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

Aparna Gupte* MIT agupte@mit.edu

Neekon Vafa† MIT nvafa@mit.edu

Vinod Vaikuntanathan‡ MIT 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 $\mathsf{poly}(n)$ samples requires time quasi-polynomial in n. Under the (conservative) polynomial hardness of LWE, we show hardness of density estimation for n^ϵ 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.

*Research supported by the Keel Foundation Undergraduate Research and Innovation Scholarship.

†Research supported by NSF fellowship DGE-1745302 and by the grants of the third author.

‡Research supported in part by DARPA under Agreement No. HR00112020023, a grant from the MIT-IBM Watson AI, a grant from Analog Devices, a Microsoft Trustworthy AI grant, and a Thornton Family Faculty Research Innovation Fellowship from MIT. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Government or DARPA.

Contents

1 Introduction

The learning with errors (LWE) problem [Reg09] is a versatile average-case problem with connections to lattices, cryptography, learning theory and game theory. Given a sequence of noisy linear equations $(\mathbf{a}, b \approx \langle \mathbf{a}, \mathbf{s} \rangle \bmod q)$ over a ring $\mathbb{Z}/q\mathbb{Z}$, the LWE problem asks to recover the secret vector $\mathbf{s}$ (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 [Pei09, BLP+13]. Ever since its formulation in 2005, LWE has unlocked a wealth of applications in cryptography ranging from fully homomorphic encryption [BV14] to attribute-based encryption [GVW15] to, most recently, succinct non-interactive argument systems for all of P [CJJ21]. 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 [KS09], and in game theory, where the hardness of LWE implies the hardness of the complexity class PPAD [JKKZ21]. Finally, LWE enjoys remarkable structural properties such as leakage-resilience [GKPV10].

Motivated by applications to learning problems, Bruna, Regev, Song and Tang [BRST21] recently introduced a continuous version of LWE which they called CLWE. (In the definition below and henceforth, $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ is the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ where the probability of a point $\mathbf{x} \in \mathbb{R}^n$ is proportional to $e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$.)

Definition 1 (CLWE Distribution [BRST21], rescaled)**.** *Let $\gamma, \beta \in \mathbb{R}$, and let $\mathcal{S}$ be a distribution over unit vectors in $\mathbb{R}^n$. Let $\mathsf{CLWE}(m, \mathcal{S}, \gamma, \beta)$ be the distribution given by sampling $\mathbf{a}_1, \cdots, \mathbf{a}_m \sim \mathcal{N}(\mathbf{0}, I_{n \times n})$, $\mathbf{w} \sim \mathcal{S}, e_1, \cdots, e_m \sim \mathcal{N}(0, \beta^2)$ and outputting*

$$\big(\mathbf{a}_i, b_i := \gamma \cdot \langle \mathbf{a}_i, \mathbf{w} \rangle + e_i \bmod 1\big)_{i=1}^m.$$

Unless otherwise specified, $\mathcal{S}$ 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 $\mathbf{w}$ given CLWE samples, whereas the decisional CLWE problem asks to distinguish between samples from the CLWE distribution and samples with standard normal $\mathbf{a}_i$ (just like the CLWE distribution) but now with independent b_i that are distributed uniformly between 0 and 1.

Bruna et al. [BRST21] 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 [Reg09, PRS17]. 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 [Pei09, BLP+13]. An 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 [BNHR22].

Our main conceptual contribution is a direct and simple reduction from LWE to CLWE. When combined with Regev [Reg09], 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. [BRST21]. As another immediate application, by combining with the classical reduction from worst-case lattice problems to LWE [BLP+13], we obtain *classical* worst-case hardness of CLWE. Our main reduction also allows us to unlock powerful structural results on LWE [GKPV10, BLP+13, Mic18, BD20] and derive improved hardness results for learning mixtures of Gaussians with $(\log n)^{1+\epsilon}$ Gaussians instead of $\Omega(\sqrt{n})$ in [BRST21] (for arbitrary $\epsilon > 0$). We now describe these results in turn.

1.1 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 [Reg09], we are given m samples of the form $(\mathbf{a}_i, b_i := \langle \mathbf{a}_i, \mathbf{s} \rangle + e_i \bmod q)$ where $\mathbf{a}_i \sim (\mathbb{Z}/q\mathbb{Z})^n$ is uniformly random, $\mathbf{s} \in \mathbb{Z}^n$ 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 σ. The decisional LWE assumption (parameterized by n, m, q and σ) 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 6)**.** *Let $\mathcal{S} = \mathcal{S}_r$ be an arbitrary distribution over $\mathbb{Z}^n$ whose support consists of vectors with ℓ_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, σ and secret distribution $\mathcal{S}$, to decisional CLWE with parameters n, m, γ and β, as long as $\sigma \gg r$.

Our main reduction, in conjunction with prior work, immediately gives us a number of corollaries. First, letting $\mathcal{S}$ be the uniform distribution on $\{-1, 1\}^n$, and invoking the hardness result for LWE with binary secrets [BLP+13, Mic18, BD20], 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 5)**.** *For*

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

there is a polynomial (in n) time reduction from standard decisional LWE in dimension ℓ, with n samples, modulus q and noise parameter σ, to decisional CLWE in dimension n with parameters γ and β, 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 [GKPV10, BLP+13, Mic18, BD20], and go much further. In particular, using a variant of the reduction of [Mic18] modified to handle k-sparse secrets (discussed further in Section 2) 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 ± 1 vectors (resp. random k-sparse vectors).

Corollary 2 (Informal Version of Corollary 6)**.** *For*

$$\gamma = O\left(\sqrt{k \cdot \log n}\right) \text{ and } \beta = O\left(\frac{\sigma\sqrt{k}}{q}\right),$$

we have a polynomial (in n) time reduction from standard decisional LWE, in dimension ℓ, with n samples, modulus q, and noise parameter σ, to decisional CLWE in dimension n with k-sparse norm-1 secrets *and parameters γ and β, as long as $k \log_2(n/k) \gg \ell \log_2(q)$ and $\sigma \gg 1$.*

Looking ahead, 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 ℓ dimensions with a modulus $q = \mathsf{poly}(\ell)$ and error parameter $\sigma = \mathsf{poly}(\ell)$, LWE is believed to be hard for algorithms that run in 2^{ℓ^ϵ} time using $m = 2^{\ell^\epsilon}$ samples, for any $\epsilon < 1$ (see, e.g. [LP11]). 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 [NIS] 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

$$\gamma = (\log n)^{\frac{1}{2}+\delta} \log\log n$$

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 LWE [BLP+13] 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 $\mathsf{poly}(n/\beta)$-approximate gapSVP in $\sqrt{n}$ dimensions, to decisional CLWE in n dimensions with $\gamma = \widetilde{\Omega}(\sqrt{n})$ and arbitrary $\beta = 1/\mathsf{poly}(n)$.*

Finally, in Appendix C, 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.

1.2 Improved Hardness of Learning Mixtures of Gaussians

Bruna, Regev, Song and Tang [BRST21] 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 [TTM+85, MP00]. Given a set of g multivariate Gaussians in n dimensions, parameterized by their means $\boldsymbol{\mu}_i \in \mathbb{R}^n$, 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 $\mathcal{N}(\boldsymbol{\mu}_i, \Sigma_i)$.

Dasgupta [Das99] 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 $\boldsymbol{\mu}_i$, Σ_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 [Das99, SK01, VW02, AM05, KSV05, DS07, BV08, KMV10, MV10, BS15, HP15, RV17, HL18, KSS18, DKS18]. Moitra and Valiant [MV10] and Hardt and Price [HP15] also show that for parameter estimation, there is an information theoretic sample-complexity lower bound of $(1/\gamma)^g$ where γ 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 $\mathbf{x}$ [FSO06]. 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 [DKS17] show a $\mathsf{poly}(n, g, 1/\epsilon)$ 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 [DKS17] shows a polynomial upper bound on sample complexity, all known algorithms for density estimation, e.g., [MV10], run in time $(n/\epsilon)^{f(g)}$ for some $f(g) \geq g$. This is polynomial-time only for constant g. Furthermore, [DKS17] shows that even density estimation of Gaussian mixtures incurs a super-polynomial lower bound in the restricted statistical query (SQ) model [Kea98, FGR$^+$17]. Explicitly, they show that any SQ algorithm giving density estimates requires $n^{\Omega(g)}$ queries to an SQ oracle of precision $n^{-O(g)}$; 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 [BRST21]. They show that being able to output a density estimate for mixtures of $g = \Omega(\sqrt{n})$ Gaussians implies 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 [BRST21] 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 [Reg09], this gives us hardness for n^ϵ Gaussians under the quantum worst-case hardness of lattice problems. This improves on [BRST21] 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.

Summary of GMM Hardness Results				
	LWE Assumption (samples, time, adv.)	Gaussian Components	Run-time	Samples
Corollary 9	$\left(\ell^{1/\epsilon}, \mathsf{poly}(\ell), \frac{1}{\mathsf{poly}(\ell)}\right)$	$O\left(n^{\epsilon/2} \cdot \log n\right)$	$n^{\omega(1)}$	$\mathsf{poly}(n)$
Corollary 8	$\left(2^{\ell^\delta}, 2^{O(\ell^\epsilon)}, \frac{1}{2^{O(\ell^\delta)}}\right)$	$O\left((\log n)^{\frac{1}{2}+\frac{1}{2\delta}} \cdot \sqrt{\log\log n}\right)$	$\Omega\left(2^{(\log n)^{\epsilon/\delta}}\right)$	$\mathsf{poly}(n)$
Corollary 8	$\left(2^{\ell^\delta}, 2^{O(\ell^\epsilon)}, \frac{1}{\mathsf{poly}(\ell)}\right)$	$O\left((\log n)^{\frac{1}{2\delta}} \cdot \log\log n\right)$	$\Omega\left(2^{(\log n)^{\epsilon/\delta}}\right)$	$\mathsf{poly}(\log n)$

Figure 1: This tables summarizes our hardness results for density estimation of GMM. Throughout, $\delta, \epsilon \in (0,1)$ are arbitrary constants with $\delta < \epsilon$, ℓ 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 ℓ with $1/\mathsf{poly}(\ell)$ advantage given $\ell^{1/\epsilon}$ samples and $\mathsf{poly}(\ell)$ time, then any algorithm solving GMM density estimation given access to $\mathsf{poly}(n)$ samples from an arbitrary Gaussian mixture with at most $O(n^{\epsilon/2} \cdot \log n)$ Gaussian components must take super-polynomial in n time.

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 [NIS]).

We can stretch this even a little further. We show the hardness of density estimation for $g = (\log n)^{1/2+\epsilon}$ Gaussians given $\mathsf{poly}(\log n)$ 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 [DKS17], is poly-logarithmic in n. In fact, we show (in Corollary 10) 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 below and more succinctly in Figure 1. The reader is referred to Section 6 for the formal proofs.

Theorem 2 (Informal Version of Corollary 8 and Corollary 9)**.** *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^ϵ Gaussian components, given $\mathsf{poly}(n)$ samples from the mixture, requires super-polynomial time in n for arbitrary constant $\epsilon > 0$.*

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

3. *For constant $\epsilon \in (0,1)$, assuming ℓ-dimensional LWE is hard to distinguish with advantage $1/\mathsf{poly}(\ell)$ in time 2^{ℓ^ϵ}, any density estimator for $\mathbb{R}^n$ that can solve arbitrary mixtures with*

at most roughly $(\log n)^{\frac{1}{2\epsilon}}$ *Gaussian components, given* $\mathsf{poly}(\log n)$ *samples from the mixture, requires super-polynomial in* n *time.*

1.3 Other Applications

Recent results have shown reductions from CLWE to other learning tasks as well, including learning a single periodic neuron [SZB21], detecting backdoors in certain models [GKVZ22], and improperly learning halfspaces in various error models [Tie22, DKMR22].[1] 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. For example, Song, Zadik and Bruna [SZB21] 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 γ. 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 [BRST21]. 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 γ and in the sparsity of the secret distribution translates similarly for the other learning tasks mentioned, namely detecting backdoors in certain models [GKVZ22] and improperly learning halfspaces in various error models [Tie22, DKMR22]. For hardness of detecting backdoors [GKVZ22], this flexibility means reducing the magnitude of undetectable backdoor perturbations (in ℓ_2 and ℓ_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]

1.4 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 on lattices to continuous LWE; and the second is an improved hardness result for the well-studied problem of learning mixtures of Gaussians. We believe much more is in store.

For one, while we show a search-to-decision reduction for discrete-secret CLWE (see Appendix C), we still do not know such a reduction for general CLWE. This is in contrast to multiple search-to-decision reductions of varying complexity and generality for the LWE problem [Reg09, MM11].

[1] More precisely, Diakonikolas, Kane, Manurangsi and Ren [DKMR22] use our techniques to reduce from LWE instead of CLWE.

[2] The Veronese map translates a k-sparse degree-d polynomial threshold function in dimension n to a $\binom{k+d}{d}$-sparse linear threshold function (*i.e.*, halfspace) in dimension $\binom{n+d}{d}$.

Secondly, while there has been some initial exploration of the cryptographic applications of the continuous LWE problem [BNHR22], constructing *qualitatively new* cryptographic primitives or *qualitatively better* cryptographic constructions is an exciting research direction. A recent example is the result of [GKVZ22] 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 [MV10] 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. [BRST21] 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 secret directions to get a solver with a similar run-time as [BRST21] but with much smaller sample complexity. See Corollary 10 for details.)

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 $\mathsf{poly}(n) \cdot 2^{g^2}$, despite what SQ lower bounds suggest [DKS17].

If we believe in the first two possibilities, a natural place to look for a *different* hard instance is [DKS17], 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.

2 Technical Overview

2.1 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 Figure 2. Given samples $(\mathbf{a}, b = \langle \mathbf{a}, \mathbf{s} \rangle + e \pmod q) \in \mathbb{Z}_q^{n+1}$, we do the following:

1. First, we turn the errors (in b) from discrete to continuous Gaussians by adding a small continuous Gaussian to the LWE samples, using the smoothing lemma [MR07].
2. Secondly, we turn the samples $\mathbf{a}$ 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 [Reg09, MR07].

Reducing Fixed-Norm LWE to CLWE (Theorem 6)			
	Samples	Secrets	Errors
Fixed-Norm LWE	$U(\mathbb{Z}_q^n)$	$\mathcal{S}$	$D_{\mathbb{Z},\sigma_1}$
Step 1 (Lemma 15)	$U(\mathbb{Z}_q^n)$	$\mathcal{S}$	D_{σ_2}
Step 2 (Lemma 16)	$U(\mathbb{T}_q^n)$	$\mathcal{S}$	D_{σ_3}
CLWE (Lemma 18)	D_1^n	$\frac{1}{r} \cdot \mathcal{S}$	D_β

Figure 2: This table shows the steps in the reduction from fixed-norm LWE to CLWE (with discrete secret distribution $\frac{1}{r} \cdot \mathcal{S}$ of unit norm; to reduce to continuous uniform unit-vector secrets, one can apply Lemma 19). All of the reductions in the table are sample preserving, dimension preserving, and advantage preserving (up to $\mathsf{negl}(\lambda)$ additive loss). To reduce from LWE with secrets $\mathbf{s} \sim U(\mathbb{Z}_q^n)$ (instead of a fixed-norm distribution), we first apply Theorem 7 and then we perform the steps above.

3. Third, we go from uniform samples $\mathbf{a}$ to Gaussian samples. Boneh, Lewi, Montgomery and Raghunathan [BLMR13] 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 $y = z \pmod 1$ and y is statistically close to a continuous Gaussian (when marginalized over $z \sim U([0,1))$). (See Lemma 17 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*.

2.2 Hardness of Gaussian Mixture Learning

Bruna et al. [BRST21] 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 [BLP+13, Mic18, BD20] to improve and generalize the hard instance of [BRST21]. 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 $\mathbf{s} \in \{+1, 0, -1\}^n$ k-sparse if it has exactly k non-zero entries. We show the following result:

Theorem 3 (Informal Version of Corollary 4)**.** *Assume LWE in dimension ℓ with n samples is hard with secrets $\mathbf{s} \sim \mathbb{Z}_q^\ell$ and errors of width σ. Then, LWE in dimension n with k-sparse secrets is hard for errors of width $O(\sqrt{k} \cdot \sigma)$, as long as $k \log_2(n/k) \gg \ell \log_2(q)$.*

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 [Mic18] to work with k-sparse secrets.[3] We note that Brakerski and Döttling [BD20] 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 $\sqrt{kmn} = k^{\omega(1)}$ for parameter settings we consider.

For a full description of the proof of Theorem 3, the reader is referred to Section 4.

3 Preliminaries

For a distribution $\mathcal{D}$, we write $x \sim \mathcal{D}$ to denote a random variable x being sampled from $\mathcal{D}$. For any $n \in \mathbb{N}$, we let $\mathcal{D}^n$ denote the n-fold product distribution, i.e. $(x_1, \ldots, x_n) \sim \mathcal{D}^n$ is generated by sampling $x_i \sim_{\text{i.i.d.}} \mathcal{D}$ 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 $\mathcal{D}_1, \mathcal{D}_2$ supported on a measurable set $\mathcal{X}$, we define the statistical distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ to be $\Delta(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2} \int_{x \in \mathcal{X}} |\mathcal{D}_1(x) - \mathcal{D}_2(x)| dx$. We say that distributions $\mathcal{D}_1, \mathcal{D}_2$ are ϵ-close if $\Delta(\mathcal{D}_1, \mathcal{D}_2) \leq \epsilon$. For a distinguisher $\mathcal{A}$ running on two distributions $\mathcal{D}_1$, $\mathcal{D}_2$, we say that $\mathcal{A}$ has advantage ϵ if

$$\left| \Pr_{x \sim \mathcal{D}_1} [\mathcal{A}(x) = 1] - \Pr_{x \sim \mathcal{D}_2} [\mathcal{A}(x) = 1] \right| = \epsilon,$$

where the probability is also over any internal randomness of $\mathcal{A}$.

We let $I_{n \times n} \in \{0,1\}^{n \times n}$ denote the $n \times n$ identity matrix. When n is clear from context, we write this simply as I. For any matrix $M \in \mathbb{R}^{m \times n}$, we let $M^\top$ be its transpose matrix, and for $\ell \in [n]$, we write $M_{[\ell]} \in \mathbb{R}^{m \times \ell}$ to denote the submatrix of M consisting of just the first ℓ columns, and we write $M_{]\ell[} \in \mathbb{R}^{m \times (n-\ell)}$ to denote the submatrix of M consisting of all but the first ℓ columns.

For any vector $\mathbf{v} \in \mathbb{R}^n$, we write $\|\mathbf{v}\|$ to mean the standard ℓ_2-norm of $\mathbf{v}$, and we write $\|\mathbf{v}\|_\infty$ to denote the ℓ_∞-norm of $\mathbf{v}$, meaning the maximum absolute value of any component. For $n \in \mathbb{N}$, we let $S^{n-1} \subset \mathbb{R}^n$ 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 $\{0, \ldots, q-1\}$.

[3] The techniques of Brakerski et al. [BLP+13], 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 [Mic18].

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 $\mathbb{T} := \mathbb{T}_1$ 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.

