Continuous LWE

Joan Bruna[*a,b,c], Oded Regev[†a], Min Jae Song[‡a], and Yi Tang[§d]

[a]Courant Institute of Mathematical Sciences, New York University, New York
[b]Center for Data Science, New York University, New York
[c]Institute for Advanced Study, Princeton
[d]Computer Science and Engineering, University of Michigan, Ann Arbor

October 27, 2020

Abstract

We introduce a continuous analogue of the Learning with Errors (LWE) problem, which we name CLWE. We give a polynomial-time quantum reduction from worst-case lattice problems to CLWE, showing that CLWE enjoys similar hardness guarantees to those of LWE. Alternatively, our result can also be seen as opening new avenues of (quantum) attacks on lattice problems. Our work resolves an open problem regarding the computational complexity of learning mixtures of Gaussians without separability assumptions (Diakonikolas 2016, Moitra 2018). As an additional motivation, (a slight variant of) CLWE was considered in the context of robust machine learning (Diakonikolas et al. FOCS 2017), where hardness in the statistical query (SQ) model was shown; our work addresses the open question regarding its computational hardness (Bubeck et al. ICML 2019).

1 Introduction

The Learning with Errors (LWE) problem has served as a foundation for many lattice-based cryptographic schemes [Pei16]. Informally, LWE asks one to solve noisy random linear equations. To be more precise, the goal is to find a secret vector $\boldsymbol{s} \in \mathbb{Z}_q^n$ given polynomially many samples of the form $(\boldsymbol{a}_i, b_i)$, where $\boldsymbol{a}_i \in \mathbb{Z}_q^n$ is uniformly chosen and $b_i \approx \langle \boldsymbol{a}_i, \boldsymbol{s} \rangle \pmod q$. In the absence of noise, LWE can be efficiently solved using Gaussian elimination. However, LWE is known to be hard assuming hardness of worst-case lattice problems such as Gap Shortest Vector Problem (GapSVP) or Shortest Independent Vectors Problem (SIVP) in the sense that there is a polynomial-time quantum reduction from these worst-case lattice problems to LWE [Reg05].

In this work, we introduce a new problem, called Continuous LWE (CLWE). As the name suggests, this problem can be seen as a continuous analogue of LWE, where equations in $\mathbb{Z}_q^n$ are replaced with vectors in $\mathbb{R}^n$ (see Figure 1). More precisely, CLWE considers noisy inner products $z_i \approx \gamma \langle \boldsymbol{y}_i, \boldsymbol{w} \rangle \pmod 1$, where the noise is drawn from a Gaussian distribution of width $\beta > 0$,

[*]This work is partially supported by the Alfred P. Sloan Foundation, NSF RI-1816753, NSF CAREER CIF 1845360, and the Institute for Advanced Study.
[†]Research supported by the Simons Collaboration on Algorithms and Geometry, a Simons Investigator Award, and by the National Science Foundation (NSF) under Grant No. CCF-1814524.
[‡]Research supported by the National Science Foundation (NSF) under Grant No. CCF-1814524.
[§]This work was done while the author was at the Courant Institute of Mathematical Sciences, New York University.

$\gamma > 0$ is a problem parameter, $\boldsymbol{w} \in \mathbb{R}^n$ is a secret unit vector, and the public vectors $\boldsymbol{y}_i \in \mathbb{R}^n$ are drawn from the standard Gaussian. Given polynomially many samples of the form $(\boldsymbol{y}_i, z_i)$, CLWE asks one to find the secret direction $\boldsymbol{w}$.

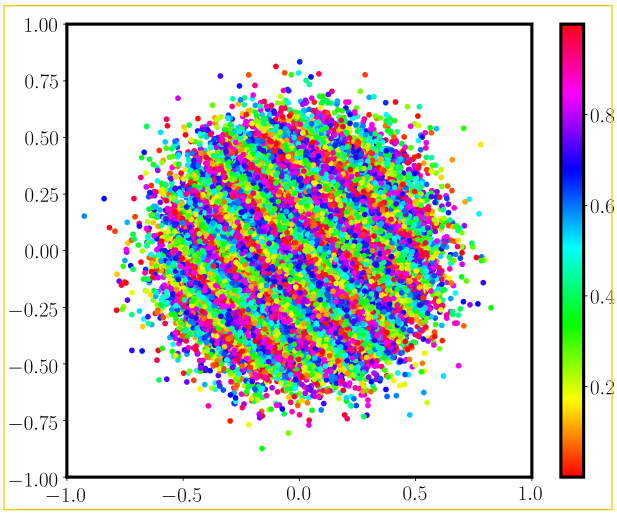

Figure 1: Scatter plot of two-dimensional CLWE samples. Color indicates the last (z) coordinate.

One can also consider a closely related homogeneous variant of CLWE (see Figure 2). This distribution, which we call homogeneous CLWE, can be obtained by essentially conditioning on $z_i \approx 0$. It is a mixture of "Gaussian pancakes" of width $\approx \beta/\gamma$ in the secret direction and width 1 in the remaining $n-1$ directions. The Gaussian components are equally spaced, with a separation of $\approx 1/\gamma$. (See Definition 2.19 for the precise statement.)

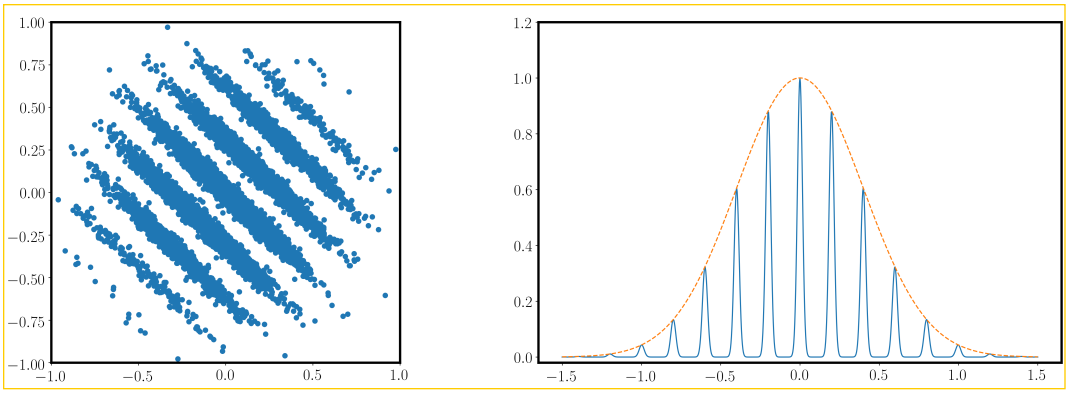

Figure 2: Left: Scatter plot of two-dimensional homogeneous CLWE samples. Right: Unnormalized probability densities of homogeneous CLWE (blue) and Gaussian (orange) along the hidden direction.

Our main result is that CLWE (and homogeneous CLWE) enjoy hardness guarantees similar to those of LWE.

Theorem 1.1 (Informal)**.** *Let n be an integer, $\beta = \beta(n) \in (0,1)$ and $\gamma = \gamma(n) \geq 2\sqrt{n}$ such that the ratio γ/β is polynomially bounded. If there exists an efficient algorithm that solves $\mathrm{CLWE}_{\beta,\gamma}$, then there exists an efficient quantum algorithm that approximates worst-case lattice problems to within polynomial factors.*

Although we defined CLWE above as a search problem of finding the hidden direction, Theorem 1.1 is actually stronger, and applies to the decision variant of CLWE in which the goal is to distinguish CLWE samples $(\boldsymbol{y}_i, z_i)$ from samples where the noisy inner product z_i is replaced by a random number distributed uniformly on $[0,1)$ (and similarly for the homogeneous variant).

Motivation: Lattice algorithms. Our original motivation to consider CLWE is as a possible approach to finding quantum algorithms for lattice problems. Indeed, the reduction above (just like the reduction to LWE [Reg05]), can be interpreted in an algorithmic way: in order to quantumly solve worst-case lattice problems, "all" we have to do is solve CLWE (classically or quantumly). The elegant geometric nature of CLWE opens up a new toolbox of techniques that can potentially be used for solving lattice problems, such as sum-of-squares-based techniques and algorithms for learning mixtures of Gaussians [MV10]. Indeed, some recent algorithms (e.g., [KKK19, RY20]) solve problems that include CLWE or homogeneous CLWE as a special case (or nearly so), yet as far as we can tell, so far none of the known results leads to an improvement over the state of the art in lattice algorithms.

To demonstrate the usefulness of CLWE as an algorithmic target, we show in Section 7 a simple moment-based algorithm that solves CLWE in time $\exp(\gamma^2)$. Even though this does not imply subexponential time algorithms for lattice problems (since Theorem 1.1 requires $\gamma > \sqrt{n}$), it is interesting to contrast this algorithm with an analogous algorithm for LWE by Arora and Ge [AG11]. The two algorithms have the same running time (where γ is replaced by the absolute noise αq in the LWE samples), and both rely on related techniques (moments in our case, powering in Arora-Ge's), yet the Arora-Ge algorithm is technically more involved than our rather trivial algorithm (which just amounts to computing the empirical covariance matrix). We interpret this as an encouraging sign that CLWE might be a better algorithmic target than LWE.

Motivation: Hardness of learning Gaussian mixtures. Learning mixtures of Gaussians is a classical problem in machine learning [Pea94]. Efficient algorithms are known for the task if the Gaussian components are guaranteed to be sufficiently well separated (e.g., [Das99, VW02, AK05, DS07, BV08, RV17, HL18, KSS18, DKS18]). Without such strong separation requirements, it is known that efficiently recovering the individual components of a mixture (technically known as "parameter estimation") is in general impossible [MV10]; intuitively, this exponential information theoretic lower bound holds because the Gaussian components "blur into each other", despite being mildly separated pairwise.

This leads to the question of whether there exists an efficient algorithm that can learn mixtures of Gaussians without strong separation requirement, not in the above strong parameter estimation sense (which is impossible), but rather in the much weaker density estimation sense, where the goal is merely to output an approximation of the given distribution's density function. See [Dia16, Moi18] for the precise statement and [DKS17] where a super-polynomial lower bound for density estimation is shown in the restricted statistical query (SQ) model [Kea98, Fel+17]. Our work provides a negative answer to this open question, showing that learning Gaussian mixtures is computationally difficult even if the goal is only to output an estimate of the density (see Proposition 5.2). It is worth noting that our hard instance has almost non-overlapping components, i.e., the pairwise

statistical distance between distinct Gaussian components is essentially 1, a property shared by the SQ-hard instance of [DKS17].

Motivation: Robust machine learning. Variants of CLWE have already been analyzed in the context of robust machine learning [Bub+19], in which the goal is to learn a classifier that is robust against adversarial examples at test time [Sze+14]. In particular, Bubeck et al. [Bub+19] use the SQ-hard Gaussian mixture instance of Diakonikolas et al. [DKS17] to establish SQ lower bounds for learning a certain binary classification task, which can be seen as a variant of homogeneous CLWE. The key difference between our distribution and that of [DKS17, Bub+19] is that our distribution has equal spacing between the "layers" along the hidden direction, whereas their "layers" are centered around roots of Hermite polynomials (the goal being to exactly match the lower moments of the standard Gaussian). The connection to lattices, which we make for the first time here, answers an open question by Bubeck et al. [Bub+19].

As additional evidence of the similarity between homogeneous CLWE and the distribution considered in [DKS17, Bub+19], we prove a super-polynomial SQ lower bound for homogeneous CLWE (even with super-polynomial precision). For $\gamma = \Omega(\sqrt{n})$, this result translates to an exponential SQ lower bound for exponential precision, which corroborates our computational hardness result based on worst-case lattice problems. The uniform spacing in the hidden structure of homogeneous CLWE leads to a simplified proof of the SQ lower bound compared to previous works, which considered non-uniform spacing between the Gaussian components. Note that computational hardness does not automatically imply SQ hardness as query functions in the SQ framework need not be efficiently computable.

Bubeck et al. [Bub+19] were also interested in a variant of the learning problem where instead of *one* hidden direction, there are $m \geq 1$ orthogonal hidden directions. So, for instance, the "Gaussian pancakes" in the $m = 1$ case above are replaced with "Gaussian baguettes" in the case $m = 2$, forming an orthogonal grid in the secret two-dimensional space. As we show in Section 9, our computational hardness easily extends to the $m > 1$ case using a relatively standard hybrid argument. The same is true for the SQ lower bound we show in Section 8 (as well as for the SQ lower bound in [DKS17, Bub+19]; the proof is nearly identical). The advantage of the $m > 1$ variant is that the distance between the Gaussian mixture components increases from $\approx 1/\gamma$ (which can be as high as $\approx 1/\sqrt{n}$ if we want our hardness to hold) to $\approx \sqrt{m}/\gamma$ (which can be as high as ≈ 1 by taking $m \approx n$). This is a desirable feature for showing hardness of robust machine learning.

Motivation: Cryptographic applications. Given the wide range of cryptographic applications of LWE [Pei16], it is only natural to expect that CLWE would also be useful for some cryptographic tasks, a question we leave for future work. CLWE's clean and highly symmetric definition should make it a better fit for some applications; its continuous nature, however, might require a discretization step due to efficiency considerations.

Analogy with LWE. As argued above, there are apparently nontrivial differences between CLWE and LWE, especially in terms of possible algorithmic approaches. However, there is undoubtedly also strong similarity between the two. In terms of parameters, the γ parameter in CLWE (density of layers) plays the role of the absolute noise level αq in LWE. And the β parameter in CLWE plays the role of the relative noise parameter α in LWE. Using this correspondence between the parameters, the hardness proved for CLWE in Theorem 1.1 is essentially identical to the one proved for LWE in [Reg05]. The similarity extends even to the noiseless case, where $\alpha = 0$

in LWE and $\beta = 0$ in CLWE. In particular, in Section 6 we present an efficient LLL-based algorithm for solving noiseless CLWE, which is analogous to Gaussian elimination for noiseless LWE.

Comparison with previous work. The CLWE problem is related to the hard problem introduced in the seminal work of Ajtai and Dwork [AD97]. Specifically, both problems involve finding a hidden direction in samples from a continuous distribution. One crucial difference, though, is in the density of the layers. Whereas in our hardness result the separation between the layers can be as large as $\approx 1/\sqrt{n}$, in Ajtai and Dwork the separation is exponentially small. This larger separation in CLWE is more than just a technicality. First, it is the reason we need to employ the quantum machinery from the LWE hardness proof [Reg05]. Second, it is nearly tight, as demonstrated by the algorithm in Section 7. Third, it is necessary for applications such as hardness of learning Gaussian mixtures. Finally, this larger separation is analogous to the main difference between LWE and earlier work [Reg04], and is what leads to the relative efficiency of LWE-based cryptography.

Acknowledgements. We thank Aravindan Vijayaraghavan and Ilias Diakonikolas for useful comments.

1.1 Technical Overview

Broadly speaking, our proof follows the iterative structure of the original LWE hardness proof [Reg05] (in fact, one might say most of the ingredients for CLWE were already present in that 2005 paper!). We also make use of some recent techniques, such as a way to reduce to decision problems directly [PRS17].

In more detail, as in previous work, our main theorem boils down to solving the following problem: we are given a $\mathrm{CLWE}_{\beta,\gamma}$ oracle and polynomially many samples from $D_{L,r}$, the discrete Gaussian distribution on L of width r,[1] and our goal is to solve $\mathrm{BDD}_{L^*,\gamma/r}$, which is the problem of finding the closest vector in the dual lattice L^* given a vector $\boldsymbol{t}$ that is within distance γ/r of L^*. (It is known that $\mathrm{BDD}_{L^*,1/r}$ can be efficiently solved even if all we are given is polynomially many samples from $D_{L,r}$, without any need for an oracle [AR05]; the point here is that the CLWE oracle allows us to extend the decoding radius from $1/r$ to γ/r.) Once this is established, the main theorem follows from previous work [PRS17, Reg05]. Very briefly, the resulting BDD solution is used in a quantum procedure to produce discrete Gaussian samples that are shorter than the ones we started with. This process is then repeated, until eventually we end up with the desired short discrete Gaussian samples. We remark that this process incurs a $\sqrt{n}$ loss in the Gaussian width (Lemma 3.4), and the reason we require $\gamma \geq 2\sqrt{n}$ is to overcome this loss.

