Discrete Gaussian Leftover Hash Lemma over Infinite Domains

Shweta Agrawal [⋆], Craig Gentry [⋆⋆], Shai Halevi [⋆ ⋆ ⋆], and Amit Sahai[†]

Abstract. The classic Leftover Hash Lemma (LHL) is one of the most useful tools in cryptography, and is often used to argue that certain distributions arising from modular subset-sums are close to uniform over some finite domain. Though extremely useful and powerful in general, the applicability of the leftover hash lemma to lattice based cryptography is limited for two reasons. First, typically the distributions we care about in lattice-based cryptography are *discrete Gaussians*, not uniform. Second, the elements chosen from these discrete Gaussian distributions lie in an infinite domain: a lattice rather than a finite field.

In this work we prove a "lattice world" analog of LHL over infinite domains, proving that certain "generalized subset sum" distributions are statistically close to well behaved discrete Gaussian distributions, even without any modular reduction. Specifically, given many vectors $\{\boldsymbol{x}_i\}_{i=1}^m$ from some lattice $L \subset \mathbb{R}^n$, we analyze the probability distribution $\sum_{i=1}^m z_i \boldsymbol{x}_i$ where the integer vector $\boldsymbol{z} \in \mathbb{Z}^m$ is chosen from a discrete Gaussian distribution. We show that when the $\boldsymbol{x}_i$'s are "random enough" and the Gaussian from which the $\boldsymbol{z}$'s are chosen is "wide enough", then the resulting distribution is statistically close to a near-spherical discrete Gaussian over the lattice L. Beyond being interesting in its own right, this "lattice-world" analog of LHL has applications for the new construction of multilinear maps [GGH12], where it is used to sample Discrete Gaussians obliviously. Specifically, given encoding of the $\boldsymbol{x}_i$'s, it is used to produce an encoding of a near-spherical Gaussian distribution over the lattice. We believe that our new lemma will have other applications, and sketch some plausible ones in this work.

1 Introduction

The Leftover Hash Lemma (LHL) is a central tool in computer science, stating that universal hash functions are good randomness extractors. In a characteristic application, the universal hash function may often be instantiated by a simple inner product function, where it is used to argue that a random linear combination of some elements (that are chosen at random and then fixed "once and for all") is statistically close to the uniform distribution over some finite domain. Though extremely useful and powerful in general, the applicability of the leftover hash lemma to lattice based cryptography is limited for two reasons. First, typically the distributions we care about in lattice-based cryptography are *discrete Gaussians*, not uniform. Second, the elements chosen from these discrete Gaussian distributions lie in an infinite domain: a lattice rather than a finite field.

The study of discrete Gaussian distributions underlies much of the advances in lattice-based cryptography over the last decade. A discrete Gaussian distribution is a distribution over some fixed lattice, in which every lattice point is sampled with probability proportional to its probability mass under a standard (n-dimensional) Gaussian distribution. Micciancio and Regev have shown in [MR07] that these distributions share many of the nice properties of their continuous counterparts, and demonstrated their usefulness for lattice-based cryptography. Since then, discrete Gaussian distributions have been used extensively in all aspects of lattice-based cryptography (most notably in the famous "Learning with Errors" problem and its variants [Reg09]). Despite their utility, we still do not understand discrete Gaussian distributions as well as we do their continuous counterparts.

A Gaussian Leftover Hash Lemma for Lattices?

The LHL has been applied often in lattice-based cryptography, but sometimes awkwardly. As an example, in the integer-based fully homomorphic encryption scheme of van Dijk et al. [vDGHV10], ciphertexts live in the lattice $\mathbb{Z}$. Roughly speaking, the public key of that scheme contains many encryptions of zero, and encryption is done by adding

[⋆] UCLA. Email: shweta@cs.ucla.edu

[⋆⋆] IBM Research. Email: craigbgentry@gmail.com

[⋆ ⋆ ⋆] IBM Research. Email: shaih@us.ibm.com

[†] UCLA. Email: sahai@cs.ucla.edu.

the plaintext value to a subset-sum of these encryptions of zero. To prove security of this encryption method, van Dijk et al. apply the left-over hash lemma in this setting, but with the cost of complicating their encryption procedure by reducing the subset-sum of ciphertexts modulo a single large ciphertext, so as to bring the scheme back in to the realm of finite rings where the leftover hash lemma is naturally applied.[1] It is natural to ask whether that scheme remains secure also without this artificial modular reduction, and more generally whether there is a more direct way to apply the LHL in settings with infinite rings.

As another example, in the recent breakthrough construction of multilinear maps [GGH12], Garg et. al. require a procedure to randomize "encodings" to break simple algebraic relations that exist between them. One natural way to achieve this randomization is by adding many random encodings of zero to the public parameters, and adding a random linear combination of these to re-randomize a given encoding (without changing the encoded value). However, in their setting, there is no way to "reduce" the encodings so that the LHL can be applied. Can they argue that the new randomized encoding yields an element from some well behaved distribution?

In this work we prove an analog of the leftover hash lemma over lattices, yielding a positive answers to the questions above. We use discrete Gaussian distributions as our notion of "well behaved" distributions. Then, for m vectors $\{\boldsymbol{x}_i\}_{i\in[m]}$ chosen "once and for all" from an n dimensional lattice $L \subset \mathbb{R}^n$, and a coefficient vector $\boldsymbol{z}$ chosen from a discrete Gaussian distribution over the integers, we give sufficient conditions under which the distribution $\sum_{i=1}^m z_i\boldsymbol{x}_i$ is "well behaved."

