Simple and Efficient Pseudorandom Generators from Gaussian Processes

Eshan Chattopadhyay
Cornell University, Ithaca, NY, USA
eshanc@cornell.edu

Anindya De
University of Pennsylvania, Philadelphia, PA, USA
anindyad@seas.upenn.edu

Rocco A. Servedio
Columbia University, New York, NY, USA
rocco@cs.columbia.edu

Abstract

We show that a very simple pseudorandom generator fools intersections of k linear threshold functions (LTFs) and arbitrary functions of k LTFs over n-dimensional Gaussian space. The two analyses of our PRG (for intersections versus arbitrary functions of LTFs) are quite different from each other and from previous analyses of PRGs for functions of halfspaces. Our analysis for arbitrary functions of LTFs establishes bounds on the Wasserstein distance between Gaussian random vectors with similar covariance matrices, and combines these bounds with a conversion from Wasserstein distance to "union-of-orthants" distance from [5]. Our analysis for intersections of LTFs uses extensions of the classical Sudakov-Fernique type inequalities, which give bounds on the difference between the expectations of the maxima of two Gaussian random vectors with similar covariance matrices.

For all values of k, our generator has seed length $O(\log n) + \mathrm{poly}(k)$ for arbitrary functions of k LTFs and $O(\log n) + \mathrm{poly}(\log k)$ for intersections of k LTFs. The best previous result, due to [14], only gave such PRGs for arbitrary functions of k LTFs when $k = O(\log \log n)$ and for intersections of k LTFs when $k = O(\frac{\log n}{\log \log n})$. Thus our PRG achieves an $O(\log n)$ seed length for values of k that are exponentially larger than previous work could achieve.

By combining our PRG over Gaussian space with an invariance principle for arbitrary functions of LTFs and with a regularity lemma, we obtain a deterministic algorithm that approximately counts satisfying assignments of arbitrary functions of k general LTFs over $\{0,1\}^n$ in time $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(k,1/\varepsilon)}$ for all values of k. This algorithm has a $\mathrm{poly}(n)$ runtime for $k = (\log n)^c$ for some absolute constant $c > 0$, while the previous best $\mathrm{poly}(n)$-time algorithms could only handle $k = O(\log \log n)$. For intersections of LTFs, by combining these tools with a recent PRG due to [28], we obtain a deterministic algorithm that can approximately count satisfying assignments of intersections of k general LTFs over $\{0,1\}^n$ in time $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(\log k,1/\varepsilon)}$. This algorithm has a $\mathrm{poly}(n)$ runtime for $k = 2^{(\log n)^c}$ for some absolute constant $c > 0$, while the previous best $\mathrm{poly}(n)$-time algorithms for intersections of k LTFs, due to [14], could only handle $k = O(\frac{\log n}{\log \log n})$.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Polynomial threshold functions, Gaussian processes, Johnson-Lindenstrauss, pseudorandom generators

Digital Object Identifier 10.4230/LIPIcs.CCC.2019.4

Related Version A full version of this paper is available at `https://eccc.weizmann.ac.il/report/2018/100/`.

Funding *Eshan Chattopadhyay*: Supported by NSF grants CCF-1412958, CCF-1849899, and the Simons foundation. Part of the work done while the author was a postdoctoral researcher at the Institute for Advanced Study, Princeton.

© Eshan Chattopadhyay, Anindya De, and
Rocco A. Servedio;
licensed under Creative Commons License CC-BY

34th Computational Complexity Conference (CCC 2019).
Editor: Amir Shpilka; Article No. 4; pp. 4:1–4:33
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Anindya De: Supported by NSF CCF 1926872 (transferred from CCF 1814706). Work done while the author was at Northwestern University supported by a start-up grant.
Rocco A. Servedio: Supported by NSF CCF 1814873, NSF CCF 1563155, and by the Simons Collaboration on Algorithms and Geometry.

Acknowledgements The authors thank Oded Regev and Li-Yang Tan for helpful discussions.