We now explain how we solve the above problem. For simplicity, assume for now that we have a *search* CLWE oracle that recovers the secret exactly. (Our actual reduction is stronger and only requires a *decision* CLWE oracle.) Let the given BDD instance be $\boldsymbol{u} + \boldsymbol{w}$, where $\boldsymbol{u} \in L^*$ and $\|\boldsymbol{w}\| = \gamma/r$. We will consider the general case of $\|\boldsymbol{w}\| \leq \gamma/r$ in Section 3. The main idea is to generate CLWE samples whose secret is essentially the desired BDD solution $\boldsymbol{w}$, which would then complete the proof. To begin, take a sample from the discrete Gaussian distribution $\boldsymbol{y} \sim D_{L,r}$ (as provided to us) and consider the inner product

$$\langle \boldsymbol{y}, \boldsymbol{u} + \boldsymbol{w} \rangle = \langle \boldsymbol{y}, \boldsymbol{w} \rangle \pmod 1 ,$$

where the equality holds since $\langle \boldsymbol{y}, \boldsymbol{u} \rangle \in \mathbb{Z}$ by definition. The $(n+1)$-dimensional vector $(\boldsymbol{y}, \langle \boldsymbol{y}, \boldsymbol{w} \rangle \bmod 1)$ is almost a CLWE sample (with parameter γ since $\gamma = r\|\boldsymbol{w}\|$ is the width of $\langle \boldsymbol{y}, \boldsymbol{w} \rangle$) — the only

[1] We actually require samples from D_{L,r_i} for polynomially many r_i's satisfying $r_i \geq r$, see Section 3.

problem is that in CLWE the $\boldsymbol{y}$'s need to be distributed according to a standard Gaussian, but here the $\boldsymbol{y}$'s are distributed according to a *discrete* Gaussian over L. To complete the transformation into bonafide CLWE samples, we add Gaussian noise of appropriate variance to both $\boldsymbol{y}$ and $\langle \boldsymbol{y}, \boldsymbol{w} \rangle$ (and rescale $\boldsymbol{y}$ so that it is distributed according to the standard Gaussian distribution). We then apply the search $\mathrm{CLWE}_{\beta,\gamma}$ oracle on these CLWE samples to recover $\boldsymbol{w}$ and thereby solve $\mathrm{BDD}_{L^*,\gamma/r}$.

As mentioned previously, our main result actually uses a *decision* CLWE oracle, which does not recover the secret $\boldsymbol{w}$ immediately. Working with this decision oracle requires some care. To that end, our proof will incorporate the "oracle hidden center" finding procedure from [PRS17], the details of which can be found in Section 3.3.

2 Preliminaries

Definition 2.1 (Statistical distance)**.** *For two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ over $\mathbb{R}^n$ with density functions ϕ_1 and ϕ_2, respectively, we define the* statistical distance *between them as*

$$\Delta(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2} \int_{\mathbb{R}^n} |\phi_1(\boldsymbol{x}) - \phi_2(\boldsymbol{x})| d\boldsymbol{x} \,.$$

We denote the statistical distance by $\Delta(\phi_1, \phi_2)$ if only the density functions are specified. Moreover, for random variables $X_1 \sim \mathcal{D}_1$ and $X_2 \sim \mathcal{D}_2$, we also denote $\Delta(X_1, X_2) = \Delta(\mathcal{D}_1, \mathcal{D}_2)$. One important fact is that applying (possibly a randomized) function cannot increase statistical distance, i.e., for random variables X, Y and function f,

$$\Delta(f(X), f(Y)) \leq \Delta(X, Y) \,.$$

We define the *advantage* of an algorithm $\mathcal{A}$ solving the decision problem of distinguishing two distributions $\mathcal{D}_n$ and $\mathcal{D}'_n$ parameterized by n as

$$\left| \Pr_{x \sim \mathcal{D}_n} [\mathcal{A}(x) = \mathrm{YES}] - \Pr_{x \sim \mathcal{D}'_n} [\mathcal{A}(x) = \mathrm{YES}] \right| .$$

Moreover, we define the *advantage* of an algorithm $\mathcal{A}$ solving the *average-case* decision problem of distinguishing two distributions $\mathcal{D}_{n,s}$ and $\mathcal{D}'_{n,s}$ parameterized by n and s, where s is equipped with some distribution $\mathcal{S}_n$, as

$$\left| \Pr_{s \sim \mathcal{S}_n} [\mathcal{A}^{\mathcal{B}_{n,s}}(1^n) = \mathrm{YES}] - \Pr_{s \sim \mathcal{S}_n} [\mathcal{A}^{\mathcal{B}'_{n,s}}(1^n) = \mathrm{YES}] \right| ,$$

where $\mathcal{B}_{n,s}$ and $\mathcal{B}_{n,s}$ are respectively the sampling oracles of $\mathcal{D}_{n,s}$ and $\mathcal{D}'_{n,s}$. We say that an algorithm $\mathcal{A}$ has *non-negligible advantage* if its advantage is a non-negligible function in n, i.e., a function in $\Omega(n^{-c})$ for some constant $c > 0$.

2.1 Lattices and Gaussians

Lattices. A *lattice* is a discrete additive subgroup of $\mathbb{R}^n$. Unless specified otherwise, we assume all lattices are full rank, i.e., their linear span is $\mathbb{R}^n$. For an n-dimensional lattice L, a set of linearly independent vectors $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ is called a *basis* of L if L is generated by the set, i.e., $L = B\mathbb{Z}^n$ where $B = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n]$. The *determinant* of a lattice L with basis B is defined as $\det(L) = |\det(B)|$; it is easy to verify that the determinant does not depend on the choice of basis.

The *dual lattice* of a lattice L, denoted by L^*, is defined as

$$L^* = \{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{Z} \text{ for all } \boldsymbol{x} \in L\} .$$

If B is a basis of L then $(B^T)^{-1}$ is a basis of L^*; in particular, $\det(L^*) = \det(L)^{-1}$.

Definition 2.2. *For an n-dimensional lattice L and $1 \le i \le n$, the i-th successive minimum of L is defined as*

$$\lambda_i(L) = \inf\{r \mid \dim(\mathrm{span}(L \cap \overline{B}(\mathbf{0}, r))) \ge i\} ,$$

where $\overline{B}(\mathbf{0}, r)$ is the closed ball of radius r centered at the origin.

We define the function $\rho_s(\boldsymbol{x}) = \exp(-\pi \|\boldsymbol{x}/s\|^2)$. Note that $\rho_s(\boldsymbol{x})/s^n$, where n is the dimension of $\boldsymbol{x}$, is the probability density of the Gaussian distribution with covariance $s^2/(2\pi) \cdot I_n$.

Definition 2.3 (Discrete Gaussian)**.** *For lattice $L \subset \mathbb{R}^n$, vector $\boldsymbol{y} \in \mathbb{R}^n$, and parameter $r > 0$, the* discrete Gaussian distribution $D_{\boldsymbol{y}+L,r}$ *on coset $\boldsymbol{y} + L$ with width r is defined to have support $\boldsymbol{y} + L$ and probability mass function proportional to ρ_r.*

For $\boldsymbol{y} = \mathbf{0}$, we simply denote the discrete Gaussian distribution on lattice L with width r by $D_{L,r}$. Abusing notation, we denote the n-dimensional *continuous Gaussian distribution* with zero mean and isotropic variance $r^2/(2\pi)$ as $D_{\mathbb{R}^n,r}$. Finally, we omit the subscript r when $r = 1$ and refer to $D_{\mathbb{R}^n}$ as the *standard* Gaussian (despite it having covariance $I_n/(2\pi)$).

Claim 2.4 ([Pei10, Fact 2.1])**.** *For any $r_1, r_2 > 0$ and vectors $\boldsymbol{x}, \boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathbb{R}^n$, let $r_0 = \sqrt{r_1^2 + r_2^2}$, $r_3 = r_1 r_2 / r_0$, and $\boldsymbol{c}_3 = (r_3/r_1)^2 \boldsymbol{c}_1 + (r_3/r_2)^2 \boldsymbol{c}_2$. Then*

$$\rho_{r_1}(\boldsymbol{x} - \boldsymbol{c}_1) \cdot \rho_{r_2}(\boldsymbol{x} - \boldsymbol{c}_2) = \rho_{r_0}(\boldsymbol{c}_1 - \boldsymbol{c}_2) \cdot \rho_{r_3}(\boldsymbol{x} - \boldsymbol{c}_3) .$$

Fourier analysis. We briefly review basic tools of Fourier analysis required later on. The Fourier transform of a function $f : \mathbb{R}^n \to \mathbb{C}$ is defined to be

$$\hat{f}(\boldsymbol{w}) = \int_{\mathbb{R}^n} f(\boldsymbol{x}) e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{w} \rangle} d\boldsymbol{x} .$$

An elementary property of the Fourier transform is that if $f(\boldsymbol{w}) = g(\boldsymbol{w} + \boldsymbol{v})$ for some $\boldsymbol{v} \in \mathbb{R}^n$, then $\hat{f}(\boldsymbol{w}) = e^{2\pi i \langle \boldsymbol{v}, \boldsymbol{w} \rangle} \hat{g}(\boldsymbol{w})$. Another important fact is that the Fourier transform of a Gaussian is also a Gaussian, i.e., $\hat{\rho} = \rho$; more generally, $\hat{\rho}_s = s^n \rho_{1/s}$. We also exploit the Poisson summation formula stated below. Note that we denote by $f(A) = \sum_{\boldsymbol{x} \in A} f(\boldsymbol{x})$ for any function f and any discrete set A.

Lemma 2.5 (Poisson summation formula)**.** *For any lattice L and any function f,*[2]

$$f(L) = \det(L^*) \cdot \hat{f}(L^*) .$$

[2] To be precise, f needs to satisfy some niceness conditions; this will always hold in our applications.

Smoothing parameter. An important lattice parameter induced by discrete Gaussian which will repeatedly appear in our work is the *smoothing parameter*, defined as follows.

Definition 2.6 (Smoothing parameter)**.** *For lattice L and real $\varepsilon > 0$, we define the* smoothing parameter $\eta_\varepsilon(L)$ *as*

$$\eta_\varepsilon(L) = \inf\{s \mid \rho_{1/s}(L^* \setminus \{\mathbf{0}\}) \leq \varepsilon\} .$$

Intuitively, this parameter is the width beyond which the discrete Gaussian distribution behaves like a continuous Gaussian. This is formalized in the lemmas below.

Lemma 2.7 ([Reg05, Claim 3.9])**.** *For any n-dimensional lattice L, vector $\boldsymbol{u} \in \mathbb{R}^n$, and $r, s > 0$ satisfying $rs/t \geq \eta_\varepsilon(L)$ for some $\varepsilon < \frac{1}{2}$, where $t = \sqrt{r^2 + s^2}$, the statistical distance between $D_{\boldsymbol{u}+L,r} + D_{\mathbb{R}^n,s}$ and $D_{\mathbb{R}^n,t}$ is at most 4ε.*

Lemma 2.8 ([PRS17, Lemma 2.5])**.** *For any n-dimensional lattice L, real $\varepsilon > 0$, and $r \geq \eta_\varepsilon(L)$, the statistical distance between $D_{\mathbb{R}^n,r} \bmod L$ and the uniform distribution over $\mathbb{R}^n/L$ is at most $\varepsilon/2$.*

Lemma 2.7 states that if we take a sample from $D_{L,r}$ and add continuous Gaussian noise $D_{\mathbb{R}^n,s}$ to the sample, the resulting distribution is statistically close to $D_{\mathbb{R}^n,\sqrt{r^2+s^2}}$, which is precisely what one gets by adding two continuous Gaussian distributions of width r and s. Unless specified otherwise, we always assume ε is negligibly small in n, say $\varepsilon = \exp(-n)$. The following are some useful upper and lower bounds on the smoothing parameter $\eta_\varepsilon(L)$.

Lemma 2.9 ([PRS17, Lemma 2.6])**.** *For any n-dimensional lattice L and $\varepsilon = \exp(-c^2 n)$,*

$$\eta_\varepsilon(L) \leq c\sqrt{n}/\lambda_1(L^*) .$$

Lemma 2.10 ([MR07, Lemma 3.3])**.** *For any n-dimensional lattice L and $\varepsilon > 0$,*

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

Lemma 2.11 ([Reg05, Claim 2.13])**.** *For any n-dimensional lattice L and $\varepsilon > 0$,*

$$\eta_\varepsilon(L) \geq \sqrt{\frac{\ln 1/\varepsilon}{\pi}} \cdot \frac{1}{\lambda_1(L^*)} .$$

Computational problems. GapSVP and SIVP are among the main computational problems on lattices and are believed to be computationally hard (even with quantum computation) for polynomial approximation factor $\alpha(n)$. We also define two additional problems, DGS and BDD.

Definition 2.12 (GapSVP)**.** *For an approximation factor $\alpha = \alpha(n)$, an instance of* GapSVP_α *is given by an n-dimensional lattice L and a number $d > 0$. In* YES *instances, $\lambda_1(L) \leq d$, whereas in* NO *instances, $\lambda_1(L) > \alpha \cdot d$.*

Definition 2.13 (SIVP)**.** *For an approximation factor $\alpha = \alpha(n)$, an instance of* SIVP_α *is given by an n-dimensional lattice L. The goal is to output a set of n linearly independent lattice vectors of length at most $\alpha \cdot \lambda_n(L)$.*

Definition 2.14 (DGS)**.** *For a function φ that maps lattices to non-negative reals, an instance of* DGS_φ *is given by a lattice L and a parameter $r \geq \varphi(L)$. The goal is to output an independent sample whose distribution is within negligible statistical distance of $D_{L,r}$.*

Definition 2.15 (BDD)**.** *For an n-dimensional lattice L and distance bound $d > 0$, an instance of* $\mathrm{BDD}_{L,d}$ *is given by a vector $\boldsymbol{t} = \boldsymbol{w} + \boldsymbol{u}$, where $\boldsymbol{u} \in L$ and $\|\boldsymbol{w}\| \leq d$. The goal is to output $\boldsymbol{w}$.*

2.2 Learning with errors

We now define the learning with errors (LWE) problem. This definition will not be used in the sequel, and is included for completeness. Let n and q be positive integers, and $\alpha > 0$ an error rate. We denote the quotient ring of integers modulo q as $\mathbb{Z}_q = \mathbb{Z}/q\mathbb{Z}$ and quotient group of reals modulo the integers as $\mathbb{T} = \mathbb{R}/\mathbb{Z} = [0, 1)$.

Definition 2.16 (LWE distribution)**.** *For integer $q \geq 2$ and vector $\boldsymbol{s} \in \mathbb{Z}_q^n$, the* LWE distribution $A_{\boldsymbol{s},\alpha}$ *over $\mathbb{Z}_q^n \times \mathbb{T}$ is sampled by independently choosing uniformly random $\boldsymbol{a} \in \mathbb{Z}_q^n$ and $e \sim D_{\mathbb{R},\alpha}$, and outputting $(\boldsymbol{a}, (\langle \boldsymbol{a}, \boldsymbol{s} \rangle / q + e) \bmod 1)$.*

Definition 2.17. *For an integer $q = q(n) \geq 2$ and error parameter $\alpha = \alpha(n) > 0$, the average-case decision problem* $\mathrm{LWE}_{q,\alpha}$ *is to distinguish the following two distributions over $\mathbb{Z}_q^n \times \mathbb{T}$: (1) the LWE distribution $A_{\boldsymbol{s},\alpha}$ for some uniformly random $\boldsymbol{s} \in \mathbb{Z}_q^n$ (which is fixed for all samples), or (2) the uniform distribution.*

2.3 Continuous learning with errors

We now define the CLWE distribution, which is the central subject of our analysis.