Oblivious Gaussian Sampler

Another application of our work is in the construction of an extremely simple *discrete Gaussian sampler* [GPV08,Pei10]. Such samplers, that sample from a spherical discrete Gaussian distribution over a lattice have been constructed by [GPV08] (using an algorithm by Klein [Kle00]) as well as Peikert [Pei10]. Here we consider a much simpler discrete Gaussian sampler (albeit a somewhat imperfect one). Specifically, consider the following sampler. In an offline phase, for $m > n$, the sampler samples a set of short vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ from L – e.g., using GPV or Peikert's algorithm. Then, in the online phase, the sampler generates $\boldsymbol{z} \in \mathbb{Z}^m$ according to a discrete Gaussian and simply outputs $\sum_{i=1}^m z_i\boldsymbol{x}_i$. 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 $\boldsymbol{z}$ and how small can we make the entries of $\boldsymbol{z}$? Ideally m would be not much larger than the dimension of the lattice and the entries of $\boldsymbol{z}$ have small variance – e.g., $\tilde{O}(\sqrt{n})$.

A very useful property of such a sampler is that it it can be used easily within an additively homomorphic scheme. Thus, it can be made *oblivious* to an explicit representation of the underlying lattice! Now, if you are given lattice points encrypted under an additively homomorphic encryption scheme, you can 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 [2].

Our Results

In this work, we obtain a discrete Gaussian version of the LHL over infinite rings. Formally, consider an n dimensional lattice L and (column) vectors $X = [\boldsymbol{x}_1|\boldsymbol{x}_2|\ldots|\boldsymbol{x}_m] \in L$. We choose $\boldsymbol{x}_i$ according to a discrete Gaussian distribution $\mathcal{D}_{L,S}$, where $\mathcal{D}_{L,S}$ is defined as follows:

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

where $\rho_{S,\boldsymbol{c}}(\boldsymbol{x}) = \exp(-\pi\|\boldsymbol{x} - \boldsymbol{c}\|^2/s^2)$ and $\rho_{S,\boldsymbol{c}}(A)$ for set A denotes $\sum_{\boldsymbol{x}\in A} \rho_{S,\boldsymbol{c}}(\boldsymbol{x})$. Let $\boldsymbol{z} \leftarrow \mathcal{D}_{\mathbb{Z}^m,s'}$. We analyze the conditions under which the vector $X \cdot \boldsymbol{z}$ is statistically close to a "near-spherical" discrete Gaussian. Formally, consider:

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

[1] Once in the realms of finite rings, one can alternatively use the generic proof of Rothblum [Rot11], which also uses the LHL.

[2] As noted by [Pei10], one can indeed generate an ellipsoidal Gaussian distribution over the lattice given a basis B by just outputting $\boldsymbol{y} \leftarrow B \cdot \boldsymbol{z}$ where $\boldsymbol{z}$ is a discrete Gaussian, but this ellipsoidal Gaussian distribution would typically be very skewed.

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 $\mathcal{D}_{L,s'X^\top}$, over L.
2. The singular values of the matrix X are of size roughly $s\sqrt{m}$, hence the shape of $\mathcal{D}_{L,s'X^\top}$ is "roughly spherical". Moreover, the "width" of $\mathcal{D}_{L,s'X^\top}$ is roughly $s's\sqrt{m} = \mathrm{poly}(n)$.

We emphasize that it is straightforward to show that the covariance matrix of $\mathcal{E}_{X,s'}$ is exactly ${s'}^2XX^\top$. 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 $XX^\top$ will typically be very "skewed" (i.e., the least singular value of $X^\top$ is typically much smaller than the largest singular value).

Our Techniques

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

The analysis of the smoothing parameter of the "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 $X^\top$ 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 entry of H is distributed as $\mathcal{N}(0,s^2)$ 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 vector $\boldsymbol{x}_i \leftarrow \mathcal{D}_{L,S}$ where S is "not too small" and "not too skewed", then with high probability the singular values of $X^\top$ are all of size roughly $s\sqrt{m}$.

Related Work

Properties of linear combinations of discrete Gaussians have been studied before in some cases by Peikert [Pei10] as well as more recently by Boneh and Freeman [BF11]. Peikert's "convolution lemma" (Thm 3.1 in [Pei10]) analyzes certain cases in which a linear combination of discrete Gaussians yields a discrete Gaussian, in the one dimensional case. More recently, Boneh and Freeman [BF11] observed that, under certain conditions, a linear combination of discrete Gaussians over a lattice is also a discrete Gaussian. However, the deviation of the Gaussian needed to achieve this are quite large. Related questions were considered by Lyubashevsky [Lyu12] where he computes the expectation of the inner product of discrete Gaussians.

Discrete Gaussian samplers have been studied by [GPV08] (who use an algorithm by [Kle00]) and [Pei10]. 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 $\mathcal{D}_{L,s,\boldsymbol{c}}$. 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. 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.

Peikert's sampler [Pei10] is elegant and its complexity is difficult to beat: the only online computation is to compute $\boldsymbol{c} - B_1\lfloor B_1^{-1}(\boldsymbol{c} - \boldsymbol{x}_2)\rceil$, where $\boldsymbol{c}$ is the center of the Gaussian, B_1 is the sampler's basis for its lattice L, and $\boldsymbol{x}_2$ 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, since our 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, at least when $\boldsymbol{c} = 0$.

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

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

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 $\boldsymbol{c} \in \mathbb{R}^n$, define the (spherical) Gaussian function on $\mathbb{R}^n$ centered at $\boldsymbol{c}$ with parameter s as $\rho_{s,\boldsymbol{c}}(\boldsymbol{x}) = \exp(-\pi \|\boldsymbol{x} - \boldsymbol{c}\|^2 / s^2)$ for all $\boldsymbol{x} \in \mathbb{R}^n$. 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 $\boldsymbol{c}$ and uniform deviation σ^2, denoted $\mathcal{N}^n(\boldsymbol{c}, \sigma^2)$, 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 $\boldsymbol{c} \in \mathbb{R}^n$, the ellipsoid Gaussian function on $\mathbb{R}^n$ centered at $\boldsymbol{c}$ with parameter S is defined by

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

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 $\mathbf{0}$.

