Continuous LWE is as Hard as LWE & Applications to Learning Gaussian Mixtures

Aparna Gupte  
MIT  
Cambridge, MA  
Email: agupte@mit.edu

Neekon Vafa  
MIT  
Cambridge, MA  
Email: nvafa@mit.edu

Vinod Vaikuntanathan  
MIT  
Cambridge, MA  
Email: vinodv@mit.edu

Abstract—We show direct and conceptually simple reductions between the classical learning with errors (LWE) problem and its continuous analog, CLWE (Bruna, Regev, Song and Tang, STOC 2021). This allows us to bring to bear the powerful machinery of LWE-based cryptography to the applications of CLWE. For example, we obtain the hardness of CLWE under the classical worst-case hardness of the gap shortest vector problem. Previously, this was known only under quantum worst-case hardness of lattice problems. More broadly, with our reductions between the two problems, any future developments to LWE will also apply to CLWE and its downstream applications.

As a concrete application, we show an improved hardness result for density estimation for mixtures of Gaussians. In this computational problem, given sample access to a mixture of Gaussians, the goal is to output a function that estimates the density function of the mixture. Under the (plausible and widely believed) exponential hardness of the classical LWE problem, we show that Gaussian mixture density estimation in $\mathbb{R}^n$ with roughly $\log n$ Gaussian components given $\text{poly}(n)$ samples requires time quasi-polynomial in $n$. Under the (conservative) polynomial hardness of LWE, we show hardness of density estimation for $n^\epsilon$ Gaussians for any constant $\epsilon > 0$, which improves on Bruna, Regev, Song and Tang (STOC 2021), who show hardness for at least $\sqrt{n}$ Gaussians under polynomial (quantum) hardness assumptions.

Our key technical tool is a reduction from classical LWE to LWE with $k$-sparse secrets where the multiplicative increase in the noise is only $O(\sqrt{k})$, independent of the ambient dimension $n$.

I. INTRODUCTION

The learning with errors (LWE) problem [1] is a versatile average-case problem with connections to lattice, cryptography, learning theory and game theory. Given a sequence of noisy linear equations $(a, b \approx \langle a, s \rangle \mod q)$ over a ring $\mathbb{Z}/q\mathbb{Z}$, the LWE problem asks to recover the secret vector $s$ (and the decisional version of the problem asks to distinguish between LWE samples and uniformly random numbers mod $q$). Starting from the seminal work of Regev, who showed that a polynomial-time algorithm for LWE will give us a polynomial-time quantum algorithm for widely studied worst-case lattice problems, there has been a large body of work showing connections between LWE and lattice problems [2], [3]. Ever since its formulation in 2005, LWE has unlocked a wealth of applications in cryptography ranging from fully homomorphic encryption [4] to attribute-based encryption [5] to, most recently, succinct non-interactive argument systems for all of P [6]. LWE-based cryptosystems lie at the center of efforts by the National Institute of Standards and Technology (NIST) to develop post-quantum cryptographic standards. LWE has also had applications to learning theory, in the form of hardness results for learning intersections of halfspaces [7], and in game theory, where the hardness of LWE implies the hardness of the complexity class PPAD [8]. Finally, LWE enjoys remarkable structural properties such as leakage-resilience [9].

Motivated by applications to learning problems, Bruna, Regev, Song and Tang [10] recently introduced a continuous version of LWE which they called CLWE. (In the definition below and henceforth, $\mathcal{N}(\mu, \Sigma)$ is the multivariate normal distribution with mean $\mu$ and covariance matrix $\Sigma$ where the probability of a point $x \in \mathbb{R}^n$ is proportional to $e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$.)

Definition 1 (CLWE Distribution [10], rescaled). Let $\gamma, \beta \in \mathbb{R}$, and let $S$ be a distribution over unit vectors in $\mathbb{R}^n$. Let $\text{CLWE}(m, S, \gamma, \beta)$ be the distribution given by sampling $a_1, \ldots, a_m \sim \mathcal{N}(0, I_{n \times n})$, $w \sim \mathcal{N}(0, \gamma I_n)$, and $b = \beta \sum_{i=1}^m w_i$. Then, $(a_i, b \approx \langle a_i, s \rangle)$ are distributed according to $\text{CLWE}(m, S, \gamma, \beta)$.
$S, e_1, \cdots, e_m \sim \mathcal{N}(0, \beta^2)$ and outputting 
\[(a_i, b_i) := \gamma \cdot \langle a_i, w \rangle + e_i \text{ mod } 1\] 
Unless otherwise specified, $S$ is taken to be the uniform distribution over all unit vectors in $\mathbb{R}^n$. We refer to $n$ as the dimension and $m$ as the number of samples.

The search CLWE problem asks to find the secret vector $w$ given CLWE samples, whereas the decisional CLWE problem asks to distinguish between samples from the CLWE distribution and samples with standard normal $a_i$ (just like the CLWE distribution) but now with independent $b_i$ that are distributed uniformly between 0 and 1.

Bruna et al. [10] showed the hardness of the CLWE problem, assuming the worst-case quantum hardness of approximate shortest vector problems on lattices (such as gapSVP and SIVP). Aside from being quantum, the reduction makes non-black-box use of the rather involved techniques from [1], [11]. A natural question is whether CLWE has a classical reduction from worst-case lattice problems, in analogy with such reductions in the context of LWE [2], [3]. Even better outcome would be if we can “piggyback” on the rich literature on worst-case to average-case reductions for LWE, without opening the box, hopefully resulting in a conceptually simple worst-case to average-case connection for CLWE. The conceptually clean way to accomplish all of this would be to come up with a direct reduction from LWE to CLWE, a problem that was explicitly posed in the recent work of Bogdanov, Noval, Hoffman and Rosen [12].

Our main conceptual contribution is a direct and simple reduction from LWE to CLWE. When combined with Regev [1], our reduction immediately gives an alternate proof of CLWE hardness assuming worst-case quantum hardness of lattice problems, reproving one of the main results of Bruna et al. [10]. As another immediate application, by combining with the classical reduction from worst-case lattice problems to LWE [3], we obtain classical worst-case hardness of CLWE. Our main reduction also allows us to unlock powerful structural results on LWE [9], [3], [13], [14] and derive improved hardness results for learning mixtures of Gaussians with $(\log n)^{1+\epsilon}$ Gaussians instead of $\Omega(\sqrt{n})$ in [10] (for arbitrary $\epsilon > 0$). We now describe these results in turn.

A. Continuous LWE is as Hard as LWE

Our main result is a direct and conceptually simple reduction from LWE to CLWE. Recall that in the decisional LWE problem [1], we are given $m$ samples of the form $(a_i, b_i) := \langle a_i, s \rangle + e_i \text{ mod } q$ where $a_i \sim (\mathbb{Z}/q\mathbb{Z})^n$ is uniformly random, $s \in \mathbb{Z}^n$ is the LWE secret vector, and the errors $e_i \sim \mathcal{N}(0, \sigma^2)$ are chosen from the one-dimensional Gaussian with standard deviation $\sigma$. The decisional LWE assumption (parameterized by $n, m, q$ and $\sigma$) postulates that these samples are computationally indistinguishable from i.i.d. samples in $(\mathbb{Z}/q\mathbb{Z})^n \times \mathbb{R}/q\mathbb{Z}$.

**Theorem 1** (Informal Version of Theorem 4). Let $S = S_r$ be an arbitrary distribution over $\mathbb{Z}^n$ whose support consists of vectors with $\ell_2$-norm exactly $r$. Then, for 
\[\gamma = \tilde{O}(r) \text{ and } \beta = O\left(\frac{\sigma}{q}\right),\] 
(where $\tilde{O}(\cdot)$ hides various poly-logarithmic factors), there is a dimension-preserving and sample-preserving polynomial-time reduction from decisional LWE, with parameters $n, m, q, \sigma$ and secret distribution $S$, to decisional CLWE with parameters $n, m, \gamma$ and $\beta$, as long as $\sigma \gg r$.

Our main reduction, in conjunction with prior work, immediately gives us a number of corollaries. First, letting $S$ be the uniform distribution on $\{-1, 1\}^n$, and invoking the hardness result for LWE with binary secrets [3], [13], [14], we obtain the following corollary. (The noise blowup of $\sqrt{n}$ in the corollary below comes from the aforementioned reductions from LWE to LWE with binary secrets.)

**Corollary 1** (Informal Version of Corollary 4). For 
\[\gamma = \tilde{O}(\sqrt{n}) \text{ and } \beta = O\left(\frac{\sigma\sqrt{n}}{q}\right),\] 
there is a polynomial (in $n$) time reduction from standard decisional LWE in dimension $\ell$, with $n$ samples, modulus $q$ and noise parameter $\sigma$, to decisional CLWE in dimension $\ell$ with parameters $\gamma$ and $\beta$, as long as $n \gg \ell \log_2(q)$ and $\sigma \gg 1$.