Definition 2.18 (CLWE distribution)**.** *For unit vector $\boldsymbol{w} \in \mathbb{R}^n$ and parameters $\beta, \gamma > 0$, define the* CLWE distribution $A_{\boldsymbol{w},\beta,\gamma}$ *over $\mathbb{R}^{n+1}$ to have density at $(\boldsymbol{y}, z)$ proportional to*

$$\rho(\boldsymbol{y}) \cdot \sum_{k \in \mathbb{Z}} \rho_\beta(z + k - \gamma \langle \boldsymbol{y}, \boldsymbol{w} \rangle) \, .$$

Equivalently, a sample $(\boldsymbol{y}, z)$ from the CLWE distribution $A_{\boldsymbol{w},\beta,\gamma}$ is given by the $(n+1)$-dimensional vector $(\boldsymbol{y}, z)$ where $\boldsymbol{y} \sim D_{\mathbb{R}^n}$ and $z = (\gamma \langle \boldsymbol{y}, \boldsymbol{w} \rangle + e) \bmod 1$ where $e \sim D_{\mathbb{R},\beta}$. The vector $\boldsymbol{w}$ is the hidden direction, γ is the density of layers, and β is the noise added to each equation. From the CLWE distribution, we can arrive at the homogeneous CLWE distribution by conditioning on $z = 0$. A formal definition is given as follows.

Definition 2.19 (Homogeneous CLWE distribution)**.** *For unit vector $\boldsymbol{w} \in \mathbb{R}^n$ and parameters $\beta, \gamma > 0$, define the* homogeneous CLWE distribution $H_{\boldsymbol{w},\beta,\gamma}$ *over $\mathbb{R}^n$ to have density at $\boldsymbol{y}$ proportional to*

$$\rho(\boldsymbol{y}) \cdot \sum_{k \in \mathbb{Z}} \rho_\beta(k - \gamma \langle \boldsymbol{y}, \boldsymbol{w} \rangle) \, . \tag{1}$$

The homogeneous CLWE distribution can be equivalently defined as a mixture of Gaussians. To see this, notice that Eq. (1) is equal to

$$\sum_{k \in \mathbb{Z}} \rho_{\sqrt{\beta^2+\gamma^2}}(k) \cdot \rho(\pi_{\boldsymbol{w}^\perp}(\boldsymbol{y})) \cdot \rho_{\beta/\sqrt{\beta^2+\gamma^2}}\left(\langle \boldsymbol{y}, \boldsymbol{w} \rangle - \frac{\gamma}{\beta^2+\gamma^2} k \right) , \tag{2}$$

where $\pi_{\boldsymbol{w}^\perp}$ denotes the projection on the orthogonal space to $\boldsymbol{w}$. Hence, $H_{\boldsymbol{w},\beta,\gamma}$ can be viewed as a mixture of Gaussian components of width $\beta/\sqrt{\beta^2+\gamma^2}$ (which is roughly β/γ for $\beta \ll \gamma$) in the secret direction, and width 1 in the orthogonal space. The components are equally spaced, with a separation of $\gamma/(\beta^2+\gamma^2)$ between them (which is roughly $1/\gamma$ for $\beta \ll \gamma$).

We remark that the integral of (1) (or equivalently, of (2)) over all $\boldsymbol{y}$ is

$$Z = \frac{\beta}{\sqrt{\beta^2+\gamma^2}} \cdot \rho\left(\frac{1}{\sqrt{\beta^2+\gamma^2}} \mathbb{Z} \right) . \tag{3}$$

This is easy to see since the integral over $\boldsymbol{y}$ of the product of the last two p terms in (2) is $\beta/\sqrt{\beta^2+\gamma^2}$ independently of k.

Definition 2.20. *For parameters $\beta, \gamma > 0$, the average-case decision problem* $\mathrm{CLWE}_{\beta,\gamma}$ *is to distinguish the following two distributions over $\mathbb{R}^n \times \mathbb{T}$: (1) the CLWE distribution $A_{\boldsymbol{w},\beta,\gamma}$ for some uniformly random unit vector $\boldsymbol{w} \in \mathbb{R}^n$ (which is fixed for all samples), or (2) $D_{\mathbb{R}^n} \times U$.*

Definition 2.21. *For parameters $\beta, \gamma > 0$, the average-case decision problem* $\mathrm{hCLWE}_{\beta,\gamma}$ *is to distinguish the following two distributions over $\mathbb{R}^n$: (1) the homogeneous CLWE distribution $H_{\boldsymbol{w},\beta,\gamma}$ for some uniformly random unit vector $\boldsymbol{w} \in \mathbb{R}^n$ (which is fixed for all samples), or (2) $D_{\mathbb{R}^n}$.*

Note that $\mathrm{CLWE}_{\beta,\gamma}$ and $\mathrm{hCLWE}_{\beta,\gamma}$ are defined as average-case problems. We could have equally well defined them to be worst-case problems, requiring the algorithm to distinguish the distributions for *all* hidden directions $\boldsymbol{w} \in \mathbb{R}^n$. The following claim shows that the two formulations are equivalent.

Claim 2.22. *For any $\beta, \gamma > 0$, there is a polynomial-time reduction from worst-case* $\mathrm{CLWE}_{\beta,\gamma}$ *to (average-case)* $\mathrm{CLWE}_{\beta,\gamma}$.

Proof. Given CLWE samples $\{(\boldsymbol{y}_i, z_i)\}_{i=1}^{\mathrm{poly}(n)}$ from $A_{\boldsymbol{w},\beta,\gamma}$, we apply a random rotation $\boldsymbol{R}$, giving us samples of the form $\{(\boldsymbol{R}\boldsymbol{y}_i, z_i\}_{i=1}^{\mathrm{poly}(n)}$. Since the standard Gaussian is rotationally invariant and $\langle \boldsymbol{y}, \boldsymbol{w} \rangle = \langle \boldsymbol{R}\boldsymbol{y}, \boldsymbol{R}^T\boldsymbol{w} \rangle$, the rotated CLWE samples are distributed according to $A_{\boldsymbol{R}^T\boldsymbol{w},\beta,\gamma}$. Since $\boldsymbol{R}$ is a random rotation, the random direction $\boldsymbol{R}^T\boldsymbol{w}$ is uniformly distributed on the sphere. □

3 Hardness of CLWE

3.1 Background and overview

In this section, we give an overview of the quantum reduction from worst-case lattice problems to CLWE. Our goal is to show that we can efficiently solve worst-case lattice problems, in particular GapSVP and SIVP, using an oracle for CLWE (and with quantum computation). We first state our main theorem, which was stated informally as Theorem 1.1 in the introduction.

Theorem 3.1. *Let $\beta = \beta(n) \in (0,1)$ and $\gamma = \gamma(n) \geq 2\sqrt{n}$ be such that γ/β is polynomially bounded. Then there is a polynomial-time quantum reduction from* $\mathrm{DGS}_{2\sqrt{n}\eta_\varepsilon(L)/\beta}$ *to* $\mathrm{CLWE}_{\beta,\gamma}$.

Using standard reductions from GapSVP and SIVP to DGS (see, e.g., [Reg05, Section 3.3]), our main theorem immediately implies the following corollary.

Corollary 3.2. *Let $\beta = \beta(n) \in (0,1)$ and $\gamma = \gamma(n) \geq 2\sqrt{n}$ such that γ/β is polynomially bounded. Then, there is a polynomial-time quantum reduction from* SIVP_α *and* GapSVP_α *to* $\mathrm{CLWE}_{\beta,\gamma}$ *for some $\alpha = \tilde{O}(n/\beta)$.*

Based on previous work, to prove Theorem 3.1, it suffices to prove the following lemma, which is the goal of this section.

Lemma 3.3. *Let $\beta = \beta(n) \in (0,1)$ and $\gamma = \gamma(n) \geq 2\sqrt{n}$ such that $q = \gamma/\beta$ is polynomially bounded. There exists a probabilistic polynomial-time (classical) algorithm with access to an oracle that solves* $\mathrm{CLWE}_{\beta,\gamma}$*, that takes as input a lattice $L \subset \mathbb{R}^n$, parameters β, γ, and $r \geq 2q \cdot \eta_\varepsilon(L)$, and* $\mathrm{poly}(n)$ *many samples from the discrete Gaussian distribution D_{L,r_i} for* $\mathrm{poly}(n)$ *parameters $r_i \geq r$ and solves* $\mathrm{BDD}_{L^*,d}$ *for $d = \gamma/(\sqrt{2}r)$.*

In other words, we can implement an oracle for $\mathrm{BDD}_{L^*,\gamma/(\sqrt{2}r)}$ using polynomially many discrete Gaussian samples and the CLWE oracle as a sub-routine. The proof of Lemma 3.3 will be given in Section 3.2 (which is the novel contribution) and Section 3.3 (which mainly follows [PRS17]).

In the rest of this subsection, we briefly explain how Theorem 3.1 follows from Lemma 3.3. This derivation is already implicit in past work [PRS17, Reg05], and is included here mainly for completeness. Readers familiar with the reduction may skip directly to Section 3.2.

The basic idea is to start with samples from a very wide discrete Gaussian (which can be efficiently sampled) and then iteratively sample from narrower discrete Gaussians, until eventually we end up with short discrete Gaussian samples, as required (see Figure 3). Each iteration consists of two steps: the first classical step is given by Lemma 3.3, allowing us to solve BDD on the dual lattice; the second step is quantum and is given in Lemma 3.4 below, which shows that solving BDD leads to sampling from narrower discrete Gaussian.

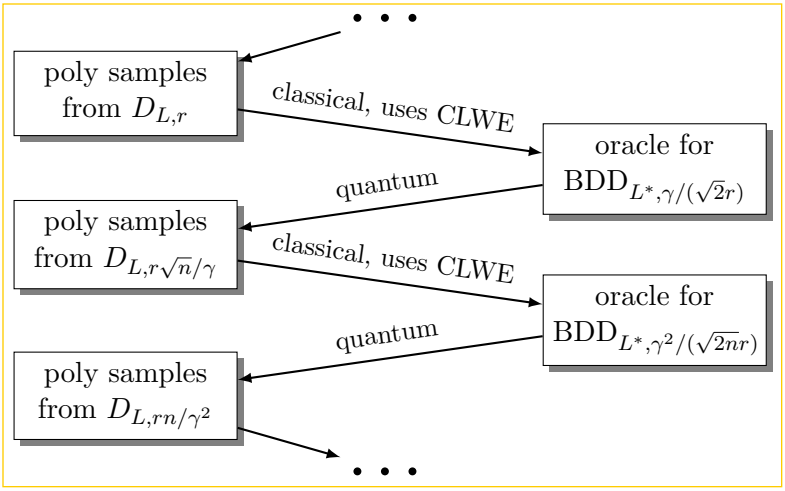

Figure 3: Two iterations of the reduction.

Lemma 3.4 ([Reg05, Lemma 3.14]). *There exists an efficient quantum algorithm that, given any n-dimensional lattice L, a number $d < \lambda_1(L^*)/2$, and an oracle that solves $\mathrm{BDD}_{L^*,d}$, outputs a sample from $D_{L,\sqrt{n}/(\sqrt{2}d)}$.*

Similar to [PRS17], there is a subtle requirement in Lemma 3.3 that we need discrete Gaussian samples from several different parameters $r' > r$. However, this is a non-issue since an oracle for $\mathrm{BDD}_{L^*,\gamma/(\sqrt{2}r)}$ also solves $\mathrm{BDD}_{L^*,\gamma/(\sqrt{2}r')}$ for any $r' \geq r$, so Lemma 3.4 in fact allows us to efficiently sample from $D_{L,r'\sqrt{n}/\gamma}$ for any $r' \geq r$.

3.2 CLWE samples from BDD

In this subsection we prove Lemma 3.5, showing how to generate CLWE samples from the given BDD instance using discrete Gaussian samples. In the next subsection we will show how to solve the BDD instance by applying the decision CLWE oracle to these samples, thereby completing the proof of Lemma 3.3.

Lemma 3.5. *There is an efficient algorithm that takes as input an m-dimensional lattice L, a vector $\boldsymbol{w}+\boldsymbol{u}$ where $\boldsymbol{u} \in L^*$, reals $r, s_1, s_2 > 0$ such that $rs_1/\sqrt{\|\boldsymbol{w}\|^2(rs_1/s_2)^2 + t^2} \geq \eta_\varepsilon(L)$ for some $\varepsilon < \frac{1}{2}$ and $t = \sqrt{r^2+s_1^2}$, and samples from $D_{L,r}$, and outputs samples that are within statistical*

distance 8ε *of the CLWE distribution* $A_{\boldsymbol{w}',\beta,\gamma}$ *for* $\boldsymbol{w}' = \boldsymbol{w}/\|\boldsymbol{w}\|$, $\beta = \|\boldsymbol{w}\|\sqrt{(rs_1/t)^2 + (s_2/\|\boldsymbol{w}\|)^2}$ *and* $\gamma = \|\boldsymbol{w}\|r^2/t$.

Proof. We start by describing the algorithm. For each $\boldsymbol{x}$ from the given samples from $D_{L,r}$, do the following. First, take the inner product $\langle \boldsymbol{x}, \boldsymbol{w} + \boldsymbol{u} \rangle$, which gives us

$$\langle \boldsymbol{x}, \boldsymbol{w} + \boldsymbol{u} \rangle = \langle \boldsymbol{x}, \boldsymbol{w} \rangle \bmod 1 \ .$$

Appending this inner product modulo 1 to the sample $\boldsymbol{x}$, we get $(\boldsymbol{x}, \langle \boldsymbol{x}, \boldsymbol{w} \rangle \bmod 1)$. Next, we "smooth out" the lattice structure of $\boldsymbol{x}$ by adding Gaussian noise $\boldsymbol{v} \sim D_{\mathbb{R}^n,s_1}$ to $\boldsymbol{x}$ and $e \sim D_{\mathbb{R},s_2}$ to $\langle \boldsymbol{x}, \boldsymbol{w} \rangle$ (modulo 1). Then, we have

$$(\boldsymbol{x} + \boldsymbol{v}, (\langle \boldsymbol{x}, \boldsymbol{w} \rangle + e) \bmod 1) \ . \tag{4}$$

Finally, we normalize the first component by t so that its marginal distribution has unit width, giving us

$$((\boldsymbol{x} + \boldsymbol{v})/t, (\langle \boldsymbol{x}, \boldsymbol{w} \rangle + e) \bmod 1) \ , \tag{5}$$

which the algorithm outputs.