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[3] $X' \in \mathbb{R}^{m \times n}$, 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 those 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 $U_{X'} = \{\|X'\boldsymbol{u}\| : \boldsymbol{u} \in \mathbb{R}^n, \|\boldsymbol{u}\| = 1\}$. 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', $X'^\top$ 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 $\sigma_n(X' + Y') \in [1 - \delta, 1 + \delta]\sigma_n(X')$.* □

[3] We use the notation X' instead of X to avoid confusion later in the text where we will instantiate $X' = X^\top$

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*

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

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 $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k \in L$ such that $L = \{\sum_{i=1}^k z_i\boldsymbol{b}_i : z_i \in \mathbb{Z}\ \forall i\}$. If we arrange the vectors $\boldsymbol{b}_i$ as the columns of a matrix $B \in \mathbb{R}^{n\times k}$ then we can write $L = \{B\boldsymbol{z} : \boldsymbol{z} \in \mathbb{Z}^k\}$. 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:*

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

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 $i \in [n]$,*

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

2.4 Gaussian Distributions over Lattices

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

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

where $\rho_{S,\boldsymbol{c}}(A)$ for set A denotes $\sum_{\boldsymbol{x}\in A}\rho_{S,\boldsymbol{c}}(\boldsymbol{x})$. In other words, the probability $\mathcal{D}_{L,S,\boldsymbol{c}}(\boldsymbol{x})$ is simply proportional to $\rho_{S,\boldsymbol{c}}(\boldsymbol{x})$, the denominator being a normalization factor. The same definitions apply to the spherical case, which is denoted by $\mathcal{D}_{L,s,\boldsymbol{c}}(\cdot)$ (with lowercase s). As before, when $\boldsymbol{c} = \boldsymbol{0}$ 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, $\boldsymbol{c} \in R^n$ a vector, and $S \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times n}$ two rank-n matrices, and denote $L' = \{B^{-1}\boldsymbol{v} : \boldsymbol{v} \in L\}$, $\boldsymbol{c'} = B^{-1}\boldsymbol{c}$, and $S' = S \times (B^\top)^{-1}$. Then the distribution $\mathcal{D}_{L,S,\boldsymbol{c}}$ is identical to the distribution induced by drawing a vector $\boldsymbol{v} \leftarrow \mathcal{D}_{L',S',\boldsymbol{c'}}$ and outputting $\boldsymbol{u} = B\boldsymbol{v}$.* □

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 $L(B)$, $\boldsymbol{v} \leftarrow \mathcal{D}_{L(B),S,\boldsymbol{c}}$, is induced by drawing an integer vector according to $\boldsymbol{z} \leftarrow \mathcal{D}_{\mathbb{Z}^n,S',\boldsymbol{c'}}$ and outputting $\boldsymbol{v} = B\boldsymbol{z}$, where $S' = S(B^{-1})^\top$ and $\boldsymbol{c'} = B^{-1}\boldsymbol{c}$.

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 $\boldsymbol{v} \leftarrow \mathcal{D}_{L,S,\boldsymbol{c}}$ is induced by drawing a vector according to the *spherical Gaussian* $\boldsymbol{u} \leftarrow \mathcal{D}_{L',s,\boldsymbol{c'}}$ and outputting $\boldsymbol{v} = \frac{1}{s}S^\top\boldsymbol{u}$, where $\boldsymbol{c'} = s(S^\top)^{-1}\boldsymbol{c}$ and $L' = \{s(S^\top)^{-1}\boldsymbol{v} : \boldsymbol{v} \in L\}$.

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 $\rho_{1/s}(L^* \setminus \{\mathbf{0}\}) \leq \epsilon$. 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 $\boldsymbol{c} \in \mathbb{R}^n$, constant $0 < \epsilon < 1$ and matrix S s.t. $\sigma_n(S) \geq \eta_\epsilon(L)$, we have that for $\boldsymbol{v} \leftarrow \mathcal{D}_{L,S,\boldsymbol{c}}$,*

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

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 $S' = sI_n$, $\boldsymbol{c'} = s(S^\top)^{-1}\boldsymbol{c}$, and $L' = \{s(S^\top)^{-1}\boldsymbol{v} : \boldsymbol{v} \in L\}$). Namely the ellipsoid Gaussian distribution $\boldsymbol{v} \leftarrow \mathcal{D}_{L,S,\boldsymbol{c}}$ is induced by drawing a vector according to the *spherical Gaussian* $\boldsymbol{u} \leftarrow \mathcal{D}_{L',s,\boldsymbol{c'}}$ and outputting $\boldsymbol{v} = \frac{1}{s}S^\top\boldsymbol{u}$.

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", $\|s(S^\top)^{-1}\boldsymbol{v}\| \leq \|\boldsymbol{v}\|$ for all $\boldsymbol{v}$. Since the lattice L' is obtained from L by "shrinking" all the vectors $\boldsymbol{v} \in L$ 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 $\boldsymbol{u} \leftarrow \mathcal{D}_{L',s,\boldsymbol{c'}}$ we get $\|\boldsymbol{u}\| \leq s\sqrt{n}$ except with probability at most $\frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n}$. Hence we can bound whp the norm of $\boldsymbol{v}$ by
$\|\boldsymbol{v}\| = \|\frac{1}{s}S^\top\boldsymbol{u}\| \leq \frac{1}{s} \cdot \sigma_1(S^\top) \cdot \|\boldsymbol{u}\| = \frac{1}{s} \cdot \sigma_1(S) \cdot s\sqrt{n} = \sigma_1(S)\sqrt{n}$.