1 Introduction

Constructing explicit pseudorandom generators (PRGs) for interesting classes of Boolean-valued functions is a fundamental problem in complexity theory which has witnessed a rich line of work. An important class of functions, which have been intensively studied from this perspective, are *linear threshold functions* (henceforth referred to as LTFs), i.e. functions of the form $f(x) = \text{sign}(\sum_{i=1}^{n} w_i x_i - \theta)$ for some $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$. LTFs arise naturally in a variety of areas including machine learning, social choice theory, circuit complexity and pseudorandomness. Through a very successful line of work [8, 26, 13], explicit PRGs have been obtained which ε-fool the class of LTFs over $\{-1,1\}^n$ with seed length $O(\log n + \log^2(1/\varepsilon))$ [26], or alternately seed length $O(\log(n/\varepsilon)(\log\log(n/\varepsilon))^2)$ [13]. For LTFs over the Gaussian distribution, [23] give an ε-PRG that fools LTFs with seed length $O(\log n + \log(1/\varepsilon)\log\log(1/\varepsilon)))$.

Given these successes in designing PRGs to fool a single LTF, a natural next goal is to develop PRGs for *intersections* of k LTFs (i.e. polytopes with k facets) or, more generally, for arbitrary Boolean functions of k LTFs. PRGs for polytopes have direct applications to central problems at the intersection of derandomization and combinatorial optimization, such as deterministic approximate volume estimation for polytopes and approximate counting of feasible solutions to 0-1 integer programs. The standard way to use a PRG for such applications is to run through the list of all seeds, and hence it is desirable to have seed length as small as possible as a joint function of n and k. In particular, a seed length of the form $O(\log n) \cdot \alpha(k, 1/\varepsilon)$ leads to a running time of $n^{O(\alpha(k,1/\varepsilon))}$, which even for constant ε is super-polynomial for any super-constant k. In contrast, a seed length of the form $O(\log n) + \alpha(k, 1/\varepsilon)$ leads to a running time of $\text{poly}(n) \cdot 2^{\alpha(k,1/\varepsilon)}$, which can be a fixed polynomial in n even for various super-constant values of k (depending on the function α).

In this paper we work over Gaussian space, and we give the first PRGs for intersections and arbitrary functions of k LTFs which have seed length of the form $O(\log n) + \alpha(k, 1/\varepsilon)$ for all k. For intersections of LTFs we achieve $\alpha(k, 1/\varepsilon) = \text{poly}(\log k, 1/\varepsilon)$, and for arbitrary functions of LTFs we achieve $\alpha(k, 1/\varepsilon) = \text{poly}(k, 1/\varepsilon)$. Thus for constant ε our seed length is $O(\log n)$ for $k = 2^{(\log n)^c}$ LTFs (for intersections) and $k = (\log n)^c$ LTFs (for arbitrary functions), where $c > 0$ is an absolute constant. Previously, such an $O(\log n)$ seed length was only known for $k = O(\log(n)/\log\log n)$ (for intersections) and $k = O(\log\log n)$ (for arbitrary functions) [14]. Thus, in both cases our PRGs achieve the (optimal) $O(\log n)$ seed length for exponentially larger values of k than was previously known.

Before stating our results in detail we recall the definition of a PRG over Gaussian space (see [18, 17, 16, 19, 23, 21]):

▶ **Definition 1** (PRGs for Boolean-valued functions over Gaussian space). *Let $\mathcal{C}$ be a class of functions from $\mathbb{R}^n$ to $\{-1,1\}$. Given $\varepsilon > 0$, a function $\mathcal{G} : \{-1,1\}^s \to \mathbb{R}^n$ is an* ε-PRG for class $\mathcal{C}$ over Gaussian space *if for every function $F \in \mathcal{C}$,*

$$|\,\boldsymbol{Pr}[F(\mathcal{G}(\mathbf{U}^{(s)})) = 1] - \boldsymbol{Pr}[F(\mathbf{G}^{(n)}) = 1]| \le \varepsilon,$$

where $\mathbf{G}^{(n)}$ *denotes* $(\mathbf{G}_1, \ldots, \mathbf{G}_n)$*, a random variable distributed according to the standard Gaussian* $\mathcal{N}(0,1)^n$*, and* $\mathbf{U}^{(s)}$ *denotes the uniform distribution on* $\{-1,1\}^s$*. The parameter* s *is called the* seed length *of* $\mathcal{G}$*.*