Definition 2 (Min-Entropy)**.** *For a discrete distribution $\mathcal{D}$ with support S, we let $H_\infty(\mathcal{D})$ denote the min-entropy of $\mathcal{D}$,*

$$H_\infty(\mathcal{D}) = -\log_2 \left(\max_{s \in S} \Pr_{x \sim \mathcal{D}}[x = s] \right).$$

Lemma 1 (Leftover Hash Lemma [HILL99])**.** *Let $\ell, n, q \in \mathbb{N}, \epsilon \in \mathbb{R}_{>0}$, and let $\mathcal{S}$ be a distribution over $\{-1, 0, 1\}^n \subseteq \mathbb{Z}_q^n$. Suppose $H_\infty(\mathcal{S}) \geq \ell \log_2(q) + 2\log_2(1/\epsilon)$. Then, the distributions given by $(A, A\mathbf{s} \pmod q)$ and $(A, \mathbf{b})$ where $A \sim \mathbb{Z}_q^{\ell \times n}$, $\mathbf{s} \sim \mathcal{S}$, $\mathbf{b} \sim \mathbb{Z}_q^\ell$ have statistical distance at most ϵ.*

3.1 Lattices and Discrete Gaussians

A rank n integer lattice is a set $\Lambda = \mathbf{B}\mathbb{Z}^n \subseteq \mathbb{Z}^d$ of all integer linear combinations of n linearly independent vectors $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$ in $\mathbb{Z}^d$. The dual lattice Λ^* of a lattice Λ is defined as the set of all vectors $\mathbf{y} \in \mathbb{R}^d$ such that $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{Z}$ for all $\mathbf{x} \in \Lambda$.

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

$$\rho_{s,\mathbf{c}}(\mathbf{x}) = \exp\left(-\pi \|(\mathbf{x} - \mathbf{c})/s\|^2\right).$$

Let $D_{s,\mathbf{c}}$ be the corresponding distribution with density at $\mathbf{x} \in \mathbb{R}^n$ given by $\rho_{s,\mathbf{c}}(x)/s^n$, namely the n-dimensional Gaussian distribution with mean $\mathbf{c}$ and covariance matrix $s^2/(2\pi) \cdot I_{n \times n}$. When $\mathbf{c} = 0$, we omit the subscript notation of $\mathbf{c}$ on ρ and D.

For an n-dimensional lattice $\Lambda \subseteq \mathbb{R}^n$ and point $\mathbf{c} \in \mathbb{R}^n$, we can define the *discrete Gaussian of width s* to be given by the mass function

$$D_{\Lambda + \mathbf{c}, s}(\mathbf{x}) = \frac{\rho_s(\mathbf{x})}{\rho_s(\Lambda + \mathbf{c})}$$

supported on $\mathbf{x} \in \Lambda + \mathbf{c}$, where by $\rho_s(\Lambda + \mathbf{c})$ we mean $\sum_{\mathbf{y} \in \Lambda} \rho_s(\mathbf{y} + \mathbf{c})$.

We now give the smoothing parameter as defined by [Reg09] and some of its standard properties.

Definition 3 ([Reg09], Definition 2.10)**.** *For an n-dimensional lattice Λ and $\epsilon > 0$, we define $\eta_\epsilon(\Lambda)$ to be the smallest s such that $\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) \leq \epsilon$.*

Lemma 2 ([Reg09], Lemma 2.12)**.** *For an n-dimensional lattice Λ 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 Λ.

Lemma 3 ([Reg09], Corollary 3.10)**.** *For any n-dimensional lattice Λ and $\epsilon \in (0, 1/2)$, $\sigma, \sigma' \in \mathbb{R}_{>0}$, and $\mathbf{z}, \mathbf{u} \in \mathbb{R}^n$, if*

$$\eta_\epsilon(\Lambda) \leq \frac{1}{\sqrt{1/(\sigma')^2 + (\|\mathbf{z}\|/\sigma)^2}},$$

then if $\mathbf{v} \sim D_{\Lambda + \mathbf{u}, \sigma'}$ and $e \sim D_\sigma$, then $\langle \mathbf{z}, \mathbf{v} \rangle + e$ has statistical distance at most 4ϵ from $D_{\sqrt{(\sigma' \|\mathbf{z}\|)^2 + \sigma^2}}$.

Lemma 4 ([MR07], Lemma 4.1)**.** *For an n-dimensional lattice Λ, $\epsilon > 0$, $\mathbf{c} \in \mathbb{R}^n$ for all $s \geq \eta_\epsilon(\Lambda)$, we have*

$$\Delta(D_{s,\mathbf{c}} \mod P(\Lambda), U(P(\Lambda))) \leq \epsilon/2,$$

where $P(\Lambda)$ is the half-open fundamental parallelepiped of Λ.

Lemma 5 ([MR07], implicit in Lemma 4.4)**.** *For an n-dimensional lattice Λ, for all $\epsilon > 0$, $\mathbf{c} \in \mathbb{R}^n$, and all $s \geq \eta_\epsilon(\Lambda)$, we have*

$$\rho_s(\Lambda + \mathbf{c}) = \rho_{s,-\mathbf{c}}(\Lambda) \in \left[\frac{1-\epsilon}{1+\epsilon}, 1\right] \cdot \rho_s(\Lambda).$$

Now we recall other facts related to lattices.

Lemma 6 ([MP13], Theorem 3)**.** *Suppose $\mathbf{v} \in \mathbb{Z}^m$ with $\gcd(\mathbf{v}) = 1$, and suppose $y_i \sim D^m_{\mathbb{Z},\sigma_i}$ for all $i \in [m]$. As long as $\sigma_i \geq \sqrt{2}\|\mathbf{v}\|_\infty \eta_{\frac{\epsilon}{2m^2}}(\mathbb{Z})$ for all $i \in [m]$, then we have $y = \sum_{i\in[m]} y_i v_i$ is $O(\epsilon)$-close to $D_{\mathbb{Z},\sigma}$ where $\sigma = \sqrt{\sum_{i\in[m]} \sigma_i^2 v_i^2}$.*

Lemma 7 ([Mic18], Lemma 2.2)**.** *For $\mathbf{w} \sim U(\mathbb{Z}_q^\ell)$, the probability that $\gcd(\mathbf{w}, q) \neq 1$ is at most $\log(q)/2^\ell$.*

Definition 4. *We say that a matrix $T \in \mathbb{Z}^{k\times m}$ is primitive if $T\mathbb{Z}^m = \mathbb{Z}^k$, i.e., if $T : \mathbb{Z}^m \to \mathbb{Z}^k$ is surjective.*

Lemma 8 ([Mic18], Lemma 2.6)**.** *For any primitive matrix $T \in \mathbb{Z}^{k\times m}$ and positive reals $\alpha, \sigma > 0$, if $TT^\top = \alpha^2 I$ and $\sigma \geq \eta_\epsilon(\ker(T))$, then $T(D_{\mathbb{Z}^m,\sigma})$ and $D_{\mathbb{Z}^k,\alpha\sigma}$ are $O(\epsilon)$-close.*

3.2 Learning with Errors

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

Definition 5 (LWE Distribution)**.** *Let $n, m, q \in \mathbb{N}$, let $\mathcal{A}$ be a distribution over $\mathbb{R}^n$, $\mathcal{S}$ be a distribution over $\mathbb{Z}^n$, and $\mathcal{E}$ be a distribution over $\mathbb{R}$. We define $\mathsf{LWE}(m, \mathcal{A}, \mathcal{S}, \mathcal{E})$ to be distribution given by sampling $\mathbf{a}_1, \cdots, \mathbf{a}_m \sim \mathcal{A}$, $\mathbf{s} \sim \mathcal{S}$, and $e_1, \cdots, e_m \sim \mathcal{E}$, and outputting $(\mathbf{a}_i, \mathbf{s}^\top \mathbf{a}_i + e_i \pmod q)$ 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.)*

We also consider the case where $\mathcal{S}$ is a distribution over $\mathbb{Z}^{n\times j}$ and $\mathcal{E}$ is a distribution over $\mathbb{R}^j$. In this case, the ouput of each sample is $(\mathbf{a}_i, S^\top \mathbf{a}_i + \mathbf{e}_i \pmod q)$, where $S \sim \mathcal{S}$ and $\mathbf{e}_i \sim \mathcal{E}$.

Definition 6 (CLWE Distribution [BRST21])**.** *Let $n, m \in \mathbb{N}, \gamma, \beta \in \mathbb{R}$, and let $\mathcal{A}$ be a distribution over $\mathbb{R}^n$ and $\mathcal{S}$ be a distribution over S^{n-1}. Let $\mathsf{CLWE}(m, \mathcal{A}, \mathcal{S}, \gamma, \beta)$ be the distribution given by sampling $\mathbf{a}_1, \cdots, \mathbf{a}_m \sim \mathcal{A}$, $\mathbf{s} \sim \mathcal{S}, e_1, \cdots, e_m \sim D_\beta$ and outputting $(\mathbf{a}_i, \gamma \cdot \langle \mathbf{a}_i, \mathbf{s}\rangle + e_i \pmod 1)$ for all $i \in [m]$. Explicitly, for one sample, the density at $(\mathbf{y}, z) \in \mathbb{R}^n \times [0, 1)$ is proportional to*

$$\mathcal{A}(\mathbf{y}) \cdot \sum_{k\in\mathbb{Z}} \rho_\beta(z + k - \gamma \cdot \langle \mathbf{y}, \mathbf{s}\rangle)$$

for fixed secret $\mathbf{s} \sim \mathcal{S}$. We refer to n as the dimension and m as the number of samples. We omit $\mathcal{S}$ if $\mathcal{S} = U(S^{n-1})$, as is standard for CLWE.

Definition 7 (hCLWE Distribution [BRST21])**.** *Let $n, m \in \mathbb{N}, \gamma, \beta \in \mathbb{R}$, and let $\mathcal{A}$ be a distribution over $\mathbb{R}^{n\times m}$ and $\mathcal{S}$ be a distribution over S^{n-1}. Let $\mathsf{hCLWE}(m, \mathcal{A}, \mathcal{S}, \gamma, \beta)$ be the the distribution $\mathsf{CLWE}(m, \mathcal{A}, \mathcal{S}, \gamma, \beta)$, but conditioned on the fact that for all samples second entries are $0 \pmod 1$.*

Explicitly, for one sample, the density at $\mathbf{y} \in \mathbb{R}^n$ is proportional to

$$\mathcal{A}(\mathbf{y}) \cdot \sum_{k\in\mathbb{Z}} \rho_\beta(k - \gamma \cdot \langle \mathbf{y}, \mathbf{s}\rangle)$$

for fixed secret $\mathbf{s} \sim \mathcal{S}$. We refer to n as the dimension and m as the number of samples. We omit $\mathcal{S}$ if $\mathcal{S} = U(S^{n-1})$, as is standard for hCLWE.

Note that the hCLWE distribution is itself a mixture of Gaussians. Explicitly, for a secret $\mathbf{s} \sim \mathcal{S}$, we can write the density of $\mathsf{hCLWE}(1, D_1, \mathbf{s}, \gamma, \beta)$ at point $\mathbf{x} \in \mathbb{R}^n$ as proportional to

$$\rho(\mathbf{x}) \cdot \sum_{k\in\mathbb{Z}} \rho_\beta(k - \gamma \cdot \langle \mathbf{s}, \mathbf{x}\rangle) = \sum_{k\in\mathbb{Z}} \rho_{\sqrt{\beta^2+\gamma^2}}(k) \cdot \rho(\pi_{\mathbf{s}^\perp}(\mathbf{x})) \cdot \rho_{\beta/\sqrt{\beta^2+\gamma^2}}\left(\langle \mathbf{s}, \mathbf{x}\rangle - \frac{\gamma}{\beta^2+\gamma^2}k\right), \quad (1)$$

where $\pi_{\mathbf{s}^\perp}(\mathbf{x})$ denotes the projection onto the orthogonal complement of $\mathbf{s}$. Thus, we can view hCLWE samples as being drawn from a mixture of Gaussians of width $\beta/\sqrt{\beta^2+\gamma^2} \approx \beta/\gamma$ in the secret direction, and width 1 in all other directions.

Definition 8 (Truncated hCLWE Distribution [BRST21])**.** *Let $n, m, g \in \mathbb{N}, \gamma, \beta \in \mathbb{R}$, and let $\mathcal{S}$ be a distribution over S^{n-1}. Let $\mathsf{hCLWE}^{(g)}(m, \mathcal{S}, \gamma, \beta)$ be the the distribution $\mathsf{hCLWE}(m, D_1^n, \mathcal{S}, \gamma, \beta)$, but restricted to the central g Gaussians, where by central g Gaussians, we mean the central g Gaussians in writing hCLWE samples as a mixture of Gaussians, as in Eq. 1. Explicitly, for secret $\mathbf{s} \sim \mathcal{S}$, the density of one sample at a point $\mathbf{x} \in \mathbb{R}^n$ is proportional to*

$$\sum_{k=-\lfloor g/2\rfloor}^{\lfloor (g-1)/2\rfloor} \rho_{\sqrt{\beta^2+\gamma^2}}(k) \cdot \rho(\pi_{\mathbf{s}^\perp}(\mathbf{x})) \cdot \rho_{\beta/\sqrt{\beta^2+\gamma^2}}\left(\langle \mathbf{s}, \mathbf{x}\rangle - \frac{\gamma}{\beta^2+\gamma^2}k\right). \quad (2)$$

Definition 9 (Density Estimation for the Gaussian Mixture Model (Definition 5.1 of [BRST21])**.** *We say that an algorithm* solves GMM density estimation *in dimension n with m samples and up to g Gaussians if, when given m samples from an arbitrary mixture of at most g Gaussian components in $\mathbb{R}^n$, the algorithm outputs some density function (as an evaluation oracle) that has statistical distance at most 10^{-3} from the true density function of the mixture, with probability at least $9/10$ (over the randomness of the samples and the internal randomness of the algorithm).*

The following theorem tells us that distinguishing a truncated version of the hCLWE Gaussian mixture from the standard Gaussian is enough to distinguish the original Gaussian mixture from the standard Gaussian. In particular, we can use density estimation to solve hCLWE since the truncated version has a finite number of Gaussians.

Theorem 4 (Proposition 5.2 of [BRST21])**.** *Let $n, m \in \mathbb{N}$, $\gamma, \beta \in \mathbb{R}_{>0}$ with $\beta < 1/32$ and $\gamma \geq 1$. Let $\mathcal{S}$ be a distribution over S^{n-1}. For sufficiently large m and for $g = 2\gamma\sqrt{\ln m/\pi}$, if there is an algorithm running in time T that distinguishes $\mathsf{hCLWE}^{(2g+1)}(m, \mathcal{S}, \gamma, \beta)$ and $D_1^{n\times m}$ with constant advantage, then there is a time $T + \mathsf{poly}(n, m)$ algorithm distinguishing $\mathsf{hCLWE}(m, D_1^n, \mathcal{S}, \gamma, \beta)$ and $D_1^{n\times m}$ with constant advantage. In particular, if there is an algorithm running in time T that solves density estimation with in dimension n with m samples and g Gaussians, then there is a time $T + \mathsf{poly}(n, m)$ algorithm distinguishing $\mathsf{hCLWE}(m, D_1^n, \mathcal{S}, \gamma, \beta)$ and $D_1^{n\times m}$ with advantage at least $1/2$.*

We also use a lemma which says that if CLWE is hard, then so is hCLWE.

Lemma 9 (Lemma 4.1 of [BRST21])**.** *Let $\delta \in (0,1)$ be an input parameter. There is a randomized $\mathsf{poly}(n, m_1, 1/\delta)$-time reduction that maps m_1 samples from $\mathsf{CLWE}(D_1^n, \mathbf{s}, \gamma, \beta)$ to $m_2 = \Omega(\delta m_1)$ samples from $\mathsf{hCLWE}(D_1^n, \mathbf{s}, \gamma, \sqrt{\beta^2 + \delta^2})$ and maps m_1 samples from $D_1^n \times U(\mathbb{T}_1)$ to m_2 samples from D_1^n, with failure probability at most $1/1000$.*

4 Hardness of k-sparse LWE

In this section, we modify the proof of [Mic18] to reduce from standard decisional LWE to a version where secrets are sparse, in the sense that they have few non-zero entries. The main changes we make to [Mic18] are that we slightly modify the gadget matrix Q and the matrix Z to handle sparse secrets (using its notation).

For completeness, we give a self-contained proof.

Definition 10. *For $k, n \in \mathbb{N}$ with $k \leq n$, let $\mathcal{S}_{n,k}$ be the subset of vectors in $\{-1, 0, +1\}^n$ with exactly k non-zero entries. We call $\mathbf{s} \in \mathbb{Z}^n$ k-*sparse *if $\mathbf{s} \in \mathcal{S}_{n,k}$.*

Lemma 10. *It holds that $H_\infty(\mathcal{S}_{n,k}) \geq k \log_2(n/k)$.*

Proof. Observe that $|\mathcal{S}_{n,k}| = \binom{n}{k} \cdot 2^k$. Using the bound $(n/k)^k \leq \binom{n}{k}$, we have

$$H_\infty(\mathcal{S}_{n,k}) \geq \log_2 \left(\left(2 \cdot \frac{n}{k} \right)^k \right) \geq k \log_2(n/k),$$

as desired. □

Our main theorem in this section is the following:

Theorem 5. *Let $q, m, n, \ell, k \in \mathbb{N}$ with $1 < k < n$, and let $\sigma, \epsilon \in \mathbb{R}_{>0}$. Suppose $\log(q)/2^\ell = \mathsf{negl}(\lambda), \sigma \geq 4\sqrt{\omega(\log\lambda) + \ln n + \ln m}$, and $k \log(n/k) \geq (\ell + 1)\log_2(q) + \omega(\log\lambda)$. Suppose there is no $T + \mathsf{poly}(n, m, \log(q), \log(\lambda))$ time distinguisher with advantage $\epsilon - \mathsf{negl}(\lambda)$ between $\mathsf{LWE}(n-1, \mathbb{Z}_q^\ell, \mathbb{Z}_q^{m\times\ell}, D_{\mathbb{Z}^m,\sigma})$ and $U(\mathbb{Z}_q^{\ell\times(n-1)} \times \mathbb{Z}_q^{m\times(n-1)})$, and further suppose there is no $T + \mathsf{poly}(n, m, \log(q), \log(\lambda))$ time distinguisher with advantage $\epsilon - \mathsf{negl}(\lambda)$ between $\mathsf{LWE}(n+1, \mathbb{Z}_q^{\ell+1}, \mathbb{Z}_q^{m\times(\ell+1)}, D_{\mathbb{Z}^m,2\sigma})$ and $U(\mathbb{Z}_q^{(\ell+1)\times(n+1)} \times \mathbb{Z}_q^{m\times(n+1)})$. Then, there is no T time distinguisher with advantage 2ϵ between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$ and $U(\mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m)$, where $\sigma' = 2\sigma\sqrt{k+1}$.*

Definition 11. *Let $n, k \in \mathbb{Z}$ with $k \leq n$. For all $i \in [n]$, we define $\mathbf{e}_i$ to be the ith standard basis column vector, i.e. having a 1 in the ith coordinate and 0s elsewhere. We then define $\mathbf{u} \in \mathbb{Z}^n$ to be $\mathbf{u} = \sum_{i=1}^k \mathbf{e}_i$, i.e. 1s in the first k coordinates and 0 elsewhere.*

Lemma 11. *There is a $\mathsf{poly}(n)$-time computable matrix $Q \in \mathbb{Z}^{n\times(2n+5)}$ such that $Q_{[n]}$ is invertible, $\mathbf{u}^\top Q_{[n]} = \mathbf{e}_1^\top$, the vector $\mathbf{v}^\top = \mathbf{u}^\top Q_{]n[} \in \mathbb{Z}^{n+5}$ satisfies $\|\mathbf{v}\|_2 = 2\sqrt{k}$ and $\|\mathbf{v}\|_\infty = 2$, and $Q_{]1[}(D_{\mathbb{Z}^{2n+4},\sigma})$ and $D_{\mathbb{Z}^n,2\sigma}$ are $\mathsf{negl}(\lambda)/t$ close as long as $\sigma \geq \sqrt{6} \cdot \sqrt{\omega(\log\lambda) + \ln n + \ln t}$ for a free parameter t.*

Proof. We use essentially the same gadget Q as in Lemma 2.7 of [Mic18], except we modify two entries of the matrix and add two columns. Specifically, we set $Q_{k,k+1} = 0$ (instead of -1),

$Q_{k,n+k+1} = 0$ (instead of 1), and add two columns to the end that are all 0 except for two entries of 1 in $Q_{k,2n+4}$ and $Q_{k,2n+5}$.