The generality of our main reduction allows us to unlock techniques from the literature on leakage-resilient cryptography, specifically results related to the robustness of the LWE assumption [9], [3], [13], [14], and go much further. In particular, using a variant of the reduction of [13] modified to handle $k$-sparse secrets (discussed further in Section II) we show the following corollary. In the corollaries, the condition $n \gg \ell \log_2 q$ (resp. $k \log_2(n/k) \gg \ell \log_2(q)$) comes from the entropy of random $\pm 1$ vectors (resp. random $k$-sparse vectors).

**Corollary 2** (Informal). For 
\[\gamma = O\left(\frac{\sqrt{k} \cdot \log n}{\sqrt{k} \cdot \log n}\right) \text{ and } \beta = O\left(\frac{\sigma\sqrt{k}}{q}\right),\]
we have a polynomial (in \( n \)) time reduction from standard decisional \( \text{LWE} \), in dimension \( \ell \), with \( n \) samples, modulus \( q \), and noise parameter \( \sigma \), to decisional \( \text{CLWE} \) in dimension \( n \) with \( k \)-sparse norm-1 secrets and parameters \( \gamma \) and \( \beta \), as long as \( k \log_2(n/k) \gg \ell \log_2(q) \) and \( \sigma \gg 1 \).

We note that Corollary 2 will help us derive improved hardness for the problem of learning mixtures of Gaussians. Towards that end, it is worth stepping back and examining how far one can push Corollary 2. The LWE problem is believed to be exponentially hard; that is, in \( \ell \) dimensions with a modulus \( q = \text{poly}(\ell) \) and error parameter \( \sigma = \text{poly}(\ell) \), LWE is believed to be hard for algorithms that run in \( 2^\ell \) time using \( m = 2^\ell \) samples, for any \( \epsilon < 1 \) (see, e.g. [15]). Breaking this sub-exponential barrier not only has wide-ranging consequences for lattice-based cryptography, but also to the ongoing NIST post-quantum standardization competition [16] where better algorithms for LWE will lead NIST to reconsider the current parameterization of LWE-based encryption and signature schemes.

Assuming such a sub-exponential hardness of LWE, we get the hardness of CLWE with \( \gamma = (\log n)^{1/2 + \delta} \log \log n \) for an arbitrarily small constant \( \delta = \delta(\epsilon) \). On the other hand, under a far more conservative polynomial-hardness assumption on LWE, we get the hardness of CLWE with \( \gamma = n^{\delta} \) for an arbitrarily small \( \delta > 0 \).

Combining our main reduction with the known classical reduction from worst-case lattice problems to \( \text{LWE} \) [3] gives us classical worst-case hardness of CLWE.

**Corollary 3** (Classical Worst-Case Hardness of CLWE, informal). There is an efficient classical reduction from worst-case \( \text{poly}(n/\beta) \)-approximate \( \text{gapSVP} \) in \( \sqrt{n} \) dimensions, to decisional \( \text{CLWE} \) in \( n \) dimensions with \( \gamma = \tilde{\Omega}(\sqrt{n}) \) and arbitrary \( \beta = 1/\text{poly}(n) \).

Finally, in the full version of our paper [17], we also show a reduction in the opposite direction, that is, from (discrete-secret) CLWE to LWE. Modulo the discrete secret requirement, this nearly completes the picture of the relationship between LWE and CLWE. In turn, our reverse reduction can be combined with the other theorems in this paper to show a search-to-decision reduction for (discrete-secret) CLWE.

**B. Improved Hardness of Learning Mixtures of Gaussians**

Bruna, Regev, Song and Tang [10] used the hardness of CLWE to deduce hardness of problems in machine learning, most prominently the hardness of learning mixtures of Gaussians. We use our improved hardness result for CLWE to show improved hardness results for learning mixtures of Gaussians. First, let us start by describing the problem of Gaussian mixture learning.

**Background on Gaussian Mixture Learning**: The problem of learning a mixture of Gaussians is of fundamental importance in many fields of science [18], [19]. Given a set of \( g \) multivariate Gaussians in \( n \) dimensions, parameterized by their means \( \mu_i \in \mathbb{R}^n \), covariance matrices \( \Sigma_i \in \mathbb{R}^{n \times n} \), and non-negative weights \( w_1, \ldots, w_g \) summing to one, the Gaussian mixture model is defined to be the distribution generated by picking a Gaussian \( i \in [g] \) with probability \( w_i \) and outputting a sample from \( \mathcal{N}(\mu_i, \Sigma_i) \).

Dasgupta [20] initiated the study of this problem in computer science. A strong notion of learning mixtures of Gaussians is that of parameter estimation, i.e. to estimate all \( \mu_i, \Sigma_i \) and \( w_i \) given samples from the distribution. If one assumes the Gaussians in the mixture are well-separated, then the problem is known to be tractable for a constant number of Gaussian components [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34]. Moitra and Valiant [28] and Hardt and Price [30] also show that for parameter estimation, there is an information theoretic sample-complexity lower bound of \((1/\gamma)^2 \) where \( \gamma \) is the separation parameter and \( g \) the number of Gaussian components.

Consequently, it makes sense to ask for a weaker notion of learning, namely density estimation, where, given samples from the Gaussian mixture, the goal is to output a “density oracle” (e.g. a circuit) that on any input \( \mathbf{x} \in \mathbb{R}^n \), outputs an estimate of the density at \( \mathbf{x} \) [35]. The statistical distance between the density estimate and the true density must be at most a parameter \( 0 \leq \epsilon \leq 1 \). The sample complexity of density estimation does not suffer from the exponential dependence in \( g \), as was the case for parameter estimation. In fact, Diakonikolas, Kane, and Stewart [36] show a \( \text{poly}(n, g, 1/\epsilon) \) upper bound on the information-theoretic sample complexity, by giving an exponential-time algorithm.

Density estimation seems to exhibit a statistical-computational trade-off. While [36] shows a polynomial upper bound on sample complexity, all known algorithms for density estimation, e.g., [28], run in time \((n/\epsilon)^{f(g)} \) for some \( f(g) \geq g \). This is polynomial-time only for constant \( g \). Furthermore, [36] shows that even density estimation of Gaussian mixtures incurs a super-polynomial lower bound in the restricted statistical query
(SQ) model [37], [38]. Explicitly, they show that any SQ algorithm giving density estimates requires $n^{\Omega(1)}$ queries to an SQ oracle of precision $n^{-O(\gamma)}$; this is super-polynomial as long as $g$ is super-constant. However, this lower bound does not say anything about arbitrary polynomial time algorithms for density estimation.

The first evidence of computational hardness of density estimation for Gaussian mixtures came from the work of Bruna, Regev, Song and Tang [10]. They show that being able to output a density estimate for mixtures of $g = \Omega(\sqrt{n})$ Gaussians implicitly a quantum polynomial-time algorithm for worst-case lattice problems. This leaves a gap between $g = O(1)$ Gaussians, which is known to be learnable in polynomial time, versus $g = \Omega(\sqrt{n})$ Gaussians, which is hard to learn. What is the true answer?

Our Results on the Hardness of Gaussian Mixture Learning: Armed with our reduction from LWE to CLWE, and leakage-resilience theorems from the literature which imply Corollaries 1 and 2, we demonstrate a rich landscape of lower-bounds for density estimation of Gaussian mixtures.

Using Corollary 1, we show a hardness result for density estimation of Gaussian mixtures that improves on [10] in two respects. First, we show hardness of density estimation for $g = n^\epsilon$ Gaussians in $n$ dimensions for any $\epsilon > 0$, assuming the polynomial-time hardness of LWE. Combined with the quantum reduction from worst-case lattice problems to LWE [1], this gives us hardness for $n^\epsilon$ Gaussians under the quantum worst-case hardness of lattice problems. This improves on [10] who show hardness for $\Omega(\sqrt{n})$ Gaussians under the same assumption. Secondly, our hardness of density estimation can be based on the classical hardness of lattice problems.

The simplicity and generality of our main reduction from LWE to CLWE gives us much more. For one, assuming the sub-exponential hardness of LWE, we show that density estimation of $g = (\log n)^{1+\epsilon}$ Gaussians cannot be done in polynomial time given a polynomial number of samples (where $\epsilon > 0$ is an arbitrarily small constant). This brings us very close to the true answer: we know that $g = O(1)$ Gaussians can be learned in polynomial time; whereas $g = (\log n)^{1+\epsilon}$ Gaussians cannot, under a standard assumption in lattice-based cryptography (indeed, one that underlies post-quantum cryptosystems that are about to be standardized by NIST [16]).

We can stretch this even a little further. We show the hardness of density estimation for $g = (\log n)^{1/2+\epsilon}$ Gaussians given poly$(\log n)$ samples (where $\epsilon > 0$ is an arbitrary constant). This may come across as a surprise: is the problem even solvable information-theoretically given such few samples? It turns out that the sample complexity of density estimation for our hard instance, and also the hard instance of [36], is poly-logarithmic in $n$. In fact, we show a quasi-polynomial time algorithm that does density estimation for our hard instance with $(\log n)^{1+2\epsilon}$ samples. In other words, this gives us a tight computational gap for density estimation for the Gaussian mixture instances we consider.