Our goal is to show that the distribution of (5) is within statistical distance 8ε of the CLWE distribution $A_{\boldsymbol{w}',\beta,\gamma}$, given by

$$(\boldsymbol{y}', (\gamma \langle \boldsymbol{y}', \boldsymbol{w}' \rangle + e') \bmod 1) \ ,$$

where $\boldsymbol{y}' \sim D_{\mathbb{R}^n}$ and $e' \sim D_{\mathbb{R},\beta}$. Because applying a function cannot increase statistical distance (specifically, dividing the first component by t and taking mod 1 of the second), it suffices to show that the distribution of

$$(\boldsymbol{x} + \boldsymbol{v}, \langle \boldsymbol{x}, \boldsymbol{w} \rangle + e) \ , \tag{6}$$

is within statistical distance 8ε of that of

$$(\boldsymbol{y}, (r/t)^2 \langle \boldsymbol{y}, \boldsymbol{w} \rangle + e') \ , \tag{7}$$

where $\boldsymbol{y} \sim D_{\mathbb{R}^n,t}$ and $e' \sim D_{\mathbb{R},\beta}$. First, observe that by Lemma 2.7, the statistical distance between the marginals on the first component (i.e., between $\boldsymbol{x} + \boldsymbol{v}$ and $\boldsymbol{y}$) is at most 4ε. It is therefore sufficient to bound the statistical distance between the second components conditioned on any fixed value $\overline{\boldsymbol{y}}$ of the first component. Conditioned on the first component being $\overline{\boldsymbol{y}}$, the second component in (6) has the same distribution as

$$\langle \boldsymbol{x} + \boldsymbol{h}, \boldsymbol{w} \rangle \tag{8}$$

where $\boldsymbol{h} \sim D_{\mathbb{R}^n,s_2/\|\boldsymbol{w}\|}$, and the second component in (7) has the same distribution as

$$\langle (r/t)^2 \overline{\boldsymbol{y}} + \boldsymbol{h}', \boldsymbol{w} \rangle \tag{9}$$

where $\boldsymbol{h}' \sim D_{\mathbb{R}^n,\beta/\|\boldsymbol{w}\|}$.

By Claim 3.6 below, conditioned on $\boldsymbol{x} + \boldsymbol{v} = \overline{\boldsymbol{y}}$, the distribution of $\boldsymbol{x}$ is $(r/t)^2 \overline{\boldsymbol{y}} + D_{L-(r/t)^2 \overline{\boldsymbol{y}},rs_1/t}$. Therefore, by Lemma 2.7, the conditional distribution of $\boldsymbol{x} + \boldsymbol{h}$ given $\boldsymbol{x} + \boldsymbol{v} = \boldsymbol{y}$ is within statistical distance 4ε of that of $(r/t)^2 \overline{\boldsymbol{y}} + \boldsymbol{h}'$. Since statistical distance cannot increase by applying a function (inner product with $\boldsymbol{w}$ in this case), (8) is within statistical distance 4ε of (9). Hence, the distribution of (6) is within statistical distance 8ε of that of (7). □

Claim 3.6. *Let $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{x} \sim D_{L,r}$ and $\boldsymbol{v} \sim D_{\mathbb{R}^n,s}$. Then, the conditional distribution of $\boldsymbol{x}$ given $\boldsymbol{y} = \overline{\boldsymbol{y}}$ is $(r/t)^2\overline{\boldsymbol{y}} + D_{L-(r/t)^2\overline{\boldsymbol{y}},rs/t}$ where $t = \sqrt{r^2+s^2}$.*

Proof. Observe that $\boldsymbol{x}$ conditioned on $\boldsymbol{y} = \overline{\boldsymbol{y}}$ is a discrete random variable supported on L. The probability of $\boldsymbol{x}$ given $\boldsymbol{y} = \overline{\boldsymbol{y}}$ is proportional to

$$\rho_r(\boldsymbol{x}) \cdot \rho_s(\overline{\boldsymbol{y}} - \boldsymbol{x}) = \rho_t(\overline{\boldsymbol{y}}) \cdot \rho_{rs/t}(\boldsymbol{x} - (r/t)^2\overline{\boldsymbol{y}}) \propto \rho_{rs/t}(\boldsymbol{x} - (r/t)^2\overline{\boldsymbol{y}}) \ ,$$

where the equality follows from Claim 2.4. Hence, the conditional distribution of $\boldsymbol{x} - (r/t)^2\boldsymbol{y}$ given $\boldsymbol{y} = \overline{\boldsymbol{y}}$ is $D_{L-(r/t)^2\overline{\boldsymbol{y}},rs/t}$. □

3.3 Solving BDD with the CLWE oracle

In this subsection, we complete the proof of Lemma 3.3. We first give some necessary background on the Oracle Hidden Center Problem (OHCP) [PRS17]. The problem asks one to search for a "hidden center" $\boldsymbol{w}^*$ using a decision oracle whose acceptance probability depends only on the distance to $\boldsymbol{w}^*$. The problem's precise statement is as follows.

Definition 3.7 (OHCP)**.** *For parameters $\varepsilon, \delta \in [0,1)$ and $\zeta \geq 1$, the $(\varepsilon,\delta,\zeta)$-*OHCP *is an approximate search problem that tries to find the "hidden" center $\boldsymbol{w}^*$. Given a scale parameter $d > 0$ and access to a randomized oracle $\mathcal{O} : \mathbb{R}^n \times \mathbb{R}^{\geq 0} \to \{0,1\}$ such that its acceptance probability $p(\boldsymbol{w},t)$ only depends on $\exp(t)\|\boldsymbol{w}-\boldsymbol{w}^*\|$ for some (unknown) "hidden center" $\boldsymbol{w}^* \in \mathbb{R}^n$ with $\delta d \leq \|\boldsymbol{w}^*\| \leq d$ and for any $\boldsymbol{w} \in \mathbb{R}^n$ with $\|\boldsymbol{w} - \boldsymbol{w}^*\| \leq \zeta d$, the goal is to output $\boldsymbol{w}$ s.t. $\|\boldsymbol{w} - \boldsymbol{w}^*\| \leq \varepsilon d$.*

Notice that OHCP corresponds to our problem since we want to solve BDD, which is equivalent to finding the "hidden" offset vector $\boldsymbol{w}^*$, using a decision oracle for $\mathrm{CLWE}_{\beta,\gamma}$. The acceptance probability of the $\mathrm{CLWE}_{\beta,\gamma}$ oracle will depend on the distance between our guess $\boldsymbol{w}$ and the true offset $\boldsymbol{w}^*$. For OHCP, we have the following result from [PRS17].

Lemma 3.8 ([PRS17], Proposition 4.4)**.** *There is a poly(κ, n)-time algorithm that takes as input a confidence parameter $\kappa \geq 20\log(n+1)$ (and the scale parameter $d > 0$) and solves $(\exp(-\kappa), \exp(-\kappa), 1+1/\kappa)$-OHCP in dimension n except with probability $\exp(-\kappa)$, provided that the oracle $\mathcal{O}$ corresponding to the OHCP instance satisfies the following conditions. For some $p(\infty) \in [0,1]$ and $t^* \geq 0$,*

1. *$p(\boldsymbol{0}, t^*) - p(\infty) \geq 1/\kappa$;*
2. *$|p(\boldsymbol{0}, t) - p(\infty)| \leq 2\exp(-t/\kappa)$ for any $t \geq 0$; and*
3. *$p(\boldsymbol{w}, t)$ is κ-Lipschitz in t for any $\boldsymbol{w} \in \mathbb{R}^n$ such that $\|\boldsymbol{w}\| \leq (1+1/\kappa)d$.*

Furthermore, each of the algorithm's oracle calls takes the form $\mathcal{O}(\cdot, i\Delta)$ for some $\Delta < 1$ that depends only on κ and n and $0 \leq i \leq \mathrm{poly}(\kappa, n)$.

The main idea in the proof of Lemma 3.8 is performing a guided random walk with advice from the decision oracle $\mathcal{O}$. The decision oracle $\mathcal{O}$ rejects a random step with high probability if it increases the distance $\|\boldsymbol{w} - \boldsymbol{w}^*\|$. Moreover, there is non-negligible probability of decreasing the distance by a factor $\exp(1/n)$ unless $\log\|\boldsymbol{w} - \boldsymbol{w}^*\| \leq -\kappa$. Hence, with sufficiently many steps, the random walk will reach $\widehat{\boldsymbol{w}}$, a guess of the hidden center, which is within $\exp(-\kappa)$ distance to $\boldsymbol{w}^*$ with high probability.

Our goal is to show that we can construct an oracle $\mathcal{O}$ satisfying the above conditions using an oracle for $\mathrm{CLWE}_{\beta,\gamma}$. Then, it follows from Lemma 3.8 that BDD with discrete Gaussian samples can be solved using an oracle for CLWE. We first state some lemmas useful for our proof. Lemma 3.9 is Babai's closest plane algorithm and Lemma 3.10 is an upper bound on the statistical distance between two one-dimensional Gaussian distributions.

Lemma 3.9 ([LLL82, Bab86]). *For any n-dimensional lattice L, there is an efficient algorithm that solves $\mathrm{BDD}_{L,d}$ for $d = 2^{-n/2} \cdot \lambda_1(L)$.*

Lemma 3.10 ([DMR18, Theorem 1.3]). *For all $\mu_1, \mu_2 \in \mathbb{R}$, and $\sigma_1, \sigma_2 > 0$, we have*

$$\Delta\big(\mathcal{N}(\mu_1, \sigma_1), \mathcal{N}(\mu_2, \sigma_2)\big) \leq \frac{3|\sigma_1^2 - \sigma_2^2|}{2\max(\sigma_1^2, \sigma_2^2)} + \frac{|\mu_1 - \mu_2|}{2\max(\sigma_1, \sigma_2)} ,$$

where $\mathcal{N}(\mu, \sigma)$ denotes the Gaussian distribution with mean μ and standard deviation σ.

Now, we prove Lemma 3.3, restated below.

Lemma 3.3. *Let $\beta = \beta(n) \in (0, 1)$ and $\gamma = \gamma(n) \geq 2\sqrt{n}$ such that $q = \gamma/\beta$ is polynomially bounded. There exists a probabilistic polynomial-time (classical) algorithm with access to an oracle that solves $\mathrm{CLWE}_{\beta,\gamma}$, that takes as input a lattice $L \subset \mathbb{R}^n$, parameters β, γ, and $r \geq 2q \cdot \eta_\varepsilon(L)$, and $\mathrm{poly}(n)$ many samples from the discrete Gaussian distribution D_{L,r_i} for $\mathrm{poly}(n)$ parameters $r_i \geq r$ and solves $\mathrm{BDD}_{L^*,d}$ for $d = \gamma/(\sqrt{2}r)$.*

Proof. Let $d' = (1 - 1/(2n)) \cdot d$. By [LM09, Corollary 2], it suffices to solve $\mathrm{BDD}_{L^*,d'}$. Let $\kappa = \mathrm{poly}(n)$ with $\kappa \geq 8qn\ell$ be such that the advantage of our $\mathrm{CLWE}_{\beta,\gamma}$ oracle is at least $1/\kappa$, where $\ell \geq 1$ is the number of samples required by the oracle.

Given as input a lattice $L \subset \mathbb{R}^n$, a parameter $r \geq 2q \cdot \eta_\varepsilon(L)$, samples from D_{L,r_i} for $1 \leq i \leq \mathrm{poly}(n)$, and a BDD instance $\boldsymbol{w}^* + \boldsymbol{u}$ where $\boldsymbol{u} \in L^*$ and $\|\boldsymbol{w}^*\| \leq d'$, we want to recover $\boldsymbol{w}^*$. Without loss of generality, we can assume that $\|\boldsymbol{w}^*\| \geq \exp(-n/2) \cdot \lambda_1(L^*) \geq (2q/r) \cdot \exp(-n/2)$ (Lemma 2.11), since we can otherwise find $\boldsymbol{w}^*$ efficiently using Babai's closest plane algorithm (Lemma 3.9).

We will use the CLWE oracle to simulate an oracle $\mathcal{O} : \mathbb{R}^n \times \mathbb{R}^{\geq 0} \to \{0, 1\}$ such that the probability that $\mathcal{O}(\boldsymbol{w}, t)$ outputs 1 ("accepts") only depends on $\exp(t)\|\boldsymbol{w} - \boldsymbol{w}^*\|$. Our oracle $\mathcal{O}$ corresponds to the oracle in Definition 3.7 with $\boldsymbol{w}^*$ as the "hidden center". We will use Lemma 3.8 to find $\boldsymbol{w}^*$.

On input $(\boldsymbol{w}, t)$, our oracle $\mathcal{O}$ receives ℓ independent samples from $D_{L,\exp(t)r}$. Then, we generate CLWE samples using the procedure from Lemma 3.5. The procedure takes as input these ℓ samples, the vector $\boldsymbol{u} + \boldsymbol{w}^* - \boldsymbol{w}$ where $\boldsymbol{u} \in L^*$, and parameters $\exp(t)r, \exp(t)s_1, s_2$. Our choice of s_1 and s_2 will be specified below. Note that the CLWE oracle requires the "hidden direction" $(\boldsymbol{w} - \boldsymbol{w}^*)/\|\boldsymbol{w} - \boldsymbol{w}^*\|$ to be uniformly distributed on the unit sphere. To this end, we apply the worst-to-average case reduction from Claim 2.22. Let $S_{\boldsymbol{w},t}$ be the resulting CLWE distribution. Our oracle $\mathcal{O}$ then calls the $\mathrm{CLWE}_{\beta,\gamma}$ oracle on $S_{\boldsymbol{w},t}^\ell$ and outputs 1 if and only if it accepts.

Using the oracle $\mathcal{O}$, we can run the procedure from Lemma 3.8 with confidence parameter κ and scale parameter d'. The output of this procedure will be some approximation $\widehat{\boldsymbol{w}}$ to the oracle's "hidden center" with the guarantee that $\|\widehat{\boldsymbol{w}} - \boldsymbol{w}^*\| \leq \exp(-\kappa)d'$. Finally, running Babai's algorithm on the vector $\boldsymbol{u} + \boldsymbol{w}^* - \widehat{\boldsymbol{w}}$ will give us $\boldsymbol{w}^*$ exactly since

$$\|\widehat{\boldsymbol{w}} - \boldsymbol{w}^*\| \leq \exp(-\kappa)d \leq \beta \exp(-\kappa)/\eta_\varepsilon(L) \leq 2^{-n}\lambda_1(L^*) ,$$

where the last inequality is from Lemma 2.9.

The running time of the above procedure is clearly polynomial in n. It remains to check that our oracle $\mathcal{O}$ (1) is a valid instance of $(\exp(-\kappa), \exp(-\kappa), 1 + 1/\kappa)$-OHCP with hidden center $\boldsymbol{w}^*$ and (2) satisfies all the conditions of Lemma 3.8. First, note that $S_{\boldsymbol{w},t}$ will be negligibly close in statistical distance to the CLWE distribution with parameters

$$\begin{aligned} \beta' &= \sqrt{(\exp(t)\|\boldsymbol{w} - \boldsymbol{w}^*\|)^2 s_1'^2 + s_2^2} , \\ \gamma' &= \exp(t)\|\boldsymbol{w} - \boldsymbol{w}^*\| r' , \end{aligned}$$

where $r' = r^2/\sqrt{r^2+s_1^2}$ and $s_1' = rs_1/\sqrt{r^2+s_1^2}$ as long as r, s_1, s_2 satisfy the conditions of Lemma 3.5. Then, we set $s_1 = r/(\sqrt{2}q)$ and choose s_2 such that

$$s_2^2 = \beta^2 - (s_1'/r')^2\gamma^2 = \beta^2 - (s_1/r)^2\gamma^2 = \beta^2/2\,.$$

Lemma 3.5 requires $rs_1/\sqrt{r^2\|\boldsymbol{w}-\boldsymbol{w}^*\|^2(s_1/s_2)^2 + r^2 + s_1^2} \geq \eta_\varepsilon(L)$. We know that $r \geq 2q\cdot\eta_\varepsilon(L)$ and $s_1 \geq \sqrt{2}\cdot\eta_\varepsilon(L)$, so it remains to determine a sufficient condition for the aforementioned inequality. Observe that for any $\boldsymbol{w}$ such that $\|\boldsymbol{w}-\boldsymbol{w}^*\| \leq d$, the condition $s_2 \geq 2d\cdot\eta_\varepsilon(L)$ is sufficient. Since $r \geq 2(\gamma/\beta)\cdot\eta_\varepsilon(L)$, this translates to $s_2 \geq \beta/(\sqrt{2})$. Hence, the transformation from Lemma 3.5 will output samples negligibly close to CLWE samples for our choice of s_1 and s_2 as long as $\|\boldsymbol{w}-\boldsymbol{w}^*\| \leq d$ (beyond the BDD distance bound d').

