the parameters as above, the smoothing parameter of $A = A(X)$ satisfies $\eta_\epsilon(A) \leq s' = 4mnw \ln(1/\epsilon)$, except with probability $2^{-O(m)}$.*

Proof. Recall that A_q is the scaled-by-q dual of M_q. By Lemma 11 we have that w.h.p. $\lambda_{n+1}(M_q) \geq q/4nw$, and from Banasczcyk's theorem (Theorem 1) we conclude that $\lambda_{m-n}(A_q) \leq 4mnw$. Hence we have $m - n$ linearly independent vectors $\vec{v}_j \in A_q$ of size below $4mnw$. We next argue that these vectors must also belong to A.

To see that they must be integer vectors, note that by definition of A_q, for every $\vec{v} \in A_q$ it holds in particular that $\vec{v} \times qI_m \in q\mathbb{Z}^m$, which means that $\vec{v} = \vec{v} \times I_m \in \mathbb{Z}^m$. To see that the $\vec{v}_j$'s are orthogonal to the the columns of X, recall that the columns of X are in M_q and the $\vec{v}_j$'s are in A_q, and therefore $X^\top \vec{v}_j \in q\mathbb{Z}^n$ for all j. On the other hand, by Lemma 3 with all but probability $2^{-O(m)}$ the rows of X are smaller than $s_1\sqrt{n}$, hence the columns are smaller than $s_1\sqrt{n}\sqrt{m}$. It thus follows that

$$\|X^\top \vec{v}_j\| \leq \|\vec{v}_j\| \cdot \|X\| \leq (4mnw) \cdot (s_1\sqrt{mn}) = 4(mn)^{1.5} s_1 w < q/2,$$

which together with $X^\top \vec{v} \equiv \vec{0} \pmod{q}$ means that we have $X^\top \vec{v}_j = \vec{0}$ (over $\mathbb{R}$, with no modular reduction). We conclude that the $\vec{v}_j$'s are integer vectors orthogonal to the columns of X, hence they belong to A.

It thus follows that all the successive minima of the rank-$(m - n)$ lattice A are bounded below $4mnw$, and Lemma 7 then says that

$$\eta_\epsilon(A) \ \leq \ 4mnw \cdot \sqrt{\frac{\ln(2(m-n)(1+1/\epsilon))}{\pi}} \ \overset{(\star)}{\leq} \ 4mnw \ln(1/\epsilon) \ = \ s'$$

(where the inequality $(\star)$ uses the fact that $1/\epsilon \gg m$). □

Putting together Lemma 10 and Corollary 3 completes the proof of Theorem 2. □

3.2 The Distribution $\mathcal{E}_{X,s'}$ Over General Lattices

Armed with Theorem 2, we turn to prove the same theorem also for general lattices.

Theorem 3. *Let L be a full-rank lattice $L \subset \mathbb{R}^n$ and B a matrix whose columns form a basis of L. Also let $M \in \mathbb{R}^{n \times n}$ be a full rank matrix, and denote $S = M(B^\top)^{-1}$, $s_1 = \sigma_1(S)$, $s_n = \sigma_n(S)$, and $w = s_1/s_n$. Finally let ϵ be negligible in n and m, s' be parameters such that $m \geq 10n \log(8(mn)^{1.5} s_1 w)$ and $s' \geq 4mnw \ln(1/\epsilon)$.*

If $s_n \geq \eta_\epsilon(\mathbb{Z}^n)$, then, when choosing the rows of an m-by-n matrix X from the ellipsoid Gaussian over L, $X \leftarrow (\mathcal{D}_{L,M})^m$, we have with all but probability $2^{-O(m)}$ over the choice of X, that the statistical distance between $\mathcal{E}_{X,s'}$ and the ellipsoid Gaussian $\mathcal{D}_{L,s'X}$ is bounded by 2ϵ.

Proof. This theorem is an immediate corollary of Theorem 2 and Fact 2. Noting that S, ϵ satisfy the conditions of Theorem 2, we conclude that when choosing the rows of an m-by-n integer matrix as $Z \leftarrow (\mathcal{D}_{\mathbb{Z}^n,S})^m$, the statistical distance between $\mathcal{E}_{Z,s'}$ and $\mathcal{D}_{\mathbb{Z}^n,s'Z}$ is at most 2ϵ.