We will give it explicitly as follows. Let the matrix $X \in \mathbb{Z}^{n\times(n-1)}$ be defined by

$$X = \begin{bmatrix} -1 & & & & & & & & \\ 1 & -1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & 1 & -1 & & & & & \\ & & & 1 & 0 & & & & \\ & & & & 1 & -1 & & & \\ & & & & & \ddots & \ddots & & \\ & & & & & & 1 & -1 & \\ & & & & & & & 1 & \end{bmatrix},$$

where the row with the abnormal 0 is the kth row. Similarly, let $Y \in \mathbb{Z}^{n\times(n-1)}$ be defined by

$$Y = \begin{bmatrix} 1 & & & & & & & & \\ 1 & 1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & 1 & 1 & & & & & \\ & & & 1 & 0 & & & & \\ & & & & 1 & 1 & & & \\ & & & & & \ddots & \ddots & & \\ & & & & & & 1 & 1 & \\ & & & & & & & 1 & \end{bmatrix},$$

where the row with the abnormal 0 is again the kth row. We then define $Q \in \mathbb{Z}^{n\times(2n+5)}$ by

$$Q = [\mathbf{e}_1, X, -\mathbf{e}_n, Y, \mathbf{e}_n, \mathbf{e}_1, \mathbf{e}_1, \mathbf{e}_k, \mathbf{e}_k].$$

First, notice that $Q_{[n]}$ is invertible, since it is upper-triangular with 1s on the diagonal. Next, notice that $\mathbf{u}^\top Q_{[n]} = \mathbf{e}_1^\top$, as $\mathbf{u}^\top \mathbf{e}_1 = 1$ and the sum of the first k entries in each column of X are all 0 by construction. We can write $\mathbf{v}^\top = \mathbf{u}^\top Q_{]n[} = [0, 2, 2, \cdots, 2, 0, \cdots, 0, 1, 1, 1, 1]$, which has ℓ_2 norm

$$\sqrt{(k-1)\cdot 2^2 + 4 \cdot 1^2} = 2\sqrt{k}.$$

It's clear to also see that $\|\mathbf{v}\|_\infty = 2$. All that is remaining to show is that $Q_{]1[}(D_{\mathbb{Z}^{2n+4},\sigma})$ and $D_{\mathbb{Z}^n,2\sigma}$ are $\mathsf{negl}(\lambda)/t$-close, which we do below.

To show that $Q_{]1[}(D_{\mathbb{Z}^{2n+4},\sigma})$ and $D_{\mathbb{Z}^n,2\sigma}$ are $\mathsf{negl}(\lambda)/t$ close, we first prove the preconditions of, and then invoke, Lemma 8. Let $T = Q_{]1[} \in \mathbb{Z}^{n\times(2n+4)}$.

First, we show that T is primitive. It suffices to show that for every standard basis column vector $\mathbf{e}_i$, there is some $\mathbf{g}_i \in \mathbb{Z}^{2n+4}$ such that $\mathbf{e}_i = T\mathbf{g}_i$. For all $j \in [2n+4]$, we define $\mathbf{f}_j$ to be the jth standard basis column vector in $\mathbb{R}^{2n+4}$. Let $\mathbf{g}_1 = \mathbf{f}_{2n+1}$, and $\mathbf{g}_{k+1} = \mathbf{f}_k$. It can be easily checked that $\mathbf{e}_1 = T\mathbf{g}_1$ and $\mathbf{e}_{k+1} = T\mathbf{g}_{k+1}$. Then, for all i such that $1 < i \leq k$ and $k+1 < i \leq n$, let

$\mathbf{g}_i = \mathbf{f}_{i-1} + \mathbf{g}_{i-1}$. Using an inductive argument, and by the construction of T, it follows that

$$\begin{aligned} T\mathbf{g}_i &= T(\mathbf{f}_{i-1} + \mathbf{g}_{i-1}) \\ &= T\mathbf{f}_{i-1} + T\mathbf{g}_{i-1} \\ &= (\mathbf{e}_i - \mathbf{e}_{i-1}) + \mathbf{e}_{i-1} \\ &= \mathbf{e}_i. \end{aligned}$$

It is easy to check that $TT^\top = 4I$. Finally, we bound the smoothing parameter of the lattice $\Lambda = \ker(T)$. Since $T \in \mathbb{Z}^{n\times(2n+4)}$ and T has full rank, its kernel Λ has dimension $n+4$. The columns of the following matrix give a basis for the lattice Λ.

$$V = \begin{bmatrix} \tilde{Y} & \mathbf{e}_1 & & -\mathbf{e}_{k-1} & \\ -\tilde{X} & -\mathbf{e}_1 & & -\mathbf{e}_{k-1} & \\ & 1 & 1 & & \\ & 1 & -1 & & \\ -\tilde{Z}_{k-1} & & & 1 & 1 \\ -\tilde{Z}_{k-1} & & & 1 & -1 \end{bmatrix} \in \mathbb{Z}^{(2n+4)\times(n+4)},$$

where we define

$$\tilde{X} = \begin{bmatrix} -1 & & & \\ 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in \mathbb{Z}^{n\times n},$$

$$\tilde{Y} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \end{bmatrix} \in \mathbb{Z}^{n\times n}, \text{ and}$$

$$\tilde{Z}_{k-1} = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \in \mathbb{Z}^{1\times n}.$$

Here $\tilde{Z}_{k-1}$ is the zero matrix except for the $(k-1)$th column which has a 1 entry. By direct computation, it is easy to see that the columns of V lie in $\ker(T)$. To see that V is a basis for $\ker(T)$, we can show that its columns are linearly independent by constructing a matrix $W \in \mathbb{Z}^{(n+4)\times(2n+4)}$ such that $WV = 2I_{(n+4)\times(n+4)}$. Indeed, we can do so in the following way. We can first define matrices

$$I_+ = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & 0 & 1 & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \in \mathbb{Z}^{n\times n}, \quad I_- = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & 0 & -1 & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \in \mathbb{Z}^{n\times n},$$

where the abnormal row is the $(k-1)$th row, and then define

$$W = \begin{bmatrix} I_+ & I_- & & & & \\ & & 1 & 1 & & \\ & & 1 & -1 & & \\ \tilde{Z}_k & -\tilde{Z}_k & & & 1 & 1 \\ & & & & 1 & -1 \end{bmatrix} \in \mathbb{Z}^{(n+4)\times(2n+4)},$$

where similarly to before, $\tilde{Z}_k \in \mathbb{Z}^{1\times n}$ is the one-hot vector with a 1 in the kth column. It is straightforward to verify that $WV = 2I_{(n+4)\times(n+4)}$, showing that the columns of V are linearly independent.

By looking at the columns of V, we have $\lambda_{n+4}(\Lambda) \le \sqrt{6}$, so by Lemma 2, we have $\eta_\epsilon(\Lambda) \le \sqrt{6}\cdot\sqrt{\omega(\log\lambda) + \ln n + \ln t} \le \sigma$, where we set $\epsilon = \mathsf{negl}(\lambda)/t$. Therefore by Lemma 8, we get that $Q_{]1[}(D_{\mathbb{Z}^{2n+4},\sigma})$ and $D_{\mathbb{Z}^n,2\sigma}$ are $\mathsf{negl}(\lambda)/t$ close if $\sigma \ge \sqrt{6}\cdot\sqrt{\omega(\log\lambda) + \ln n + \ln t}$. □

Lemma 12. *There is a* $\mathsf{poly}(n)$ *time algorithm that on input* $\mathbf{z} \in \mathcal{S}_{n,k}$ *outputs a matrix* $Z \in \mathbb{Z}^{n\times n}$ *(as a function of* $\mathbf{z}$*) that satisfies the following properties:*

- Z *is a permutation matrix with signs, i.e. a permutation matrix where the non-zero entries could be* ± 1 *instead of just* 1*,*
- $Z = Z^\top = Z^{-1}$*, and*
- $Z\mathbf{z} = \mathbf{u}$.

Proof. We can define Z as follows. Let

$$T_{\le k} = \{i \in [k] : z_i \ne 0\}, \qquad T_{>k} = \{i \in [n]\setminus[k] : z_i \ne 0\},$$
$$T^*_{\le k} = \{i \in [k] : z_i = 0\}, \qquad T^*_{>k} = \{i \in [n]\setminus[k] : z_i = 0\}.$$

Intuitively, $T_{\le k}$ and $T_{>k}$ partition the non-zero coordinates of $\mathbf{z}$ based on whether they lie in the first k coordinates, and $T^*_{\le k}$ and $T^*_{>k}$ partition the zero-coordinates of $\mathbf{z}$ based on whether they lie in the first k coordinates. Note that by k-sparsity of $\mathbf{z}$, we have

$$|T_{>k}| = k - |T_{\le k}| = |[k]\setminus T_{\le k}| = |T^*_{\le k}|.$$

Therefore, we can choose an arbitrary bijection $f : T_{>k} \to T^*_{\le k}$.

For all $i \in T_{\le k}$, we set $Z_{i,i} = z_i \in \{+1,-1\}$. For all $i \in T^*_{>k}$, we set $Z_{i,i} = 1$. For all $i \in T_{>k}$, we set $Z_{f(i),i} = z_i \in \{+1,-1\}$ and $Z_{i,f(i)} = Z_{f^{-1}(f(i)),f(i)} = z_i \in \{+1,-1\}$. We set all other entries of Z to be 0. It's clear from this definition that $Z = Z^\top$.

First, observe that Z is a signed permutation matrix. For all $i \in T_{\le k}\cup T^*_{>k}$, Z is the identity map up to signs (on basis vectors $\mathbf{e}_i$), and for all $i \in T_{>k}$, Z consists of signed transpositions $Z\mathbf{e}_i = z_i\mathbf{e}_{f(i)}$ and $Z\mathbf{e}_{f(i)} = z_i\mathbf{e}_{f^{-1}(f(i))} = z_i\mathbf{e}_i$. Therefore, Z is a signed permutation matrix, and furthermore we have also shown $Z^2 = I_{n\times n}$. Therefore, $Z = Z^{-1}$.

Lastly, we show $Z\mathbf{z} = \mathbf{u}$. We can decompose $\mathbf{z}$ as $\mathbf{z} = \mathbf{z}_{\le k} + \mathbf{z}_{>k}$ in the natural way by considering the non-zero coordinates of $\mathbf{z}$ on $[k]$ and $[n]\setminus[k]$ respectively. We then have

$$Z\mathbf{z} = Z(\mathbf{z}_{\le k} + \mathbf{z}_{>k}) = Z\mathbf{z}_{\le k} + Z\mathbf{z}_{>k} = 1_{T_{\le k}} + 1_{T^*_{\le k}} = \mathbf{u},$$

as desired. □

Definition 12. *We define a randomized mapping φ as follows. Let Q be as defined in Lemma 11. We sample $\mathbf{z} \sim \mathcal{S}_{n,k}$, $\mathbf{s} \sim \mathbb{Z}_q^m$, $\mathbf{a} \sim \mathbb{Z}_q^{n-1}$, $\mathbf{e} \sim D_{\mathbb{Z}^m,2\sigma}$, $G \sim D_{\mathbb{Z}^{m\times(n+5)},\sigma}$. Let $Z \in \mathbb{Z}^{n\times n}$ be as defined in Lemma 12 as a function of $\mathbf{z}$. On input $B \in \mathbb{Z}_q^{m\times(n-1)}$, we define*

$$\varphi(B; \mathbf{z}, \mathbf{s}, \mathbf{a}, \mathbf{e}, G) = \left[\left[\mathbf{s}, \mathbf{s}\cdot\mathbf{a}^\top + B, G\right] Q^\top Z, \mathbf{s} + \mathbf{e}\right].$$

First, we show that φ maps $B \sim U(\mathbb{Z}_q^{m\times(n-1)})$ to $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$.

Lemma 13. *Assume the same hypothesis as Theorem 5. For $B \sim U(\mathbb{Z}_q^{m\times(n-1)})$, we have $\varphi(B)$ and $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$ are $\mathsf{negl}(\lambda)$-close.*

Proof. We fix $\mathbf{a} \in \mathbb{Z}_q^{n-1}, \mathbf{z} \in S_{n,k}$ and we argue that $\varphi(B)$ maps to $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathbf{z}, D_{\mathbb{Z},\sigma'})$, i.e. the LWE distribution with secret $\mathbf{z}$. Averaging over $\mathbf{a}$ and $\mathbf{z}$ gives the desired result.

First, we show that $X = \left[\left[\mathbf{s}, \mathbf{s}\cdot\mathbf{a}^\top + B, G\right] Q^\top Z\right]$ looks uniform. By construction, $[\mathbf{s}, \mathbf{s}\cdot\mathbf{a}^\top + B]$ has distribution $U(\mathbb{Z}_q^{m\times n})$, by using the independent randomness of $\mathbf{s}$ and B. We can write

$$X = [\mathbf{s}, \mathbf{s}\cdot\mathbf{a}^\top + B]Q_{[n]}^\top Z + GQ_{]n[}^\top Z.$$

Since $Q_{[n]}$ and Z are invertible, by a one-time pad argument, we have $X \sim U(\mathbb{Z}_q^{m\times n})$, independent of G and e.

Now, we have to argue that the conditional distribution on $\mathbf{x} = \mathbf{s} + \mathbf{e}$ is equal to $X\mathbf{z} + \mathbf{e}'$ for some Gaussian noise $\mathbf{e}'$. We can directly write

$$\begin{aligned}
\mathbf{x} - X\mathbf{z} &= \mathbf{s} + \mathbf{e} - ([\mathbf{s}, \mathbf{s}\cdot\mathbf{a}^\top + B]Q_{[n]}^\top Z + GQ_{]n[}^\top Z)\mathbf{z} \\
&= \mathbf{s} + \mathbf{e} - [\mathbf{s}, \mathbf{s}\cdot\mathbf{a}^\top + B]Q_{[n]}^\top \mathbf{u} - GQ_{]n[}^\top \mathbf{u} \\
&= \mathbf{s} + \mathbf{e} - [\mathbf{s}, \mathbf{s}\cdot\mathbf{a}^\top + B]\mathbf{e}_1 - G\mathbf{v} \\
&= \mathbf{e} - G\mathbf{v},
\end{aligned}$$

where we use the fact that $Z\mathbf{z} = \mathbf{u}$, $\mathbf{u}^\top Q_{[n]} = \mathbf{e}_1^\top$ and $\mathbf{u}^\top Q_{]n[} = \mathbf{v}^\top$.

For all $j \in [m]$, let $\mathbf{g}_j \in \mathbb{Z}^{n+5}$ be the jth row of G. For each entry (row) $\tilde{e}_j$ of $\mathbf{e} - G\mathbf{v}$, we can write $\tilde{e}_j = e_j - \mathbf{g}_j^\top \mathbf{v} = \langle [e_j, \mathbf{g}_j], [1, -\mathbf{v}] \rangle$ and apply Lemma 6 with the vector $\mathbf{v}' = [1, -\mathbf{v}]$ to argue that $\tilde{e}_j$ is $O(\epsilon)$-close to $D_{\mathbb{Z},\sigma'}$ with $\sigma' = \sqrt{(2\sigma)^2 + \sum_{i\in[n+5]}(\sigma v_i)^2} = \sigma\sqrt{4 + \|\mathbf{v}\|_2^2} = 2\sigma\sqrt{k+1}$, as long as $\sigma \geq \sqrt{2}\|\mathbf{v}\|_\infty \eta_{\epsilon/(2(n+6)^2)}(\mathbb{Z})$. Now, using the triangle inequality over all m rows to get overall statistical distance $\mathsf{negl}(\lambda)$, we can set $\epsilon = \mathsf{negl}(\lambda)/m$, for which

$$\sigma \geq \sqrt{2} \cdot 2 \cdot \eta_{\mathsf{negl}(\lambda)/(mn^2)}(\mathbb{Z})$$

is sufficient. By Lemma 2, this holds as long as $\sigma \geq 4\sqrt{\ln m + \ln n + \omega(\log\lambda)}$, which we are given. □

Next, we show φ maps the standard LWE (with matrices as secrets) to standard LWE in slightly different dimensions, very much following the proof of Claim 3.3 of [Mic18].

Lemma 14. *Assume the same hypothesis as Theorem 5. Let $\mathcal{D}_1$ denote the distribution of $SA + E \pmod q$, where $A \sim U(\mathbb{Z}_q^{\ell\times(n-1)})$, $S \sim U(\mathbb{Z}_q^{m\times\ell})$, $E \sim D_{\mathbb{Z},\sigma}^{m\times(n-1)}$. Let $\mathcal{D}_2$ denote the distribution of $\hat{S}\hat{A} + \hat{E} \pmod q$, where $\hat{A} \sim U(\mathbb{Z}_q^{(\ell+1)\times(n+1)})$, $\hat{S} \sim U(\mathbb{Z}_q^{m\times(\ell+1)})$, $\hat{E} \sim D_{\mathbb{Z},2\sigma}^{m\times(n+1)}$. Then, $\varphi(\mathcal{D}_1)$ is $\mathsf{negl}(\lambda)$-close to $\mathcal{D}_2$.*

The proof goes exactly as in Claim 3.3 of [Mic18]. The only differences are in our matrices Q, Z, and our distribution of secrets $\mathbf{z} \sim \mathcal{S}_{n,k}$. The full differences are as follows.

- While our Z is different, since $Z = Z^\top$ is a permutation matrix with signs, it still holds that $Z \cdot D^n_{\mathbb{Z},2\sigma} = D^n_{\mathbb{Z},2\sigma}$ due to symmetry.
- We have $Q_{]1[}(D^{2n+4}_{\mathbb{Z},\sigma})$ is $\mathsf{negl}(\lambda)/m$-close to $D^n_{\mathbb{Z},2\sigma}$ by Lemma 11.
- The probability that $\mathbf{w}$ (in their notation) is not primitive is at most $\log(q)/2^\ell = \mathsf{negl}(\lambda)$, as desired.
- When applying leftover hash lemma (Lemma 1), the min-entropy of $\mathbf{z} \sim \mathcal{S}_{n,k}$ is now at least $k \log_2(n/k)$. Thus, we require $k \log_2(n/k) \geq (\ell+1)\log_2(q) + \omega(\log \lambda)$ instead of $n \geq (\ell+1)\log_2(q) + \omega(\log m)$.

For completeness, we provide a self-contained proof, exactly following Claim 3.3 of [Mic18].