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 $\rho_{S,\boldsymbol{c}}(L) \in [\frac{1-\epsilon}{1+\epsilon}, 1] \cdot \rho_S(L)$.* □

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 $\boldsymbol{x} \leftarrow \mathcal{D}_{L,s}$ from the spherical discrete Gaussian on L and $\boldsymbol{y} \leftarrow \mathcal{N}^n(0, r^2/2\pi)$ from a continuous Gaussian, and outputting $\boldsymbol{z} = \boldsymbol{x}+\boldsymbol{y}$. Then for any point $\boldsymbol{u} \in \mathbb{R}^n$, the probability *density $\mathcal{R}_{L,r,s}(\boldsymbol{u})$ is close to the probability density under the spherical continuous Gaussian $\mathcal{N}^n(0, t^2/2\pi)$ upto a factor of $\frac{1-\epsilon}{1+\epsilon}$:*

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

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 $E(\boldsymbol{u})$, we have

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

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* $\boldsymbol{v} \in L$, then the probability mass of T is not very different than the probability mass of $T - \boldsymbol{v} = \{\boldsymbol{u} - \boldsymbol{v} : \boldsymbol{u} \in T\}$. 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 $\boldsymbol{v} \in L$, it holds that $\mathcal{D}_{L,s}(T) - \mathcal{D}_{L,s}(T - \boldsymbol{v}) \le \frac{\mathsf{erf}(q(1+4/c)/2)}{\mathsf{erf}(2q)} \cdot \frac{1+\epsilon}{1-\epsilon}$, where $q = \|v\|\sqrt{\pi}/s$.*

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

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

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

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

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

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

In more detail, denoting $r = s/c$ we consider drawing $\boldsymbol{u} \leftarrow \mathcal{D}_{L,s}$ and $\boldsymbol{z} \leftarrow \mathcal{N}^n(0, r^2/2\pi)$, and setting $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{z}$. 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 $\boldsymbol{w}$ 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 $\boldsymbol{v}$. When expressed in this basis, we get the event $\boldsymbol{z} \in H'_{\boldsymbol{v}}$ exactly when the coefficient in the direction of $\boldsymbol{v}$ (which is distributed close to the 1-diemsnional Gaussian $\mathcal{N}(0, t^2/2\pi)$) exceeds $\|\boldsymbol{v}(1 + \frac{4}{c})/2\|$ in magnitude. Hence we have

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

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

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

Putting the last two bounds together, we get

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

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

One useful special case of Lemma 6 is when $c = 100$ (say) and $\|\boldsymbol{v}\| \approx s$, where we get a bound $\mathcal{D}_{L,s}(T) - \mathcal{D}_{L,s}(T - \boldsymbol{v}) \leq \frac{\mathsf{erf}(0.52\sqrt{\pi})}{\mathsf{erf}(2\sqrt{\pi})} \cdot \frac{1+\epsilon}{1-\epsilon} \approx 0.81$. We note that when $\frac{\|\boldsymbol{v}\|}{s} \to 0$, 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 $H'_{\boldsymbol{v}}$ and $H''_{\boldsymbol{v}}$ thicker, e.g. $H''_{\boldsymbol{v}} = H_{\boldsymbol{v}}$ and $H'_{\boldsymbol{v}} = 2H_{\boldsymbol{v}}$). 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 $\boldsymbol{v} \in L$, it holds that $\mathcal{D}_{L,S}(T) - \mathcal{D}_{L,S}(T - \boldsymbol{v}) \leq \frac{\mathsf{erf}(q(1+4/c)/2)}{\mathsf{erf}(2q)} \cdot \frac{1+\epsilon}{1-\epsilon}$, 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 Our Discrete Gaussian LHL

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 $s' > \Omega(\text{poly}(n)\log(1/\epsilon))$. 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, $\boldsymbol{x}_i \leftarrow \mathcal{D}_{L,s}$.[4]

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

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

Our goal is to prove that $\mathcal{E}_{X,s'}$ is close to the ellipsoid Gaussian $\mathcal{D}_{L,s'X^\top}$, over L. We begin by proving that the singular values of $X^\top$ are all roughly of the size $s\sqrt{m}$[5].

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

[5] Since we eventually apply the following lemmas to $X^\top$, we will use $X^\top$ in the statement of the lemmas for consistency at the risk of notational clumsiness.

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 $X^\top$ independently at random from a spherical discrete Gaussian on L with parameter $s > 2K\eta_\epsilon(L)$, $X^\top \leftarrow (\mathcal{D}_{L,s})^m$, we have*

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

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 $Z = X^\top + Y$. 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 columns of Z and a continuous 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 $X^\top = Z - Y$, we can apply Fact 1 (with $\delta = 1/2$) to conclude that $\sigma_n(X^\top) \geq (1 - \frac{1}{2})s\sqrt{2\pi m}/C > s\sqrt{2\pi m}/K$ and $\sigma_n(X^\top) \leq (1 + \frac{1}{2})2sC\sqrt{2\pi m} \leq sK\sqrt{2\pi m}$. In summary, we have

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

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 $\mathcal{D}_{L,s'X^\top}$. 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 columns 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 columns of an n-by-m 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 $\mathcal{D}_{\mathbb{Z}^n,s'X^\top}$ is bounded by 2ϵ.

The rest of this subsection is devoted to proving Theorem 2. We begin by showing that with overwhelming probability, the columns 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 columns of an n-by-m matrix X independently at random from $\mathcal{D}_{\mathbb{Z}^n,S}$ we get $X \cdot \mathbb{Z}^m = \mathbb{Z}^n$ with all but probability $2^{-O(m)}$.*

Proof. Consider choosing the columns one by one, and we show that (a) as long as the current columns 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 columns 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 columns $\boldsymbol{x}_i$ of X: At any step i we consider only the "short vectors" among the previous $\boldsymbol{x}_i$'s, namely $X_{i-1} \overset{\text{def}}{=} \{\boldsymbol{x}_j : j < i, \|\boldsymbol{x}_j\| \leq s\sqrt{n}\}$.

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 $\|\boldsymbol{x}_i\| \leq s\sqrt{n}$ and $\boldsymbol{x}_i$ 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 $\|\boldsymbol{x}_i\| \leq s\sqrt{n}$ and $\boldsymbol{x}_i$ falls outside that sub-lattice, and $\chi_i = 0$ otherwise;
3. Else (if $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}$ $\mathbb{Z}$-span the entire $\mathbb{Z}^n$), we defined $\chi_i = 1$.