Since $S_{\boldsymbol{w},t}$ is negligibly close to the CLWE distribution, the acceptance probability $p(\boldsymbol{w},t)$ of $\mathcal{O}$ only depends on $\exp(t)\|\boldsymbol{w}-\boldsymbol{w}^*\|$. Moreover, by assumption $\|\boldsymbol{w}^*\| \geq \exp(-n/2)\cdot(2q/r) \geq \exp(-\kappa)d'$. Hence, $\mathcal{O}, \kappa, d'$ correspond to a valid instance of $(\exp(-\kappa), \exp(-\kappa), 1+1/\kappa)$-OHCP with "hidden center" $\boldsymbol{w}^*$.

Next, we show that $p(\boldsymbol{w},t)$ of $\mathcal{O}$ satisfies all three conditions of Lemma 3.8 with $p(\infty)$ taken to be the acceptance probability of the CLWE oracle on samples from $D_{\mathbb{R}^n}\times U$. Item 1 of Lemma 3.8 follows from our assumption that our $\mathrm{CLWE}_{\beta,\gamma}$ oracle has advantage $1/\kappa$, and by our choice of r, s_1, and s_2, when $t^* = \log(\gamma/(\|\boldsymbol{w}^*\|r')) > \log(\sqrt{2})$, the generated CLWE samples satisfy $\gamma'(t^*) = \gamma$ and $\beta'(t^*) = \beta$. Hence, $p(\boldsymbol{0},t^*) - p(\infty) \geq 1/\kappa$.

We now show that Item 2 holds, which states that $|p(\boldsymbol{0},t) - p(\infty)| \leq 2\exp(-t/\kappa)$ for any $t > 0$. We will show that $S_{\boldsymbol{0},t}$ converges exponentially fast to $D_{\mathbb{R}^n}\times U$ in statistical distance. Let $f(\boldsymbol{y},z)$ be the probability density of $S_{\boldsymbol{0},t}$. Then,

$$\begin{aligned}\Delta(S_{\boldsymbol{0},t}, D_{\mathbb{R}^n}\times U) &= \frac{1}{2}\int |f(z|\boldsymbol{y}) - U(z)|\rho(\boldsymbol{y})d\boldsymbol{y}dz \\ &= \frac{1}{2}\int\Big(\int |f(z|\boldsymbol{y}) - U(z)|dz\Big)\rho(\boldsymbol{y})d\boldsymbol{y}\,.\end{aligned}$$

Hence, it suffices to show that the conditional density of z given $\boldsymbol{y}$ for $S_{\boldsymbol{0},t}$ converges exponentially fast to the uniform distribution on $\mathbb{T}$. Notice that the conditional distribution of z given $\boldsymbol{y}$ is the Gaussian distribution with width parameter $\beta' \geq \exp(t)\|\boldsymbol{w}^*\|r/(2q) \geq \exp(t-n/2)$, where we have used our assumption that $\|\boldsymbol{w}^*\| \geq (2q/r)\cdot\exp(-n/2)$. By Lemma 2.9 applied to $\mathbb{Z}$, we know that β' is larger than $\eta_\varepsilon(\mathbb{Z})$ for $\varepsilon = \exp(-\exp(2t-n))$. Hence, one sample from this conditional distribution is within statistical distance ε of the uniform distribution by Lemma 2.8. By the triangle inequality applied to ℓ samples,

$$\Delta\Big(S_{\boldsymbol{0},t}^\ell, (D_{\mathbb{R}^n}\times U)^\ell\Big) \leq \min(1, \ell\exp(-\exp(2t-n))) \leq 2\exp(-t/\kappa)\,,$$

where in the last inequality, we use the the fact that we can choose κ to be such that $2\exp(-t/\kappa) \geq 1$ unless $t \geq \kappa/2$. And when $t \geq \kappa/2 \geq 4qn\ell$, we have $\ell\exp(-\exp(2t-n)) \ll \exp(-t/\kappa)$.

It remains to verify Item 3, which states that $p(\boldsymbol{w},t)$ is κ-Lipschitz in t for any $\|\boldsymbol{w}\| \leq (1+1/\kappa)d' \leq d$. We show this by bounding the statistical distance between $S_{\boldsymbol{w},t_1}$ and $S_{\boldsymbol{w},t_2}$ for $t_1 \geq t_2$. With a slight abuse in notation, let $f_{t_i}(\boldsymbol{y},z)$ be the probability density of $S_{\boldsymbol{w},t_i}$ and let (β_i,γ_i) be the corresponding CLWE distribution parameters. For simplicity, also denote the hidden direction by $\boldsymbol{w}' = (\boldsymbol{w}-\boldsymbol{w}^*)/\|\boldsymbol{w}-\boldsymbol{w}^*\|$. Then,

$$
\begin{aligned}
\Delta(f_{t_1}, f_{t_2}) &= \frac{1}{2}\int \Big(\int |f_{t_1}(z|\boldsymbol{y}) - f_{t_2}(z|\boldsymbol{y})| dz \Big) \rho(\boldsymbol{y}) d\boldsymbol{y} \\
&= \int \Delta\Big(\mathcal{N}(\gamma_1 \langle \boldsymbol{y}, \boldsymbol{w}' \rangle, \beta_1/\sqrt{2\pi}), \mathcal{N}(\gamma_2 \langle \boldsymbol{y}, \boldsymbol{w}' \rangle, \beta_2/\sqrt{2\pi})\Big) \rho(\boldsymbol{y}) d\boldsymbol{y} \\
&\le \frac{1}{2}\int \Big(3(1-(\beta_2/\beta_1)^2) + \sqrt{2\pi}(\gamma_1 - \gamma_2)/\beta_1 \cdot |\langle \boldsymbol{y}, \boldsymbol{w}' \rangle| \Big) \cdot \rho(\boldsymbol{y}) d\boldsymbol{y} \qquad (10) \\
&\le \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\rho}[M(\boldsymbol{y})] \cdot \Big(1 - \exp(-2(t_1 - t_2))\Big) \text{ where } M(\boldsymbol{y}) = \frac{1}{2}\Big(3 + 2\sqrt{\pi} q \cdot |\langle \boldsymbol{y}, \boldsymbol{w}' \rangle|\Big) \\
&\le \mathop{\mathbb{E}}_{\boldsymbol{y}\sim\rho}[M(\boldsymbol{y})] \cdot 2(t_1 - t_2) \qquad (11) \\
&\le (\kappa/\ell) \cdot (t_1 - t_2)\ , \qquad (12)
\end{aligned}
$$

where (10) follows from Lemma 3.10, (11) uses the fact that $1 - \exp(-2(t_1 - t_2)) \le 2(t_1 - t_2)$, and (12) uses the fact that $\mathbb{E}_{\boldsymbol{y}\sim\rho}[M(\boldsymbol{y})] \le 4q \le \kappa/(2\ell)$. Using the triangle inequality over ℓ samples, the statistical distance between $S^{\ell}_{\boldsymbol{w},t_1}$ and $S^{\ell}_{\boldsymbol{w},t_2}$ is at most

$$
\min(1, \ell \cdot (\kappa/\ell)(t_1 - t_2)) \le \kappa(t_1 - t_2)\ .
$$

Therefore, $p(\boldsymbol{w}, t)$ is κ-Lipschitz in t. □

4 Hardness of Homogeneous CLWE

In this section, we show the hardness of homogeneous CLWE by reducing from CLWE, whose hardness was established in the previous section. The main step of the reduction is to transform CLWE samples to homogeneous CLWE samples using rejection sampling (Lemma 4.1).

Consider the samples $(\boldsymbol{y}, z) \sim A_{\boldsymbol{w},\beta,\gamma}$ in $\mathrm{CLWE}_{\beta,\gamma}$. If we condition $\boldsymbol{y}$ on $z = 0 \pmod 1$ then we get exactly samples $\boldsymbol{y} \sim H_{\boldsymbol{w},\beta,\gamma}$ for $\mathrm{hCLWE}_{\beta,\gamma}$. However, this approach is impractical as $z = 0 \pmod 1$ happens with probability 0. Instead we condition $\boldsymbol{y}$ on $z \approx 0 \pmod 1$ somehow. One can imagine that the resulting samples $\boldsymbol{y}$ will still have a "wavy" probability density in the direction of $\boldsymbol{w}$ with spacing $1/\gamma$, which accords with the picture of homogeneous CLWE. To avoid throwing away too many samples, we will do rejection sampling with some small "window" $\delta = 1/\operatorname{poly}(n)$. Formally, we have the following lemma.

Lemma 4.1. *There is a $\operatorname{poly}(n, 1/\delta)$-time probabilistic algorithm that takes as input a parameter $\delta \in (0,1)$ and samples from $A_{\boldsymbol{w},\beta,\gamma}$, and outputs samples from $H_{\boldsymbol{w},\sqrt{\beta^2+\delta^2},\gamma}$.*

Proof. Without loss of generality assume that $\boldsymbol{w} = \boldsymbol{e}_1$. By definition, the probability density of sample $(\boldsymbol{y}, z) \sim A_{\boldsymbol{w},\beta,\gamma}$ is

$$
p(\boldsymbol{y}, z) = \frac{1}{\beta} \cdot \rho(\boldsymbol{y}) \cdot \sum_{k\in\mathbb{Z}} \rho_\beta(z + k - \gamma y_1)\ .
$$

Let $g : \mathbb{T} \to [0,1]$ be the function $g(z) = g_0(z)/M$, where $g_0(z) = \sum_{k\in\mathbb{Z}} \rho_\delta(z + k)$ and $M = \sup_{z\in\mathbb{T}} g_0(z)$. We perform rejection sampling on the samples $(\boldsymbol{y}, z)$ with acceptance probability $\Pr[\text{accept} \mid \boldsymbol{y}, z] = g(z)$. We remark that $g(z)$ is efficiently computable (see [Bra+13, Section 5.2]).

The probability density of outputting $\boldsymbol{y}$ and accept is

$$\begin{aligned}\int_{\mathbb{T}} p(\boldsymbol{y},z)g(z)dz &= \frac{\rho(\boldsymbol{y})}{\beta M}\cdot\int_{\mathbb{T}}\sum_{k_1,k_2\in\mathbb{Z}}\rho_\beta(z+k_1-\gamma y_1)\rho_\delta(z+k_2)dz\\ &= \frac{\rho(\boldsymbol{y})}{\beta M}\cdot\int_{\mathbb{T}}\sum_{k,k_2\in\mathbb{Z}}\rho_{\sqrt{\beta^2+\delta^2}}(k-\gamma y_1)\rho_{\beta\delta/\sqrt{\beta^2+\delta^2}}\Big(z+k_2+\frac{\delta^2(k-\gamma y_1)}{\beta^2+\delta^2}\Big)dz\\ &= \frac{\delta}{M\sqrt{\beta^2+\delta^2}}\cdot\rho(\boldsymbol{y})\cdot\sum_{k\in\mathbb{Z}}\rho_{\sqrt{\beta^2+\delta^2}}(k-\gamma y_1)\ ,\end{aligned}$$

where the second equality follows from Claim 2.4. This shows that the conditional distribution of $\boldsymbol{y}$ upon acceptance is indeed $H_{\boldsymbol{e}_1,\sqrt{\beta^2+\delta^2},\gamma}$. Moreover, a byproduct of this calculation is that the expected acceptance probability is $\Pr[\text{accept}] = Z\delta/(M\sqrt{\beta^2+\delta^2})$, where, according to Eq. (3),

$$\begin{aligned}Z &= \sqrt{\frac{\beta^2+\delta^2}{\beta^2+\delta^2+\gamma^2}}\cdot\rho_{\sqrt{\beta^2+\delta^2+\gamma^2}}(\mathbb{Z})\\ &= \sqrt{\beta^2+\delta^2}\cdot\rho_{1/\sqrt{\beta^2+\delta^2+\gamma^2}}(\mathbb{Z})\\ &\geq \sqrt{\beta^2+\delta^2}\ ,\end{aligned}$$

and the second equality uses Lemma 2.5. Observe that

$$\begin{aligned}g_0(z) &= \sum_{k\in\mathbb{Z}}\rho_\delta(z+k)\\ &\leq 2\cdot\sum_{k=0}^{\infty}\rho_\delta(k)\\ &< 2\cdot\sum_{k=0}^{\infty}\exp(-\pi k) < 4\end{aligned}$$

since $\delta < 1$, implying that $M \leq 4$. Therefore, $\Pr[\text{accept}] \geq \delta/4$, and so the rejection sampling procedure has $\mathrm{poly}(n, 1/\delta)$ expected running time. □

The above lemma reduces CLWE to homogeneous CLWE with slightly worse parameters. Hence, homogeneous CLWE is as hard as CLWE. Specifically, combining Theorem 3.1 (with β taken to be $\beta/\sqrt{2}$) and Lemma 4.1 (with δ also taken to be $\beta/\sqrt{2}$), we obtain the following corollary.

Corollary 4.2. *For any $\beta = \beta(n) \in (0,1)$ and $\gamma = \gamma(n) \geq 2\sqrt{n}$ such that γ/β is polynomially bounded, there is a polynomial-time quantum reduction from $\mathrm{DGS}_{2\sqrt{2n}\eta_\varepsilon(L)/\beta}$ to $\mathrm{hCLWE}_{\beta,\gamma}$.*

5 Hardness of Density Estimation for Gaussian Mixtures

In this section, we prove the hardness of density estimation for k-mixtures of n-dimensional Gaussians by showing a reduction from homogeneous CLWE. This answers an open question regarding its computational complexity [Dia16, Moi18]. We first formally define density estimation for Gaussian mixtures.

Definition 5.1 (Density estimation of Gaussian mixtures)**.** *Let $\mathcal{G}_{n,k}$ be the family of k-mixtures of n-dimensional Gaussians. The problem of* density estimation *for $\mathcal{G}_{n,k}$ is the following. Given $\delta > 0$ and sample access to an unknown $P \in \mathcal{G}_{n,k}$, with probability $9/10$, output a hypothesis distribution Q (in the form of an evaluation oracle) such that $\Delta(Q, P) \le \delta$.*

For our purposes, we fix the precision parameter δ to a very small constant, say, $\delta = 10^{-3}$. Now we show a reduction from $\mathrm{hCLWE}_{\beta,\gamma}$ to the problem of density estimation for Gaussian mixtures. Corollary 4.2 shows that $\mathrm{hCLWE}_{\beta,\gamma}$ is hard for $\gamma \ge 2\sqrt{n}$ (assuming worst-case lattice problems are hard). Hence, by taking $\gamma = 2\sqrt{n}$ and $g(n) = O(\log n)$ in Proposition 5.2, we rule out the possibility of a $\mathrm{poly}(n,k)$-time density estimation algorithm for $\mathcal{G}_{n,k}$ under the same hardness assumption.