These results are summarized more precisely and succinctly in Fig. 1. The reader is referred to the full version of our paper for the proofs [17].

Theorem 2 (Informal). We give the following lower bounds for GMM density estimation based on LWE assumptions of varying strength.

1) Assuming standard polynomial hardness of LWE, any density estimator for $\mathbb{R}^n$ that can solve arbitrary mixtures with at most $n^\epsilon$ Gaussian components, given $\text{poly}(n)$ samples from the mixture, requires super-polynomial time in $n$ for arbitrary constant $\epsilon > 0$.

2) For constant $\epsilon \in (0, 1)$, assuming $\ell$-dimensional LWE is hard to distinguish with advantage $1/2^\ell$ in time $2^{\ell^\epsilon}$, any density estimator for $\mathbb{R}^n$ that can solve arbitrary mixtures with at most roughly $(\log n)^{1/2+\epsilon}$ Gaussian components, given $\text{poly}(n)$ samples from the mixture, requires super-polynomial in $n$.

3) For constant $\epsilon \in (0, 1)$, assuming $\ell$-dimensional LWE is hard to distinguish with advantage $1/\text{poly}(\ell)$ in time $2^{\ell^\epsilon}$, any density estimator for $\mathbb{R}^n$ that can solve arbitrary mixtures with at most roughly $(\log n)^{1/2+\epsilon}$ Gaussian components, given $\text{poly}(\log n)$ samples from the mixture, requires super-polynomial in $n$ time.

C. Other Applications

Recent results have shown reductions from CLWE to other learning tasks as well, including learning a single periodic neuron [39], detecting backdoors in certain models [40], and improperly learning halfspaces in various error models [41], [42]. Our main result allows these results to be based on the hardness of LWE instead of CLWE.

In fact, we mention that our reduction can be used to show further hardness of the above learning tasks.

1More precisely, Diakonikolas, Kane, Manurangsi and Ren [42] use our techniques to reduce from LWE instead of CLWE.
For example, Song, Zadik and Bruna [39] directly show CLWE-hardness of learning single periodic neurons, i.e., neural networks with no hidden layers and a periodic activation function $\varphi(t) = \cos(2\pi \gamma t)$ with frequency $\gamma$. Our reduction from LWE to CLWE shows that this hardness result can be based directly on LWE instead of worst-case lattice assumptions, as done in [10]. Furthermore, our results expand the scope of their reduction in two ways:

1) Their reduction shows hardness of learning periodic neurons with frequency $\gamma \geq \sqrt{n}$, while ours, based on exponential hardness of LWE, applies to frequencies almost as small as $\gamma = \log n$, which covers a larger class of periodic neurons.

2) Second, the hardness of $k$-sparse CLWE from (standard) LWE shows that even learning sparse features (instead of features drawn from the unit sphere $S^{n-1}$) is hard under LWE for appropriate parameter settings.

This flexibility in $\gamma$ and in the sparsity of the secret distribution translates similarly for the other learning tasks mentioned, namely detecting backdoors in certain models [40] and improperly learning halfspaces in various error models [41], [42]. For hardness of detecting backdoors [40], this flexibility means reducing the magnitude of undetectable backdoor perturbations (in $\ell_2$ and $\ell_0$ norms). For hardness of learning halfspaces, this flexibility means that agnostically learning noisy halfspaces is hard even if the optimal halfspace is now sparse.2

### D. Perspectives and Future Directions

The main technical contribution of our paper is a reduction from the learning with errors (LWE) problem to its continuous analog, CLWE. A powerful outcome of our reduction is the fact that one can now bring to bear powerful tools from the study of the LWE problem to the study of continuous LWE and its downstream applications. We show two such examples in this paper: the first is a classical worst-case to average-case reduction from the approximate shortest vector problem to LWE [1], [43]. Secondly, while there has been some initial exploration of the cryptographic applications of the continuous LWE problem [12], constructing qualitatively new cryptographic primitives or qualitatively better cryptographic constructions is an exciting research direction. A recent example is the result of [40] who show use the hardness of CLWE to undetectably backdoor neural networks.

Finally, in terms of the hardness of learning mixtures of Gaussians, the question remains: what is the true answer? The best algorithms for learning mixtures of Gaussians [28] run in polynomial time only for a constant number of Gaussians. We show hardness (under a plausible setting of LWE) for roughly $\sqrt{\log n}$ Gaussians.

In our hard instance, the Gaussian components live on a line, and indeed a one-dimensional lattice. For such Gaussians, we know from Bruna et al. [10] that there exists an algorithm running in time roughly $2^{O(g^2)}$, which becomes almost polynomial at the extremes of our parameter settings. Thus, we show the best lower bound possible for our hard instance. (In fact, for our hard instance, we can afford to enumerate over all sparse

| LWE Assumption (samples, time, adv.) | Number of Gaussian Components | Run-time | Number of Samples |
|--------------------------------------|-------------------------------|----------|------------------|
| $\ell^{1/\epsilon}, \text{poly}(\ell), \frac{1}{\text{poly}(\ell)}$ | $O(n^{s/2} \cdot \log n)$ | $\log^{\Omega(1)}(n)$ | $\text{poly}(n)$ |
| $2^{\ell^{s}}, 2^{\text{poly}(\ell^s)}, \frac{1}{\text{poly}(\ell^s)}$ | $O\left(\left(\log n\right)^{2+\frac{s}{\ell}} \cdot \sqrt{\log \log n}\right)$ | $\Omega\left(2^\left(\log n\right)^{s/\ell}\right)$ | $\text{poly}(n)$ |
| $\ell^{s}, 2^{\text{poly}(\ell^s)}, \frac{1}{\text{poly}(\ell^s)}$ | $O\left(\left(\log n\right)^{2s} \cdot \log \log n\right)$ | $\Omega\left(2^\left(\log n\right)^{s/\ell}\right)$ | $\text{poly}(\log n)$ |

Fig. 1. This tables summarizes our hardness results for density estimation of GMM. Throughout, $\delta, \epsilon \in (0, 1)$ are arbitrary constants with $\delta < \epsilon$, $\ell$ is the dimension of LWE, and the Gaussians live in $\mathbb{R}^n$. “Adv.” stands for the advantage of the LWE distinguisher. As an example, the first row says for an arbitrary constant $0 < \epsilon < 1$, assuming standard, decisional LWE has no solver in dimension $\ell$ with $1/\text{poly}(\ell)$ advantage given $\ell^{s/\epsilon}$ samples and $\text{poly}(\ell)$ time, then any algorithm solving GMM density estimation given access to $\text{poly}(n)$ samples from an arbitrary Gaussian mixture with at most $O(n^{s/2} \cdot \log n)$ Gaussian components must take super-polynomial in $n$ time.
secret directions to get a solver with a similar run-time as [10] but with much smaller sample complexity.)

There remain three possibilities:

- There is a different hard instance for learning any super-constant number of Gaussians in polynomial time, and hardness can be shown by reduction from lattice problems; or
- There is a different hard instance for learning any super-constant number of Gaussians in polynomial time, but lattice problems are not the source of hardness; or
- We live in algorithmica, where the true complexity of Gaussian mixture learning is better than \( n^{f(g)} \) and looks perhaps more like \( \text{poly}(n) \cdot 2^g \), despite what SQ lower bounds suggest [36].

If we believe in the first two possibilities, a natural place to look for a different hard instance is [36], who consider a family of \( g \) Gaussian pancakes centered at the roots of a Hermite polynomial. This allows them to match the first \( 2g - 1 \) moments with that of the standard Gaussian. A tantalizing open problem is to try and prove hardness for their distribution for all algorithms, not just SQ algorithms, possibly under some cryptographic assumptions or perhaps even lattice assumptions.

### II. Technical Overview

#### A. From Fixed-Norm LWE to CLWE

The goal of our main theorem (Theorem 1) is to reduce from the fixed-norm LWE problem to CLWE. This involves a number of transformations, succinctly summarized in Fig. 2. Given samples \( (a, b) = (a, s) + e \ (\text{mod } q) \in \mathbb{Z}_q^{n+1} \), we do the following:

1. First, we turn the errors \( (b) \) from discrete to continuous Gaussians by adding a small continuous Gaussian to the LWE samples, using the smoothing lemma [44].
2. Secondly, we turn the samples \( a \) from discrete to continuously uniform over the torus by doing the same thing, namely adding a continuous Gaussian noise, and once again invoking appropriate smoothing lemmas from [1], [44].
3. Third, we go from uniform samples \( a \) to Gaussian samples. Boneh, Lewi, Montgomery and Raghnathan [45] give a general reduction from \( U(\mathbb{Z}_q^n) \) samples to “coset-sampleable” distributions, and as one example, they show how to reduce discrete uniform samples to discrete Gaussian samples, at the cost of a \( \log q \) multiplicative overhead in the dimension, which is unavoidable information-theoretically. We improve this reduction and circumvent this lower bound in the continuous version by having no overhead in the dimension, i.e. the dimension of both samples are the same. The key ingredient to this improvement is a simple Gaussian pre-image sampling algorithm, which on input \( z \sim U([0,1]) \), outputs \( y \) such that \( y = z \ (\text{mod } 1) \) and \( y \) is statistically close to a continuous Gaussian (when marginalized over \( z \sim U([0,1]) \)). (See Lemma 7 for a more precise statement.)
4. This finishes up our reduction! The final thing to do is to scale down the secret and randomly rotate it to ensure that it is a uniformly random unit vector.