It is clear from the definition of the χ_i's that $\sum_{i=1}^m \chi_i \geq n$ implies that the $\boldsymbol{x}_i$'s $\mathbb{R}$-span all of of $\mathbb{R}^n$. Moreover we claim that if $\sum_{i=1}^m \chi_i \geq n(\log(s\sqrt{n}) + 1)$ then the $\boldsymbol{x}_i$'s must $\mathbb{Z}$-span the entire lattice $\mathbb{Z}^n$. To see this, consider the first n vectors $\boldsymbol{x}_i$ 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 $\boldsymbol{x}_i$ do not yet $\mathbb{Z}$-span the entire integer lattice, any subsequent $\boldsymbol{x}_i$ 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 $\boldsymbol{x}_i$'s must $\mathbb{Z}$-span the entire integer lattice. We therefore have

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

It is left to lower-bound the last expression. We claim that regardless of the previous $\boldsymbol{x}_{i'}$'s for $i' < i$, we always have $\Pr[\chi_i = 1] \geq 1/4$. 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 $\boldsymbol{e}_j$ 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 $T_{i-1} - \boldsymbol{e}_j$ is disjoint from T_{i-1}. Since $\|\boldsymbol{e}_j\| = 1$ and $s > \eta_\epsilon(\mathbb{Z}^n) \geq \sqrt{n}$, then Corollary 2 (with $c = 9$) says that

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

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

$$\Pr[\chi_i = 1] \geq \Pr[\boldsymbol{x}_i \notin T_{i-1} \text{ and } \|\boldsymbol{x}_i\| \leq \sqrt{n}] \;\geq\; \Pr[\boldsymbol{x}_i \notin T_{i-1}] - \frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n}$$
$$\geq 0.3 - \frac{1+\epsilon}{1-\epsilon} \cdot 2^{-n} \;>\; 0.25$$

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 $\boldsymbol{e}_j$ 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 columns 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 rows of X, and for any $\boldsymbol{z} \in \mathbb{Z}^n$ we denote by $A_{\boldsymbol{z}} = A_{\boldsymbol{z}}(X)$ the $\boldsymbol{z}$ coset of A:

$$A = A(X) \overset{\text{def}}{=} \{\boldsymbol{v} \in \mathbb{Z}^m : X \cdot \boldsymbol{v} = 0\} \text{ and } A_{\boldsymbol{z}} = A_{\boldsymbol{z}}(X) \overset{\text{def}}{=} \{\boldsymbol{v} \in \mathbb{Z}^m : X \cdot \boldsymbol{v} = \boldsymbol{z}\}.$$

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

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 $\mathcal{D}_{\mathbb{Z}^n,s'X^\top}$. 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 $\boldsymbol{v}$), then $\mathcal{E}_{X,s'}$ and $\mathcal{D}_{\mathbb{Z}^n,s'X^\top}$ must be close.

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

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

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 $\boldsymbol{z} \in \mathbb{Z}^n$. The probability mass assigned to $\boldsymbol{z}$ 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 $\boldsymbol{v} \in \mathbb{Z}^m$ for which $X \cdot \boldsymbol{v} = \boldsymbol{z}$. In other words, this is exactly the probability mass assigned by $\mathcal{D}_{\mathbb{Z}^m,s'}$ to the coset $A_{\boldsymbol{z}}$. Below let $T = T(X) \subseteq \mathbb{R}^m$ be the linear subspace containing the lattice A, and $T_{\boldsymbol{z}} = T_{\boldsymbol{z}}(X) \subseteq \mathbb{R}^m$ be the affine subspace containing the coset $A_{\boldsymbol{z}}$:

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

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

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

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

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

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

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

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

As a last step, recall that $\boldsymbol{u}_{\boldsymbol{z}} = Y^\top \boldsymbol{z}$ where $YY^\top = (XX^\top)^{-1}$. Thus we have

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

Putting everything together we get

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

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

The smoothing parameter of $\boldsymbol{A}$. 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 rows of X. We extend A to a full-rank lattice as follows: First we extend the rows space of X, by throwing in also the scaled standard unit vectors $q\boldsymbol{e}_i$ for the integer parameter q mentioned above ($q \ge 8(mn)^{1.5}s_1 w$). That is, we let $M_q = M_q(X)$ be the full-rank m-dimensional lattice spanned by the rows of X and the vectors $q\boldsymbol{e}_i$,

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

(where we idenfity $\mathbb{Z}_q$ 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.,

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

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

Obviously all the rows 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 rows of X is of size at least $q/4nw$.

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

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

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

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

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

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

Hence except with probability $m2^{-m}$ all the entries of $X^\top \boldsymbol{z}$ are smaller than $q/2$ in magnitude, which means that $[X^\top \boldsymbol{z}]_q = X^\top \boldsymbol{z}$, and so $[X^\top \boldsymbol{z}]_q$ belongs to the row 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 $[X^\top \boldsymbol{z}]_q$ for all the "small" $\boldsymbol{z}$'s belong to the row space of X.

We next turn to analyzing "large" $\boldsymbol{z}$'s. Fix one "large" vector $\boldsymbol{z}$, and for that vector define the set of "bad" vectors $\boldsymbol{x} \in \mathbb{Z}^n$, i.e. the ones for which $|[\langle \boldsymbol{z}, \boldsymbol{x}\rangle]_q| < q/4nw$ (and the other vectors $\boldsymbol{x} \in \mathbb{Z}^n$ are "good"). Observe that if $\boldsymbol{x}$ 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 $\mu = \min\big(\lceil s_n \rceil,\ \lfloor q/|2z_{\max}|\rfloor\big)$, since

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

(The last two inequalities follow since $q/2nw < \mu|z_{\max}| \leq q/2$ and $|[\langle \boldsymbol{z}, \boldsymbol{x}\rangle]_q| < q/4nw$.) Hence the injunction $\boldsymbol{x} \mapsto \boldsymbol{x} + \mu \boldsymbol{e}_{i_{\max}}$ maps "bad" $\boldsymbol{x}$'es to "good" $\boldsymbol{x}$'es. Moreover, since the $\boldsymbol{x}$'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 $\boldsymbol{x}$ differs from the total mass of the "good" vectors $\boldsymbol{x} + \mu \boldsymbol{e}_{i_{\max}}$ by at most 0.81. It follows that when choosing $\boldsymbol{x} \leftarrow \mathcal{D}_{\mathbb{Z}^n,S}$, we have $\Pr_{\boldsymbol{x}}\left[|[\langle \boldsymbol{z}, \boldsymbol{x}\rangle]_q| < q/4nw\right] \leq (1{+}0.81)/2 < 0.91$. Thus the probability that all the entries of $[X^\top \boldsymbol{z}]_q$ 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 $\|[X^\top \boldsymbol{z}]_q\| < q/4nw$ 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 $\boldsymbol{z} \in \mathbb{Z}_q^n$ for which $[X^\top \boldsymbol{z}]_q$ is linearly independent of the rows of X and yet $|[X^\top \boldsymbol{z}]_q| < q/4nw$.