Proposition 5.2. *Let $\beta = \beta(n) \in (0, 1/32)$, $\gamma = \gamma(n) \ge 1$, and $g(n) \ge 4\pi$. For $k = 2\gamma\sqrt{g(n)/\pi}$, if there is an $\exp(g(n))$-time algorithm that solves density estimation for $\mathcal{G}_{n,2k+1}$, then there is a $O(\exp(g(n)))$-time algorithm that solves $\mathrm{hCLWE}_{\beta,\gamma}$.*

Proof. We apply the density estimation algorithm $\mathcal{A}$ to the unknown given distribution P. As we will show below, with constant probability, it outputs a density estimate f that satisfies $\Delta(f, P) < 2\delta = 2{\cdot}10^{-3}$ (and this is even though $H_{\boldsymbol{w},\beta,\gamma}$ has infinitely many components). We then test whether $P = D_{\mathbb{R}^n}$ or not using the following procedure. We repeat the following procedure $m = 1/(6\sqrt{\delta})$ times. We draw $\boldsymbol{x} \sim D_{\mathbb{R}^n}$ and check whether the following holds

$$\frac{f(\boldsymbol{x})}{D(\boldsymbol{x})} \in [1 - \sqrt{\delta}, 1 + \sqrt{\delta}] \,, \tag{13}$$

where D denotes the density of $D_{\mathbb{R}^n}$. We output $P = D_{\mathbb{R}^n}$ if Eq. (13) holds for all m independent trials and $P = H_{\boldsymbol{w},\beta,\gamma}$ otherwise. Since $\Delta(H_{\boldsymbol{w},\beta,\gamma}, D_{\mathbb{R}^n}) > 1/2$ (Claim 5.3), it is not hard to see that this test solves $\mathrm{hCLWE}_{\beta,\gamma}$ with probability at least $2/3$ (see [RS09, Observation 24] for a closely related statement). Moreover, the total running time is $O(\exp(g(n)))$ since this test uses a constant number of samples.

If $P = D_{\mathbb{R}^n}$, it is obvious that $\mathcal{A}$ outputs a close density estimate with constant probability since $D_{\mathbb{R}^n} \in \mathcal{G}_{n,2k+1}$. It remains to consider the case $P = H_{\boldsymbol{w},\beta,\gamma}$. To this end, we observe that $H_{\boldsymbol{w},\beta,\gamma}$ is close to a $(2k+1)$-mixture of Gaussians. Indeed, by Claim 5.4 below,

$$\Delta(H_{\boldsymbol{w},\beta,\gamma}, H^{(k)}) \le 2\exp(-\pi \cdot k^2/(\beta^2 + \gamma^2)) < 2\exp(-\pi \cdot k^2/(2\gamma^2)) \,,$$

where $H^{(k)}$ is the distribution given by truncating $H_{\boldsymbol{w},\beta,\gamma}$ to the $(2k+1)$ central mixture components. Hence, the statistical distance between the joint distribution of $\exp(g(n))$ samples from $H_{\boldsymbol{w},\beta,\gamma}$ and that of $\exp(g(n))$ samples from $H^{(k)}$ is bounded by

$$2\exp(-\pi \cdot k^2/(2\gamma^2)) \cdot \exp(g(n)) = 2\exp(-g(n)) \le 2\exp(-4\pi) \,.$$

Since the two distributions are statistically close, a standard argument shows that $\mathcal{A}$ will output f satisfying $\Delta(f, H_{\boldsymbol{w},\beta,\gamma}) \le \Delta(f, H^{(k)}) + \Delta(H^{(k)}, H_{\boldsymbol{w},\beta,\gamma}) < 2\delta$ with constant probability. □

Claim 5.3. *Let $\beta = \beta(n) \in (0, 1/32)$ and $\gamma = \gamma(n) \ge 1$. Then,*

$$\Delta(H_{\boldsymbol{w},\beta,\gamma}, D_{\mathbb{R}^n}) > 1/2 \,.$$

Proof. Let $\gamma' = \sqrt{\beta^2 + \gamma^2} > \gamma$. Let $\boldsymbol{y} \in \mathbb{R}^n$ be a random vector distributed according to $H_{\boldsymbol{w},\beta,\gamma}$. Using the Gaussian mixture form of (2), we observe that $\langle \boldsymbol{y}, \boldsymbol{w} \rangle \bmod \gamma/\gamma'^2$ is distributed according to $D_{\beta/\gamma'} \bmod \gamma/\gamma'^2$. Since statistical distance cannot increase by applying a function (inner product

with $\boldsymbol{w}$ and then applying the modulo operation in this case), it suffices to lower bound the statistical distance between $D_{\beta/\gamma'} \bmod \gamma/\gamma'^2$ and $D \bmod \gamma/\gamma'^2$, where D denotes the 1-dimensional standard Gaussian.

By Chernoff, for all $\zeta > 0$, at least $1-\zeta$ mass of $D_{\beta/\gamma'}$ is contained in $[-a\cdot(\beta/\gamma'), a\cdot(\beta/\gamma')]$, where $a = \sqrt{\log(1/\zeta)}$. Hence, $D_{\beta/\gamma'} \bmod \gamma/\gamma'^2$ is at least $1 - 2a\beta\gamma'/\gamma - \zeta$ far in statistical distance from the uniform distribution over $\mathbb{R}/(\gamma/\gamma'^2)\mathbb{Z}$, which we denote by U. Moreover, by Lemma 2.8 and Lemma 2.9, $D \bmod \gamma/\gamma'^2$ is within statistical distance $\varepsilon/2 = \exp(-\gamma'^4/\gamma^2)/2$ from U. Therefore,

$$\begin{aligned}\Delta(D_{\beta/\gamma'} \bmod \gamma/\gamma'^2, D \bmod \gamma/\gamma'^2) &\geq \Delta(D_{\beta/\gamma'} \bmod \gamma/\gamma'^2, U) - \Delta(U, D \bmod \gamma/\gamma'^2) \\ &\geq 1 - 2a\beta\gamma'/\gamma - \zeta - \varepsilon/2 \\ &> 1 - 2\sqrt{2}a\beta - \zeta - \exp(-\gamma^2)/2 \\ &> 1/2\ ,\end{aligned} \tag{14}$$

where we set $\zeta = \exp(-2)$ and use the fact that $\beta \leq 1/32$ and $\gamma \geq 1$ in (14). □

Claim 5.4. *Let $\beta = \beta(n) \in (0,1), \gamma = \gamma(n) \geq 1$, and $k \in \mathbb{Z}^+$. Then,*

$$\Delta(H_{\boldsymbol{w},\beta,\gamma}, H^{(k)}) \leq 2\exp(-\pi \cdot k^2/(\beta^2+\gamma^2))\ ,$$

where $H^{(k)}$ is the distribution given by truncating $H_{\boldsymbol{w},\beta,\gamma}$ to the central $(2k+1)$ mixture components.

Proof. We express $H_{\boldsymbol{w},\beta,\gamma}$ in its Gaussian mixture form given in Eq. (2) and define a random variable X taking on values in $\mathbb{Z}$ such that the probability of $X = i$ is equal to the probability that a sample comes from the i-th component in $H_{\boldsymbol{w},\beta,\gamma}$. Then, we observe that $H^{(k)}$ is the distribution given by conditioning on $|X| \leq k$. Since X is a discrete Gaussian random variable with distribution $D_{\mathbb{Z},\sqrt{\beta^2+\gamma^2}}$, we observe that $\Pr[|X| > k] \leq \varepsilon := 2\exp(-\pi \cdot k^2/(\beta^2+\gamma^2))$ by [MP12, Lemma 2.8]. Since conditioning on an event of probability $1-\varepsilon$ cannot change the statistical distance by more than ε, we have

$$\Delta(H_{\boldsymbol{w},\beta,\gamma}, H^{(k)}) \leq \varepsilon\ .$$

□

6 LLL Solves Noiseless CLWE

The noiseless CLWE problem ($\beta = 0$) can be solved in polynomial time using LLL. This applies both to the homogeneous and the inhomogeneous versions, as well as to the search version. The argument can be extended to the case of exponentially small $\beta > 0$.

The key idea is to take samples $(\boldsymbol{y}_i, z_i)$, and find integer coefficients $c_1, \ldots, c_m$ such that $\boldsymbol{y} = \sum_{i=1}^m c_i \boldsymbol{y}_i$ is short, say $\|\boldsymbol{y}\| \ll 1/\gamma$. By Cauchy-Schwarz, we then have that $\gamma\langle \boldsymbol{y}, \boldsymbol{w}\rangle = \sum_{i=1}^m c_i z_i$ over the reals (not modulo 1!). This is formalized in Theorem 6.2. We first state Minkowski's Convex Body Theorem, which we will use in the proof of our procedure.

Lemma 6.1 ([Min10])**.** *Let L be a full-rank n-dimensional lattice. Then, for any centrally-symmetric convex set S, if $\mathrm{vol}(S) > 2^n \cdot |\det(L)|$, then S contains a non-zero lattice point.*

Theorem 6.2. *Let $\gamma = \gamma(n)$ be a polynomial in n. Then, there exists a polynomial-time algorithm for solving $\mathrm{CLWE}_{0,\gamma}$.*

Proof. Take $n+1$ CLWE samples $\{(\boldsymbol{y}_i, z_i)\}_{i=1}^{n+1}$ and consider the matrix

$$Y = \begin{bmatrix} \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_n & \boldsymbol{y}_{n+1} \\ 0 & \cdots & 0 & \delta \end{bmatrix},$$

where $\delta = 2^{-3n^2}$.

Consider the lattice L generated by the columns of Y. Since $\boldsymbol{y}_i$'s are drawn from the Gaussian distribution, L is full rank. By Hadamard's inequality, and the fact that with probability exponentially close to 1, $\|\boldsymbol{y}_i\| \le \sqrt{n}$ for all i, we have

$$|\det(L)| \le \delta \cdot n^{n/2} < 2^{-2n^2}.$$

Now consider the n-dimensional cube S centered at $\mathbf{0}$ with side length 2^{-n}. Then, $\mathrm{vol}(S) = 2^{-n^2}$, and by Lemma 6.1, L contains a vector $\boldsymbol{v}$ satisfying $\|\boldsymbol{v}\|_\infty \le 2^{-n}$ and so $\|\boldsymbol{v}\|_2 \le \sqrt{n} \cdot 2^{-n}$. Applying the LLL algorithm [LLL82] gives us an integer combination of the columns of Y whose length is within $2^{(n+1)/2}$ factor of the shortest vector in L, which will therefore have ℓ_2 norm less than $\sqrt{n} \cdot 2^{-(n-1)/2}$. Let $\boldsymbol{y}$ be the corresponding combination of the $\boldsymbol{y}_i$ vectors (which is equivalently given by the first n coordinates of the output of LLL) and $z \in (-1/2, 1/2]$ a representative of the corresponding integer combination of the z_i mod 1. Then, we have $\|\boldsymbol{y}\|_2 \le \sqrt{n} \cdot 2^{-(n-1)/2}$ and therefore we obtain the linear equation $\gamma \cdot \langle \boldsymbol{y}, \boldsymbol{w} \rangle = z$ over the reals (without mod 1).

We now repeat the above procedure n times, and recover $\boldsymbol{w}$ by solving the resulting n linear equations. It remains to argue why the n vectors $\boldsymbol{y}$ we collect are linearly independent. First, note that the output $\boldsymbol{y}$ is guaranteed to be a non-zero vector since with probability 1, no integer combination of the Gaussian distributed $\boldsymbol{y}_i$ is $\mathbf{0}$. Next, note that LLL is equivariant to rotations, i.e., if we rotate the input basis then the output vector will also be rotated by the same rotation. Moreover, spherical Gaussians are rotationally invariant. Hence, the distribution of the output vector $\boldsymbol{y} \in \mathbb{R}^n$ is also rotationally invariant. Therefore, repeating the above procedure n times will give us n linearly independent vectors. □

7 Subexponential Algorithm for Homogeneous CLWE

For $\gamma = o(\sqrt{n})$, the covariance matrix will reveal the discrete structure of homogeneous CLWE, which will lead to a subexponential time algorithm for the problem. This clarifies why the hardness results of homogeneous CLWE do not extend beyond $\gamma \ge 2\sqrt{n}$.

We define *noiseless homogeneous CLWE distribution* $H_{\boldsymbol{w},\gamma}$ as $H_{\boldsymbol{w},\beta,\gamma}$ with $\beta = 0$. We begin with a claim that will allow us to focus on the noiseless case.

Claim 7.1. *By adding Gaussian noise $D_{\mathbb{R}^n,\beta/\gamma}$ to $H_{\boldsymbol{w},\gamma}$ and then rescaling by a factor of $\gamma/\sqrt{\beta^2+\gamma^2}$, the resulting distribution is $H_{\boldsymbol{w},\tilde{\beta},\tilde{\gamma}}$, where $\tilde{\gamma} = \gamma/\sqrt{1+(\beta/\gamma)^2}$ and $\tilde{\beta} = \tilde{\gamma}(\beta/\gamma)$.*[3]

Proof. Without loss of generality, suppose $\boldsymbol{w} = \boldsymbol{e}_1$.

Let $\boldsymbol{z} \sim H_{\boldsymbol{w},\gamma} + D_{\mathbb{R}^n,\beta/\gamma}$ and $\tilde{\boldsymbol{z}} = \gamma \boldsymbol{z}/\sqrt{\beta^2+\gamma^2}$. It is easy to verify that the marginals density of $\boldsymbol{z}$ on subspace $\boldsymbol{e}_1^\perp$ will simply be p. Hence we focus on calculating the density of z_1 and $\tilde{z}_1$. The

[3] Equivalently, in terms of the Gaussian mixture representation of Eq. (2), the resulting distribution has layers spaced by $1/\sqrt{\gamma^2+\beta^2}$ and of width $\beta/\sqrt{\gamma^2+\beta^2}$.

density can be computed by convolving the probability densities of $H_{\boldsymbol{w},\gamma}$ and $D_{\mathbb{R}^n,\beta/\gamma}$ as follows.

$$\begin{aligned} H_{\boldsymbol{w},\gamma} * D_{\mathbb{R}^n,\beta/\gamma}(z_1) &\propto \sum_{k\in\mathbb{Z}} \rho(k/\gamma)\cdot\rho_{\beta/\gamma}(z_1-k/\gamma) \\ &= \rho_{\sqrt{\beta^2+\gamma^2}/\gamma}(z_1)\cdot\sum_{k\in\mathbb{Z}}\rho_{\beta/\sqrt{\beta^2+\gamma^2}}\Big(k/\gamma-\frac{\gamma^2}{\beta^2+\gamma^2}z_1\Big) \\ &= \rho(\tilde{z}_1)\cdot\sum_{k\in\mathbb{Z}}\rho_{\tilde{\beta}}\Big(k-\tilde{\gamma}\tilde{z}_1\Big)\,, \end{aligned}$$

where the second to last equality follows from Claim 2.4. This verifies that the resulting distribution is indeed $H_{\boldsymbol{w},\tilde{\beta},\tilde{\gamma}}$. □

Claim 7.1 implies an homogeneous CLWE distribution with $\beta > 0$ is equivalent to a noiseless homogeneous CLWE distribution with independent Gaussian noise added. We will first analyze the noiseless case and then derive the covariance of noisy (i.e., $\beta > 0$) case by adding independent Gaussian noise and rescaling.

Lemma 7.2. *Let $\Sigma \succ 0$ be the covariance matrix of the n-dimensional noiseless homogeneous CLWE distribution $H_{\boldsymbol{w},\gamma}$ with $\gamma \geq 1$. Then,*

$$\Big\|\Sigma - \frac{1}{2\pi}I_n\Big\| \geq \gamma^2\exp(-\pi\gamma^2)\,,$$

where $\|\cdot\|$ denotes the spectral norm.