We note that up until the final scaling down and re-randomization step, our reduction is secret-preserving.

#### B. Hardness of Gaussian Mixture Learning

Bruna et al. [10] show that a homogeneous version of CLWE, called hCLWE, has a natural interpretation as a certain distribution of mixtures of Gaussians. They show that any distinguisher between the hCLWE distribution and the standard multivariate Gaussian is enough to solve CLWE. Therefore, an algorithm for density estimation for Gaussian mixtures, which is a harder problem than distinguishing between that mixture and the standard Gaussian, implies a solver for CLWE. The condition that \( g > \sqrt{n} \) is a consequence of their reduction from worst-case lattice problems.

Our direct reduction from LWE to CLWE opens up a large toolkit of techniques that were developed in LWE-based cryptography. In this work, we leverage tools from leakage-resilient cryptography [3], [13], [14] to improve and generalize the hard instance of [10]. The key observation is that the number of Gaussians \( g \) in the mixture at the end of the day roughly corresponds...
to the norm of the secrets in LWE. Thus, the hardness of LWE with low-norm secrets will give us the hardness of Gaussian mixture learning with a small number of Gaussians.

Indeed, we achieve this by reducing LWE to \( k \)-sparse LWE. We call a vector \( s \in \{+1, 0, -1\}^n \) \( k \)-sparse if it has exactly \( k \) non-zero entries. We show the following result:

**Theorem 3** (Informal). Assume LWE in dimension \( \ell \) with \( n \) samples is hard with secrets \( s \sim \mathbb{Z}_q^\ell \) and errors of width \( \sigma \). Then, LWE in dimension \( n \) with \( k \)-sparse secrets is hard for errors of width \( \delta \mathcal{O}(\sqrt{\mathcal{K} \cdot \sigma}) \), as long as \( k \log(n/k) \geq \ell \log_2(q) \) and \( \sigma \geq 1 \).

It turns out that for our purposes, the quantitative tightness of our theorem is important. Namely, we require that the blowup in the noise depends polynomially only on \( k \) and not on other parameters. Roughly speaking, the reason is that if we have a blow-up factor of \( r \), for our LWE assumption, we need \( q/\sigma \gg r \) for the resulting CLWE distribution to be meaningful. For our parameter settings, if \( r \) depends polynomially on the dimension \( n \) (the dimension of the ambient space for the Gaussians) or the number of samples \( m \), then we require sub-exponentially large modulus-to-noise ratio in our LWE assumption, which is a notably stronger assumption. Indeed, the noise blow-up factor of the reduction we achieve and use is \( O(\sqrt{k}) \).

Our proof of this theorem uses a variant of the proof of [13] to work with \( k \)-sparse secrets.\(^3\) We note that Brakerski and Döttling [14] give a general reduction from LWE to LWE with arbitrary secret distributions with large enough entropy, but the noise blowup when applying their results directly to \( k \)-sparse secrets is roughly \( \sqrt{kmn} = k^{o(1)} \) for parameter settings we consider.

For a full description of the proof of Theorem 3, the reader is referred to the full version of our paper [17].

**Organization**

In Section III, we introduce the necessary technical ingredients for the rest of the paper. In Section IV, we prove our main result: we first prove Theorem 4, the reduction from fixed-norm LWE to CLWE, and then we finally prove Corollary 4, a full reduction from LWE to CLWE. We defer details and remaining proofs to the full version of our paper [17].

---

\(^3\)The techniques of Brakerski et al. [3], who show the hardness of binary secret LWE, can also be easily modified to prove \( k \)-sparse hardness, but the overall reduction is somewhat more complex. For this reason, we choose to show how to modify the reduction of [13].

---

**III. Preliminaries**

For a distribution \( D \), we write \( x \sim D \) to denote a random variable \( x \) being sampled from \( D \). For any \( n \in \mathbb{N} \), we let \( D^n \) denote the \( n \)-fold product distribution, i.e. \( (x_1, \ldots, x_n) \sim D^n \) is generated by sampling \( x_i \overset{\text{i.i.d.}}{\sim} D \) independently. For any finite set \( S \), we write \( U(S) \) to denote the discrete uniform distribution over \( S \); we abuse notation and write \( x \sim S \) to denote \( x \sim U(S) \). For any continuous set \( S \), we write \( U(S) \) to denote the continuous uniform distribution over \( S \) (i.e. having support \( S \) and constant density); we also abuse notation and write \( x \sim S \) to denote \( x \sim U(S) \).

For distributions \( D_1, D_2 \) supported on a measurable set \( X \), we define the statistical distance between \( D_1 \) and \( D_2 \) to be \( \Delta(D_1, D_2) = \frac{1}{2} \int_{x \in X} |D_1(x) - D_2(x)| dx \). We say that distributions \( D_1, D_2 \) are \( \epsilon \)-close if \( \Delta(D_1, D_2) \leq \epsilon \). For a distinguisher \( A \) running on two distributions \( D_1, D_2 \), we say that \( A \) has advantage \( \epsilon \) if

\[
\Pr_{x \sim D_1} [A(x) = 1] - \Pr_{x \sim D_2} [A(x) = 1] = \epsilon,
\]

where the probability is also over any internal randomness of \( A \).

For any vector \( v \in \mathbb{R}^n \), we write \( \|v\| \) to mean the standard \( \ell_2 \)-norm of \( v \). For \( n \in \mathbb{N} \), we let \( S^{n-1} \subseteq \mathbb{R}^n \) denote the \( (n - 1) \)-dimensional sphere embedded in \( \mathbb{R}^n \), or equivalently the set of unit vectors in \( \mathbb{R}^n \). By \( \mathbb{Z}_q \), we refer to the ring of integers modulo \( q \), represented by \( \{0, \ldots, q - 1\} \). By \( \mathbb{T}_q \), we refer to the set \( \mathbb{R}/q\mathbb{Z} = [0, q) \subseteq \mathbb{R} \) where addition (and subtraction) is taken modulo \( q \) (i.e. \( \mathbb{T}_q \) is the torus scaled up by \( q \)). We denote \( \mathbb{T} \equiv \mathbb{T}_1 \) to be the standard torus. By taking a real number mod \( q \), we refer to taking its representative as an element of \( \mathbb{T}_q \) in \([0, q)\) unless stated otherwise.

**A. Lattices and Discrete Gaussians**

A rank \( n \) integer lattice is a set \( \Lambda = B\mathbb{Z}^n \subseteq \mathbb{Z}^d \) of all integer linear combinations of \( n \) linearly independent vectors \( B = [b_1, \ldots, b_n] \in \mathbb{Z}^d \). The dual lattice \( \Lambda^* \) of a lattice \( \Lambda \) is defined as the set of all vectors \( y \in \mathbb{R}^d \) such that \( \langle x, y \rangle \in \mathbb{Z} \) for all \( x \in \Lambda \).

For arbitrary \( x \in \mathbb{R}^n \) and \( c \in \mathbb{R}^n \), we define the Gaussian function

\[
\rho_{s,c}(x) = \exp(-\pi \|x - c\|^2).
\]

Let \( D_{s,c} \) be the corresponding distribution with density at \( x \in \mathbb{R}^n \) given by \( \rho_{s,c}(x)/s^n \), namely the \( n \)-dimensional Gaussian distribution with mean \( c \) and covariance matrix \( s^2/(2\pi) \cdot I_{n \times n} \), where \( I_{n \times n} \) is the \( n \times n \) identity matrix. When \( c = 0 \), we omit the subscript notation of \( c \) on \( \rho \) and \( D \).
For an $n$-dimensional lattice $\Lambda \subseteq \mathbb{R}^n$ and point $c \in \mathbb{R}^n$, we can define the 
 discrete Gaussian of width $s$ to be given by the mass function
\[
D_{\Lambda+c,s}(x) = \frac{\rho_s(x)}{\rho_s(\Lambda + c)}
\]
supported on $x \in \Lambda + c$, where by $\rho_s(\Lambda + c)$ we mean
$\sum_{y \in \Lambda} \rho_s(y + c)$. We now give the smoothing parameter as defined by [1] and some of its standard properties.

**Definition 2 ([1], Definition 2.10).** For an $n$-dimensional lattice $\Lambda$ and $\epsilon > 0$, we define $\eta_\epsilon(\Lambda)$ to be the smallest $s$ such that $\rho_{1/s}(\Lambda^* \setminus \{0\}) \leq \epsilon$.