Proof of Lemma 14. Let $B \sim \mathcal{D}_1$. Let $Y = [\mathbf{s}, \mathbf{s}\mathbf{a}^\top + B]$. By linearity, we can decompose Y as $Y = Y_s + Y_e$, where $Y_s = [\mathbf{s}, \mathbf{s}\mathbf{a}^\top + SA]$ and $Y_e = [\mathbf{0}, E]$. Similarly, we can write

$$\varphi(B) = \left[\left[\mathbf{s}, \mathbf{s}\cdot\mathbf{a}^\top + B, G\right] Q^\top Z, \mathbf{s} + \mathbf{e}\right] = [X_s, \mathbf{s}] + [X_e, \mathbf{e}],$$

where $X_s = Y_s Q^\top_{[n]} Z$ and $X_e = [Y_e, G]Q^\top Z = [E, G]Q^\top_{]1[}Z$. Our goal is to now show that $[X_s, \mathbf{s}]$ is statistically close to $\hat{S}\hat{A}$, and that $[X_e, \mathbf{e}]$ is statistically close to $\hat{E}$, where $\hat{S}\hat{A} + \hat{E}$ is a sample from $\mathcal{D}_2$. If this holds, then $\varphi(B)$ is statistically close to $\hat{S}\hat{A} + \hat{E}$, which completes the proof.

First, let us look at $[X_e, \mathbf{e}]$. Note that $\mathbf{e}$ is a discrete Gaussian vector of width 2σ independent of everything else, so the last column has the desired distribution. Furthermore, note that E and G have entries that are discrete Gaussian of width σ, so $[E, G] \sim D^{m\times(2n+4)}_{\mathbb{Z},\sigma}$. By Lemma 11, setting $t = m$, we can use the triangle inequality over all m rows to get that $[E, G]Q^\top_{]1[}$ is $\mathsf{negl}(\lambda)$ close to $D^{m\times n}_{\mathbb{Z},2\sigma}$ as long as $\sigma \geq \sqrt{6}\sqrt{\omega(\log\lambda) + \ln n + \ln m}$. Since Z is a signed permutation, by symmetry, we then know that $X_e = [E, G]Q^\top_{]1[}Z$ is $\mathsf{negl}(\lambda)$ close to $D^{m\times n}_{\mathbb{Z},2\sigma}$, and thus $[X_e, \mathbf{e}]$ is $\mathsf{negl}(\lambda)$ close to $D^{m\times(n+1)}_{\mathbb{Z},2\sigma}$, which is the same distribution as $\hat{E}$. Note that this depends only on $\mathbf{e}, G$, and E.

To finish, we look at $[X_s, \mathbf{s}]$. We now define

$$\hat{S} = \left[\mathbf{s}, S\right] W^{-1} \in \mathbb{Z}_q^{m\times(\ell+1)},$$

where W is a uniformly random invertible matrix over $\mathbb{Z}_q^{(\ell+1)\times(\ell+1)}$. Since W is invertible, using the randomness of S and $\mathbf{s}$, $\hat{S}$ is uniformly random independently of W. Next, we define

$$\hat{A} = WHQ^\top_{[n]}Z^\top[I_{n\times n}, \mathbf{z}] \in \mathbb{Z}_q^{(\ell+1)\times(n+1)}, \text{ where}$$
$$H = \begin{bmatrix} 1 & \mathbf{a}^\top \\ \mathbf{0} & A \end{bmatrix} \in \mathbb{Z}_q^{(\ell+1)\times n}.$$

Note that we have the identity $Q^\top_{[n]}Z^\top\mathbf{z} = Q^\top_{[n]}Z\mathbf{z} = Q^\top_{[n]}\mathbf{u} = \mathbf{e}_1$ by Lemmas 12 and 11, as well as the identity $\hat{S}WH = [\mathbf{s}, S]H = Y_s$. Therefore,

$$\hat{S}\hat{A} = \hat{S}WHQ^\top_{[n]}Z^\top[I_{n\times n}, \mathbf{z}] = Y_sQ^\top_{[n]}Z^\top[I_{n\times n}, \mathbf{z}] = [Y_sQ^\top_{[n]}Z, Y_s\mathbf{e}_1] = [X_s, \mathbf{s}],$$

as desired.

Now, we have to show that $\hat{S}$ and $\hat{A}$ have the correct distributions. We have already shown that $\hat{S}$ has the correct distribution (only depending on S and $\mathbf{s}$), so it suffices to show that $\hat{A}$ has the correct distribution given S and $\mathbf{s}$, using the randomness of $A, \mathbf{a}, W$ and $\mathbf{z}$. First, let's look at the matrix WH. Let $\mathbf{w}$ be the first column of W. The first column of WH will be exactly $\mathbf{w}$. Since W is a uniformly random invertible matrix, $\mathbf{w}$ is distributed uniformly among all primitive vectors in $\mathbb{Z}_q^{\ell+1}$, i.e. so that $\gcd(\mathbf{w}, q) = 1$. By Lemma 7, as long as $\log(q)/2^\ell = \mathsf{negl}(\lambda)$, which we have assumed, then the distribution of $\mathbf{w}$ is $\mathsf{negl}(\lambda)$-close to uniform over $\mathbb{Z}_q^{\ell+1}$. The remaining columns of WH will be $W \begin{bmatrix} \mathbf{a}^\top \\ A \end{bmatrix}$, which by using the uniform randomness of $\mathbf{a}$ and A, and the invertibility of W, will be uniformly random and independent of $\mathbf{w}$. Therefore, $WH \in \mathbb{Z}_q^{(\ell+1)\times n}$ is $\mathsf{negl}(\lambda)$-close to uniformly random. Now, since $Q_{[n]}^\top$ and $Z^\top$ are invertible, we have $WHQ_{[n]}^\top Z^\top$ is $\mathsf{negl}(\lambda)$-close to uniform, independently of $\mathbf{z}$. Let $A' = WHQ_{[n]}^\top Z^\top$, which we have just shown is $\mathsf{negl}(\lambda)$-close to uniform, independently of $\mathbf{z}$. Note that

$$\hat{A} = A'[I_{n\times n}, \mathbf{z}] = [A', A'\mathbf{z}].$$

Applying the leftover hash lemma (Lemma 1) and Lemma 10, since $k\log_2(n/k) \geq (\ell+1)\log_2(q) + \omega(\log\lambda)$, we know $\hat{A}$ is $\mathsf{negl}(\lambda)$-close to uniform, independently of $\hat{S}$ and $\hat{E}$. This completes the proof that $\varphi(\mathcal{D}_1)$ and $\mathcal{D}_2$ are $\mathsf{negl}(\lambda)$-close. □

With the above claims, we are ready to prove the main theorem of this section.

Proof of Theorem 5. We will show the contrapositive. Suppose we have a T-time distinguisher between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$ and $U(\mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m) = U(\mathbb{Z}_q^{m\times(n+1)})$ with advantage 2ϵ.

We have two cases. Suppose that this distinguisher distinguishes between $U(\mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m) = U(\mathbb{Z}_q^{m\times(n+1)})$ and $\mathcal{D}_2$ as given in Lemma 14, with advantage ϵ. Then, we have a T time distinguisher between $\mathsf{LWE}(n+1, \mathbb{Z}_q^{\ell+1}, \mathbb{Z}_q^{m\times(\ell+1)}, D_{\mathbb{Z}^m,2\sigma})$ and $U(\mathbb{Z}_q^{(\ell+1)\times(n+1)} \times \mathbb{Z}_q^{m\times(n+1)})$ where we simply discard the samples, i.e. the first part in $\mathbb{Z}_q^{(\ell+1)\times(n+1)}$ (the matrix $\hat{A}$).

Now, for the second case, suppose that this distinguisher does not distinguish between $U(\mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m) = U(\mathbb{Z}_q^{m\times(n+1)})$ and $\mathcal{D}_2$ with advantage ϵ. Then, we have a T-time distinguisher between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$ and $\mathcal{D}_2$ with advantage $\geq 2\epsilon - \epsilon = \epsilon$ by the triangle inequality. Now, we can use this distinguisher to distinguish $\mathsf{LWE}(n-1, \mathbb{Z}_q^\ell, \mathbb{Z}_q^{m\times\ell}, D_{\mathbb{Z}^m,2\sigma})$ and $U(\mathbb{Z}_q^{\ell\times(n-1)} \times \mathbb{Z}_q^{m\times(n-1)})$ by once again discarding the samples, i.e. the first part in $\mathbb{Z}_q^{\ell\times(n-1)}$ (the matrix A), and then by applying φ to the remaining part in $\mathbb{Z}_q^{m\times(n-1)}$. Now, using Lemmas 13 and 14, the resulting distributions coming out of φ when given $U(\mathbb{Z}_q^{m\times(n-1)})$ and $\mathcal{D}_1$ will be $\mathsf{negl}(\lambda)$-close to $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$ and $\mathcal{D}_2$, respectively. Thus, our assumed distingiusher will be correct, where the only runtime increase is in the randomized transformation φ, taking time $\mathsf{poly}(n, m, \log(q), \log(\lambda))$. □

Now, we state a simpler version of Theorem 5 that is easier to use.

Corollary 4. *Suppose* $\log(q)/2^\ell = \mathsf{negl}(\lambda), \sigma \geq 4\sqrt{\omega(\log\lambda) + \ln n + \ln m}$, *and* $k\log_2(n/k) \geq (\ell+1)\log_2(q) + \omega(\log\lambda)$. *Then, if* $\mathsf{LWE}(n, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ *and* $U(\mathbb{Z}_q^{\ell\times n} \times \mathbb{Z}_q^n)$ *have no* $T + \mathsf{poly}(n, m, q, \lambda)$ *time distinguisher with advantage* ϵ, *then* $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$ *and* $U(\mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m)$ *have no* T*-time distinguisher with advantage* $2\epsilon m + \mathsf{negl}(\lambda)$, *where* $\sigma' = 2\sigma\sqrt{k+1}$.

Proof. If $\mathsf{LWE}(n, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{\ell \times n} \times \mathbb{Z}_q^n)$ cannot be distinguished with advantage ϵ, then by a hybriding argument, the version where the secrets are matrices (with dimension m instead of 1) cannot be distinguished with advantage ϵm. Then, applying Theorem 5, $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$ and $U(\mathbb{Z}_q^{n \times m} \times \mathbb{Z}_q^m)$ cannot be distinguished with advantage $2\epsilon m + \mathsf{negl}(\lambda)$, where we reparameterize to absorb small additive factors, with the observation that LWE is harder when the dimension and noise grow, and easier when the number of samples grows. □

5 Reducing LWE to CLWE

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

Theorem 6 (Fixed-Norm LWE to CLWE)**.** *Let $r \in \mathbb{R}_{\geq 1}$, and let $\mathcal{S}$ be an arbitrary distribution over $\mathbb{Z}^n$ where all elements in the support of $\mathcal{S}$ have ℓ_2 norm r. Then, for*

$$\gamma = r \cdot \sqrt{\ln(m) + \ln(n) + \omega(\log \lambda)}, \text{ and}$$
$$\beta = O\left(\frac{\sigma}{q}\right),$$

if there is no $T + \mathsf{poly}(n, m, \log(q), \log(\sigma), \log(\lambda))$ time distinguisher between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{n \times m} \times \mathbb{Z}_q^m)$ with advantage at least $\epsilon - \mathsf{negl}(\lambda)$, then there is no T-time distinguisher between $\mathsf{CLWE}(m, D_1^n, \frac{1}{r} \cdot \mathcal{S}, \gamma, \beta)$ and $D_1^{n \times m} \times U(\mathbb{T}^m)$ with advantage ϵ, as long as $\sigma \geq 3r\sqrt{\ln(m) + \ln(n) + \omega(\log \lambda)}$.

See Figure 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 $\mathsf{negl}(\lambda)$ factors. We also remark that to keep the theorem general, the final distribution is not exactly the CLWE distribution, as the secret distribution is $\frac{1}{r} \cdot \mathcal{S}$ instead of $U(S^{n-1})$. However, using Lemma 19, it is straightforward to reduce from $\frac{1}{r} \cdot \mathcal{S}$ secrets to $U(S^{n-1})$ secrets.

This reduction goes via a series of transformations, which we briefly outline below:

1. Starting from standard decisional LWE, with samples $\mathbf{a} \sim U(\mathbb{Z}_q^n)$, (fixed) secret $\mathbf{s} \sim \mathcal{S}$ (where the support of $\mathcal{S}$ has fixed norm), and errors $e \sim D_{\mathbb{Z},\sigma}$, we convert discrete Gaussian errors e to continuous Gaussian errors $e \sim D_{\sigma_2}$ for σ_2 slightly larger than σ.

2. We convert discrete uniform samples $\mathbf{a} \sim U(\mathbb{Z}_q^n)$ to continuous uniform samples $\mathbf{a} \sim U(\mathbb{T}_q^n)$ with errors from D_{σ_3}, where σ_3 is slightly larger than σ_2.

3. We convert uniform $\mathbf{a} \sim U(\mathbb{T}_q^n)$ to Gaussian $\mathbf{a} \sim D_1^n$; viewing it as a CLWE distribution, we scale such the secret $\mathbf{s}$ is a unit vector (i.e. $\mathbf{s} \sim \frac{1}{r} \cdot \mathcal{S}$), $\gamma \approx r$, and the noise distribution becomes D_β where $\beta = \sigma_3/q$.

Setting of parameters. If we start with dimension n and m samples with error width σ:

1. After the first step, we get $\sigma_2 = O(\sigma)$, as long as $\sigma \geq 2\sqrt{\ln m + \omega(\log \lambda)}$.
2. After the second step, we get $\sigma_3 = O(\sigma_2) = O(\sigma)$, as long as $\sigma_2 \geq 3r\sqrt{\ln n + \ln m + \omega(\log \lambda)}$.
3. After the third step, we get $\gamma = r \cdot \sqrt{\ln n + \ln m + \omega(\log \lambda)}$ and $\beta = \sigma_3/q = O(\sigma/q)$.

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 15. *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 $\mathcal{S}$ over $\mathbb{Z}^n$, suppose there is no distinguisher between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m)$ running in time $T + \mathsf{poly}(m, n, \log(q), \log(\sigma))$. Then, there is no T-time distinguisher $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\sigma'})$ and $U(\mathbb{Z}_q^{n\times m}) \times U(\mathbb{T}_q^m)$ with an additive $\mathsf{negl}(\lambda)$ advantage loss, where*

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

Proof. We run our original distinguisher for $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\sigma'})$ and $U(\mathbb{Z}_q^{n\times m}) \times U(\mathbb{T}_q^m)$. For every sample $(\mathbf{a}, b)$ (from either $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\mathbb{Z},\sigma})$ or $U(\mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m)$), we sample a continuous Gaussian $e' \sim D_{\sigma''}$ where σ'' will be set later, and send $(\mathbf{a}, b + e' \pmod q)$ to the distinguisher.

By Lemma 4, we know that the distribution of $e' \pmod 1$ has statistical distance at most ϵ to $U([0,1))$ as long as $\sigma'' \geq \eta_\epsilon(\mathbb{Z})$. Therefore, if we are given samples from $U(\mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m)$, due to symmetry of $b \sim \mathbb{Z}_q$, we can set $\epsilon = \lambda^{-\omega(1)}/m$ to have $b + e' \pmod q$ look $\mathsf{negl}(\lambda)/m$-close to $\mathbb{T}_q$, making it look like samples from $U(\mathbb{Z}_q^{n\times m}) \times U(\mathbb{T}_q^m)$.

If we are given samples from $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\mathbb{Z},\sigma})$, then the second component can be seen as having noise $e + e'$, where $e \sim D_{\mathbb{Z},\sigma}$ and $e' \sim D_{\sigma''}$. Applying Lemma 3, as long as $1/\sqrt{1/\sigma^2 + 1/(\sigma'')^2} \geq \eta_\epsilon(\mathbb{Z})$, then $e + e'$ will look $O(\epsilon)$-close to $D_{\sqrt{\sigma^2+(\sigma'')^2}}$. Thus, as long as $\sigma, \sigma'' \geq \sqrt{2} \cdot \eta_\epsilon(\mathbb{Z})$, it all goes through, as taking errors mod q (i.e. in $\mathbb{T}_q$ instead of $\mathbb{R}$) can only decrease statistical distance. Now, applying Lemma 2, we can set $\epsilon = \lambda^{-\omega(1)}/m$ and $\sigma'' = \sqrt{4\ln(m) + \omega(\log\lambda)}$, 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 $\mathsf{negl}(\lambda)$-closeness of all samples. □

Step 2: Converting discrete to continuous samples. Now, we convert discrete uniform samples $\mathbf{a} \sim \mathbb{Z}_q^n$ to continuous uniform samples $\mathbf{a} \sim \mathbb{T}_q^n$.

Lemma 16. *Let $n, m, q \in \mathbb{N}$, $\sigma \in \mathbb{R}$. Let $\mathcal{S}$ 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 $T+\mathsf{poly}(m, n, \log(q), \log(\sigma))$-time distinguisher between the distributions $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_\sigma)$ and $U(\mathbb{Z}_q^{n\times m}) \times U(\mathbb{T}_q^m)$. Then, there is no T-time distinguisher between the distributions $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_{\sigma'})$ and $U(\mathbb{T}_q^{n\times m} \times \mathbb{T}_q^m)$ with an additive $\mathsf{negl}(\lambda)$ advantage loss, where we set

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

Proof. We run our distinguisher for $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_{\sigma'})$ and $U(\mathbb{T}_q^{n\times m} \times \mathbb{T}_q^m)$. Let $\epsilon = \mathsf{negl}(\lambda)/m$, and let $\sigma'' \geq \sqrt{2} \cdot \eta_\epsilon(\mathbb{Z}^n)$. For each sample $(\mathbf{a}, b)$ (from either $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_\sigma)$ or $U(\mathbb{Z}_q^{n\times m}) \times U(\mathbb{T}_q^m)$), we sample a continuous Gaussian $\mathbf{a}' \sim (D_{\sigma''})^n$ and send $(\mathbf{a} + \mathbf{a}' \pmod q, b)$ to the distinguisher. By Lemma 4, we know that the distribution of $\mathbf{a}' \pmod 1$ has statistical distance at most $\epsilon = \mathsf{negl}(\lambda)/m$

to $U([0,1)^n)$. Thus, by symmetry over $\mathbf{a} \sim (\mathbb{Z}_q)^n$, the distribution of $\mathbf{a} + \mathbf{a}' \pmod q$ will be $\mathsf{negl}(\lambda)/m$-close to uniform over $(\mathbb{T}_q)^n$. Therefore, by the triangle inequality, if we are given samples from $U(\mathbb{Z}_q^{n\times m}) \times U(\mathbb{T}_q^m)$, the reduction gives samples to the distinguisher that are $\mathsf{negl}(\lambda)$-close to $U(\mathbb{T}_q^{n\times m} \times \mathbb{T}_q^m)$.

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

$$(\mathbf{a}+\mathbf{a}', \langle \mathbf{a}, \mathbf{s}\rangle + e) = (\mathbf{a}+\mathbf{a}', \langle \mathbf{a}+\mathbf{a}', \mathbf{s}\rangle + e - \langle \mathbf{a}', \mathbf{s}\rangle) = (\mathbf{a}+\mathbf{a}', \langle \mathbf{a}+\mathbf{a}', \mathbf{s}\rangle + e'),$$

where we define

$$e' = e - \langle \mathbf{a}', \mathbf{s}\rangle$$

over $\mathbb{R}$. Conditioned on $\mathbf{a}+\mathbf{a}' \mod q$, $\mathbf{a}'$ is a discrete Gaussian distributed according to $D_{\mathbb{Z}^n+(\mathbf{a}+\mathbf{a}'),\sigma''}$. By Lemma 3, as long as $\sigma \geq r\sigma''$, the distribution of e' is $O(\epsilon) = \mathsf{negl}(\lambda)/m$ close to $D_{\sigma'}$, where

$$\sigma' = \sqrt{\sigma^2 + r^2(\sigma'')^2}.$$

Averaging the distribution of e' over $\mathbf{s}$ will not change the distribution over e', as all secrets $\mathbf{s}$ have fixed norm r. Therefore, if we are given the m samples from $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_\sigma)$, the reduction gives us samples $\mathsf{negl}(\lambda)$-close to $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_{\sigma'})$, as desired.