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

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

which together with $X \cdot \boldsymbol{v} \equiv \boldsymbol{0} \pmod{q}$ means that we have $X \cdot \boldsymbol{v}_j = \boldsymbol{0}$ (over $\mathbb{R}$, with no modular reduction). We conclude that the $\boldsymbol{v}_j$'s are integer vectors orthogonal to the rows 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 columns of an n-by-m 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 $\mathcal{D}_{L,s'X^\top}$ 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 columns of an n-by-m integer matrix as $Z \leftarrow (\mathcal{D}_{\mathbb{Z}^n,S})^m$, the statistical distance between $\mathcal{E}_{Z,s'}$ and $\mathcal{D}_{\mathbb{Z}^n,s'Z^\top}$ is at most 2ϵ.

Letting $X^\top = BZ^\top$, 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 $\mathcal{D}_{L,s'X^\top} = B \times \mathcal{D}_{\mathbb{Z}^n,s'Z^\top}$. 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 $\mathcal{D}_{L,s'X^\top}$.

4 Applications

In this section, we discuss the application of our discrete Gaussian LHL in the construction of multilinear maps from lattices [GGH12]. This construction uses our LHL crucially for randomizing certain encodings. Below we provide a detailed description of how our lemma is used. This construction is also illustrative in providing an example of a canonical setting where our lemma should be useful.

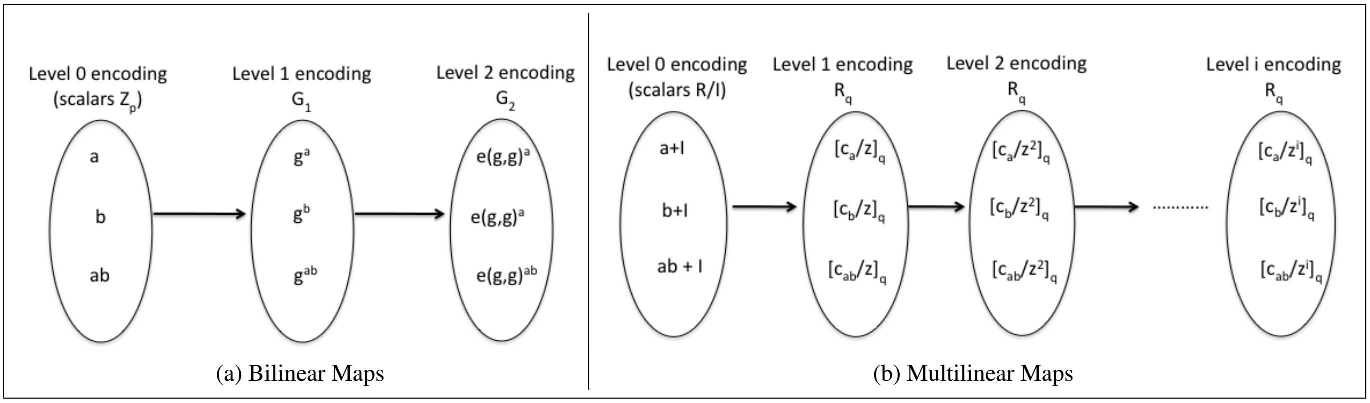
High Level overview of GGH Construction. To begin, we provide a high level overview of the construction. Our description here is necessarily high level and skips many important details, we refer the reader to [GGH12] for a complete description. Recall that Bilinear groups consist of two groups G_1 and G_2 along with a map $e : G_1 \times G_1 \to G_2$ such that $e(x^a, y^b) = e(x, y)^{ab}$. A canonical hard problem over Bilinear groups is the Discrete Log problem where given g, g^a, one is asked to compute a. In [GGH12], the authors view $a \to g^a$ as an "encoding" of a that satisfies (at a high level), the following properties:

1. The encoding is easy to compute in the forward direction and hard to invert.
2. Encoding is additively homomorphic as well as one-time multiplicatively homomorphic (via the pairing).
3. Given encodings of two elements, it is easy to test whether the underlying scalars are equal: $a = b$ if $g^a = g^b$.
4. Given encodings, it is hard to test more complicated relations between the underlying scalars. For example, given the tuple (x, y, z) where $x, y, z \in G_1$ encode a, b, c respectively, i.e. $(x, y, z) = (g^a, g^b, g^c)$ and $w \in G_2$, test if w encodes abc, i.e. test if $w = e(g, g)^{abc}$?

In [GGH12], the authors construct encodings that approximately satisfy (and generalize) the above properties from lattices. See figure 1 for a high level schematic. In their setting, $R = \mathbb{Z}[x]/f(x)$ is a polynomial ring for cyclotomic $f(x)$ and $R_q = R/qR$ for some large q. Let $\boldsymbol{g} \in R_q$ be a small element and $\mathcal{I} = (\boldsymbol{g})$ be a principal ideal over R generated by $\boldsymbol{g}$. The scalars they encode are elements of the quotient ring $R/\mathcal{I}$ (so if $|R/\mathcal{I}| = p$, the elements can be represented by $0, \ldots p-1$) and the source and target groups are R_q.