**Lemma 1 ([1], Lemma 2.12).** For an $n$-dimensional lattice $\Lambda$ and $\epsilon > 0$, we have
\[
\eta_\epsilon(\Lambda) \leq \sqrt{\frac{\ln(2n(1 + 1/\epsilon))}{\pi}} \cdot \lambda_n(\Lambda).
\]
Here $\lambda_i(\Lambda)$ is defined as the minimum length of the longest vector in a set of $i$ linearly independent vectors in $\Lambda$.

**Lemma 2 ([1], Corollary 3.10).** For any $n$-dimensional lattice $\Lambda$ and $\epsilon \in (0, 1/2)$, $\sigma, \sigma' \in \mathbb{R}_{>0}$, and $z, u \in \mathbb{R}^n$, if
\[
\eta_\epsilon(\Lambda) \leq \frac{1}{\sqrt{1/(\sigma')^2 + (\|z\|/\sigma)^2}},
\]
then if $v \sim D_{\Lambda + u, \sigma'}$ and $\epsilon \sim D_\sigma$, then $\langle z, v \rangle + \epsilon$ has statistical distance at most $4\epsilon$ from $D_{\Lambda \setminus \left\{0\right\}}$.

**Lemma 3 ([44], Lemma 4.1).** For an $n$-dimensional lattice $\Lambda$, $\epsilon > 0$, $c \in \mathbb{R}^n$, for all $s \geq \eta_\epsilon(\Lambda)$, we have
\[
\Delta(D, \epsilon \mod P(\Lambda), U(P(\Lambda))) \leq \epsilon/2,
\]
where $P(\Lambda)$ is the half-open fundamental parallelepiped of $\Lambda$.

**Lemma 4 ([44], implicit in Lemma 4.4).** For an $n$-dimensional lattice $\Lambda$, for all $\epsilon > 0$, $c \in \mathbb{R}^n$, and all $s \geq \eta_\epsilon(\Lambda)$, we have
\[
\rho_s(\Lambda + c) = \rho_{s-c}(\Lambda) \in \left[\frac{1 - \epsilon}{1 + \epsilon}, 1\right] \cdot \rho_s(\Lambda).
\]

### B. Learning with Errors

Throughout, we work with decisional versions of LWE and CLWE.

**Definition 3 (LWE Distribution).** Let $n, m, q \in \mathbb{N}$, let $A$ be a distribution over $\mathbb{R}^n$, $S$ be a distribution over $\mathbb{Z}^n$, and $E$ be a distribution over $\mathbb{R}$. We define $\text{LWE}(m, A, S, E)$ to be distribution given by sampling $a_1, \ldots, a_m \sim A$, $s \sim S$, and $e_1, \ldots, e_m \sim E$, and outputting $(a_i, s^\top a_i + e_i \mod q)$ for all $i \in [m]$. We refer to $n$ as the dimension and $m$ as the number of samples. (The modulus $q$ is suppressed from notation for brevity as it will be clear from context.)

**Definition 4 (CLWE Distribution [10]).** Let $n, m \in \mathbb{N}, \gamma, \beta \in \mathbb{R}$, and let $A$ be a distribution over $\mathbb{R}^n$ and $S$ be a distribution over $\mathbb{S}$. Let $\text{CLWE}(m, A, S, \gamma, \beta)$ be the distribution given by sampling $a_1, \ldots, a_m \sim A$, $s \sim S$, $e_1, \ldots, e_m \sim D_\gamma$ and outputting $(a_i, \gamma \cdot (a_i, s) + e_i \mod 1)$ for all $i \in [m]$. Explicitly, for one sample, the density at $(y, z) \in \mathbb{R}^n \times [0, 1)$ is proportional to
\[
\mathcal{A}(y) \cdot \sum_{k \in \mathbb{Z}} \rho_\beta(z + k - \gamma \cdot (y, s))
\]
for fixed secret $s \sim S$. We refer to $n$ as the dimension and $m$ as the number of samples. We omit $S$ if $S = U(S)$, as is standard for CLWE.

### IV. Reducing LWE to CLWE

Our main result in this section is a reduction from fixed-norm LWE to CLWE:

**Theorem 4 (Fixed-Norm LWE to CLWE).** Let $r \in \mathbb{R}_{\geq 1}$, and let $S$ be an arbitrary distribution over $\mathbb{Z}^n$ where all elements in the support of $S$ have $\ell_2$ norm $r$. Then, for
\[
\gamma = r \cdot \sqrt{\ln(m) + \ln(n) + \omega(\log \lambda)},
\]
\[
\beta = O\left(\frac{\sigma}{q}\right),
\]
if there is no $T + \poly(n, m, \log(q), \log(\sigma), \log(\lambda))$ time distinguisher between $\text{LWE}(m, \mathbb{Z}_q^n, S, D_{\gamma, \sigma})$ and $U(\mathbb{Z}_q^n \times \mathbb{Z}_m^n)$ with advantage at least $\epsilon - \negl(\lambda)$, then there is no $T$-time distinguisher between $\text{CLWE}(m, D_n, \frac{1}{2r} \cdot S, \gamma, \beta)$ and $D_n^{1 \times m} \times U(\mathbb{S}^m)$ with advantage $\epsilon$, as long as $\sigma \geq 3r\sqrt{\ln(m) + \ln(n) + \omega(\log \lambda)}$.

See Fig. 2 for a summary of the steps. We note that the dimension and number of samples remains the same in this reduction, and the advantage stays the same up to additive $\negl(\lambda)$ factors. We also remark that to keep the theorem general, the final distribution is not exactly the CLWE distribution, as the secret distribution is $\frac{1}{2r} \cdot S$ instead of $U(S)$. However, using Lemma 9, it is straightforward to reduce from $\frac{1}{2r} \cdot S$ secrets to $U(S)$ secrets.

This reduction goes via a series of transformations, which is briefly outlined in Section II.

**Step 1:** Converting discrete errors to continuous
errors. First, we make the error distribution statistically close to a continuous Gaussian instead of a discrete Gaussian. Essentially, all we do is add a small continuous Gaussian noise to the second component and argue that this makes the noise look like a continuous Gaussian instead of a discrete one.

This sort of reduction is standard in the literature, but we provide it here for completeness.

**Lemma 5.** Let \( n, m, q \in \mathbb{N} \), \( \sigma \in \mathbb{R}_{>0} \), and suppose \( \sigma > \sqrt{4 \ln m + \omega(\log \lambda)} \). For any distribution \( S \) over \( \mathbb{Z}^n \), suppose there is no distinguisher between \( \text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma}) \) and \( U(\mathbb{Z}_q^n \times \mathbb{Z}_q^{m}) \) running in time \( T + \text{poly}(m, n, \log(q), \log(\sigma)) \). Then, there is no \( T \)-time distinguisher \( \text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma}) \) and \( U(\mathbb{Z}_q^n \times \mathbb{Z}_q^{m}) \) with an additive \( \omega(\log \lambda) \) advantage loss, where

\[
\sigma' = \sqrt{\sigma^2 + 4 \ln(m) + \omega(\log \lambda)} = O(\sigma).
\]

**Proof.** We run our original distinguisher for \( \text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma'}) \) and \( U(\mathbb{Z}_q^n \times \mathbb{Z}_q^{m}) \). For every sample \((a, b)\) (from either \( \text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma}) \) or \( U(\mathbb{Z}_q^n \times \mathbb{Z}_q^{m}) \)), we sample a continuous Gaussian \( \epsilon' \sim D_{\sigma'} \) where \( \sigma' \) will be set later, and send \((a, b + \epsilon' \mod q)\) to the distinguisher.

By Lemma 3, we know that the distribution of \( \epsilon' \) (mod 1) has statistical distance at most \( \epsilon = \omega(\log \lambda)/m \) as long as \( \sigma' \geq \eta_\epsilon(Z) \). Therefore, if we are given samples from \( U(\mathbb{Z}_q^n \times \mathbb{Z}_q^{m}) \), due to symmetry of \( b \sim \mathbb{Z}_q \), we can set \( \epsilon = \omega^{(1)}(m) \) to have \( b + \epsilon' \mod q \) look \( \omega(\log \lambda) \)-close to \( T_q \), making it look samples from \( U(\mathbb{Z}_q^n \times \mathbb{Z}_q^{m}) \).

If we are given samples from \( \text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma}) \), then the second component can be seen as having noise \( e + \epsilon' \), where \( e \sim D_{\sigma}, \epsilon' \sim D_{\sigma'} \). Applying Lemma 2, as long as \( 1/\sqrt{\sigma^2 + 1/(\sigma'^2)} \geq \eta_\epsilon(Z) \), then \( e + \epsilon' \) will look \( \omega(\log \lambda) \)-close to \( T_q \). Thus, as long as \( \sigma' \geq \sqrt{2} \cdot \eta_\epsilon(Z) \), it all goes through, as taking \( \epsilon \mod \sigma \) (i.e. in \( T_q \) instead of \( \mathbb{R} \)) can only decrease statistical distance. Now, applying Lemma 1, we can set \( \epsilon = \omega^{(1)}(m) \) and \( \sigma'' \geq \sqrt{4 \ln(m) + \omega(\log \lambda)} \), and as long as \( \sigma > \sqrt{4 \ln(m) + \omega(\log \lambda)} \), all goes through. Now, doing the triangle inequality over all \( m \) samples, we get \( \omega(\log \lambda) \)-closeness of all samples.