To set parameters, we choose $\sigma'' = 3\sqrt{\ln n + \ln m + \omega(\log\lambda)}$ to ensure that $\sigma'' \geq \sqrt{2}\cdot\eta_{\mathsf{negl}(\lambda)/m}(\mathbb{Z}^n)$. 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 17. *Let $t \in \mathbb{R}_{>0}$ be a parameter. There is a $\mathsf{poly}(n, \log(t), \log(\lambda))$-time algorithm such that on input $\mathbf{z} \in \mathbb{T}_1^n$, the algorithm outputs some $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathbf{z} \pmod 1$. Moreover, if $\mathbf{z} \sim U(\mathbb{T}_1^n)$, then the distribution on the outputs $\mathbf{y}$ is $\mathsf{negl}(\lambda)/t$-close to D_τ^n, 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 $\mathbf{y}$ 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 $z \sim \mathbb{T}_1$, then y is $\mathsf{negl}(\lambda)/(tn)$-close to D_τ. We sample

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

which can be sampled efficiently (see e.g. [BLP+13], Section 5.1 of full version), where we have $\mathsf{negl}(\lambda)/(tn)$ 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 $\mathsf{negl}(\lambda)/(tn)$-close to D_τ when $z \sim U(\mathbb{T}_1)$. Note that for fixed $z \in [0,1)$, we can write the generalized PDF of $D_{\mathbb{Z}+z,\tau}$ as

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

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

$$\begin{aligned}
D_{\mathbb{Z}+U([0,1)),\tau}(x) &= \int_0^1 1 \cdot D_{\mathbb{Z}+z,\tau}(x) \cdot dz \\
&= \int_0^1 \delta(x - z \mod 1) \cdot \frac{\rho_\tau(x)}{\rho_\tau(\mathbb{Z}+z)} dz \\
&= \frac{\rho_\tau(x)}{\rho_\tau(\mathbb{Z}+x)} \\
&\in \left[1, \frac{1+\epsilon}{1-\epsilon}\right] \cdot \frac{\rho_\tau(x)}{\rho_\tau(\mathbb{Z})} \\
&\propto \left[1, \frac{1+\epsilon}{1-\epsilon}\right] \cdot \rho_\tau(x),
\end{aligned}$$

where the inclusion comes from Lemma 5. Therefore, a standard calculation shows that the statistical distance between $D_{\mathbb{Z}+U([0,1)),\tau}$ and D_τ is at most $O(\epsilon)$. Setting $\epsilon = \lambda^{-\omega(1)}/(t \cdot n)$, we need to take $\tau \geq \eta_{\lambda^{-\omega(1)}/(t\cdot n)}(\mathbb{Z})$, which we can do by setting $\tau = \sqrt{\ln n + \ln t + \omega(\log \lambda)}$ by Lemma 2.

□

Lemma 18. *Let $n, m, q \in \mathbb{N}, \sigma, r, \gamma \in \mathbb{R}$. Let $\mathcal{S}$ be a distribution over $\mathbb{Z}^n$ where all elements in the support have fixed norm r. Suppose there is no $T + \mathsf{poly}(n, m, \log(q), \log(\lambda))$ time distinguisher between the distributions $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_\sigma)$ and $U(\mathbb{T}_q^{n\times m} \times \mathbb{T}_q^m)$. Then, there is no T-time distinguisher between the distributions $\mathsf{CLWE}(m, D_1^n, \frac{1}{r} \cdot \mathcal{S}, \gamma, \beta)$ and $D_1^{n\times m} \times U(\mathbb{T}_1^m)$ with an additive advantage loss of $\mathsf{negl}(\lambda)$, where*

$$\begin{aligned}
\gamma &= r \cdot \sqrt{\ln n + \ln m + \omega(\log \lambda)}, \\
\beta &= \frac{\sigma}{q}.
\end{aligned}$$

Proof. We run the distinguisher for $\mathsf{CLWE}(m, D_1^n, \frac{1}{r} \cdot \mathcal{S}, \gamma, \beta)$ and $D_1^{n\times m} \times U(\mathbb{T}_1^m)$. For each sample $(\mathbf{a}, b)$ from either $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_\sigma)$ or $U(\mathbb{T}_q^{n\times m} \times \mathbb{T}_q^m)$, we invoke Lemma 17 on $\mathbf{a}/q$ with parameter $t = m$ to get some $\mathbf{y} \in \mathbb{R}^n$ with statistical distance $\mathsf{negl}(\lambda)/m$ from D_τ^n such that $\mathbf{y} = \mathbf{a}/q \pmod 1$, where $\tau = \sqrt{\ln n + \ln m + \omega(\log \lambda)}$. We then send $(\mathbf{y}/\tau, b/q)$ to the distinguisher. Let $\gamma = r \cdot \tau$, $\mathbf{y}' = \mathbf{y}/\tau$, $\mathbf{s}' = \mathbf{s}/r$, and $e' = e/q$. If $(\mathbf{a}, b)$ is a sample from $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_\sigma)$, then for secret $\mathbf{s} \sim \mathcal{S}$, since $\mathbf{s} \in \mathbb{Z}^n$, we have

$$\begin{aligned}
(\mathbf{y}/\tau, b/q) = (\mathbf{y}', \langle \mathbf{a}/q, \mathbf{s}\rangle + e/q \pmod 1) &= (\mathbf{y}', \langle \mathbf{y}, \mathbf{s}\rangle + e' \pmod 1) \\
&= (\mathbf{y}', r \cdot \tau \cdot \langle \mathbf{y}', \mathbf{s}/r\rangle + e' \pmod 1) \\
&= (\mathbf{y}', \gamma \cdot \langle \mathbf{y}', \mathbf{s}'\rangle + e' \pmod 1)
\end{aligned}$$

where this is now $\mathsf{negl}(\lambda)/m$ close to a sample from $\mathsf{CLWE}(m, D_1^n, \frac{1}{r} \cdot \mathcal{S}, \gamma, \beta)$, as $\mathbf{y}' \sim D_1^n$, $\mathbf{s}' \sim \frac{1}{r} \cdot \mathcal{S}$, and $e' \sim D_{\sigma/q} = D_\beta$. Applying this reduction to $U(\mathbb{T}_q^{n\times m} \times \mathbb{T}_q^m)$ clearly gives us a statistically close sample to $D_1^{n\times m} \times U(\mathbb{T}_1^m)$ by Lemma 17 and the triangle inequality over all m samples. □

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 ([BRST21], Claim 2.22). For completeness, we provide a proof.

Lemma 19 ([BRST21], Claim 2.22)**.** *Let $n, m \in \mathbb{N}$, and let $\beta \in \mathbb{R}_{>0}$. Let $\mathcal{S}$ be a distribution over $\mathbb{R}^n$ of fixed norm 1. There is no T-time distinguisher between the distributions $\mathsf{CLWE}(m, D_1^n, \gamma, \beta)$ and $D_1^{n\times m} \times U(\mathbb{T}_1^m)$, assuming there is no $T+\mathsf{poly}(n, m)$ time distinguisher between the distributions $\mathsf{CLWE}(m, D_1^n, \mathcal{S}, \gamma, \beta)$ and $D_1^{n\times m} \times U(\mathbb{T}_1^m)$. 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) $\mathcal{S}$.*

Note that while we do not use Lemma 19 in proving Theorem 6, we do use the lemma in subsequent sections.

Proof. We run the distinguisher for $\mathsf{CLWE}(m, D_1^n, \gamma, \beta)$ and $D_1^{n\times m} \times U(\mathbb{T}_1^m)$. Let $R \in \mathbb{R}^{n\times n}$ be a uniformly random rotation matrix in $\mathbb{R}^n$, fixed for all samples. When giving the distinguisher a sample, we get $(\mathbf{a}, b)$ from either $\mathsf{CLWE}(m, D_1^n, \mathcal{S}, \gamma, \beta)$ or $D_1^{n\times m} \times U(\mathbb{T}_1^m)$, and send $(R\mathbf{a}, b)$ to the distinguisher. If $(\mathbf{a}, b)$ is drawn from $\mathsf{CLWE}(m, D_1^n, \mathcal{S}, \gamma, \beta)$, then we have

$$\begin{aligned}(R\mathbf{a}, b) = (R\mathbf{a}, \gamma\langle\mathbf{a}, \mathbf{s}\rangle + e \pmod 1) &= (R\mathbf{a}, \gamma\langle R\mathbf{a}, R\mathbf{s}\rangle + e \pmod 1)\\ &= (\mathbf{a}', \gamma\langle\mathbf{a}', \mathbf{w}\rangle + e \pmod 1),\end{aligned}$$

for $\mathbf{a} \sim D_1^n$, $\mathbf{s} \sim \mathcal{S}$, and $e \sim D_\beta$, where we set $\mathbf{a}' = R\mathbf{a}$ and $\mathbf{w} = R\mathbf{s}$ (fixed for all samples). For an arbitrary rotation R, since the distribution on $\mathbf{a}$ is spherically symmetric, we have $\mathbf{a}' = R\mathbf{a} \sim D_1^n$, independently of R. For a random rotation matrix R, for arbitrary $\mathbf{s}$, we have that $\mathbf{w} = R\mathbf{s}$ is a uniformly random unit vector in $\mathbb{R}^n$. Since this holds for arbitrary $\mathbf{s}$, this also holds when averaging over the distribution $\mathbf{s} \sim \mathcal{S}$. If $(\mathbf{a}, b)$ is drawn from $D_1^{n\times m} \times U(\mathbb{T}_1^m)$, then $(R\mathbf{a}, b)$ is drawn identically to $(\mathbf{a}, b)$, since the distribution on $\mathbf{a}' = R\mathbf{a}$ is spherically symmetric. Thus, the reduction maps the distributions perfectly. □

Now, we are ready to prove the main theorem of this section, Theorem 6.

Proof of Theorem 6. Throughout this proof, when we refer to distinguishing probability, we omit additive $\mathsf{negl}(\lambda)$ terms for simplicity.

Suppose there is no distinguisher with advantage ϵ between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m)$. Then, by Lemma 15, there is no ϵ-distinguisher between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\sigma_2})$ and $U(\mathbb{Z}_q^{n\times m}) \times U(\mathbb{T}_q^m)$, where $\sigma_2 = O(\sigma)$, as long as $\sigma \geq \sqrt{4\ln(m) + \omega(\log\lambda)}$, which it is by our assumption on σ. Then, by Lemma 16, there is no ϵ-distinguisher between $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_{\sigma_3})$ and $U(\mathbb{T}_q^{n\times m} \times \mathbb{T}_q^m)$, 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 18, there is no ϵ-distinguisher between $\mathsf{CLWE}(m, D_1^n, \frac{1}{r}\cdot\mathcal{S}, \gamma, \beta)$ and $D_1^{n\times m} \times U(\mathbb{T}_1^m)$, 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. □

5.1 Full Reduction from LWE to CLWE

Now, to reduce from standard decisional LWE where the secret is drawn uniformly over $\mathbb{Z}_q^n$ instead of a fixed-norm distribution, we need to somehow reduce standard LWE to some version where the norm is fixed. We show two ways to do this:

1. In Corollary 5, we use a reduction from LWE to binary-secret LWE [Mic18] (i.e. Section 4 but without sparsity) to bridge this gap.
2. In Appendix A, we give another (perhaps simpler) reduction, but we reduce to search CLWE instead of decisional CLWE. (As a result of Appendix C, we get an indirect search-to-decision reduction for discrete-secret CLWE that can be applied here.)

In this section, we show the first approach.

Theorem 7 ([Mic18], Theorem 3.1 and Lemma 2.9)**.** *Let $q, \ell, n, m \in \mathbb{Z}$, $\sigma \in \mathbb{R}$. There is no T-time algorithm has advantage ϵ in distinguishing $\mathsf{LWE}(m, \mathbb{Z}_q^{n+1}, \{+1,-1\}^{n+1}, D_{\mathbb{Z},\sigma'})$ and $U(\mathbb{Z}_q^{(n+1)\times m} \times \mathbb{Z}_q^m)$, assuming there is no time $T+\mathsf{poly}(\ell, n, \log(q), \log(\lambda))$ algorithm with advantage $(\epsilon-\mathsf{negl}(\lambda))/(2m)$ in distinguishing $\mathsf{LWE}(n+1, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{\ell\times(n+1)} \times \mathbb{Z}_q^{n+1})$, as long as $\log(q)/2^\ell = \mathsf{negl}(\lambda), \sigma \geq 4\sqrt{\omega(\log\lambda) + \ln n + \ln m}$, $n \geq 2\ell\log_2 q + \omega(\log\lambda)$, and $\sigma' = 2\sigma\sqrt{n+1}$.*

Remark 2. *Note that we phrase the parameter requirements differently here than is done in [Mic18], mainly because we want to delink the security parameter from n. Explicitly:*

- *The requirements $q \leq 2^{\mathsf{poly}(n)}$ and $\ell \geq \omega(\log n)$ in [Mic18] are needed only to make sure that the first row of a primitive matrix is close to uniform over $\mathbb{Z}_q$. Indeed, Lemma 2.2 of [Mic18] shows the statistical distance is at most $\log(q)/2^\ell$. Thus, the requirement $\log(q)/2^\ell = \mathsf{negl}(\lambda)$ is sufficient.*
- *We require $\sigma \geq 4\sqrt{\omega(\log\lambda) + \ln n + \ln m}$ instead of $\sigma \geq \omega(\sqrt{\log n})$ for various triangle inequalities to go through to get $\mathsf{negl}(\lambda)$ overall statistical distance.*

Now, we are ready to give a proof of Corollary 5.

Corollary 5 (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 ϵ between $\mathsf{CLWE}(m, D_1^n, \gamma, \beta)$ and $D_1^{n\times m} \times U(\mathbb{T}^m)$, assuming there is no $T + \mathsf{poly}(\ell, n, m, \log(q), \log(\sigma), \log(\lambda))$ time distinguisher with advantage $(\epsilon - \mathsf{negl}(\lambda))/(2m)$ between $\mathsf{LWE}(m, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{\ell\times m} \times \mathbb{Z}_q^m)$, 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 $\log(q)/2^\ell = \mathsf{negl}(\lambda)$, $n \geq 2\ell\log_2 q + \omega(\log\lambda)$, and $\sigma \geq C \cdot \sqrt{\ln m + \omega(\log\lambda)}$ for some universal constant C.

Remark 3. *Note that for reasonable parameter settings of CLWE (namely where $\beta \ll 1$), we require $q/\sigma \gg \sqrt{n}$.*

Proof of Corollary 5. Suppose there is no distinguisher with advantage $(\epsilon - \mathsf{negl}(\lambda))/(2m)$ between $\mathsf{LWE}(m, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{\ell \times m} \times \mathbb{Z}_q^m)$. Then, since $n < m$ and more samples can only help, there is no distinguisher with advantage $(\epsilon - \mathsf{negl}(\lambda))/(2m)$ between $\mathsf{LWE}(n, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{\ell \times n} \times \mathbb{Z}_q^n)$. Then, by Theorem 7, there is no distinguisher between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \{+1, -1\}^n, D_{\mathbb{Z},\sigma_1})$ and $U(\mathbb{Z}_q^{n \times m} \times \mathbb{Z}_q^m)$ with advantage ϵ, where $\sigma_1 = 2\sigma\sqrt{n+1}$, and all other sufficient conditions are met by the hypotheses of the corollary. (From here on out, we omit additive $\mathsf{negl}(\lambda)$ terms in the distinguishing probability for simplicity.)

Now, since the secrets all have fixed norm $\sqrt{n}$, we can apply Theorem 6. Then, for parameters

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

and

$$\beta = O\left(\frac{\sigma_1}{q}\right) = O\left(\frac{\sigma\sqrt{n}}{q}\right),$$

there is no distinguisher between $\mathsf{CLWE}(m, D_1^n, \frac{1}{\sqrt{n}}\{+1, -1\}^n, \gamma, \beta)$ and $D_1^{n \times m} \times U(\mathbb{T}^m)$, as long as

$$\sigma_1 = 2\sigma\sqrt{n+1} \geq 3\sqrt{n}\sqrt{\ln(m) + \ln(n) + \omega(\log \lambda)},$$

which indeed holds as long as $\sigma \geq C \cdot \sqrt{\ln(m) + \omega(\log \lambda)}$ for some universal constant C.

Lastly, we make the secret direction for the CLWE distribution a completely random unit vector in $\mathbb{R}^n$ via Lemma 19. This has no effect on any of the parameters, so this means there is no distinguisher between $\mathsf{CLWE}(m, D_1^n, \gamma, \beta)$ and $D_1^{n \times m} \times U(\mathbb{T}^m)$, as desired. □

5.2 Hardness of Sparse CLWE

In this subsection, we take advantage of our reduction from LWE to k-sparse LWE to reduce LWE to a k-sparse version of CLWE with a very similar proof to that of Corollary 5. Later on, the main benefit of this reduction is that in the resulting CLWE distribution, γ will be small, which will result in a family of GMM instances, each with a small number of Gaussians.

Corollary 6 (Reduction from LWE to k-sparse CLWE)**.** *Suppose* $\log(q)/2^\ell = \mathsf{negl}(\lambda), \sigma \geq 2 \cdot \sqrt{\ln n + \ln m + \omega(\log \lambda)}$*, and* $k \log_2(n/k) \geq (\ell + 1)\log_2(q) + \omega(\log \lambda)$*. Then, for parameters*

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

and

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

for some universal constant C*, if* $\mathsf{LWE}(n, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ *and* $U(\mathbb{Z}_q^{\ell \times n} \times \mathbb{Z}_q^n)$ *have no* $T + \mathsf{poly}(n, m, \log(q), \log(\sigma), \log(\lambda))$ *time distinguisher with advantage* ϵ*, then* $\mathsf{CLWE}(m, D_1^n, \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}, \gamma, \beta)$ *and* $D_1^{n \times m} \times U(\mathbb{T}^m)$ *have no* T*-time distinguisher with advantage* $2\epsilon m + \mathsf{negl}(\lambda)$*.*

Proof. By Corollary 4, we know there is no distinguisher with advantage $2\epsilon m + \mathsf{negl}(\lambda)$ between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}_{n,k}, D_{\mathbb{Z},\sigma'})$ and $U(\mathbb{Z}_q^{n \times m} \times \mathbb{Z}_q^m)$, where $\sigma' = 2\sigma\sqrt{k+1}$. Now, applying Theorem 6, since all secret vectors have norm $\sqrt{k}$, for parameters

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

and

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

there is no T-time distinguisher with advantage $2\epsilon m + \mathsf{negl}(\lambda)$ between $\mathsf{CLWE}(m, D_1^n, \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}, \gamma, \beta)$ and $D_1^{n\times m} \times U(\mathbb{T}^m)$, as long as

$$\sigma' = 2\sigma\sqrt{k+1} \geq 3\sqrt{k}\sqrt{\ln(m) + \ln(n) + \omega(\log\lambda)},$$

which indeed holds by our assumption on σ. □

5.3 Classical Hardness of CLWE

With our reduction from fixed-norm LWE to CLWE, we can now show that worst-case lattice problems reduce *classically* to CLWE, whereas Corollary 3.2 of [BRST21] gives a *quantum* reduction from worst-case lattice problems to CLWE. This now essentially follows from the following theorem due to [BLP+13]:

Theorem 8 (Theorem 1.1 of [BLP+13], informal)**.** *There is an efficient classical reduction from (worst-case) $\sqrt{n}$-dimensional* gapSVP *to decisional* LWE *in dimension n with modulus $q = \mathsf{poly}(n)$.*

Given Theorem 8, we can now prove Corollary 3.