Letting $X = BZ$, we get by Fact 2 that choosing the columns of Z from $\mathcal{D}_{\mathbb{Z}^n,S}$ induces the distribution $\mathcal{D}_{L,M}$ on the columns of X. Also multiplying the output of both distributions $\mathcal{E}_{Z,s'}$ and $\mathcal{D}_{\mathbb{Z}^n,s'Z}$ by B, we have that $\mathcal{E}_{X,s'} = B \times \mathcal{E}_{Z,s'}$ and $\mathcal{D}_{L,s'X} = B \times \mathcal{D}_{\mathbb{Z}^n,s'Z}$. Since the distance between $\mathcal{E}_{Z,s'}$ and $\mathcal{D}_{\mathbb{Z}^n,s'Z}$ is at most 2ϵ, then so is the distance between $\mathcal{E}_{X,s'}$ and $\mathcal{D}_{L,s'X}$. □

References

[ABB10] Shweta Agrawal, Dan Boneh, and Xavier Boyen. Efficient lattice (H)IBE in the standard model. In *EUROCRYPT*, pages 553–572, 2010.

[ABV+12] Shweta Agrawal, Xavier Boyen, Vinod Vaikuntanathan, Panagiotis Voulgaris, and Hoeteck Wee. Functional encryption for threshold functions (or fuzzy IBE) from lattices. In *PKC*, 2012.

[AFV11] Shweta Agrawal, David Mandell Freeman, and Vinod Vaikuntanathan. Functional encryption for inner product predicates from learning with errors. In *Asiacrypt*, 2011.

[Ban93] Wojciech Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(4):625–635, 1993.

[BF11] Dan Boneh and David Mandall Freeman. Homomorphic signatures for polynomial functions. In *Eurocrypt*, 2011.

[BGV12] Zvika Brakerski, Craig Gentry, and Vinod Vaikuntanathan. Fully homomorphic encryption without bootstrapping. In *ITCS*, pages 97–106, 2012.

[Boy10] Xavier Boyen. Lattice mixing and vanishing trapdoors: A framework for fully secure short signatures and more. In *Public Key Cryptography*, pages 499–517, 2010.

[BPR12] Abhishek Banerjee, Chris Peikert, and Alon Rosen. Pseudorandom functions and lattices. In *EUROCRYPT*, pages 719–737, 2012.

[BV11a] Zvika Brakerski and Vinod Vaikuntanathan. Fully homomorphic encryption from ring-LWE and security for key dependent messages. In *CRYPTO*, pages 505–524, 2011.

[BV11b] Zvika Brakerski and Vinod Vaikuntanathan. Efficient fully homomorphic encryption from (standard) LWE. In *FOCS*, pages 97–106, 2011.

[CHKP10] David Cash, Dennis Hofheinz, Eike Kiltz, and Chris Peikert. Bonsai trees, or how to delegate a lattice basis. In *EUROCRYPT*, pages 523–552, 2010.

[Gen09] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *STOC*, pages 169–178, 2009.

[Gen10] Craig Gentry. Toward basing fully homomorphic encryption on worst-case hardness. In *CRYPTO*, pages 116–137, 2010.

[GGH12] Sanjam Garg, Craig Gentry, and Shai Halevi. Candidate multilinear maps from ideal lattices and applications. Cryptology ePrint Archive, Report 2012/610, 2012. `http://eprint.iacr.org/`.

[GPV08] Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In Cynthia Dwork, editor, *STOC*, pages 197–206. ACM, 2008.

[HILL99] Johan Hastad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, March 1999.

[ILL89] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, STOC '89, pages 12–24, New York, NY, USA, 1989. ACM.

[LPRTJ05] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Advances in Mathematics*, 195(2), 2005.

[MR04] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. In *FOCS*, pages 372–381. IEEE Computer Society, 2004.

[MR07] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM J. Computing*, 37(1):267–302, 2007.

[Pei10] Chris Peikert. An efficient and parallel gaussian sampler for lattices. In *Crypto*, 2010.

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *JACM*, 56(6), 2009.

[Rot11] Ron Rothblum. Homomorphic encryption: From private-key to public-key. In *TCC*, pages 219–234, 2011.

[Tao12] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.

[vDGHV10] Marten van Dijk, Craig Gentry, Shai Halevi, and Vinod Vaikuntanathan. Fully homomorphic encryption over the integers. In *EUROCRYPT*, pages 24–43, 2010.