**Step 2: Converting discrete to continuous samples.**

Now, we convert discrete uniform samples \( a \sim \mathbb{Z}_q^n \) to continuous uniform samples \( a \sim T_q^n \).

**Lemma 6.** Let \( n, m, q \in \mathbb{N} \), \( \sigma \in \mathbb{R} \). Let \( S \) be a distribution over \( \mathbb{Z}^n \) where all elements in the support have fixed norm \( r \), and suppose that

\[
\sigma \geq 3r \sqrt{\ln n + \ln m + \omega(\log \lambda)}.
\]

Suppose there is no \( T + \text{poly}(m, n, \log(q), \log(\sigma)) \)-time distinguisher between the distributions \( \text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma}) \) and \( U(\mathbb{Z}_q^n \times \mathbb{Z}_q^{m}) \). Then, there is no \( T \)-time distinguisher between the distributions \( \text{LWE}(m, T_q^n, S, D_{\sigma}) \) and \( U(\mathbb{Z}_q^n \times T_q^m) \) with an additive \( \omega(\log \lambda) \) advantage loss, where we set

\[
\sigma' = \sqrt{\sigma^2 + 9r^2(\ln n + \ln m + \omega(\log \lambda))} = O(\sigma).
\]

**Proof.** We run our original distinguisher for \( \text{LWE}(m, T_q^n, S, D_{\sigma}) \) and \( U(T_q^n \times T_q^m) \). Let \( \epsilon = \omega(\log \lambda)/m \), and let \( \sigma'' \geq \sqrt{2} \cdot \eta_\epsilon(Z) \). For each sample \((a, b)\) (from either \( \text{LWE}(m, T_q^n, S, D_{\sigma}) \) or \( U(T_q^n \times T_q^m) \)), we sample a continuous Gaussian \( a' \sim D_{\sigma''} \) and send \((a + a' \mod q, b)\) to the distinguisher. By Lemma 3, we know that the distribution of \( a' \) (mod 1) has statistical distance at most \( \epsilon = \omega(\log \lambda)/m \) to \( U([0, 1]^n) \). Thus, by symmetry over \( a \sim (\mathbb{Z}_q^n)^n \), the distribution of \( a + a' \) (mod \( q \)) will be \( \omega(\log \lambda)/m \)-close to uniform over \( (T_q^n)^n \). Therefore, by the triangle inequality, if we are given samples from \( U(\mathbb{Z}_q^n \times T_q^m) \), the reduction gives samples to the distinguisher that are \( \omega(\log \lambda) \)-close to \( U(T_q^n \times T_q^m) \).

If we are given samples from \( \text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma}) \), then the reduction gives us (taking everything mod \( q \))

\[
(a + a', \langle a, s \rangle + e) = (a + a', (a + a', s) + e - \langle a', s \rangle) = (a + a', (a + a', s) + e'),
\]

where we define \( e' = e - \langle a', s \rangle \) over \( \mathbb{R} \). Conditioned on \( a + a' \mod q \), \( a' \) is a discrete Gaussian distributed according to \( D_{\omega^{(1)} + \omega^{(1)} \cdot s} \). By Lemma 2, as long as \( \sigma \geq \omega(\log \lambda)/m \), the distribution of \( e' \) is \( O(e) = \omega(\log \lambda)/m \)-close to \( D_{\sigma''} \), where \( \sigma' = \sqrt{\sigma^2 + 9r^2(\sigma''^2)} \). Averaging the distribution of \( e' \) over \( s \) will not change the distribution over \( e' \), as all secrets \( s \) have fixed norm \( r \). Therefore, if we are given the \( m \) samples from \( \text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma}) \), the reduction gives us samples \( \omega(\log \lambda) \)-close to \( \text{LWE}(m, T_q^n, S, D_{\sigma}) \), as desired.

To set parameters, we choose \( \sigma'' = 3r \sqrt{\ln n + \ln m + \omega(\log \lambda)} \) to ensure that \( \sigma'' \geq \sqrt{2} \cdot \eta_\epsilon(Z) \). This gives

\[
\sigma' = \sqrt{\sigma^2 + 9r^2(\ln n + \ln m + \omega(\log \lambda))},
\]

along with the requirement that

\[
\sigma \geq r \sigma'' = 3r \sqrt{\ln n + \ln m + \omega(\log \lambda)}.
\]

**Step 3: Converting uniform to Gaussian samples.**
Lemma 7. Let $t \in \mathbb{R}_{>0}$ be a parameter. There is a $\text{poly}(n, \log(t), \log(\lambda))$-time algorithm such that on input $z \in T_1^n$, the algorithm outputs some $y \in \mathbb{R}^n$ such that $y = z \pmod{1}$. Moreover, if $z \sim U(T_1^n)$, then the distribution on the outputs $y$ is $\text{neg}(\lambda)/t$-close to $D_\tau$, where $\tau = \sqrt{\ln n + \ln t + \omega(\log \lambda)}$.

Remark 1. In the discrete setting, there is in some sense a necessary multiplicative $\Omega(\log q)$ overhead in the dimension due to entropy arguments, but the above shows that we can overcome that barrier in the continuous case.

Proof. We give each coordinate of $y$ separately. By the triangle inequality, it suffices to show how to sample $y \in \mathbb{R}$ such that $y = z \pmod{1}$ and such that if $z \sim T_1$, then $y$ is $\text{neg}(\lambda)/(tn)$-close to $D_\tau$. We sample

$$y \sim D_{\mathbb{Z} + z, \tau},$$

which can be sampled efficiently (see e.g. [3], Section 5.1 of full version), where we have $\text{neg}(\lambda)/(tn)$ statistical distance between $y$ and $D_{\mathbb{Z} + z, \tau}$, and always satisfy $y \in \mathbb{Z} + z$. Since $y \in \mathbb{Z} + z$, it follows that $y = z \pmod{1}$.

Now, we need to argue that the distribution of $y$ looks $\text{neg}(\lambda)/(tn)$-close to $D_\tau$ when $z \sim U(T_1)$. Note that for fixed $z \in [0, 1)$, we can write the generalized PDF of $D_{\mathbb{Z} + z, \tau}$ as

$$D_{\mathbb{Z} + z, \tau}(x) = \delta(x - z \pmod{1}) \cdot \frac{\rho_\tau(x)}{\rho_\tau(\mathbb{Z} + z)}$$

for arbitrary $x \in \mathbb{R}$, where $\delta(\cdot)$ is the Dirac delta function. Thus, as long as $\tau \geq \eta_\varepsilon(\mathbb{Z})$ (for $\varepsilon$ set later), the density of the marginal distribution $D_{\mathbb{Z} + z, \tau}$ where $z \sim U([0, 1))$ is given by

$$D_{\mathbb{Z} + U([0, 1)), \tau}(x) = \int_0^1 \cdot D_{\mathbb{Z} + z, \tau}(x) \cdot dz = \int_0^1 \delta(x - z \pmod{1}) \cdot \frac{\rho_\tau(x)}{\rho_\tau(\mathbb{Z} + z)} \cdot dz$$

$$= \frac{\rho_\tau(x)}{\rho_\tau(\mathbb{Z} + x)}$$

$$\in \left[1, \frac{1 + \varepsilon}{1 - \varepsilon}\right] \cdot \frac{\rho_\tau(x)}{\rho_\tau(\mathbb{Z})} \cdot \left[1, \frac{1 + \varepsilon}{1 - \varepsilon}\right] \cdot \rho_\tau(x),$$

where the inclusion comes from Lemma 4. Therefore, a standard calculation shows that the statistical distance between $D_{\mathbb{Z} + U([0, 1)), \tau}$ and $D_\tau$ is at most $O(\varepsilon)$. Setting $\varepsilon = \lambda^{-\omega(1)/(t-n)}$, we need to take $\tau \geq \eta_\lambda^{-\omega(1)/(t-n)}(\mathbb{Z})$, which we can do by setting $\tau = \sqrt{\ln n + \ln t + \omega(\log \lambda)}$ by Lemma 1. \hfill \Box