1.1 Our results and comparison to prior work

Our PRG results. The following are our main PRG theorems:

▶ **Theorem 2** (Fooling arbitrary functions of LTFs)**.** *There is an explicit PRG which* ε*-fools any Boolean function of* k *LTFs* $g(h_1, \ldots, h_k)$ *over* $\mathcal{N}(0,1)^n$*, for any* $\varepsilon > 0$ *and any* k*, with seed length*

$$O\left(\log n + \mathrm{poly}(k, 1/\varepsilon)\right).$$

This seed length is not far from the best possible in terms of its dependence on both n and k, as it is not difficult (see Appendix A) to establish a seed length lower bound for this class of $\max\{\lfloor \log n \rfloor, k\} = \Omega(k + \log n)$ for any $1 \leq k \leq n$.

In the special case when the combining function g is an AND, we get an exponential improvement in the seed length dependence on k:[1]

▶ **Theorem 3** (Fooling intersections of LTFs)**.** *There is an explicit PRG which* ε*-fools any intersection of* k *LTFs over* $\mathcal{N}(0,1)^n$*, for any* $\varepsilon > 0$ *and any* k*, with seed length*

$$O\left(\log n + \mathrm{poly}(\log k, 1/\varepsilon)\right).$$

Here too the seed length is not far from best possible for a broad range of parameters; we note that the above-mentioned lower bound of $\log n$ even when $k = 1$ implies a seed length lower bound of $\Omega(\log n)$, which is $\Omega(\log n + \log k)$ for any $k \leq \mathrm{poly}(n)$ (the most interesting regime for Theorem 3).

For arbitrary functions of k LTFs, Theorem 2 is the first result which gives a seed length of $O(\log n)$ for $k = (\log n)^c$, and for intersections of k LTFs Theorem 3 is the first result which gives a seed length of $O(\log n)$ for $k = 2^{(\log n)^c}$. As mentioned earlier and discussed in more detail below, an optimal seed length of $O(\log n)$ was previously only known [14] for exponentially smaller values of k in both settings. Below we briefly review prior results on explicit PRGs for these classes, starting with intersections of LTFs.

The most directly comparable prior result for intersections of k LTFs is the main result of [28], which gives a PRG for intersections of k LTFs over $\{-1,1\}^n$ with seed length $\log(n) \cdot \mathrm{poly}(\log k, 1/\varepsilon)$. (Such a PRG directly implies a PRG for intersections of k LTFs over Gaussian space with the same seed length via a standard reduction.) The [28] PRG builds on a PRG due to Harsha et al. [16] which has seed length $\log(n) \cdot \mathrm{poly}(\log k, 1/\varepsilon)$ for intersections of sufficiently regular LTFs; the [16] PRG in turn is similar to a PRG construction of Meka and Zuckerman [26] (for a single LTF) in which the basic idea is to (pseudorandomly) hash the coordinates into buckets and use ℓ-wise independence for coordinates hashed to the same bucket. The analysis of the [28] PRG combines a range of technical ingredients including an invariance principle for polytopes that [16] establish, combinatorial PRGs for depth-2 circuits, and new Littlewood-Offord type theorems for polytopes.

[1] We note that a weak form of Theorem 2, with a seed length of $O\left(\log n + \mathrm{poly}(2^k, 1/\varepsilon)\right)$, follows immediately from Theorem 3 just by setting its error parameter to be $\varepsilon/2^k$ and observing that any function of k LTFs is a union of at most 2^k many disjoint intersections of k LTFs. However, this is exponentially worse than we achieve in Theorem 2 above.