Proof. Without loss of generality, let $\boldsymbol{w} = \boldsymbol{e}_1$. Then $H_{\boldsymbol{w},\gamma} = D_L \times D_{\mathbb{R}^{n-1}}$ where L is the one-dimensional lattice $(1/\gamma)\mathbb{Z}$. Then, $\Sigma = \mathrm{diag}(\mathbb{E}_{x\sim D_L}[x^2], \frac{1}{2\pi}, \ldots, \frac{1}{2\pi})$, so it suffices to show that

$$\Big|\mathop{\mathbb{E}}_{x\sim D_L}[x^2] - \frac{1}{2\pi}\Big| \geq \gamma^2\exp(-\pi\gamma^2)\,.$$

Define $g(x) = x^2\cdot\rho(x)$. The Fourier transform of ρ is itself; the Fourier transform of g is given by

$$\widehat{g}(y) = \Big(\frac{1}{2\pi} - y^2\Big)\rho(y)\,.$$

By definition and Poisson's summation formula (Lemma 2.5), we have

$$\begin{aligned} \mathop{\mathbb{E}}_{x\sim D_L}[x^2] &= \frac{g(L)}{\rho(L)} \\ &= \frac{\det(L^*)\cdot\widehat{g}(L^*)}{\det(L^*)\cdot\rho(L^*)} = \frac{\widehat{g}(L^*)}{\rho(L^*)}\,, \end{aligned}$$

where $L^* = ((1/\gamma)\mathbb{Z})^* = \gamma\mathbb{Z}$. Combining this with the expression for $\widehat{g}$, we have

$$\begin{aligned} \Big|\mathop{\mathbb{E}}_{x\sim D_L}[x^2] - \frac{1}{2\pi}\Big| &= \frac{\sum_{y\in L^*} y^2\rho(y)}{1+\rho(L^*\setminus\{0\})} \\ &\geq \gamma^2\exp(-\pi\gamma^2)\,, \end{aligned}$$

where we use the fact that for $\gamma \geq 1$,

$$\rho(\gamma\mathbb{Z}\setminus\{0\}) \leq \rho(\mathbb{Z}\setminus\{0\}) < 2\sum_{k=1}^{\infty}\exp(-\pi k) = \frac{2\exp(-\pi)}{1-\exp(-\pi)} < 1\,.$$

□

Combining Claim 7.1 and Lemma 7.2, we get the following corollary for the noisy case.

Corollary 7.3. *Let $\Sigma \succ 0$ be the covariance matrix of m-dimensional homogeneous CLWE distribution $H_{\boldsymbol{w},\beta,\gamma}$ with $\gamma \geq 1$ and $\beta > 0$. Then,*

$$\left\| \Sigma - \frac{1}{2\pi} I_n \right\| \geq \gamma^2 \exp(-\pi(\beta^2 + \gamma^2)) ,$$

where $\|\cdot\|$ denotes the spectral norm.

Proof. Using Claim 7.1, we can view samples from $H_{\boldsymbol{w},\beta,\gamma}$ as samples from $H_{\boldsymbol{w},\gamma'}$ with independent Gaussian noise of width β'/γ' added and rescaled by $\gamma'/\sqrt{\beta'^2 + \gamma'^2}$, where β', γ' are given by

$$\begin{aligned} \beta' &= \beta\sqrt{1 + (\beta/\gamma)^2} , \\ \gamma' &= \sqrt{\beta^2 + \gamma^2} . \end{aligned}$$

Let Σ be the covariance of $H_{\boldsymbol{w},\beta,\gamma}$ and let Σ_0 be the covariance of $H_{\boldsymbol{w},\gamma'}$. Since the Gaussian noise added to $H_{\boldsymbol{w},\gamma'}$ is independent and $\beta'/\gamma' = \beta/\gamma$,

$$\Sigma = \frac{1}{1 + (\beta/\gamma)^2} \left(\Sigma_0 + \frac{(\beta/\gamma)^2}{2\pi} I_n \right) .$$

Hence,

$$\begin{aligned} \left\| \Sigma - \frac{1}{2\pi} I_n \right\| &= \frac{1}{1 + (\beta/\gamma)^2} \left\| \left(\Sigma_0 + \frac{(\beta/\gamma)^2}{2\pi} I_n \right) - \frac{1 + (\beta/\gamma)^2}{2\pi} I_n \right\| \\ &= \frac{1}{1 + (\beta/\gamma)^2} \left\| \Sigma_0 - \frac{1}{2\pi} I_n \right\| \\ &\geq \gamma^2 \exp(-\pi(\beta^2 + \gamma^2)) . \end{aligned}$$

where the last inequality follows from Lemma 7.2. □

We use the following lemma, which provides an upper bound on the error in estimating the covariance matrix by samples. The sub-gaussian norm of a random variable Y is defined as $\|Y\|_{\psi_2} = \inf\{t > 0 \mid \mathbb{E}[\exp(Y^2/t^2)] \leq 2\}$ and that of an m-dimensional random vector $\boldsymbol{y}$ is defined as $\|\boldsymbol{y}\|_{\psi_2} = \sup_{\boldsymbol{u} \in \mathbb{S}^{n-1}} \|\langle \boldsymbol{y}, \boldsymbol{u} \rangle\|_{\psi_2}$.

Lemma 7.4 ([Ver18, Theorem 4.6.1])**.** *Let A be an $m \times n$ matrix whose rows A_i are independent, mean zero, sub-gaussian isotropic random vectors in $\mathbb{R}^n$. Then for any $u \geq 0$ we have*

$$\left\| \frac{1}{m} A^T A - I_n \right\| \leq K^2 \max(\delta, \delta^2) \quad \text{where } \delta = C\left(\sqrt{\frac{n}{m}} + \frac{u}{\sqrt{m}} \right) ,$$

with probability at least $1 - 2e^{-u^2}$ for some constant $C > 0$. Here, $K = \max_i \|A_i\|_{\psi_i}$.

Combining Corollary 7.3 and Lemma 7.4, we have the following theorem for distinguishing homogeneous CLWE distribution and Gaussian distribution.

Theorem 7.5. *Let $\gamma = n^{\varepsilon}$, where $\varepsilon < 1/2$ is a constant, and let $\beta = \beta(n) \in (0, 1)$. Then, there exists an $\exp(O(n^{2\varepsilon}))$-time algorithm that solves $\mathrm{hCLWE}_{\beta,\gamma}$.*

Proof. Our algorithm takes m samples from the unknown input distribution P and computes the sample covariance matrix $\Sigma_m = (1/m)A^T A$, where A's rows are the samples, and its eigenvalues $\mu_1, \ldots, \mu_n$. Then, it determines whether P is a homogeneous CLWE distribution or not by testing that

$$\left|\mu_i - \frac{1}{2\pi}\right| \leq \frac{1}{2} \cdot \gamma^2 \exp(-\pi(\beta^2 + \gamma^2)) \quad \text{for all } i \in [n] \,.$$

The running time of this algorithm is $O(n^2 m) = \exp(O(n^{2\varepsilon}))$. To show correctness, we first consider the case $P = D_{\mathbb{R}^n}$. The standard Gaussian distribution satisfies the conditions of Lemma 7.4 (after rescaling by $1/(2\pi)$). Hence, the eigenvalues of Σ_m will be within distance $O(\sqrt{n/m})$ from $1/(2\pi)$ with high probability.

Now consider the case $P = H_{\boldsymbol{w},\beta,\gamma}$. We can assume $\boldsymbol{w} = \boldsymbol{e}_1$ without loss of generality since eigenvalues are invariant under rotations. Denote by $\boldsymbol{y}$ a random vector distributed according to $H_{\boldsymbol{w},\beta,\gamma}$ and $\sigma^2 = \mathbb{E}_{\boldsymbol{y} \sim H_{\boldsymbol{w},\beta,\gamma}}[y_1^2]$. The covariance of P is given by

$$\Sigma = \begin{pmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & \frac{1}{2\pi} I_{n-1} \end{pmatrix} . \tag{15}$$

Now consider the sample covariance Σ_m of P and denote by $\sigma_m^2 = \boldsymbol{w}^T \Sigma_m \boldsymbol{w} = (1/m) \sum_{i=1}^m A_{i1}^2$. Since A_{i1}'s are sub-gaussian random variables [MP12, Lemma 2.8], $\sigma_m^2 - \sigma^2$ is a sum of m independent, mean-zero, sub-exponential random variables. For $m = \omega(n)$, Bernstein's inequality [Ver18, Corollary 2.8.3] implies that $|\sigma_m^2 - \sigma^2| = O(\sqrt{n/m})$ with high probability. By Corollary 7.3, we know that

$$\left|\sigma^2 - \frac{1}{2\pi}\right| \geq \gamma^2 \exp(-\pi(\beta^2 + \gamma^2)) \,.$$

Hence, if we choose $m = \exp(c\gamma^2)$ with some sufficiently large constant c, then Σ_m will have an eigenvalue that is noticeably far from $1/(2\pi)$ with high probability. □

8 SQ Lower Bound for Homogeneous CLWE

Statistical Query (SQ) algorithms [Kea98] are a restricted class of algorithms that are only allowed to query expectations of functions of the input distribution without directly accessing individual samples. To be more precise, SQ algorithms access the input distribution indirectly via the STAT(τ) oracle, which given a query function f and data distribution D, returns a value contained in the interval $\mathbb{E}_{x \sim D}[f(x)] + [-\tau, \tau]$ for some precision parameter τ.

In this section, we prove SQ hardness of distinguishing homogeneous CLWE distributions from the standard Gaussian. In particular, we show that SQ algorithms that solve homogeneous CLWE require super-polynomial number of queries even with super-polynomial precision. This is formalized in Theorem 8.1.

Theorem 8.1. *Let $\beta = \beta(n) \in (0,1)$ and $\gamma = \gamma(n) \geq \sqrt{2}$. Then, any (randomized) SQ algorithm with precision $\tau \geq 4 \cdot \exp(-\pi \cdot \gamma^2/4)$ that successfully solves* hCLWE$_{\beta,\gamma}$ *with probability $\eta > 1/2$ requires at least $(2\eta - 1) \cdot \exp(cn) \cdot \tau^2 \beta^2 / (4\gamma^2)$ statistical queries of precision τ for some constant $c > 0$.*

Note that when $\gamma = \Omega(\sqrt{n})$ and $\gamma/\beta = \mathrm{poly}(n)$, even exponential precision $\tau = \exp(-O(n))$ results in a query lower bound that grows as $\exp(\Omega(n))$. This establishes an unconditional hardness

result for SQ algorithms in the parameter regime $\gamma = \Omega(\sqrt{n})$, which is consistent with our computational hardness result based on worst-case lattice problems. The uniform spacing in homogeneous CLWE distributions gives us tight control over their pairwise correlation (see definition in (16)), which leads to a simple proof of the SQ lower bound.

We first provide some necessary background on the SQ framework. We denote by $\mathcal{B}(\mathcal{U}, D)$ the decision problem in which the input distribution P either equals D or belongs to $\mathcal{U}$, and the goal of the algorithm is to identify whether $P = D$ or $P \in \mathcal{U}$. For our purposes, D will be the standard Gaussian $D_{\mathbb{R}^n}$ and $\mathcal{U}$ will be a finite set of homogeneous CLWE distributions. Abusing notation, we denote by $D(x)$ the density of D. Following [Fel+17], we define the *pairwise correlation* between two distributions P, Q relative to D as

$$\chi_D(P,Q) := \mathbb{E}_{\boldsymbol{x}\sim D}\left[\left(\frac{P(\boldsymbol{x})}{D(\boldsymbol{x})} - 1\right)\cdot\left(\frac{Q(\boldsymbol{x})}{D(\boldsymbol{x})} - 1\right)\right] = \mathbb{E}_{\boldsymbol{x}\sim D}\left[\frac{P(\boldsymbol{x})Q(\boldsymbol{x})}{D(\boldsymbol{x})^2}\right] - 1\,. \tag{16}$$

Lemma 8.2 below establishes a lower bound on the number of statistical queries required to solve $\mathcal{B}(\mathcal{U}, D)$ in terms of pairwise correlation between distributions in $\mathcal{U}$.

Lemma 8.2 ([Fel+17, Lemma 3.10])**.** *Let D be a distribution and $\mathcal{U}$ be a set of distributions both over a domain X such that for any $P, Q \in \mathcal{U}$*

$$|\chi_D(P,Q)| \le \begin{cases} \delta & \text{if } P = Q \\ \varepsilon & \text{otherwise} \end{cases}.$$

Let $\tau \ge \sqrt{2\varepsilon}$. Then, any (randomized) SQ algorithm that solves $\mathcal{B}(\mathcal{U}, D)$ with success probability $\eta > 1/2$ requires at least $(2\eta - 1)\cdot|\mathcal{U}|\cdot\tau^2/(2\delta)$ queries to $\mathrm{STAT}(\tau)$.

The following proposition establishes a tight upper bound on the pairwise correlation between homogeneous CLWE distributions. To deduce Theorem 8.1 from Lemma 8.2 and Proposition 8.3, we take a set of unit vectors $\mathcal{U}$ such that any two distinct vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{U}$ satisfy $|\langle \boldsymbol{v}, \boldsymbol{w}\rangle| \le 1/\sqrt{2}$, and identify it with the set of homogeneous CLWE distributions $\{H_{\boldsymbol{w},\beta,\gamma}\}_{\boldsymbol{w}\in\mathcal{U}}$. A standard probabilistic argument shows that such a $\mathcal{U}$ can be as large as $\exp(\Omega(n))$, which proves Theorem 8.1.

Proposition 8.3. *Let $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ be unit vectors and let $H_{\boldsymbol{v}}, H_{\boldsymbol{w}}$ be n-dimensional homogeneous CLWE distributions with parameters $\gamma \ge 1, \beta \in (0,1)$, and hidden direction $\boldsymbol{v}$ and $\boldsymbol{w}$, respectively. Then, for any $\alpha \ge 0$ that satisfies $\gamma^2(1-\alpha^2) \ge 1$,*

$$|\chi_D(H_{\boldsymbol{v}}, H_{\boldsymbol{w}})| \le \begin{cases} 2(\gamma/\beta)^2 & \text{if } \boldsymbol{v} = \boldsymbol{w} \\ 8\exp(-\pi\cdot\gamma^2(1-\alpha^2)) & \text{if } |\langle \boldsymbol{v}, \boldsymbol{w}\rangle| \le \alpha \end{cases}.$$

Proof. We will show that computing $\chi_D(H_{\boldsymbol{v}}, H_{\boldsymbol{w}})$ reduces to evaluating the Gaussian mass of two lattices L_1 and L_2 defined below. Then, we will tightly bound the Gaussian mass using Lemma 2.5 and Lemma 2.10, which will result in upper bounds on $|\chi_D(H_{\boldsymbol{v}}, H_{\boldsymbol{w}})|$. We define L_1 and L_2 by specifying their bases B_1 and B_2, respectively.

$$B_1 = \frac{1}{\sqrt{\beta^2+\gamma^2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$B_2 = \frac{1}{\sqrt{\beta^2+\gamma^2}}\begin{pmatrix} 1 & 0 \\ -\frac{\alpha\gamma^2}{\zeta\sqrt{\beta^2+\gamma^2}} & \frac{\sqrt{\beta^2+\gamma^2}}{\zeta} \end{pmatrix},$$

where $\zeta = \sqrt{(\beta^2+\gamma^2) - \alpha^2\gamma^4/(\beta^2+\gamma^2)}$. Then the basis of the dual lattice L_1^* and L_2^* is B_1^{-T} and B_2^{-T}, respectively. Note that $\lambda_2(L_1)^2 = 1/(\beta^2+\gamma^2)$ and that the two columns of B_2 have the same norm, and so

$$\begin{aligned}\lambda_2(L_2)^2 &\le \frac{1}{\beta^2+\gamma^2} \cdot \max\left\{1 + \frac{\alpha^2\gamma^4}{\zeta^2(\beta^2+\gamma^2)}, \frac{\beta^2+\gamma^2}{\zeta^2}\right\} \\ &= \frac{1}{\zeta^2} \qquad (17) \\ &\le \frac{1}{\gamma^2(1-\alpha^2)} \,. \qquad (18)\end{aligned}$$