Lemma 8. Let $n, m, q \in \mathbb{N}, \sigma, r, \gamma, \beta \in \mathbb{R}$. Let $S$ be a distribution over $\mathbb{Z}^n$ where all elements in the support have fixed norm $r$. Suppose there is no $T + \text{poly}(n, m, \log(q), \log(\lambda))$ time distinguisher between the distributions $\text{LWE}(m, T_q^n, S, \gamma, \beta)$ and $D_1^{n \times m} \times U(T_q^n)$. Then, there is no $T$-time distinguisher between the distributions $\text{CLWE}(m, D_1^n, \frac{1}{r} \cdot S, \gamma, \beta)$ and $D_1^{n \times m} \times U(T_q^n)$ with an additive advantage loss of $\text{neg}(\lambda)$, where

$$\gamma = r \cdot \frac{\sqrt{\ln n + \ln m + \omega(\log \lambda)}}{q},$$

$$\beta = \frac{\sigma}{q}.$$

Proof. We run the distinguisher for $\text{CLWE}(m, D_1^n, \frac{1}{r} \cdot S, \gamma, \beta)$ and $D_1^{n \times m} \times U(T_q^n)$ for each sample $(a, b)$ from either $\text{LWE}(m, T_q^n, S, D_\sigma)$ or $U(T_q^n \times T_q^n)$, we invoke Lemma 7 on $a/q$ with parameter $t = m$ to get some $y \in \mathbb{R}^n$ with statistical distance $\text{neg}(\lambda)/m$ from $D_\sigma$ such that $y = a/q (\pmod{1})$, where $\tau = \sqrt{\ln n + \ln m + \omega(\log \lambda)}$. We then send $(y/\tau, b/q)$ to the distinguisher. Let $\gamma = r \cdot \tau$, $y' = y/\tau$, $s' = s/\tau$, and $e' = e/q$. If $(a, b)$ is a sample from $\text{LWE}(m, T_q^n, S, D_\sigma)$, then for secret $s \sim S$, since $s \in \mathbb{Z}^n$, we have

$$(y/\tau, b/q) = (y', (a/q, s) + e/\tau (\pmod{1})) = (y', r \cdot \tau \cdot (y', s/\tau) + e' (\pmod{1})) = (y', \gamma \cdot (y', s') + e' (\pmod{1}))$$

where this is now $\text{neg}(\lambda)/m$ close to a sample from $\text{CLWE}(m, D_1^n, \frac{1}{r} \cdot S, \gamma, \beta)$, as $y' \sim D_1^n$, $s' \sim \frac{1}{r} \cdot S$, and $e' \sim D_{\sigma/q} = D_{\beta}$. Applying this reduction to $U(T_q^n \times T_q^n)$ clearly gives us a statistically close sample to $D_1^{n \times m} \times U(T_q^n)$ by Lemma 7 and the triangle inequality over all $m$ samples. \hfill \Box

Step 4 (optional): Converting the secret to a random direction. The distribution on the secret as given above is not uniform over the sphere, so if desired, one can apply the worst-case to average-case reduction for CLWE ([10], Claim 2.22). Note that while we do not use Lemma 9 in proving Theorem 4, we do use it in the proof of Corollary 4.

Lemma 9 ([10], Claim 2.22). Let $n, m \in \mathbb{N}$, and let $\beta \in \mathbb{R}_{>0}$. Let $S$ be a distribution over $\mathbb{R}^n$ of fixed norm 1. There is no $T$-time distinguisher between the distributions $\text{LWE}(m, D_1^n, \gamma, \beta)$ and $D_1^{n \times m} \times U(T_q^n)$, assuming there is no $T + \text{poly}(n, m)$ time distinguisher between the distributions $\text{CLWE}(m, D_1^n, S, \gamma, \beta)$ and $D_1^{n \times m} \times U(T_q^n)$. \hfill \Box
$D_1^{n \times m} \times U(\mathbb{T}_m^n)$. That is, we can reduce CLWE to CLWE to randomize the secret to be a uniformly random unit vector instead of drawn from (possibly discrete) $S$.

Now, we are ready to prove the main theorem, Theorem 4.

Proof of Theorem 4. Throughout this proof, when we refer to distinguishing probability, we omit additive $\text{negl}(\lambda)$ terms for simplicity.

Suppose there is no distinguisher with advantage $\epsilon$ between $\text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma_2})$ and $U(\mathbb{Z}_q^n) \times U(\mathbb{T}_m^n)$, Then, by Lemma 5, there is no $\epsilon$-distinguisher between $\text{LWE}(m, \mathbb{Z}_q^n, S, D_{\sigma_2})$ and $U(\mathbb{Z}_q^{n \times m}) \times U(\mathbb{T}_m^n)$, where $\sigma_2 = O(\sigma)$, as long as $\sigma \geq 4\lambda \ln(m) + \omega(\log \lambda)$, which it is by our assumption on $\sigma$. Then, by Lemma 6, there is no $\epsilon$-distinguisher between $\text{LWE}(m, \mathbb{T}_m^n, S, D_{\sigma_3})$ and $U(\mathbb{T}_m^n) \times U(\mathbb{T}_m^n)$, where $\sigma_3 = O(\sigma_2) = O(\sigma)$, which holds as long as $\sigma_2 \geq 3r\sqrt{\ln(m) + \ln(n)} + \omega(\log \lambda)$, which it does because $\sigma_2 \geq \sigma \geq 3r\sqrt{\ln(m) + \ln(n)} + \omega(\log \lambda)$. Now, by Lemma 8, there is no $\epsilon$-distinguisher between $\text{CLWE}(m, D_1^{n \times m} \cdot S, \gamma, \beta)$ and $D_1^{n \times m} \times U(\mathbb{T}_m^n)$, where

$$\gamma = r \cdot \sqrt{\ln(m) + \ln(n)} + \omega(\log \lambda),$$

and

$$\beta = \frac{\sigma_3}{q} = O\left(\frac{\sigma}{q}\right),$$

as desired. □

A. Full Reduction from LWE to CLWE

We use Lemma 2.9 and Theorem 3.1 of [13] to reduce standard decisional LWE with secrets $s \sim U(\mathbb{Z}_q^n)$ to fixed-norm LWE. In particular, we reduce to LWE with secrets $s \sim U(\{-1, 1\}^n)$ of norm $\sqrt{n}$. Finally, we combine this with Theorem 4 and Lemma 9 to prove our main result.

Corollary 4 (Full Reduction from LWE to CLWE). Let $q, \ell, n, m \in \mathbb{N}$ with $m > n$, and let $\gamma, \beta, \sigma, \epsilon \in \mathbb{R}_{>0}$. There is no $T$-time distinguisher with advantage $\epsilon$ between $\text{CLWE}(m, D_1^{n \times m} \cdot \gamma, \beta)$ and $D_1^{n \times m} \times U(\mathbb{T}_m^n)$, assuming there is no $T + \text{poly}(\ell, n, m, \log(q), \log(\lambda))$ time distinguisher with advantage $(\epsilon - \text{negl}(\lambda))/(2m)$ between $\text{LWE}(m, \mathbb{Z}_q^{\ell \times n} \cdot \mathbb{Z}_q^{\ell \cdot m})$ and $U(\mathbb{T}_m^n) \times U(\mathbb{T}_m^n)$, for

$$\gamma = O\left(\sqrt{n} \cdot \sqrt{\ln(m) + \omega(\log \lambda)}\right),$$

$$\beta = O\left(\frac{\sigma\sqrt{n}}{q}\right),$$

as long as $\log(q)/2^{\ell} = \text{negl}(\lambda), n \geq 2^{\ell}\log_2 q + \omega(\log \lambda)$, and $\sigma \geq C \cdot \sqrt{\ln(m) + \omega(\log \lambda)}$ for some universal constant $C$.

Acknowledgment

We thank the anonymous reviewers for their valuable feedback on previous versions of our write-up, especially one reviewer for pointing out an error in a previous version of our full paper [17].

References

[1] O. Regev, “On lattices, learning with errors, random linear codes, and cryptography,” Journal of the ACM (JACM), vol. 56, no. 6, pp. 1–40, 2009. 1, 2, 4, 5, 6, 8

[2] C. Peikert, “Public-key cryptosystems from the worst-case shortest vector problem: extended abstract,” in Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009, M. Mitzenmacher, Ed. ACM, 2009, pp. 333–342. [Online]. Available: https://doi.org/10.1145/1536414.1536461 1, 2

[3] Z. Brakerski, A. Langlois, C. Peikert, O. Regev, and D. Stehlé, “Classical hardness of learning with errors,” in Proceedings of the forty-fifth annual ACM symposium on Theory of computing, 2013, pp. 575–584. 1, 2, 3, 6, 7, 10

[4] Z. Brakerski and V. Vaikuntanathan, “Efficient fully homomorphic encryption from (standard) $\text{LWE}$,” SIAM J. Comput., vol. 43, no. 2, pp. 831–871, 2014. [Online]. Available: https://doi.org/10.1137/120868669 1

[5] S. Gorbunov, V. Vaikuntanathan, and H. Wei, “Attribute-based encryption for circuits,” J. ACM, vol. 62, no. 6, pp. 45:1–45:33, 2015. [Online]. Available: http://doi.acm.org/10.1145/2824233 1

[6] A. R. Choudhuri, A. Jain, and Z. Jin, “Snargs for $\mathsf{SNP}$ from $\mathsf{LWE}$,” in 62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022. IEEE, 2021, pp. 68–79. [Online]. Available: https://doi.org/10.1109/FOCS52979.2021.00016 1