PRGs for intersection of LTFs were also studied by Gopalan, O'Donnell, Wu, and Zuckerman [14], Diakonikolas, Kane and Nelson [9], and recently by Servedio and Tan [29]. These results give PRGs with respect to the uniform distribution on the Boolean cube (in fact, the PRG in [14] fools arbitrary product distributions). For general k, the seed length of the PRG in [14] for intersection of k LTFs is $O((\log n + k\log(k/\varepsilon)) \cdot \log(k/\varepsilon))$. This linear dependence of the seed length on k is far from optimal; for example, if $k \geq n$ then their result does not yield a non-trivial PRG. For the special case when k/ε is at most $\mathrm{poly}(\log n)$, [14] achieves the better seed length of $O(\log n + k\log(k/\varepsilon))$. Thus, for $k = O(\log n/\log\log n)$, the [14] seed length is $O(\log n)$.

The work of Diakonikolas et al. [9] achieves a similar polynomial dependence on k in the seed length of their PRG (more precisely, they achieve seed length $O(\log n \cdot \mathrm{poly}(k, 1/\varepsilon))$, and their PRG works also for intersections of k degree-2 polynomial threshold functions). The work of Servedio and Tan [29] achieves seed length with polylogarithmic dependence on k, but only gives a good bound against intersections of LTFs with small integer weights. In more detail, if each of the k LTFs in the intersection has all its weights w_i being integers in $[-t, t]$, then the PRG in [29] has seed length $\mathrm{poly}(\log n, \log k, t, 1/\varepsilon)$. The parameter t for an LTF can in general be exponential in n (and in fact, for a random LTF, t is exponential in n with high probability), and hence the [29] result is of interest only for intersections of low-weight LTFs.

Turning to arbitrary functions of k LTFs, we observe that (as indicated in the earlier footnote) any PRG for intersections of k LTFs can be used to fool arbitrary functions of k LTFs by setting its accuracy parameter to $\varepsilon/2^k$. If the seed length of the PRG has an inverse polynomial dependence on the accuracy parameter (as in our result) then this simple approach does not yield a very good seed length, but [14] used essentially this approach to obtain a PRG that fools any function of k LTFs with seed length $O((k^2 + k\log(1/\varepsilon) + \log n) \cdot (k + \log(1/\varepsilon)))$. In the special case when $k \cdot 2^k/\varepsilon$ is at most $\mathrm{poly}(\log n)$, they achieve a better seed length of $O(k^2 + k\log(1/\varepsilon) + \log n)$, which is $O(\log n)$ for constant ε and $k = O(\log\log n)$.

Our results on deterministic approximate counting. By combining our new PRGs with invariance principles and a (multi-)regularity lemma, we obtain deterministic algorithms which approximately count the number of satisfying assignments to intersections or arbitrary functions of k arbitrary LTFs over $\{-1, 1\}^n$. (Note that such algorithms, unlike PRGs, are non-oblivious, i.e. they can "inspect" the particular LTFs which comprise the input to the problem.)

▶ **Theorem 4** (Deterministic approximate counting for arbitrary functions of k LTFs over $\{-1, 1\}^n$)**.** *There is a deterministic algorithm which, given as input k LTFs $h_1, \ldots, h_k$ over $\{-1, 1\}^n$, an explicit function $g : \{-1, 1\}^k \to \{-1, 1\}$ and an error parameter $\varepsilon > 0$, runs in* $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(k,1/\varepsilon)}$ *time and outputs a value $\tilde{v} \in [0, 1]$ such that $|\tilde{v} - v| \leq \varepsilon$, where v is the fraction of points in $\{-1, 1\}^n$ that satisfy $g(h_1, \ldots, h_k)$.*

For intersections of LTFs, by combining our approach with the [28] PRG we can get an exponentially better runtime dependence on k:

▶ **Theorem 5** (Deterministic approximate counting for intersections of k LTFs over $\{-1, 1\}^n$)**.** *There is a deterministic algorithm which, given as input k LTFs $h_1, \ldots, h_k$ over $\{-1, 1\}^n$ and an error parameter $\varepsilon > 0$, runs in* $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(\log k,1/\varepsilon)}$ *time and outputs a value $\tilde{v} \in [0, 1]$ such that $|\tilde{v} - v| \leq \varepsilon$, where v is the fraction of points in $\{-1, 1\}^n$ that satisfy $h_1 \wedge \cdots \wedge h_k$.*