Proof Sketch of Corollary 3. One way to approach this (with slightly worse parameters) is to directly combine Theorem 8 and Corollary 5. However, to be less wasteful, we briefly describe below how to optimize the reduction by bypassing LWE with $U(\mathbb{Z}_q^n)$ secrets and working instead with just binary secrets. In fact, Theorem 8 uses a definition of LWE with continuous noise, so one has to be a bit careful regardless.

At a very high level, we combine Theorem 8 and Theorem 6, but modified (in a small way) so that the LWE distribution resulting from Theorem 8 has fixed norm. We modify their reduction as follows:

- We observe that their modulus switching reduction, Corollary 3.2, preserves the secret distribution $U(\{0,1\}^n)$. The last step of their reduction, just after applying Corollary 3.2, reduces this secret distribution, $U(\{0,1\}^n)$, to $U(\mathbb{Z}_q^n)$ by a standard random self-reduction. This has the the effect of going back from binary LWE to standard LWE to finish the reduction. In our case, we remove this final reduction and keep the secret distribution binary.

- Furthermore, throughout the reduction, we substitute $U(\{0,1\}^n)$ secrets with $U(\{+1,-1\}^n)$ secrets. To do this, we modify Theorem 4.1 in [BLP+13] to handle $U(\{+1,-1\}^n)$ secrets. Their proof of Theorem 4.1 is general in that it only requires the secret distribution to be efficiently samplable, have enough high min-entropy as needed to apply the leftover hash lemma, have norm at most $\sqrt{n}$, and have small "quality" (see Definition 4.5 of [BLP+13]). Since the quality of $U(\{+1,-1\}^n)$ can be bounded above by 2, it is easy to see that $U(\{+1,-1\}^n)$ satisfies all of these conditions. (We note there are other ways to make this change; Theorem 7, due to [Mic18], shows a reduction from $U(\mathbb{Z}_q^n)$ with $U(\{+1,-1\}^n)$ with very similar parameters. In fact, if q is odd, $\{+1,-1\}^n$ secrets and $\{0,1\}^n$ secrets have straightforward reductions to each other, as shown in Lemma 2.12 and Lemma 2.13 of [Mic18].) The only reason we make

this change is that the secrets will now have fixed norm $\sqrt{n}$ (instead of norm at most $\sqrt{n}$), which will allow us to use our fixed-norm LWE to CLWE reduction. Lastly, Corollary 3.2 of [BLP+13] simply requires an upper bound on the norm of the secret distribution, so the same result holds for $U(\{+1,-1\}^n)$ secrets.

Therefore, we have a (classical) reduction from worst-case lattice problems in dimension $\sqrt{n}$ to (decisional) LWE in dimension n with $q = \mathsf{poly}(n)$, with secret distribution $\mathcal{S} = U(\{+1,-1\}^n)$ and continuous Gaussian errors. Thus, we can just use Lemma 16 and Lemma 18 to reduce to CLWE with $r = \sqrt{n}$. If desired, one can use Lemma 19 to make the secret distribution $U(S^{n-1})$ instead of $U(\frac{1}{\sqrt{n}}\{+1,-1\}^n)$. (The exact parameter dependencies come from combining Theorem 2.16 of [BLP+13], Theorem 2.17 of [BLP+13], Theorem 4.1 of [BLP+13], Corollary 3.2 of [BLP+13], Lemma 16, Lemma 18, and optionally Lemma 19.) □

6 Hardness of Density Estimation for Mixtures of Gaussians

Now, using tools from the previous sections, we reduce LWE to density estimation for mixtures of Gaussians, using similar ideas as [BRST21]. Our machinery from the previous sections now allows us to give a fine-grained version of hardness of learning mixtures of Gaussians.

Lemma 20 (Reducing LWE to GMM via k-sparse CLWE)**.** *Suppose* $\log(q)/2^\ell = o(1), \sigma \geq 10 \cdot \sqrt{\ln n + \ln m}$, $k \log_2(n/k) \geq (\ell+1)\log_2(q) + \omega(1)$, $q = \omega(\sigma\sqrt{k})$ *and* $q \leq m^2$. *Then, for*

$$g = O\left(\sqrt{k\ln(m)} \cdot \sqrt{\ln(m) + \ln(n)}\right),$$

if $\mathsf{LWE}(n, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ *and* $U(\mathbb{Z}_q^{\ell\times n} \times \mathbb{Z}_q^n)$ *have no* $T+\mathsf{poly}(n,m,q)$ *time distinguisher with advantage* $\Omega(1/m^3)$, *then density estimation for GMM in dimension* n *with* g *Gaussian components and* m *samples has no* T*-time solver.*

Proof. In short, this follows by composing the reductions from LWE to k-sparse CLWE (Corollary 6), from CLWE to hCLWE (Lemma 9), and from hCLWE to density estimation for mixtures of Gaussians (Theorem 4).

As used in Corollary 6, let $\beta = \Theta(\sigma\sqrt{k}/q) = o(1)$, and let m' denote the number of CLWE samples. In anticipation of applying Lemma 9 in reducing CLWE to hCLWE with $\delta = \beta$, we set

$$m' = \Theta\left(\frac{m}{\beta}\right) = \Theta\left(\frac{mq}{\sigma\sqrt{k}}\right) < m^3,$$

where the final inequality holds (for, say, sufficiently large values of m) since $q \leq m^2$, $k \geq 1$, and $\sigma \geq 10\sqrt{\ln n + \ln m} = \omega(1)$. Since $m' < m^3$, it follows that $\ln(m') < \ln(m^3) = 3\ln(m)$.

To reduce LWE to k-sparse CLWE, we apply Corollary 6 with $\epsilon = 1/(6m')$. Since we have the conditions $\log(q)/2^\ell = o(1)$, $k\log_2(n/k) \geq (\ell+1)\log_2(q) + \omega(1)$, and

$$\sigma \geq 10\sqrt{\ln n + \ln m} > 3\sqrt{\ln n + \ln m'} > 2\sqrt{\ln n + \ln m' + \omega(1)},$$

one can choose sufficiently small $\lambda = \omega(1)$ to satisfy the conditions of Corollary 6 such that the $\mathsf{negl}(\lambda)$ additive term in the advantage loss is at most $1/100$ and such that

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

Corollary 6 then implies that there is no T-time distinguisher with advantage

$$2\epsilon m' + \frac{1}{100} < \frac{2}{5}$$

between $\mathsf{CLWE}(m', D_1^n, \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}, \gamma, \beta)$ and $D_1^{n\times m'} \times U\left(\mathbb{T}^{m'}\right)$. By Lemma 9, we reduce m' samples of CLWE to m samples of hCLWE with parameter $\delta = \beta$, so that $\beta' = \sqrt{2}\beta$ and $\gamma' = \gamma$, at the cost of $\mathsf{poly}(n, m, 1/\beta) = \mathsf{poly}(n, m, q)$ time and $1/1000$ additional failure probability. Then, by Theorem 4, there is no GMM learner for

$$g = 4\gamma\sqrt{\ln(m)/\pi} + 1 = O\left(\sqrt{k\ln(m)} \cdot \sqrt{\ln(m) + \ln(n)}\right)$$

Gaussians, as long as $\beta' < 1/32$, which holds in our case as $\beta' = \sqrt{2}\beta = o(1)$. □

Now, we set parameters and invoke Lemma 20.

Corollary 7. *Suppose* $10\sqrt{\ln(m) + \ln(n)} \le \sigma$, $\omega(\sigma\sqrt{k}) \le q \le \mathsf{poly}(\ell)$, $k\log_2(n/k) = (1 + \Theta(1))\ell\log_2(q)$, $q \le m^2$, *and* $m \le \mathsf{poly}(n)$, *and suppose that* $\mathsf{LWE}(n, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ *and* $U(\mathbb{Z}_q^{\ell\times n} \times \mathbb{Z}_q^n)$ *have no* $T(\ell) + \mathsf{poly}(n)$ *time distinguisher with advantage at least* $\Omega(1/m^3)$. *Then, there is no algorithm solving density estimation in dimension* n *with* m *samples for* g *Gaussians, where*

$$g = O\left(\sqrt{k \cdot \log(m) \cdot \log(n)}\right).$$

Proof. First, since $q \le \mathsf{poly}(\ell)$, we have $\log(q)/2^\ell \le O(\log(\ell)/2^\ell) = o(1)$. Thus, we can invoke Lemma 20. This gives

$$g = O\left(\sqrt{k\ln(m)} \cdot \sqrt{\ln(m) + \ln(n)}\right) = O\left(\sqrt{k \cdot \log(m) \cdot \log(n)}\right),$$

as $\ln(m) = O(\log n)$ by our assumption that $m \le \mathsf{poly}(n)$. □

Corollary 8. *Let* $\epsilon, \delta \in (0, 1)$ *be arbitrary constants with* $\delta < \epsilon$ *and let* $n = 2^{\ell^\delta}$. *Assuming*

$$\mathsf{LWE}\left(2^{\ell^\delta}, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma}\right)$$

has no $2^{O(\ell^\epsilon)}$ *time distinguisher from* $U\left(\mathbb{Z}_q^{\ell\times 2^{\ell^\delta}} \times \mathbb{Z}_q^{2^{\ell^\delta}}\right)$ *with advantage at least* $\Omega\left(m\left(2^{\ell^\delta}\right)^{-3}\right)$, *where* $\sigma = \ell^{1/2}$ *and* $q = \ell^2$, *then there is no algorithm solving density estimation for* g *Gaussians in* $\mathbb{R}^n$ *with* $m = m(n)$ *samples in time* $2^{\log_2(n)^{\epsilon/\delta}}$, *where* $\ell \le m(n) \le \mathsf{poly}(n)$ *and* $g = O\left((\log n)^{1/(2\delta)} \cdot \sqrt{\log(m(n))} \cdot \sqrt{\log\log n}\right)$.

In particular, for the number of GMM samples $m(n)$ *satisfying* $\ell \le m(n) \le \mathsf{poly}(\log(n)) = \mathsf{poly}(\ell)$, *we have* $g = O((\log n)^{1/(2\delta)} \cdot \log\log n)$ *assuming there is no* $1/\mathsf{poly}(\ell)$*-advantage distinguisher for LWE, and for* $m(n)$ *satisfying* $\ell \le m(n) = \mathsf{poly}(n) = 2^{O(\ell^\delta)}$, *we have* $g = O((\log n)^{1/2+1/(2\delta)} \cdot \sqrt{\log\log n})$ *assuming there is no* $1/2^{O(\ell^\delta)}$*-advantage distinguisher for LWE.*

Proof. We set $n = 2^{\ell^\delta}$ and $k = 4\ell^{1-\delta}\log_2(\ell)$ in Corollary 7. Let us first confirm that all the hypotheses of Corollary 7 hold. First, observe that

$$10\sqrt{\ln n + \ln m} = O(\ell^{\delta/2}) = o(\ell^{1/2}) \le \sigma.$$

We also have $q = \ell^2 = \omega(\ell^{1/2}\cdot\ell) \ge \omega(\sigma\cdot\sqrt{k})$, as needed. Further, we have

$$k\log_2(n/k) = (4-o(1))\ell^{1-\delta}\log_2(\ell)\ell^\delta = (4-o(1))\ell\log_2(\ell) = (2-o(1))\ell\log_2(q),$$

and lastly, $q = \ell^2 \le m^2$, as needed. If we have a

$$2^{\ell^\epsilon} = 2^{\log_2(n)^{\epsilon/\delta}}$$

time distinguisher for the mixture of Gaussians, we get a $2^{\ell^\epsilon} + \mathsf{poly}(n) = 2^{O(\ell^\epsilon)}$ time algorithm for LWE. The number of Gaussians becomes

$$\begin{aligned} g = O\left(\sqrt{k\cdot\log m\cdot\log n}\right) &= O\left(\sqrt{\ell^{1-\delta}\cdot\log(\ell)\cdot\log(n)\cdot\log(m)}\right) \\ &= O\left((\log_2 n)^{\frac{1-\delta}{2\delta}}\cdot\sqrt{\log\log n}\cdot\sqrt{\log n}\cdot\sqrt{\log m}\right) \\ &= O\left((\log n)^{\frac{1}{2\delta}}\cdot\sqrt{\log m}\cdot\sqrt{\log\log n}\right), \end{aligned}$$

as desired. □

We give another setting of parameters where the number of Gaussian components in the mixture is larger, but assumption on LWE is weaker.

Corollary 9. *Let $\alpha > 1$ be an arbitrary constant. Assuming $\mathsf{LWE}(n, \mathbb{Z}_q^\ell, \mathbb{Z}_q^\ell, D_{\mathbb{Z},\sigma})$ and $U(\mathbb{Z}_q^{\ell\times n}\times\mathbb{Z}_q^n)$ have no $T(\ell)+\mathsf{poly}(n)$ time distinguisher with advantage $\Omega(1/m^3)$ where $n = \ell^\alpha$, $\sigma = \ell^{1/2}$ and $q = \ell^2$, then there is no algorithm solving density estimation for mixtures of g Gaussians with m samples in time $T(\ell) = T(n^{1/\alpha})$, where $g = O\left(n^{1/(2\alpha)}\cdot\log n\right)$ and $\ell \le m \le \mathsf{poly}(n) = \mathsf{poly}(\ell)$.*

In particular, if $T(\ell) = \mathsf{poly}(\ell)$, then assuming the LWE problem is hard to distinguish for $\mathsf{poly}(\ell)$-time algorithms with advantage $1/\mathsf{poly}(\ell)$, then density estimation cannot be solved in $\mathsf{poly}(n)$ time with $\mathsf{poly}(n) \ge \ell$ samples for $g = n^{\Omega(1)}$ Gaussians.

Proof. We set $k = 4\ell/(\alpha-1) = 4n^{1/\alpha}/(\alpha-1)$ and apply Corollary 7. Observe that

$$\begin{aligned} k\log_2(n/k) = \frac{4\ell}{\alpha-1}\cdot\log_2\left(\frac{\ell^\alpha}{4\ell/\alpha}\right) &= \frac{4\ell}{\alpha-1}\cdot((\alpha-1)\log_2(\ell) - O(1)) \\ &= 4\ell\log_2(\ell) - O(\ell) \\ &= 2\ell\log_2(q) - O(\ell) \\ &= (1+\Theta(1))\ell\log_2(q), \end{aligned}$$

as necessary. Let us see that the other hypotheses of Corollary 7 hold. We have

$$10\sqrt{\ln n + \ln m} = O\left(\sqrt{\log\ell}\right) = o(\sigma).$$

Also observe that $q = \ell^2 \ge \omega(\ell^{1/2}\cdot\ell) \ge \omega(\sigma\cdot\sqrt{k})$ and $q = \ell^2 \le m^2$.

If we have a time $T(n^{1/\alpha}) = T(\ell)$ distinguisher for hCLWE, we get a time $T(\ell)+\mathsf{poly}(n)$ time distinguisher for LWE. The number of Gaussian components becomes

$$g = O\left(\sqrt{k\cdot\log(m)\cdot\log(n)}\right) = O\left(n^{1/(2\alpha)}\cdot\log(n)\right).$$

□

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 Lemma 9.

References

[ACPS09] Benny Applebaum, David Cash, Chris Peikert, and Amit Sahai. Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In *Annual International Cryptology Conference*, pages 595–618. Springer, 2009. 35, 36, 40

[AM05] Dimitris Achlioptas and Frank McSherry. On spectral learning of mixtures of distributions. In *International Conference on Computational Learning Theory*, pages 458–469. Springer, 2005. 4

[Ban93] Wojciech Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(1):625–635, 1993. 36

[BD20] Zvika Brakerski and Nico Döttling. Hardness of LWE on general entropic distributions. In Anne Canteaut and Yuval Ishai, editors, *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*, volume 12106 of *Lecture Notes in Computer Science*, pages 551–575. Springer, 2020. 2, 8, 9

[BLMR13] Dan Boneh, Kevin Lewi, Hart William Montgomery, and Ananth Raghunathan. Key homomorphic prfs and their applications. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I*, volume 8042 of *Lecture Notes in Computer Science*, pages 410–428. Springer, 2013. 8

[BLP+13] Zvika Brakerski, Adeline Langlois, Chris Peikert, Oded Regev, and Damien Stehlé. Classical hardness of learning with errors. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 575–584, 2013. 1, 2, 3, 8, 9, 22, 27, 28, 35, 36, 40

[BNHR22] Andrej Bogdanov, Miguel Cueto Noval, Charlotte Hoffmann, and Alon Rosen. Public-key encryption from continuous LWE. *IACR Cryptol. ePrint Arch.*, page 93, 2022. 1, 7

[BRST21] Joan Bruna, Oded Regev, Min Jae Song, and Yi Tang. Continuous LWE. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 694–707, 2021. 1, 2, 3, 4, 6, 7, 8, 11, 12, 13, 24, 27, 28, 37

[BS15] Mikhail Belkin and Kaushik Sinha. Polynomial learning of distribution families. *SIAM Journal on Computing*, 44(4):889–911, 2015. 4

[BV08] S Charles Brubaker and Santosh S Vempala. Isotropic pca and affine-invariant clustering. In *Building Bridges*, pages 241–281. Springer, 2008. 4

[BV14] Zvika Brakerski and Vinod Vaikuntanathan. Efficient fully homomorphic encryption from (standard) LWE. *SIAM J. Comput.*, 43(2):831–871, 2014. 1

[CJJ21] Arka Rai Choudhuri, Abhihsek Jain, and Zhengzhong Jin. Snargs for $\mathcal{P}$ from LWE. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 68–79. IEEE, 2021. 1

[Das99] Sanjoy Dasgupta. Learning mixtures of gaussians. In *40th Annual Symposium on Foundations of Computer Science, FOCS '99, 17-18 October, 1999, New York, NY, USA*, pages 634–644. IEEE Computer Society, 1999. 4

[DKMR22] Ilias Diakonikolas, Daniel M Kane, Pasin Manurangsi, and Lisheng Ren. Cryptographic hardness of learning halfspaces with massart noise. *arXiv preprint arXiv:2207.14266*, 2022. 6

[DKS17] Ilias Diakonikolas, Daniel M Kane, and Alistair 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)*, pages 73–84. IEEE, 2017. 4, 5, 7

[DKS18] Ilias Diakonikolas, Daniel M Kane, and Alistair 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*, pages 1047–1060, 2018. 4

[DS07] Sanjoy Dasgupta and Leonard J Schulman. A probabilistic analysis of em for mixtures of separated, spherical gaussians. *Journal of Machine Learning Research*, 8:203–226, 2007. 4

[FGR+17] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2):1–37, 2017. 4

[FSO06] Jon Feldman, Rocco A Servedio, and Ryan O'Donnell. Pac learning axis-aligned mixtures of gaussians with no separation assumption. In *International Conference on Computational Learning Theory*, pages 20–34. Springer, 2006. 4

[GKPV10] Shafi Goldwasser, Yael Tauman Kalai, Chris Peikert, and Vinod Vaikuntanathan. Robustness of the learning with errors assumption. 2010. 1, 2

[GKVZ22] Shafi Goldwasser, Michael P Kim, Vinod Vaikuntanathan, and Or Zamir. Planting undetectable backdoors in machine learning models. *arXiv preprint arXiv:2204.06974*, 2022. 6, 7

[GVW15] Sergey Gorbunov, Vinod Vaikuntanathan, and Hoeteck Wee. Attribute-based encryption for circuits. *J. ACM*, 62(6):45:1–45:33, 2015. 1