[7] A. R. Klivans and A. A. Sherstov, “Cryptographic hardness for learning intersections of halfspaces,” J. Comput. Syst. Sci., vol. 79, no. 1, pp. 2–12, 2009. [Online]. Available: https://doi.org/10.1016/j.jcss.2008.07.008 1

[8] R. Jawale, Y. T. Kalai, A. Jain, and A. R. Choudhuri, “Robustness of the learning with errors assumption,” in STOC ’21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021, S. Khuller and V. V. Williams, Eds. ACM, 2021, pp. 708–721. [Online]. Available: https://doi.org/10.1145/3406325.3451055 1

[9] S. Goldwasser, Y. T. Kalai, C. Peikert, and V. Vaikuntanathan, “Robustness of the learning with errors assumption,” 2010. 1, 2

[10] J. Bruna, O. Regev, M. J. Song, and Y. Tang, “Continuous LWE,” in Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, 2021, pp. 694–707. 1, 2, 3, 4, 5, 6, 8, 10

[11] C. Peikert, O. Regev, and N. Stephens-Davidowitz, “Pseudorandomness of ring-lwe for any ring and modulus,” in Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017, H. Hatami, P. McKenzie, and V. King, Eds. ACM, 2017, pp. 461–473. [Online]. Available: https://doi.org/10.1145/3055399.3055489 2

[12] A. Bogdanov, M. C. Noval, C. Hoffmann, and A. Rosen, “Public-key encryption from continuous LWE,” IACR Cryptol. ePrint Arch., p. 93, 2022. [Online]. Available: https://eprint.iacr.org/2022/093 2, 5

[13] D. Micciancio, “On the hardness of learning with errors with binary secrets,” Theory Comput., vol. 14, no. 1, pp. 1–17, 2018. [Online]. Available: https://doi.org/10.4086/toc.2018.v014a013 2, 6, 7, 11
[14] Z. Brakerski and N. Döttling, “Hardness of LWE on general entropic distributions,” in Advances in Cryptology - EUROCRYPT 2020 - 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020, Proceedings, Part II, ser. Lecture Notes in Computer Science, A. Canteaut and Y. Ishai, Eds., vol. 12106. Springer, 2020, pp. 551–575. [Online]. Available: https://doi.org/10.1007/978-3-030-45724-2_19 2, 6, 7

[15] R. Lindner and C. Peikert, “Better key sizes (and attacks) for lwe-based encryption,” in Cryptographers’ Track at the RSA Conference. Springer, 2011, pp. 319–339. 3

[16] NIST, “Post-quantum cryptography standardization,” https://csrc.nist.gov/Projects/Post-Quantum-Cryptography. 3, 4

[17] A. Gupte, N. Vafa, and V. Vaikuntanathan, “Continuous lwe is as hard as lwe & applications to learning gaussian mixtures,” arXiv preprint arXiv:2204.02550, 2022. 3, 4, 7

[18] D. Titterington, P. Titterington, S. M. A. Smith, U. Makov, and J. W. . Sons, Statistical Analysis of Finite Mixture Distributions, ser. Applied section. Wiley, 1985. [Online]. Available: https://books.google.com/books?id=hZ0QQAIAAJ

[19] G. J. McLachlan and D. Peel, Finite mixture models. Wiley series in Probability and Statistics. 2000. 3

[20] S. Dasgupta, “Learning mixtures of gaussians,” in 40th Annual Symposium on Foundations of Computer Science, FOCS ’99, 17-18 October, 1999, New York, NY, USA. IEEE Computer Society, 1999, pp. 634–644. [Online]. Available: https://doi.org/10.1109/SFFCS.1999.814639 3

[21] A. Sanjeev and R. Kannan, “Learning mixtures of arbitrary gaussians,” in Proceedings of the thirty-third annual ACM symposium on Theory of computing, 2001, pp. 247–257. 3

[22] S. Vempala and G. Wang, “A spectral algorithm for learning mixtures of distributions,” in The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings. IEEE, 2002, pp. 113–122. 3

[23] D. Achlioptas and F. McSherry, “On spectral learning of mixtures of distributions,” in International Conference on Computational Learning Theory. Springer, 2005, pp. 458–469. 3

[24] R. Kannan, H. Sulsamian, and S. Vempala, “The spectral method for general mixture models,” in International Conference on Computational Learning Theory. Springer, 2005, pp. 444–457. 3

[25] S. Dasgupta and L. J. Schulman, “A probabilistic analysis of em for mixtures of separated, spherical gaussians,” Journal of Machine Learning Research, vol. 8, pp. 203–226, 2007. 3

[26] S. C. Brubaker and S. S. Vempala, “Isotropic pca and affine-invariant clustering,” in Building Bridges. Springer, 2008, pp. 241–281. 3

[27] A. T. Kalai, A. Moitra, and G. Valiant, “Efficiently learning mixtures of two gaussians,” in Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010, L. J. Schulman, Ed. ACM, 2010, pp. 553–562. [Online]. Available: https://doi.org/10.1145/1806689.1806765 3

[28] A. Moitra and G. Valiant, “Settling the polynomial learnability of mixtures of gaussians,” in 2010 IEEE 51st Annual Symposium on Foundations of Computer Science. IEEE, 2010, pp. 93–102. 3, 5

[29] M. Belkin and K. Sinha, “Polynomial learning of distribution families,” SIAM Journal on Computing, vol. 44, no. 4, pp. 889–911, 2015. 3

[30] M. Hardt and E. Price, “Tight bounds for learning a mixture of two gaussians,” in Proceedings of the forty-seventh annual ACM symposium on Theory of computing, 2015, pp. 753–760. 3

[31] O. Regev and A. Vijayaraghavan, “On learning mixtures of well-separated gaussians,” in 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS). IEEE, 2017, pp. 85–96. 3

[32] S. B. Hopkins and J. Li, “Mixture models, robustness, and sum of squares proofs,” in Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, 2018, pp. 1021–1034. 3

[33] P. K. Kothari, J. Steinhardt, and D. Steurer, “Robust moment estimation and improved clustering via sum of squares,” in Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, 2018, pp. 1035–1046. 3

[34] I. Diakonikolas, D. M. Kane, and A. Stewart, “List-decodable robust mean estimation and learning mixtures of spherical gaussians,” in Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, 2018, pp. 1047–1060. 3

[35] J. Feldman, R. A. Servedio, and R. O’Donnell, “Pac learning axis-aligned mixtures of gaussians with no separation assumption,” in International Conference on Computational Learning Theory. Springer, 2006, pp. 20–34. 3

[36] I. Diakonikolas, D. M. Kane, and A. Stewart, “Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures,” in 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS). IEEE, 2017, pp. 73–84. 3, 4, 6

[37] M. Kears, “Efficient noise-tolerant learning from statistical queries,” Journal of the ACM (JACM), vol. 45, no. 6, pp. 983–1006, 1998. 4

[38] V. Feldman, E. Grigorescu, L. Reyzin, S. S. Vempala, and Y. Xiao, “Statistical algorithms and a lower bound for detecting planted cliques,” Journal of the ACM (JACM), vol. 64, no. 2, pp. 1–37, 2017. 4

[39] M. J. Song, I. Zadik, and J. Bruna, “On the cryptographic hardness of learning single periodic neurons,” in Advances in Neural Information Processing Systems, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 29.602–29.615. [Online]. Available: https://proceedings.neurips.cc/paper/2021/file/f78688f6b645507f413ade54e230355ac-Paper.pdf 4, 5

[40] S. Goldwasser, M. P. Kim, V. Vaikuntanathan, and O. Zamir, “Planting undetectable backdoors in machine learning models,” arXiv preprint arXiv:2204.06974, 2022. 4, 5

[41] S. Tiegel, “Hardness of agnostically learning halfspaces from worst-case lattice problems,” arXiv preprint arXiv:2207.14030, 2022. 4, 5

[42] I. Diakonikolas, D. M. Kane, P. Manurangsi, and L. Ren, “Cryptographic hardness of learning halfspaces with massart noise,” arXiv preprint arXiv:2207.14266, 2022. 4, 5

[43] D. Micciancio and P. Mol, “Pseudorandom knapsacks and the sample complexity of LWE search-to-decision reductions,” in Advances in Cryptology - CRYPTO 2011 - 31st Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011, Proceedings, ser. Lecture Notes in Computer Science, P. Rogaway, Ed., vol. 6841. Springer, 2011, pp. 465–484. [Online]. Available: https://doi.org/10.1007/978-3-642-22792-9_ 26 5

[44] D. Micciancio and O. Regev, “Worst-case to average-case reductions based on gaussian measures,” SIAM Journal on Computing, vol. 37, no. 1, pp. 267–302, 2007. 6, 8

[45] D. Boneh, K. Lewi, H. W. Montgomery, and A. Raghunathan, “Key homomorphic prfs and their applications,” in Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013, Proceedings, Part I, ser. Lecture Notes in Computer Science, R. Canetti and J. A. Garay, Eds., vol. 8042. Springer, 2013, pp. 410–428. [Online]. Available: https://doi.org/10.1007/978-3-642-40041-4_23 6