We are not aware of prior results on deterministic approximate counting for intersections (or arbitrary functions) of k LTFs which run faster than simply enumerating over the seeds of a PRG. Thus Theorem 4 gives the first deterministic algorithm that runs in *fixed* poly(n) runtime even for k which is polylogarithmic in n; as indicated earlier, given the previous state of the art on PRGs for arbitrary functions of k LTFs for $k = \omega(\log \log n)$ prior algorithms would have a running time of at least $n^{\text{poly}(k)}$. Similarly, Theorem 5 gives the first deterministic algorithm that runs in *fixed* poly(n) runtime even for $k = 2^{(\log n)^{\Omega(1)}}$. The previous state of the art on PRGs for intersection of k LTFs for $k = \omega(\log n / \log \log n)$ would have a running time of at least $n^{\text{poly}(\log k)}$ (such a running time is obtained simply by enumerating over the seeds of the [28] PRG).

A key ingredient in the proof of Theorem 4 is an invariance principle for *arbitrary functions* of k LTFs, analogous to the main structural result of [16] which is an invariance principle for *intersections* of k LTFs. Such an invariance principle was proved in [14], and we provide an alternate proof in Appendix C (which is very different from the proofs of the invariance principles in [16, 14]). We believe this could be of independent interest. We elaborate on this, still at a conceptual level, in Section 3 and give full details in Section 7.

A straightforward approach to Theorem 5 using only the multi-regularity lemma and an invariance principle would have a running time which is exponential in k because the number of leaves in the decision tree constructed by the multi-regularity lemma is exponential in k. We achieve a quasi-polynomial dependence on k by exploiting additional structure in the decision tree (specifically, that it is a so-called "junta decision tree" in which the same variable occurs at each node of any given depth). Intuitively, this makes it possible for us to use the [28] PRG on the space of all variables occurring in the decision tree (to "pseudorandomly sample" leaves of the decision tree and use only those to construct an accurate estimate of the overall desired probability). Since the size of this variable space, roughly speaking, is $m = \tilde{O}(k)$ (crucially with no dependence on n), the [28] PRG's seed length in this context (of intersections of k LTFs over m variables) is $\log(m) \cdot \text{poly}(\log k, 1/\varepsilon) = \text{poly}(\log k, 1/\varepsilon)$, which leads to our overall final running time of $\text{poly}(n) \cdot 2^{\text{poly}(\log k, 1/\varepsilon)}$.

2 Our PRG and a high-level overview of its analysis

We use the same simple PRG construction to obtain both of our PRG results (Theorems 2 and 3); the two results are obtained by instantiating the parameters in two different ways. We describe this PRG below with general parameters; the precise parameter settings we use for each class (intersections versus arbitrary functions of k LTFs) will be made clear in the course of the respective analyses.

An idealized version of our PRG is as follows:

1. Let $\mathbf{G}^{(d)}$ be an $\mathcal{N}(0,1)^d$ Gaussian (which we view as a column vector).
2. Let $\mathbf{A} \in \mathbb{R}^{d \times n}$ be a pseudorandom Johnson-Lindenstrauss matrix drawn from the distribution of pseudorandom $d \times n$ JL-matrices given by the work of Kane, Meka and Nelson [20] (more details on this will be given below).
3. A draw from our generator Gen is $\mathbf{Z} := \mathbf{A}^{\mathsf{T}} \mathbf{G}^{(d)}$ (note that this is a vector in $\mathbb{R}^n$).

The actual PRG differs from the above-described idealized version because using finitely many bits it is not possible to generate a draw from the continuous $\mathbf{G}^{(d)}$ distribution with perfect fidelity. So in Step 1 the actual PRG uses a discrete approximation of each coordinate of $\mathbf{G}^{(d)}$ (we explain precisely what is meant by this in Appendix B); let $\hat{\mathbf{G}}^{(d)}$ denote the resulting distribution over $\mathbb{R}^d$. For clarity of exposition, the main analysis in the paper will be carried out for a "perfect" Gaussian $\mathbf{G}^{(d)}$, i.e. we will analyze the idealized PRG and show that it is an $O(\varepsilon)$-PRG for each of our two classes of interest (intersections and arbitrary functions of k LTFs). Appendix B shows that, for each of these two classes, if the idealized generator (which uses $\mathbf{G}^{(d)}$) is an $O(\varepsilon)$-PRG, then so is the actual generator which uses $\hat{\mathbf{G}}^{(d)}$.