The encoding of element $\boldsymbol{s} + \mathcal{I} \in R/\mathcal{I}$ is given by $[\boldsymbol{c}/\boldsymbol{z}]_q$ for some small $\boldsymbol{c} \in \boldsymbol{s} + \mathcal{I}$ and some (uniformly) randomly chosen, fixed, hidden $\boldsymbol{z} \in R_q$. Note that this encoding right away satisfies property (2): the encoding is additively homomorphic – $[\boldsymbol{c}_1/\boldsymbol{z} + \boldsymbol{c}_2/\boldsymbol{z}]_q = [(\boldsymbol{c}_1 + \boldsymbol{c}_2)/\boldsymbol{z}]_q$ where $\boldsymbol{c}_1 + \boldsymbol{c}_2$ is short and $\boldsymbol{c}_1 + \boldsymbol{c}_2 \in \boldsymbol{s}_1 + \boldsymbol{s}_2 + \mathcal{I}$ as well as multiplicatively homomorphic $[\boldsymbol{c}_1/\boldsymbol{z}]_q \times [\boldsymbol{c}_2/\boldsymbol{z}]_q = [\boldsymbol{c}_1\boldsymbol{c}_2/\boldsymbol{z}]_q$ where $\boldsymbol{c}_1\boldsymbol{c}_2$ is short and $\boldsymbol{c}_1\boldsymbol{c}_2 \in \boldsymbol{s}_1\boldsymbol{s}_2 + \mathcal{I}$. Moreover, this immediately generalizes to the multilinear setting, where a level k encoding of a coset $\boldsymbol{s} + \mathcal{I}$ is of the form $[\boldsymbol{c}/\boldsymbol{z}^k]_q$ for short $\boldsymbol{c} \in \boldsymbol{s} + \mathcal{I}$ and $[\boldsymbol{c}_1/\boldsymbol{z}^i]_q[\boldsymbol{c}_2/\boldsymbol{z}^j]_q = [\boldsymbol{c}_1\boldsymbol{c}_2/\boldsymbol{z}^{i+j}]_q$ where $\boldsymbol{c}_1\boldsymbol{c}_2 \in \boldsymbol{s}_1\boldsymbol{s}_2 + \mathcal{I}$ and are short. To satisfy property (3), i.e. testing whether $\boldsymbol{u}_1, \boldsymbol{u}_2$ encode the same coset entails testing whether $[\boldsymbol{u}_1 - \boldsymbol{u}_2]_q$ encodes 0. The authors enable this by providing a "zero testing parameter" in the public key: for level k, publish $\boldsymbol{v}_k = [\boldsymbol{h}\boldsymbol{z}^k/\boldsymbol{g}]_q$ where $\boldsymbol{h}$ is somewhat short. Then, to test equality, one simply tests if $[\boldsymbol{u}_1 - \boldsymbol{u}_2]_q \cdot \boldsymbol{v}_k$ is short mod q. Note that if $[\boldsymbol{u}_1 - \boldsymbol{u}_2]_q$ encodes 0 at level k, then $[\boldsymbol{u}_1 - \boldsymbol{u}_2]_q = [\boldsymbol{c}/\boldsymbol{z}^k]_q$ where $\boldsymbol{c} \in I$, hence $\boldsymbol{c} = \boldsymbol{c}'\boldsymbol{g}$ for some c'. Then, $[\boldsymbol{u}_1 - \boldsymbol{u}_2]_q \cdot \boldsymbol{v}_k \mod q = [\boldsymbol{c}'\boldsymbol{h}]_q$ is a short element for an appropriate setting of parameters.

Next, we come to property (1): a natural method to enable efficient encoding given public parameters (which hide $\boldsymbol{z}$) is to publish an encoding of 1, i.e. $\boldsymbol{y}_1 = [\boldsymbol{a}_1/\boldsymbol{z}]_q$ where $a_1 \in \mathcal{I} + 1$ and have the encoder pick a short element in his chosen coset $\boldsymbol{c} \in \boldsymbol{s} + \mathcal{I}$ and set the encoding as $\boldsymbol{c} \cdot \boldsymbol{y}_1$. Then translating from level i to level $i + 1$ is $\boldsymbol{u}_{i+1} = [\boldsymbol{u}_i \cdot \boldsymbol{y}]_q$. However this simple encoding is certainly not hard to decode: just dividing by $\boldsymbol{y}_1$ suffices! Moreover property (4) is also not satisfied; several complicated relations are easy to test algebraically. For example, a typical application would need to generate encoding of a random s, t and st , say $\boldsymbol{u}_s, \boldsymbol{u}_t$ and $\boldsymbol{u}_{st}$ so that $(\boldsymbol{u}_s, \boldsymbol{u}_t, \boldsymbol{u}_{st})$ is hard to distinguish from an encoding of s, t, r where r is random, without the appropriate zero testing parameter. However, the simple encoding



(a) Bilinear Maps

(b) Multilinear Maps

Fig. 1. Viewing multilinear maps as graded encodings.

presented above does not achieve this – the procedure would draw random short $\boldsymbol{c}_s, \boldsymbol{c}_t$, from $\boldsymbol{s}+\mathcal{I}, \boldsymbol{t}+\mathcal{I}$ respectively, compute $\boldsymbol{c}_{st} = \boldsymbol{c}_s\boldsymbol{c}_t$ and encode $\boldsymbol{u}_s = [\boldsymbol{c}_s\boldsymbol{y}_1]_q$, $\boldsymbol{u}_t = [\boldsymbol{c}_t\boldsymbol{y}_1]_q$ and $\boldsymbol{u}_{st} = [\boldsymbol{c}_{st}\boldsymbol{y}_1]_q$. But then an adversary can simply check if $\boldsymbol{u}_{st} = \boldsymbol{u}_s\boldsymbol{u}_t/\boldsymbol{y}_1$.