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999. 10

[HL18] Samuel B Hopkins and Jerry Li. Mixture models, robustness, and sum of squares proofs. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1021–1034, 2018. 4

[HP15] Moritz Hardt and Eric Price. Tight bounds for learning a mixture of two gaussians. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 753–760, 2015. 4

[JKKZ21] Ruta Jawale, Yael Tauman Kalai, Dakshita Khurana, and Rachel Yun Zhang. Snargs for bounded depth computations and PPAD hardness from sub-exponential LWE. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 708–721. ACM, 2021. 1

[Kea98] Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998. 4

[KMV10] Adam Tauman Kalai, Ankur Moitra, and Gregory Valiant. Efficiently learning mixtures of two gaussians. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 553–562. ACM, 2010. 4

[KS09] Adam R. Klivans and Alexander A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. *J. Comput. Syst. Sci.*, 75(1):2–12, 2009. 1

[KSS18] Pravesh K Kothari, Jacob Steinhardt, and David Steurer. Robust moment estimation and improved clustering via sum of squares. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1035–1046, 2018. 4

[KSV05] Ravindran Kannan, Hadi Salmasian, and Santosh Vempala. The spectral method for general mixture models. In *International Conference on Computational Learning Theory*, pages 444–457. Springer, 2005. 4

[LP11] Richard Lindner and Chris Peikert. Better key sizes (and attacks) for lwe-based encryption. In *Cryptographers' Track at the RSA Conference*, pages 319–339. Springer, 2011. 3

[Mic18] Daniele Micciancio. On the hardness of learning with errors with binary secrets. *Theory Comput.*, 14(1):1–17, 2018. 2, 8, 9, 11, 13, 17, 18, 25, 27

[MM11] Daniele Micciancio and Petros Mol. Pseudorandom knapsacks and the sample complexity of LWE search-to-decision reductions. In Phillip Rogaway, editor, *Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011. Proceedings*, volume 6841 of *Lecture Notes in Computer Science*, pages 465–484. Springer, 2011. 6, 40

[MP00] G. J. McLachlan and D. Peel. *Finite mixture models*. Wiley Series in Probability and Statistics, 2000. 3

[MP12] Daniele Micciancio and Chris Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 700–718. Springer, 2012. 40

[MP13] Daniele Micciancio and Chris Peikert. Hardness of sis and lwe with small parameters. In *Annual Cryptology Conference*, pages 21–39. Springer, 2013. 11

[MR07] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM Journal on Computing*, 37(1):267–302, 2007. 7, 11, 36

[MR09] Daniele Micciancio and Oded Regev. Lattice-based cryptography. In *Post-quantum cryptography*, pages 147–191. Springer, 2009. 35

[MV10] Ankur Moitra and Gregory Valiant. Settling the polynomial learnability of mixtures of gaussians. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 93–102. IEEE, 2010. 4, 7

[NIS] NIST. Post-quantum cryptography standardization. `https://csrc.nist.gov/Projects/Post-Quantum-Cryptography`. 3, 5

[Pei09] Chris Peikert. Public-key cryptosystems from the worst-case shortest vector problem: extended abstract. In Michael Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 333–342. ACM, 2009. 1, 40

[Pei10] Chris Peikert. An efficient and parallel gaussian sampler for lattices. In *Annual Cryptology Conference*, pages 80–97. Springer, 2010. 40

[PRS17] Chris Peikert, Oded Regev, and Noah Stephens-Davidowitz. Pseudorandomness of ring-lwe for any ring and modulus. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 461–473. ACM, 2017. 1

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM (JACM)*, 56(6):1–40, 2009. 1, 2, 4, 6, 7, 10

[RV17] Oded Regev and Aravindan Vijayaraghavan. On learning mixtures of well-separated gaussians. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 85–96. IEEE, 2017. 4

[SK01] Arora Sanjeev and Ravi Kannan. Learning mixtures of arbitrary gaussians. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 247–257, 2001. 4

[SZB21] Min Jae Song, Ilias Zadik, and Joan Bruna. On the cryptographic hardness of learning single periodic neurons. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 29602–29615. Curran Associates, Inc., 2021. 6

[Tie22] Stefan Tiegel. Hardness of agnostically learning halfspaces from worst-case lattice problems. *arXiv preprint arXiv:2207.14030*, 2022. 6

[TTM+85] D.M. Titterington, P.S.D.M. Titterington, S.A.F. M, A.F.M. Smith, U.E. Makov, and John Wiley & Sons. *Statistical Analysis of Finite Mixture Distributions*. Applied section. Wiley, 1985. 3

[VW02] Santosh Vempala and Grant Wang. A spectral algorithm for learning mixtures of distributions. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 113–122. IEEE, 2002. 4

A Alternate Reduction from LWE to CLWE

In this section, we propose an alternate reduction from LWE to CLWE than that of Corollary 5. We note that we reduce to *search* CLWE, and not decisional CLWE. Here is a brief outline to the steps of this reduction:

1. First, we start with the standard search version of LWE (dimension n, mod q, and noise $D_{\mathbb{Z},\sigma}$).
2. Then, we reduce to the (search) "Hermite normal form" of LWE, where the secret is drawn from the *error distribution* instead of uniform over $\mathbb{Z}_q^n$ (with a small additive blowup in the number of samples), following [ACPS09, MR09] (and the more refined analysis by [BLP+13]).
3. Since the secrets are now short, we know there is some (small) $r \approx \sigma\sqrt{n}$ for which non-negligibly often, secrets will have ℓ_2 norm *exactly* r. The reduction in this step is the trivial reduction, but crucially uses the fact that this is a *search* reduction.
4. Since the secrets now have fixed (and small) norm, we use Theorem 6 to reduce to CLWE, slightly modified to be a search reduction (as opposed to a decision reduction).

Explicitly, we have the following theorem.

Theorem 9 (Alternate Reduction from LWE to CLWE)**.** *Suppose there exists no algorithm running in time $T + \mathsf{poly}(n, m, \log(\lambda), \log(q))$ that outputs $\mathbf{s}$ with probability ϵ when given $(A, \mathbf{s}^\top A + \mathbf{e} \pmod q)$, where $A \sim U(\mathbb{Z}_q^{n\times m})$, $\mathbf{s} \sim U(\mathbb{Z}_q^n)$, and $\mathbf{e} \sim D_{\mathbb{Z}^m,\sigma}$. Suppose $q \le 2^{2^{O(n)}}$ and $\sigma \ge 2\sqrt{\ln(n)}$. Then, there is no T-time algorithm outputting $\mathbf{s}'$ with probability at least $(\epsilon + 2^{-n}) \cdot 2\sigma^2 n + \mathsf{negl}(\lambda)$ when given $(A', \gamma \cdot (\mathbf{s}')^\top A' + \mathbf{e} \pmod 1)$, where $A' \sim (D_1)^{n\times m'}$, $\mathbf{s}' \sim \frac{1}{r}\mathcal{S}$, $\mathbf{e} \sim D_\beta^{m'}$, where $\mathcal{S}$ is the set of all vectors in $\mathbb{Z}^n$ with norm exactly r, for some $r = O(\sigma\sqrt{n})$, and where*

$$\begin{aligned} m' &= m + O(n), \\ \gamma &= O\left(\sigma\sqrt{n} \cdot \sqrt{\ln(m) + \ln(n) + \omega(\log\lambda)}\right), \\ \beta &= O\left(\frac{\sigma\sqrt{n} \cdot \sqrt{\ln(m) + \ln(n) + \omega(\log\lambda)}}{q}\right). \end{aligned}$$

Note that one can reduce to secret distribution $\mathbf{s}' \sim U(S^{n-1})$ if desired, by using (a search version of) Lemma 19.

Proof. First, we invoke Lemma 2 of [ACPS09] to turn the secret distribution from $\mathbb{Z}_q^n$ into $D_{\mathbb{Z}^n,\sigma}$. The only loss in parameters we get is the number of samples, which becomes $m' = m + O(n) + O(\log\log q)$, as $O(n) + O(\log\log q)$ samples suffice to efficiently find n linearly independent vectors over $\mathbb{Z}_q$, with probability of failure at most 2^{-n}. (See Claim 2.13 of the full version of [BLP+13].) Since $q \leq 2^{2^{O(n)}}$, $O(\log\log q)$ can be absorbed into $O(n)$, making $m' = m + O(n)$. This means there is no solver for $\mathbf{s}$ with probability at least $\epsilon + 2^{-n}$ for this "normal form" of LWE (i.e. with $\mathbf{s} \sim D_{\mathbb{Z}^n,\sigma}$).

Now, we observe that there exists some $r \leq \sigma\sqrt{n}$ such that non-negligibly often, secrets $\mathbf{s} \sim D_{\mathbb{Z}^n,\sigma}$ will have ℓ_2 norm exactly r. To see this, we use the proof of Lemma 4.4 in [MR07] (ultimately based on Lemma 1.5 of [Ban93]) to see that the probability that $\|\mathbf{s}\| \geq \sigma\sqrt{n}$ is at most 2^{-n} for $\mathbf{s} \sim D_{\mathbb{Z}^n,\sigma}$. Conditioned on $\|\mathbf{s}\| \leq \sigma\sqrt{n}$, since $\|\mathbf{s}\|^2$ is a non-negative integer, we know it must take on at most $\sigma^2 \cdot n + 1$ different values. Let $\mathcal{S}_r$ be the set of all vectors in $\mathbb{Z}^n$ with ℓ_2 norm exactly $r \in \mathbb{R}$. What we have just shown is that there exists some $r \leq \sigma\sqrt{n}$ such that $\Pr_{\mathbf{s}\sim D_{\mathbb{Z}^n,\sigma}}[\mathbf{s} \in \mathcal{S}_r] \geq (1 - 2^{-n})/(\sigma^2 \cdot n + 1) \geq 1/(2\sigma^2 n)$. From here on out, we now fix r to be such an r. Moreover, it is easy to see that $r \geq \sigma$, as the only way for the norm to be below σ is if all coordinates of the discrete Gaussian have magnitude at most σ, which happens with exponentially small probability in n.

Therefore, if we have some solver with success probability $(\epsilon + 2^{-n}) \cdot 2\sigma^2 n$ for LWE with $\mathbf{s} \sim \mathcal{S}$, then that same solver has success probability at least $\epsilon + 2^{-n}$ for LWE with $\mathbf{s} \sim D_{\mathbb{Z}^n,\sigma}$. Therefore, we now know there is no solver for LWE with secret distribution $\mathcal{S}$ with success probability $(\epsilon + 2^{-n}) \cdot 2\sigma^2 n$.

Before invoking Theorem 6, we simply increase the width of the noise, as the requirement on the width of the noise is large for the reduction to go through. Specifically, we set $\sigma' = 3r\sqrt{\ln(m') + \ln(n) + \omega(\log\lambda)} \geq \sigma$, where the inequality comes from the fact that $r \geq \sigma$. We can achieve this reduction by simply adding noise.

Now, we directly invoke Theorem 6, as our requirement on σ' is now satisfied. While the reduction is formally a decisional reduction, the proof also works in the search setting. In fact, throughout the reduction, the secret remains the same, up to scaling by r. This implies there is no solver with success probability at least $(\epsilon + 2^{-n}) \cdot 2\sigma^2 n + \mathsf{negl}(\lambda)$, for parameters

$$\gamma = r \cdot \sqrt{\ln(m') + \ln(n) + \omega(\log\lambda)} \leq O\left(\sigma\sqrt{n} \cdot \sqrt{\ln(m) + \ln(n) + \omega(\log\lambda)}\right),$$
$$\beta = O\left(\frac{\sqrt{(\sigma')^2 + r^2(\ln(m') + \ln(n) + \omega(\log\lambda)}}{q}\right) = O\left(\frac{\sigma\sqrt{n} \cdot \sqrt{\ln(m) + \ln(n) + \omega(\log\lambda)}}{q}\right).$$

□

B Low-Sample Algorithm for Sparse hCLWE

Theorem 10. *Let* $m = 5k\log_2(n)/\log_2(1/(\beta\sqrt{k}))$. *Suppose* $\gamma \geq 2\sqrt{k(\ln n + \ln m)}$ *and* $\log_2(1/(\beta\sqrt{k})) = \omega(\log\log m)$. *Then, there is a* $O\left(m \cdot \mathsf{poly}(n) \cdot 2^k \binom{n}{k}\right)$*-time algorithm using* m *samples that learns the parameters for GMM when restricted to (*m *sample) mixtures* $D_1^{n\times m}$ *and* $\mathsf{hCLWE}^{(g)}(m, D_1^n, \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}, \gamma, \beta)$ *for any fixed* $g \geq C \cdot \gamma \cdot \sqrt{\log m}$ *for some universal constant* C. *That is, we view all the parameters as fixed, and the algorithm either learns the correct secret* $\mathbf{s} \sim \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}$

(and thus the corresponding $\mathsf{hCLWE}^{(g)}$ *distribution), or knows that the distribution is* D_1^n*, with success probability at least* $9/10$ *in both cases.*

Remark 4. *This theorem can be generalized for other settings of* β, γ*, but we state it this way because it suffices for our purposes. It also works for the setting of non-truncated hCLWE.*

Remark 5. *While the runtime of this algorithm is similar to the algorithm solving hCLWE given in Theorem 7.5 of [BRST21] as applied in a black-box way, the sample complexity needed here is* $\ll k \log_2(n)$*, as opposed to roughly* $2^{O(\gamma^2)} = n^{\Omega(k)}$*.*

Algorithm 1: Low Sample algorithm for $\mathsf{hCLWE}^{(g)}$

Input: Sampling oracle to distribution $\mathcal{D}$.
Output: $\mathbf{s}$ to indicate $\mathcal{D} = \mathsf{hCLWE}^{(g)}$ with secret $\mathbf{s}$, and 0 for $\mathcal{D} = D_1^n$.

Draw m samples $\mathbf{a}_1, \ldots, \mathbf{a}_m \sim \mathcal{D}$.
for $\mathbf{s} \in \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}$ **do**
  Compute $f_{\mathbf{s}}(\mathbf{a}_i) = \langle \mathbf{a}_i, \mathbf{s} \rangle \mod \gamma/(\sqrt{k} \cdot \gamma'^2)$ for all $i \in [m]$.
  if $f_{\mathbf{s}}(\mathbf{a}_i) \in [-a\beta/\gamma', a\beta/\gamma']$ *for all* $i \in [m]$ **then**
    return $\mathbf{s}$.
return 0.

Proof. Let $t := |\mathcal{S}_{n,k}| = \binom{n}{k} \cdot 2^k$ denote the number of k-sparse $\{-1, 0, +1\}$-secrets. For the sake of this proof, we take the representatives of $\mathbb{T}_q$ to be in the interval $[-q/2, q/2)$. Further, let $\gamma' = \sqrt{\gamma^2 + \beta^2}$ and $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{s} \in \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}$. We define $f_{\mathbf{s}} : \mathbb{R}^n \to \mathbb{T}_{\gamma/(\sqrt{k}\cdot\gamma'^2)}$ by

$$f_{\mathbf{s}}(\mathbf{a}) := \langle \mathbf{a}, \mathbf{s} \rangle \mod \gamma/(\sqrt{k} \cdot \gamma'^2).$$

We use the main idea in the proof of Claim 5.3 in [BRST21] to give an algorithm that finds the correct secret $\mathbf{s}$, if it exists, or report that none exists if the distribution is D_1^n. Given m samples $\mathbf{a}_1, \ldots, \mathbf{a}_m$ from an unknown distribution $\mathcal{D}$, we compute $f_{\mathbf{s}}(\mathbf{a}_i)$ for all possible secret directions $\mathbf{s} \in \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}$ and for all samples $i \in [m]$. This takes time $O(m \cdot t \cdot \mathsf{poly}(n))$, where we allow $\mathsf{poly}(n)$ time to take numbers mod $\gamma/(\sqrt{k}\gamma'^2)$. If there is some $\mathbf{s}$ such that $f_{\mathbf{s}}(\mathbf{a}_i)$ is small for all samples $i \in [m]$, then we output $\mathbf{s}$, and otherwise we guess $\mathcal{D} = D_1^m$.

Now, suppose that the input distribution is $\mathcal{D} = \mathsf{hCLWE}^{(g)}(m, D_1^n, \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}, \gamma, \beta)$. Let $\mathbf{s}^*$ be the randomly sampled but fixed secret direction. Then for all the m samples $\mathbf{a}_i$, we have that $\langle \mathbf{s}^*, \mathbf{a}_i \rangle \pmod{\gamma/(\gamma'^2)}$ is distributed as $D_{\beta/\gamma'} \mod \gamma/\gamma'^2$. This can be seen from Equation 2. (As an aside, note that by Claim 5.3 of [BRST21] this holds even when the input distribution is not truncated, that is, $\mathcal{D} = \mathsf{hCLWE}(m, D_1^n, \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}, \gamma, \beta)$.) Now, supposing for simplicity that k is a perfect square, we can take this mod $\gamma/(\sqrt{k} \cdot \gamma'^2)$ to get that $f_{\mathbf{s}^*}(\mathbf{a}_i)$ is distributed as $D_{\beta/\gamma'} \mod \gamma/(\sqrt{k} \cdot \gamma'^2)$. (In case k is not a perfect square, we can take $\gamma/\left(\left\lceil\sqrt{k}\right\rceil \cdot \gamma'^2\right)$ as the modulus instead.)

For a parameter $\delta > 0$ specified later, let $a = \sqrt{\ln(1/\delta)}$. By a standard Chernoff bound, the probability mass of $D_{\beta/\gamma'}$ that is outside the interval $[-a\beta/\gamma', a\beta/\gamma']$ is at most δ. Taking a union

bound over the m samples $\mathbf{a}_i$,

$$\Pr\left[\exists i \in [m] \text{ s.t. } f_{\mathbf{s}^*}(\mathbf{a}_i) \notin [-a\beta/\gamma', a\beta/\gamma']\right] \leq m\delta = \frac{1}{100}, \tag{3}$$

when setting $\delta = 1/(100m)$.