High level idea of our generator. The Johnson-Lindenstrauss (JL) transform is one of the most important tools in high-dimensional data analysis. In a nutshell, for any error parameter ε, the JL transform gives a family $\mathcal{D}$ of $d \times n$ matrices such that for $\mathbf{A} \sim \mathcal{D}$ and any k unit vectors $W^1, \ldots, W^k \in \mathbb{R}^n$, with high probability, the following holds: For all $0 \le i, j \le k$, $\|\mathbf{A}W^i - \mathbf{A}W^j\|_2 = (1 \pm \varepsilon)\|W^i - W^j\|_2$ (where $W^0 = 0$). Crucially, one can obtain this guarantee with $d = O(\varepsilon^{-2} \log k)$.

We can reinterpret the guarantee of the JL transform in the following way: Let $\mathbf{A} \sim \mathcal{D}$ and consider the two distributions $\mathbf{Z} := \mathbf{A}^T \cdot \mathbf{G}^{(d)}$ and $\mathbf{Z}' = \mathbf{G}^{(n)}$. Let $W \in \mathbb{R}^{k \times n}$ be the $k \times n$ matrix whose rows are $W^1, \ldots, W^k$. Then, for any $\vec{\theta}$, the distributions $\mathbf{X} = W \cdot \mathbf{Z} - \vec{\theta}$ and $\mathbf{Y} = W \cdot \mathbf{Z}' - \vec{\theta}$ (i) are both Gaussian distributions over $\mathbb{R}^k$, (ii) have the same mean, and (iii) are such that the two $k \times k$ covariance matrices $\mathbf{Cov}(\mathbf{X})$ and $\mathbf{Cov}(\mathbf{Y})$ differ pointwise by at most ε. Let us define the affine function $f : \mathbb{R}^n \to \mathbb{R}^k$ as $f(z) = Wz - \vec{\theta}$. Then, the guarantee of the JL transform is that $\mathsf{Cov}(f(\mathbf{Z})) \approx_\varepsilon \mathsf{Cov}(f(\mathbf{Z}'))$; we may loosely view this guarantee as showing that the generator above *fools the covariance.*

The above perspective leads to the insight which motivates our work, which is essentially the following: since both $\mathbf{X}$ and $\mathbf{Y}$ are Gaussians, which are completely determined by their means and covariances, *other interesting tests besides the covariance may reasonably be expected to be fooled by (a pseudorandom version of) the Johnson-Lindenstrauss transform.* In this paper we consider tests of the form $h(\mathrm{sign}(f(z)_1), \ldots, \mathrm{sign}(f(z)_k))$, where h may be any function from $\{-1,1\}^k$ to $\{-1,1\}$ (we will also specialize to the case where h is an AND) and $f(z)_i$ denotes the i^{th} coordinate of $f(z)$. In other words, we are interested in fooling functions (given by h) of k LTFs (given by $\mathrm{sign}(f(z)_1), \ldots, \mathrm{sign}(f(z)_k)$). As we show in this paper, for a suitable choice of d (depending on whether h is arbitrary or is an AND) our generator can indeed fool all functions of the above form.

Seed length of our PRG. In order to state the seed length of our generator we first need to identify all of the relevant parameters. In Step 1, for each of our two results we will take $d = O(\log(k/\delta')/\varepsilon'^2)$ where ε' is a parameter that will be discussed below; as mentioned above each coordinate of $\hat{\mathbf{G}}^{(d)}$ will be a discrete approximation of an $\mathcal{N}(0,1)$ Gaussian. In Step 2, the KMN distribution over pseudorandom $d \times n$ JL-matrices has two additional parameters, which we denote ε' and δ' (see Section 4.2 for details.)