Randomizing the encodings. To break these simple algebraic relations, the authors include in the public parameters also the "randomizers" $\boldsymbol{x}_i$ which are just random encodings of zero. Namely $\boldsymbol{x}_i = [\boldsymbol{b}_i/\boldsymbol{z}]_q$ where the $\boldsymbol{b}_i$'s are short elements in $\mathcal{I}$. Denote by X the matrix with the vectors $\boldsymbol{x}_i$ as columns, namely $X = (\boldsymbol{x}_1|\dots|\boldsymbol{x}_m)$. Denote by B the matrix with the numerators $\boldsymbol{b}_i$ as columns, i.e., $B = (\boldsymbol{b}_1|\dots|\boldsymbol{b}_m)$.

Then, they use the $\boldsymbol{x}_i$'s to randomize level-one encodings: Given $\boldsymbol{u}' = [\boldsymbol{c}'/\boldsymbol{z}]_q$ with appropriate noise-bound $\|\boldsymbol{c}'\| < \gamma$, they draw an m-vector of integer coefficients $\boldsymbol{r} \leftarrow D_{\mathbb{Z}^m,\sigma^*}$ for large enough σ^* (e.g. $\sigma^* = 2^\lambda\gamma$ where λ is the security parameter), and output

$$\boldsymbol{u} \;:=\; [\boldsymbol{u}' + X\boldsymbol{r}]_q \;=\; [\boldsymbol{u}' + \sum_{i=1}^m r_i\boldsymbol{x}_i]_q \;(=\; [\frac{\boldsymbol{c}' + \sum_i r_i\boldsymbol{b}_i}{\boldsymbol{z}}]_q).$$

We write $B\boldsymbol{r}$ as a shorthand for $\sum_i r_i\boldsymbol{b}_i$ and similarly $X\boldsymbol{r}$ as a shorthand for $\sum_i r_i\boldsymbol{x}_i$.

Since all the $\boldsymbol{b}_i$'s are in the ideal $\mathcal{I}$, then obviously $\boldsymbol{c}' + \sum_i r_i\boldsymbol{b}_i$ is in the same coset of $\mathcal{I}$ as $\boldsymbol{c}'$ itself. Moreover since $\|\boldsymbol{b}_i\| < \text{poly}(n)$ then $\|B\boldsymbol{r}\| < \sigma^*\text{poly}(m,n)$. If indeed $\|\boldsymbol{c}'\| < \gamma$, then $\|\boldsymbol{c}' + B\boldsymbol{r}\| < \gamma + \sigma^*\text{poly}(m,n)$. Now, the [GGH12] can claim that the distribution of $\boldsymbol{u}$ is nearly independent of original $\boldsymbol{u'}$ conditioned on its coset. If the $\boldsymbol{b}_i$'s are chosen from a wide enough spherical distribution then our Gaussian LHL (Thm 3) allows them to conclude that $B\boldsymbol{r}$ is close to a wide ellipsoid Gaussian. With appropriate choice of σ^* the "width" of that distribution is much larger than the original $\boldsymbol{c'}$, hence the distribution of $\boldsymbol{c}' + B\boldsymbol{r}$ is nearly independent of $\boldsymbol{c'}$, conditioned on the coset it belongs to.

5 Discussion

Unlike the classic LHL, our lattice version of LHL is less than perfect – instead of yielding a perfectly spherical Gaussian, it only gives us an approximately spherical one, i.e. $\mathcal{D}_{L,s'X^\top}$. Here approximately spherical means that all the singular values of the matrix $X^\top$ are within a small, constant sized interval. It is therefore natural to ask: 1) Can we do better and obtain a perfectly spherical Gaussian? 2) Is an approximately spherical Gaussian sufficient for cryptographic applications?

First let us consider whether we can make the Gaussian perfectly spherical. Indeed, as the number of lattice vectors m grows larger, we expect the greatest and least singular value of the discrete Gaussian matrix X to converge – this would imply that as $m \to \infty$, the linear combination $\sum_{i=1}^m z_i\boldsymbol{x}_i$ does indeed behave like a spherical Gaussian. While we do not prove this, we refer the reader to [RV10] for intuitive evidence. However, the focus of this work is small

m (such as $m = 2n$) suitable for applications, in which case we cannot hope for the same. Discrete Gaussians over infinite rings just do not behave that way, and one way to view our work is getting a handle on their behavior.

This leads to the second question: is approximately spherical good enough? This depends on the application. We have already seen that it is sufficient for GGH encodings [GGH12], where a canonical, wide-enough, but non-spherical Gaussian is used to "drown out" an initial encoding, and send it to a canonical distribution of encodings that encode the same value. Our LHL shows that one can sample from such a canonical approximate Gaussian distribution without using the initial Gaussian samples "wastefully".

On the other hand, we caution the reader that if the application requires the basis vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ to be kept secret (such as when the basis is a trapdoor), then one must carefully consider whether our Gaussian sampler can be used safely. This is because, as demonstrated by [NR09] and [DN12], lattice applications where the basis is desired to be secret can be broken completely even if partial information about the basis is leaked. In an application where the trapdoor is available explicitly and oblivious sampling is not needed, it is safer to use the samplers of [GPV08] or [Pei10] to sample a perfectly spherical Gaussian that is statistically independent of the trapdoor.

References

[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.

[DN12] Léo Ducas and Phong Q. Nguyen. Learning a zonotope and more: Cryptanalysis of ntrusign countermeasures. In *ASIACRYPT*, pages 433–450, 2012.

[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.

[Kle00] Philip Klein. Finding the closest lattice vector when it's unusually close. In *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, SODA '00, pages 937–941, 2000.

[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.

[Lyu12] Vadim Lyubashevsky. Lattice signatures without trapdoors. In *Proceedings of the 31st Annual international conference on Theory and Applications of Cryptographic Techniques*, EUROCRYPT'12, Berlin, Heidelberg, 2012. Springer-Verlag.

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

[NR09] Phong Q. Nguyen and Oded Regev. Learning a parallelepiped: Cryptanalysis of ggh and&#x00a0;ntru signatures. *J. Cryptol.*, 22(2):139–160, April 2009.

[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.

[RV10] Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *International Congress of Mathematicans*, 2010.

[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.