We still have to argue that for this $\mathcal{D}$ that no other $\mathbf{s} \neq \mathbf{s}^*$ passes the test. To see this, fix some $\mathbf{s} \neq \mathbf{s}^*$. Let $\mathbf{z} = \mathbf{s} - \mathbf{s}^* \in \frac{1}{\sqrt{k}} \cdot \{-2, -1, 0, 1, 2\}^n$. Since $\mathbf{s} \in \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}$ and $\mathbf{s} \neq \mathbf{s}^*$, it follows that $\|\mathbf{z}\| \geq \sqrt{2}/\sqrt{k}$. We have

$$\langle \mathbf{a}_i, \mathbf{s} \rangle = \langle \mathbf{a}_i, \mathbf{s}^* + \mathbf{z} \rangle = \langle \mathbf{a}_i, \mathbf{s}^* \rangle + \langle \mathbf{a}_i, \mathbf{z} \rangle. \tag{4}$$

On its own, for fixed $\mathbf{z}$, $\langle \mathbf{a}_i, \mathbf{z} \rangle$ is distributed according to $D_{\|\mathbf{z}\|}$, which is from a discrete Gaussian wider than $D_{1/\sqrt{k}}$. By Lemma 4 (and the union bound over m samples), it follows that the distribution of $(\langle \mathbf{a}_i, \mathbf{z} \rangle \pmod{\gamma/(\sqrt{k} \cdot \gamma'^2)})_{i \in [m]}$ is $m \cdot \exp(-\gamma'^4/\gamma^2)/2$-close to $U(\mathbb{T}_{\gamma/(\sqrt{k} \cdot \gamma'^2)})^m$. Therefore,

$$\begin{aligned}
\Pr_{\mathbf{a}_i}\left[(f_{\mathbf{z}}(\mathbf{a}_i))_{i \in [m]} \in [-2a\beta/\gamma', 2a\beta/\gamma']^m\right] &\leq \Delta\left(D_1^m \mod \gamma/(\sqrt{k}\gamma'^2), U\left(\mathbb{T}_{\gamma/(\sqrt{k}\gamma'^2)}\right)^m\right) + \left(4a\beta\sqrt{k} \cdot \frac{\gamma'}{\gamma}\right)^m \\
&\leq m \exp(-\gamma^2)/2 + \left(4a\beta\sqrt{k} \cdot \frac{\gamma'}{\gamma}\right)^m \\
&\leq m \exp(-\gamma^2)/2 + \left(8a\beta\sqrt{k}\right)^m.
\end{aligned}$$

Since this was for a particular secret $\mathbf{s} \neq \mathbf{s}^*$, we can union bound over all $\mathbf{s} \neq \mathbf{s}^*$ to see that

$$\Pr_{\mathbf{a}_i}\left[\exists \mathbf{s} \neq \mathbf{s}^* \text{ s.t. } (f_{\mathbf{z}}(\mathbf{a}_i))_{i \in [m]} \in [-2a\beta/\gamma', 2a\beta/\gamma']^m\right] \leq t \cdot m \exp(-\gamma^2)/2 + t \cdot \left(8 \cdot a\beta\sqrt{k}\right)^m. \tag{5}$$

Note that if $f_{\mathbf{z}}(\mathbf{a}_i) \notin [-2a\beta/\gamma', 2a\beta/\gamma']$ and $f_{\mathbf{s}^*}(\mathbf{a}_i) \in [-a\beta/\gamma', a\beta/\gamma']$, then by equation (4), it follows that $f_{\mathbf{s}}(\mathbf{a}_i) \notin [-a\beta/\gamma', a\beta/\gamma']$. Thus, equations (3) and (5) fully characterize the "bad" events, as if both events do not happen, then $\mathbf{s}^*$ passes the test, and no other $\mathbf{s} \neq \mathbf{s}^*$ passes the test. Therefore, if samples are from the hCLWE distribution, then the probability of failure is at most

$$\frac{1}{100} + t \cdot m \exp(-\gamma^2)/2 + t \cdot \left(8a\beta\sqrt{k}\right)^m.$$

We first analyze the middle term. Since $\gamma \geq 2\sqrt{k(\ln n + \ln m)}$, we have

$$t \cdot m \cdot \exp(-\gamma^2)/2 \leq \frac{2^k \cdot n^k \cdot m}{2 \cdot m^{4k} \cdot n^{4k}} < \frac{1}{100}$$

for large enough m and k. For the last term, we have

$$\begin{aligned}
t \cdot \left(8a\beta\sqrt{k}\right)^m = t \cdot \left(8 \cdot \ln(100m) \cdot \beta\sqrt{k}\right)^m &= t \cdot \left(2^{3+\log_2(\ln(100m))-\log_2(1/(\beta\sqrt{k}))}\right)^m \\
&\leq t \cdot 2^{-(m/2)\cdot\log_2(1/(\beta\sqrt{k}))} \\
&\leq 2^{k+k\log_2(n)-(m/2)\cdot\log_2(1/(\beta\sqrt{k}))} \\
&\leq 2^{-k\log_2(n)} \\
&\leq \frac{1}{100}
\end{aligned}$$

for sufficiently large k, n, where we have used our hypothesis on $\log_2(1/(\beta\sqrt{k}))$ and choice of m. Thus, for the hCLWE distribution, we output the correct secret $\mathbf{s}^*$ with probability at least $19/20$.

Now, suppose we are given samples from D_1^n. For any fixed $\mathbf{s} \in \frac{1}{\sqrt{k}}\mathcal{S}_{n,k}$, we have $\langle \mathbf{a}_i, \mathbf{s} \rangle \sim D_1$, independently of $\mathbf{s}$. By Lemma 4 and Lemma 2,

$$\Delta(D_1^m \mod \gamma/(\sqrt{k}\cdot\gamma'^2), \mathbb{T}^m_{\gamma/(\sqrt{k}\cdot\gamma'^2)}) \le m\exp(-\gamma'^4 k/\gamma^2)/2 \le m\exp(-\gamma^2)/2.$$

Therefore, by a simpler analysis than the one above, we have

$$\Pr_{\mathbf{a}_i}\left[\exists \mathbf{s} \in \mathcal{S}_{n,k} \text{ s.t. } (f_{\mathbf{s}}(\mathbf{a}_i))_{i\in[m]} \in [-a\beta/\gamma', a\beta/\gamma']^m\right] \le t\cdot m\exp(-\gamma^2)/2 + t\cdot\left(8\cdot a\beta\sqrt{k}\right)^m, \quad (6)$$

which we have previously bounded above by $2/100$. This completes the proof, as we will output 0 (to indicate $\mathcal{D} = D_1^n$) with probability at least $19/20$ in this case. □

Now, we combine Theorem 10 and Corollary 8 to get the following tightness for the mixtures of Gaussians we consider.

Corollary 10. *Following the notation of Corollary 8, there is an algorithm solving for the parameters for GMM, when restricted to D_1^n and hCLWE, using $m = O(\ell) = O\left((\log n)^{1/\delta}\right)$ samples and time $2^{O\left((\log n)^{1/\delta}\log\log n\right)}$.*

Proof. We apply Theorem 10. If we trace β in the proof of Corollary 8, we see that

$$\beta\sqrt{k} = O\left(\frac{\sigma k}{q}\right) = O\left(\frac{\sqrt{\ell}\cdot\ell}{\ell^2}\right) = O\left(\ell^{-1/2}\right).$$

Therefore, $\log(1/(\beta\sqrt{k})) = \Omega(\log\ell)$, which implies

$$m = \frac{5k\log_2(n)}{\log_2(1/\beta\sqrt{k})} = O\left(\frac{\ell^{1-\delta}\cdot\log(\ell)\cdot\ell^\delta}{\log(\ell)}\right) = O(\ell) = O((\log n)^{1/\delta}),$$

and thus that $\log(1/(\beta\sqrt{k})) = \omega(\log\log m)$. For the runtime, observe that

$$\begin{aligned} m\cdot\mathsf{poly}(n)\cdot 2^k\binom{n}{k} \le m\cdot n^{O(k)} \le m\cdot 2^{O(\log(n)\ell^{1-\delta}\log(\ell))} &\le \mathsf{poly}(\log n)\cdot 2^{O(\log(n)^{1/\delta}\log\log n)} \\ &= 2^{O(\log(n)^{1/\delta}\log\log n)}, \end{aligned}$$

as desired. □

C Reduction from CLWE to LWE

Here, we show a reversed version of Theorem 6, i.e. a reduction from discrete-secret CLWE to fixed-norm LWE. Note that this gives a reduction only from discrete-secret CLWE to LWE, and not CLWE with secrets $\mathbf{s} \sim U(S^{n-1})$ to LWE.

Theorem 11 (CLWE to LWE)**.** *Let $r \in \mathbb{R}_{\geq 1}$, and let $\mathcal{S}$ be an arbitrary distribution over $\mathbb{Z}^n$ where all elements in the support of $\mathcal{S}$ have ℓ_2 norm r. Then, for*

$$\gamma = r \cdot \sqrt{\ln(m) + \ln(n) + \omega(\log \lambda)}, \text{ and}$$
$$\sigma = O(\beta \cdot q),$$

if there is no $T + \mathsf{poly}(n, m, \log(q), \log(\lambda))$ time distinguisher between $\mathsf{CLWE}(m, D_1^n, \frac{1}{r} \cdot \mathcal{S}, \gamma, \beta)$ and $D_1^{n \times m} \times U(\mathbb{T}^m)$ with advantage at least $\epsilon - \mathsf{negl}(\lambda)$, then there is no T-time distinguisher between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_\sigma)$ and $U(\mathbb{Z}_q^{n \times m} \times \mathbb{T}_q^m)$ with advantage ϵ, as long as $\beta \cdot q \geq 3r\sqrt{\ln(m) + \ln(n) + \omega(\log \lambda)}$.

Note that we reduce to a continuous-error version of LWE. Using standard techniques (see Theorem 3.1 of [Pei10]), this can be reduced to discrete Gaussian errors. Similarly, the final LWE distribution can be made to have secrets $U(\mathbb{Z}_q^n)$ by a standard random self-reduction. Lastly, the proof of Theorem 11 preserves the secret vector up to scaling, so it also is a reduction between the search versions of the problems.

With this reduction from discrete-secret CLWE to LWE, we get a search-to-decision reduction for discrete-secret CLWE. This can be obtained immediately by combining (the search version of) Theorem 11, standard search-to-decision reductions for LWE (see [Pei09, ACPS09, MM11, MP12, BLP+13]), and Theorem 5 (or Theorem 6 if the LWE search-to-decision reduction preserves the norm of the secret). We leave open the question of whether there is a more direct search-to-decision reduction for CLWE.

The steps of this proof are essentially just versions of Lemma 18 and Lemma 16 but in the reverse directions. We give these "reversed" lemmas below.

Lemma 21 (Reverse of Lemma 18)**.** *Let $n, m, q \in \mathbb{N}, \sigma, r, \gamma \in \mathbb{R}$. Let $\mathcal{S}$ be a distribution over $\mathbb{Z}^n$ where all elements in the support have fixed norm r. Suppose there is no $T + \mathsf{poly}(n, m, \log(\lambda), \log(q))$ time distinguisher between the distributions $\mathsf{CLWE}(m, D_1^n, \frac{1}{r} \cdot \mathcal{S}, \gamma, \beta)$ and $D_1^{n \times m} \times U(\mathbb{T}_1^m)$. Then, there is no T-time distinguisher between the distributions $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_\sigma)$ and $U(\mathbb{T}_q^{n \times m} \times \mathbb{T}_q^m)$, with an additive advantage loss of $\mathsf{negl}(\lambda)$, where*

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

Proof. Suppose we have one sample $(\mathbf{a}, b)$, from either $\mathsf{CLWE}(m, D_1^n, \frac{1}{r} \cdot \mathcal{S}, \gamma, \beta)$ or $D_1^{n \times m} \times U(\mathbb{T}_1^m)$. Now, consider the sample $(\mathbf{a} \cdot \tau \cdot q \pmod q, b \cdot q \pmod q)$, where $\tau = \sqrt{\ln(n) + \ln(m) + \omega(\log \lambda)}$ and $\gamma = r \cdot \tau$. Let $\mathbf{s} \sim \frac{1}{r} \cdot S$ be the CLWE secret, and let $\mathbf{a}' = \mathbf{a} \cdot \tau \cdot q$, let $e' = e \cdot q$, and let $\mathbf{s}' = r \cdot \mathbf{s} \in \mathbb{Z}^n$. If $(\mathbf{a}, b)$ is from the CLWE distribution, then we have (taking all components mod q)

$$\begin{aligned}
(\mathbf{a} \cdot \tau \cdot q, b \cdot q) &= (\mathbf{a}', (\gamma \cdot \langle \mathbf{s}, \mathbf{a} \rangle + e) \cdot q) \\
&= (\mathbf{a}', \gamma \cdot \langle \mathbf{s}, q \cdot \mathbf{a} \rangle + e') \\
&= (\mathbf{a}', \frac{\gamma}{r \cdot \tau} \cdot \langle r \cdot \mathbf{s}, \tau \cdot q \cdot \mathbf{a} \rangle + e') \\
&= (\mathbf{a}', \langle r \cdot \mathbf{s}, \tau \cdot q \cdot \mathbf{a} \rangle + e') \\
&= (\mathbf{a}', \langle \mathbf{s}', \mathbf{a}' \rangle + e') \\
&= (\mathbf{a}', \langle \mathbf{s}', \mathbf{a}' \pmod q \rangle + e').
\end{aligned}$$

Note that $\mathbf{s}' = \mathbf{s} \cdot r \sim \mathcal{S}$ and $e' = e \cdot q \sim D_{\beta \cdot q}$, so $\mathbf{s}'$ and $\mathbf{e}'$ have the right distribution. Lastly, $\mathbf{a}' = \tau \cdot q \cdot \mathbf{a} \sim D^n_{\tau \cdot q}$, so by Lemma 4, $\mathbf{a}'$ (mod q) is $\mathsf{negl}(\lambda)/m$-close to $\mathbb{T}^n_q$ as long as $q \cdot \tau \geq \eta_{\mathsf{negl}(\lambda)/m}(q \cdot \mathbb{Z}^n)$, which holds by Lemma 2 by construction of τ. Thus, taking the triangle inequality over all m samples, the resulting distribution is $\mathsf{negl}(\lambda)$-close to $\mathsf{LWE}(m, \mathbb{T}^n_q, \mathcal{S}, D_\sigma)$ where $\sigma = \beta \cdot q$.

Lastly, if $(\mathbf{a}, b)$ is from the null distribution, then clearly $b \cdot q \sim \mathbb{T}_q$, and by the same argument as above, $\mathbf{a}'$ (mod q) is $\mathsf{negl}(\lambda)/m$-close to $\mathbb{T}^n_q$, which by the triangle inequality, implies the resulting distribution is $\mathsf{negl}(\lambda)$-close to $U(\mathbb{T}^{n \times m}_q \times \mathbb{T}^m_q)$, as desired. □

Lemma 22 (Reverse of Lemma 16)**.** *Let $n, m, q \in \mathbb{N}$, $\sigma \in \mathbb{R}$. Let $\mathcal{S}$ 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 $T+\mathsf{poly}(m, n, \log(\lambda), \log(q))$-time distinguisher between the distributions $\mathsf{LWE}(m, \mathbb{T}^n_q, \mathcal{S}, D_\sigma)$ and $U(\mathbb{T}^{n \times m}_q \times \mathbb{T}^m_q)$. Then, there is no T-time distinguisher between the distributions $\mathsf{LWE}(m, \mathbb{Z}^n_q, \mathcal{S}, D_{\sigma'})$ and $U(\mathbb{Z}^{n \times m}_q \times \mathbb{T}^m_q)$ with an additive $\mathsf{negl}(\lambda)$ advantage loss, where we set

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

Proof. Suppose we are given a sample $(\mathbf{a}, b)$ from either $\mathsf{LWE}(m, \mathbb{T}^n_q, \mathcal{S}, D_\sigma)$ or $U(\mathbb{T}^{n \times m}_q \times \mathbb{T}^m_q)$. Let $\mathbf{a}' \sim D_{\mathbb{Z}^n - \mathbf{a}, \tau}$, where $\tau = \sqrt{\ln(n) + \ln(m) + \omega(\log \lambda)}$. Let $\mathbf{a}'' = \mathbf{a} + \mathbf{a}'$ (mod q), and observe that $\mathbf{a}'' = \mathbf{a} + \mathbf{a}'$ (mod q) $\in \mathbb{Z}^n_q$, as $\mathbf{a}'$ is supported on $\mathbb{Z}^n - \mathbf{a}$. Now, consider the sample $(\mathbf{a}'', b)$. Let $\mathbf{s} \sim \mathcal{S}$ be the LWE secret. If this is from the LWE distribution, we have

$$\begin{aligned}
(\mathbf{a}'', b) &= (\mathbf{a}'', \langle \mathbf{s}, \mathbf{a} \rangle + e) \\
&= (\mathbf{a}'', \langle \mathbf{s}, \mathbf{a}'' - \mathbf{a}' \rangle + e) \\
&= (\mathbf{a}'', \langle \mathbf{s}, \mathbf{a}'' \rangle - \langle \mathbf{s}, \mathbf{a}' \rangle + e) \\
&= (\mathbf{a}'', \langle \mathbf{s}, \mathbf{a}'' \rangle + e'),
\end{aligned}$$

where we define $e' = e - \langle \mathbf{s}, \mathbf{a}' \rangle$. First, let's analyze the distribution of e'. By applying Lemma 3, since $\mathbf{s}$ has norm r, we know that e' is $\mathsf{negl}(\lambda)/m$ close to $D_{\sigma'}$ where

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

as long as

$$\eta_{\mathsf{negl}(\lambda)/m}(\mathbb{Z}^n) \leq \frac{1}{\sqrt{1/\tau^2 + (r/\sigma)^2}},$$

which holds if $\tau, \sigma/r \geq \sqrt{2} \cdot \eta_{\mathsf{negl}(\lambda)/m}(\mathbb{Z}^n)$, which it does by Lemma 2 and construction of τ and our condition on σ. Therefore, by the triangle inequality over all m samples, the errors look $\mathsf{negl}(\lambda)$-close to $D_{\sigma'}$.

Now, we consider the distribution of $\mathbf{a}'' = \mathbf{a} + \mathbf{a}'$ (mod q) $\in \mathbb{Z}^n_q$. Note that the lattice $\mathbb{Z}^n + \mathbf{a}$ depends only on $\mathbf{a}$ (mod 1), so given $\mathbb{Z}^n + \mathbf{a}$, the conditional distribution on $\mathbf{a}$ is $U(\mathbb{Z}^n_q + (\mathbf{a} \pmod 1))$. Therefore, by a one-time pad argument, the distribution of $\mathbf{a} + \mathbf{a}'$ is exactly $U(\mathbb{Z}^n_q)$, even conditioned on $\mathbf{a}'$. Therefore, $(\mathbf{a}'', \langle \mathbf{s}, \mathbf{a}'' \rangle + e')$ looks $\mathsf{negl}(\lambda)/m$-close to a sample from $\mathsf{LWE}(m, \mathbb{Z}^n_q, \mathcal{S}, D_{\sigma'})$, which by the triangle inequality over m samples, makes the resulting distribution $\mathsf{negl}(\lambda)$-close to $\mathsf{LWE}(m, \mathbb{Z}^n_q, \mathcal{S}, D_{\sigma'})$.

Lastly, suppose $(\mathbf{a}, b)$ is from the null distribution. Then by (a simpler version of) the above argument, the distribution of $\mathbf{a}''$ is given by $\mathbf{a}'' \sim U(\mathbb{Z}^n_q)$, making $(\mathbf{a}, b) \sim U(\mathbb{Z}^n_q \times \mathbb{T}_q)$, as desired. □

Now we are ready to prove Theorem 11.

Proof of Theorem 11. Suppose there is no $T + \mathsf{poly}(n, m, \log(q), \log(\lambda))$ time distinguisher between $\mathsf{CLWE}(m, D_1^n, \frac{1}{r} \cdot \mathcal{S}, \gamma, \beta)$ and $D_1^{n \times m} \times U(\mathbb{T}^m)$ with advantage at least $\epsilon - \mathsf{negl}(\lambda)$. Then, by Lemma 21, there is no distinguisher between $\mathsf{LWE}(m, \mathbb{T}_q^n, \mathcal{S}, D_\sigma)$ and $U(\mathbb{T}_q^{n \times m} \times \mathbb{T}_q^m)$, where $\gamma = r\sqrt{\ln(n) + \ln(m) + \omega(\log \lambda)}$ and $\sigma = \beta \cdot q$. Then, by Lemma 22, there is no T-time distinguisher between $\mathsf{LWE}(m, \mathbb{Z}_q^n, \mathcal{S}, D_{\sigma'})$ and $U(\mathbb{Z}_q^{n \times m} \times \mathbb{T}_q^m)$ with advantage ϵ, where $\sigma' = O(\sigma)$, as long as $\sigma = \beta \cdot q \geq 3r\sqrt{\ln n + \ln m + \omega(\log \lambda)}$. □