Now define the density ratio $a(t) := H(t)/D(t)$, where D is the standard Gaussian and H is the marginal distribution of homogeneous CLWE with parameters β, γ along the hidden direction. We immediately obtain

$$a(t) = \frac{1}{Z}\sum_{k\in\mathbb{Z}} \rho_{\beta/\gamma}(t - k/\gamma) \,, \qquad (19)$$

where $Z = \int_{\mathbb{R}} \rho(t) \cdot \sum_{k\in\mathbb{Z}} \rho_{\beta/\gamma}(t-k/\gamma)dt$. By Eq. (3), Z is given by

$$Z = \frac{\beta}{\sqrt{\beta^2+\gamma^2}} \cdot \rho\left(\frac{1}{\sqrt{\beta^2+\gamma^2}}\mathbb{Z}\right) .$$

Moreover, we can express Z^2 in terms of the Gaussian mass of (L_1) as

$$Z^2 = \frac{\beta^2}{\beta^2+\gamma^2} \cdot \rho(L_1) \,.$$

$\chi_D(H_{\boldsymbol{v}}, H_{\boldsymbol{w}})$ can be expressed in terms of $a(t)$ as

$$\chi_D(H_{\boldsymbol{v}}, H_{\boldsymbol{w}}) = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim D}\Big[a(\langle \boldsymbol{x}, \boldsymbol{w}\rangle) \cdot a(\langle \boldsymbol{x}, \boldsymbol{v}\rangle)\Big] - 1 \,. \qquad (20)$$

Without loss of generality, assume $\boldsymbol{v} = \boldsymbol{e}_1$ and $\boldsymbol{w} = \alpha\boldsymbol{e}_1 + \xi\boldsymbol{e}_2$, where $\xi = \sqrt{1-\alpha^2}$. We first compute the pairwise correlation for $\boldsymbol{v} \neq \boldsymbol{w}$. For notational convenience, we denote by $\varepsilon = 8 \cdot \exp(-\pi \cdot \gamma^2(1-\alpha^2))$.

$$
\begin{aligned}
\chi_D(H_{\boldsymbol{v}}, H_{\boldsymbol{w}}) + 1 &= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim D}\Big[a(x_1)\cdot a(\alpha x_1 + \xi x_2)\Big] \\
&= \frac{1}{Z^2}\sum_{k,\ell\in\mathbb{Z}}\int\int \rho_\beta(\gamma x_1 - k)\cdot\rho_\beta((\gamma\alpha x_1 + \gamma\xi x_2) - \ell)\cdot\rho(x_1)\cdot\rho(x_2)dx_1dx_2 \\
&= \frac{1}{Z^2}\cdot\frac{\beta}{\sqrt{(\gamma\xi)^2+\beta^2}}\sum_{k,\ell\in\mathbb{Z}}\int \rho_\beta(\gamma x_1 - k)\cdot\rho(x_1)\cdot\rho_{\sqrt{1+\beta^2/(\gamma\xi)^2}}(\ell/(\gamma\xi) - (\alpha/\xi)x_1)dx_1 \\
&= \frac{1}{Z^2}\cdot\frac{\beta}{\sqrt{(\gamma\xi)^2+\beta^2}}\cdot\frac{\beta\sqrt{(\gamma\xi)^2+\beta^2}}{\zeta\sqrt{\beta^2+\gamma^2}}\sum_{k,\ell\in\mathbb{Z}}\rho_{\sqrt{\beta^2+\gamma^2}}(k)\cdot\rho_\zeta\Big(\ell - \gamma^2\alpha\cdot k/(\beta^2+\gamma^2)\Big) \\
&= \frac{\sqrt{\beta^2+\gamma^2}}{\zeta}\cdot\frac{\sum_{k,\ell\in\mathbb{Z}}\rho_{\sqrt{\beta^2+\gamma^2}}(k)\cdot\rho_\zeta\Big(\ell - \gamma^2\alpha\cdot k/(\beta^2+\gamma^2)\Big)}{\rho(L_1)} \\
&= \frac{\sqrt{\beta^2+\gamma^2}}{\zeta}\cdot\frac{\rho(L_2)}{\rho(L_1)} \\
&= \frac{\sqrt{\beta^2+\gamma^2}}{\zeta}\cdot\frac{\det(L_2^*)}{\det(L_1^*)}\cdot\frac{\rho(L_2^*)}{\rho(L_1^*)} \\
&= \frac{\rho(L_2^*)}{\rho(L_1^*)} \qquad (21)\\
&\in \left[\frac{1}{1+\varepsilon}, 1+\varepsilon\right],
\end{aligned}
$$

In (21), we used the Poisson summation formula (Lemma 2.5). The last line follows from (18) and Lemma 2.10, which implies that for any 2-dimensional lattice L satisfying $\lambda_2(L) \le 1$,

$$\rho(L^* \setminus \{\mathbf{0}\}) \le 8\exp(-\pi/\lambda_2(L)^2)\,. \qquad (22)$$

Now consider the case $\boldsymbol{v} = \boldsymbol{w}$. Using (17), we get an upper bound $\lambda_2(L_2) \le 1/\beta$ when $\alpha = 1$. It follows that $\lambda_2((\beta/\gamma)L_2) \le 1/\gamma \le 1$. Hence,

$$
\begin{aligned}
\chi_D(H_{\boldsymbol{v}}, H_{\boldsymbol{v}}) + 1 &= \frac{\sqrt{\beta^2+\gamma^2}}{\zeta}\cdot\frac{\rho(L_2)}{\rho(L_1)} \\
&\le \frac{\sqrt{\beta^2+\gamma^2}}{\zeta}\cdot\frac{\rho((\beta/\gamma)L_2)}{\rho(L_1)} \\
&= \frac{\sqrt{\beta^2+\gamma^2}}{\zeta}\cdot\frac{\det((\gamma/\beta)L_2^*)}{\det(L_1^*)}\cdot\frac{\rho((\gamma/\beta)L_2^*)}{\rho(L_1^*)} \\
&= \frac{\gamma^2}{\beta^2}\cdot\frac{\rho((\gamma/\beta)L_2^*)}{\rho(L_1^*)} \qquad (23)\\
&\le 2(\gamma/\beta)^2\,. \qquad (24)
\end{aligned}
$$

where we used Lemma 2.5 in (23) and in (24), we used (22) and the fact that $\lambda_2((\beta/\gamma)L_2) \le 1$ to deduce $\rho((\gamma/\beta)L_2^* \setminus \{\mathbf{0}\}) \le 1$. □

9 Extension of Homogeneous CLWE to $m \ge 1$ Hidden Directions

In this section, we generalize the hardness result to the setting where the homogeneous CLWE distribution has $m \ge 1$ hidden directions. The proof is a relatively standard hybrid argument.

Definition 9.1 (*m*-Homogeneous CLWE distribution)**.** *For $0 \le m \le n$, matrix $\boldsymbol{W} \in \mathbb{R}^{n \times m}$ with orthonormal columns $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m$, and $\beta, \gamma > 0$, define the* m-homogeneous CLWE distribution *$H_{\boldsymbol{W},\beta,\gamma}$ over $\mathbb{R}^n$ to have density at $\boldsymbol{y}$ proportional to*

$$\rho(\boldsymbol{y}) \cdot \prod_{i=1}^{m} \sum_{k \in \mathbb{Z}} \rho_\beta(k - \gamma \langle \boldsymbol{y}, \boldsymbol{w}_i \rangle) \ .$$

Note that the 0-homogeneous CLWE distribution is just $D_{\mathbb{R}^n}$ regardless of β and γ.

Definition 9.2. *For parameters $\beta, \gamma > 0$ and $1 \le m \le n$, the average-case decision problem $\mathrm{hCLWE}^{(m)}_{\beta,\gamma}$ is to distinguish the following two distributions over $\mathbb{R}^n$: (1) the m-homogeneous CLWE distribution $H_{\boldsymbol{W},\beta,\gamma}$ for some matrix $\boldsymbol{W} \in \mathbb{R}^{n \times m}$ (which is fixed for all samples) with orthonormal columns chosen uniformly from the set of all such matrices, or (2) $D_{\mathbb{R}^n}$.*

Lemma 9.3. *For any $\beta, \gamma > 0$ and positive integer $m = m(n)$ such that $m \le n$ and $n - m = \Omega(n^c)$ for some constant $c > 0$, if there exists an efficient algorithm that solves $\mathrm{hCLWE}^{(m)}_{\beta,\gamma}$ with non-negligible advantage, then there exists an efficient algorithm that solves $\mathrm{hCLWE}_{\beta,\gamma}$ with non-negligible advantage.*

Proof. Suppose $\mathcal{A}$ is an efficient algorithm that solves $\mathrm{hCLWE}^{(m)}_{\beta,\gamma}$ with non-negligible advantage in dimension m. Then consider the following algorithm $\mathcal{B}$ that uses $\mathcal{A}$ as an oracle and solves $\mathrm{hCLWE}_{\beta,\gamma}$ in dimension $n' = n - m + 1$.

1. Input: n'-dimensional samples, drawn from either $\mathrm{hCLWE}_{\beta,\gamma}$ or $D_{\mathbb{R}^{n'}}$;
2. Choose $0 < i < m - 1$ uniformly at random;
3. Append $m - 1 = n - n'$ coordinates to the given samples, where the first i appended coordinates are drawn from $H_{\boldsymbol{I}_i,\beta,\gamma}$ (with $\boldsymbol{I}_i$ denoting the rank-i identity matrix) and the rest of the coordinates are drawn from $D_{\mathbb{R}^{m-i-1}}$;
4. Rotate the augmented samples using a uniformly random rotation from the orthogonal group $O(n)$;
5. Call $\mathcal{A}$ with the samples and output the result.

As $n = O(n'^{1/c})$, $\mathcal{B}$ is an efficient algorithm. Moreover, the samples passed to $\mathcal{A}$ are effectively drawn from either $\mathrm{hCLWE}^{(i+1)}_{\beta,\gamma}$ or $\mathrm{hCLWE}^{(i)}_{\beta,\gamma}$. Therefore the advantage of $\mathcal{B}$ is at least $1/m$ fraction of the advantage of $\mathcal{A}$, which would be non-negligible (in terms of m, and thus also in terms of n') as well. □

Combining Corollary 4.2 and Lemma 9.3, we obtain the following corollary.

Corollary 9.4. *For any $\beta = \beta(n) \in (0, 1)$ and $\gamma = \gamma(n) \ge 2\sqrt{n}$ such that γ/β is polynomially bounded, and positive integer $m = m(n)$ such that $m \le n$ and $n - m = \Omega(n^c)$ for some constant $c > 0$, there is a polynomial-time quantum reduction from $\mathrm{DGS}_{2\sqrt{2n}\eta_\varepsilon(L)/\beta}$ to $\mathrm{hCLWE}^{(m)}_{\beta,\gamma}$.*

References

[AD97] M. Ajtai and C. Dwork. A public-key cryptosystem with worst-case/average-case equivalence. *STOC*. 1997, pp. 284–293.

[AG11] S. Arora and R. Ge. New algorithms for learning in presence of errors. *ICALP*. 2011, pp. 403–415.

[AK05] S. Arora and R. Kannan. Learning mixtures of separated nonspherical Gaussians. *Ann. Appl. Probab.* 15: 1A (2005), pp. 69–92.

[AR05] D. Aharonov and O. Regev. Lattice problems in NP $\cap$ CoNP. *J. ACM* 52: 5 (2005), pp. 749–765.

[Bab86] L. Babai. On Lovász' lattice reduction and the nearest lattice point problem. *Combinatorica* 6: 1 (1986), pp. 1–13.

[Bra+13] Z. Brakerski, A. Langlois, C. Peikert, O. Regev, and D. Stehlé. Classical hardness of learning with errors. *STOC*. 2013, pp. 575–584.

[Bub+19] S. Bubeck, Y. T. Lee, E. Price, and I. Razenshteyn. Adversarial examples from computational constraints. *ICML*. Vol. 97. 2019, pp. 831–840.

[BV08] S. C. Brubaker and S. Vempala. Isotropic PCA and affine-invariant clustering. *FOCS*. 2008, pp. 551–560.

[Das99] S. Dasgupta. Learning mixtures of Gaussians. *FOCS*. 1999, p. 634.

[Dia16] I. Diakonikolas. Learning structured distributions. *Handbook of Big Data*. 2016, pp. 267–284.

[DKS17] I. Diakonikolas, D. M. Kane, and A. Stewart. Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures. *FOCS*. 2017, pp. 73–84.

[DKS18] I. Diakonikolas, D. M. Kane, and A. Stewart. List-decodable robust mean estimation and learning mixtures of spherical Gaussians. *STOC*. 2018, pp. 1047–1060.

[DMR18] L. Devroye, A. Mehrabian, and T. Reddad. The total variation distance between high-dimensional Gaussians. 2018. arXiv: 1810.08693.

[DS07] S. Dasgupta and L. Schulman. A probabilistic analysis of EM for mixtures of separated, spherical Gaussians. *JMLR* 8 (2007), pp. 203–226.

[Fel+17] V. Feldman, E. Grigorescu, L. Reyzin, S. S. Vempala, and Y. Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* 64: 2 (2017).

[HL18] S. B. Hopkins and J. Li. Mixture models, robustness, and sum of squares proofs. *STOC*. 2018, pp. 1021–1034.

[Kea98] M. Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM* 45: 6 (1998), pp. 983–1006.

[KKK19] S. Karmalkar, A. Klivans, and P. Kothari. List-decodable linear regression. *NeurIPS*. 2019, pp. 7425–7434.

[KSS18] P. K. Kothari, J. Steinhardt, and D. Steurer. Robust moment estimation and improved clustering via sum of squares. *STOC*. 2018, pp. 1035–1046.

[LLL82] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. en. *Mathematische Annalen* 261: 4 (1982), pp. 515–534.

[LM09] V. Lyubashevsky and D. Micciancio. On bounded distance decoding, unique shortest vectors, and the minimum distance problem. *CRYPTO*. 2009, pp. 577–594.

[Min10] H. Minkowski. Geometrie der Zahlen. B.G. Teubner, 1910.

[Moi18] A. Moitra. Algorithmic aspects of machine learning. Cambridge University Press, 2018.

[MP12] D. Micciancio and C. Peikert. Trapdoors for lattices: simpler, tighter, faster, smaller. *EUROCRYPT*. 2012, pp. 700–718.

[MR07] D. Micciancio and O. Regev. Worst-case to average-case reductions based on Gaussian measures. *SIAM J. Comput.* 37: 1 (2007), pp. 267–302.

[MV10] A. Moitra and G. Valiant. Settling the polynomial learnability of mixtures of Gaussians. *FOCS*. 2010, pp. 93–102.

[Pea94] K. Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A* 185 (1894), pp. 71–110.

[Pei10] C. Peikert. An efficient and parallel Gaussian sampler for lattices. *CRYPTO*. 2010, pp. 80–97.

[Pei16] C. Peikert. A decade of lattice cryptography. *Foundations and Trends in Theoretical Computer Science* 10: 4 (2016), pp. 283–424.

[PRS17] C. Peikert, O. Regev, and N. Stephens-Davidowitz. Pseudorandomness of ring-LWE for any ring and modulus. *STOC*. 2017, pp. 461–473.

[Reg04] O. Regev. New lattice-based cryptographic constructions. *J. ACM* 51: 6 (2004), pp. 899–942.

[Reg05] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. *STOC*. 2005, pp. 84–93.

[RS09] R. Rubinfeld and R. A. Servedio. Testing monotone high-dimensional distributions. *Random Structures & Algorithms* 34: 1 (2009), pp. 24–44.

[RV17] O. Regev and A. Vijayaraghavan. On learning mixtures of well-separated Gaussians. *FOCS*. 2017, pp. 85–96.

[RY20] P. Raghavendra and M. Yau. List decodable learning via sum of squares. *SODA*. 2020, pp. 161–180.

[Sze+14] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. *ICLR*. 2014.

[Ver18] R. Vershynin. High-dimensional probability: an introduction with applications in data science. Cambridge University Press, 2018.

[VW02] S. Vempala and G. Wang. A spectral algorithm for learning mixtures of distributions. *FOCS*. 2002, p. 113.