For the first step, as we show in Appendix B, a total of $O(d \log(kd/\varepsilon))$ many random bits suffice to generate a draw from $\hat{\mathbf{G}}^{(d)}$. For the second step, as we discuss in Section 4.2, a pseudorandom $d \times n$ JL-matrix with parameters ε', δ' can be drawn from the KMN distribution using $O(\log n + \log(1/\delta') \cdot \log(\log(1/\delta')/\varepsilon'))$ bits of randomness. So the overall seed length of our PRG is

$$\begin{aligned}
&O(d \log(kd/\varepsilon)) + O(\log n + \log(1/\delta') \cdot \log(\log(1/\delta')/\varepsilon')) \\
&= O\left(\frac{\log(k/\delta')}{\varepsilon'^2} \cdot (\log k + \log\log(k/\delta') + \log(1/(\varepsilon'\varepsilon))) + \log n\right).
\end{aligned}$$

As we will see in Section 4.2, we will always take δ' to be ε, so the seed length of our generator is

$$O\left(\frac{\log(k/\varepsilon)}{\varepsilon'^2} \cdot (\log k + \log\log(k/\varepsilon) + \log(1/(\varepsilon'\varepsilon))) + \log n\right). \tag{1}$$

We will instantiate the parameter ε' to one specific value (a function of k and ε) in Section 5 for arbitrary functions of LTFs, and to another specific value in Section 6 for intersections of LTFs, thus obtaining the seed lengths claimed in Theorems 2 and 3.

In the rest of this section we give an overview of the analyses of our PRGs. While the same PRG gives both our results, the analyses are quite different for the two classes we consider (arbitrary functions of LTFs and intersections of LTFs). We first sketch the (simpler) analysis for fooling arbitrary functions of LTFs.

2.1 An outline of our analysis for fooling arbitrary functions of LTFs

We start by recalling some definitions which are useful for our overview. An *orthant* of $\mathbb{R}^k$ is a subset $O \subset \mathbb{R}^k$ of the form

$$O = \{x \in \mathbb{R}^k : \mathrm{sign}(x_i) = b_i, i = 1, \dots, k\} \qquad \text{for some } (b_1, \dots, b_k) \in \{-1,1\}^k.$$

Given two random variables $\mathbf{X}, \mathbf{Y}$ over $\mathbb{R}^k$, the *quadratic Wasserstein distance* $\mathcal{W}_2(\mathbf{X}, \mathbf{Y})$ between $\mathbf{X}$ and $\mathbf{Y}$ is defined to be

$$\mathcal{W}_2(\mathbf{X}, \mathbf{Y}) = \inf_{(\widehat{\mathbf{X}}, \widehat{\mathbf{Y}})} (\mathbf{E}[\|\widehat{\mathbf{X}} - \widehat{\mathbf{Y}}\|^2])^{1/2},$$

where the infimum is taken over all couplings $(\widehat{\mathbf{X}}, \widehat{\mathbf{Y}})$ of $\mathbf{X}$ and $\mathbf{Y}$.[2]

Now we can present our overview. Our goal is to show that our PRG ε-fools every function of the form $g(h_1(x), \dots, h_k(x)) : \mathbb{R}^n \to \{-1,1\}$, where $g : \{-1,1\}^k \to \{-1,1\}$ is arbitrary and each $h_i : \mathbb{R}^n \to \{-1,1\}$ is an LTF, relative to the standard Gaussian distribution. This is equivalent to showing the following: for any unit vectors $W^1, \dots, W^k \in \mathbb{R}^n$ and any $\vec{\theta} = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$, taking W to be the $k \times n$ matrix whose rows are $W^1, \dots, W^k$ and taking $\mathcal{O}$ to be any union of orthants over $\mathbb{R}^k$, we have

$$\left| \mathop{\mathbf{Pr}}_{\mathbf{Z} \leftarrow \mathsf{Gen}}[W\mathbf{Z} - \vec{\theta} \in \mathcal{O}] - \mathop{\mathbf{Pr}}_{\mathbf{G}^{(n)} \leftarrow \mathcal{N}(0,1)^n}[W\mathbf{G}^{(n)} - \vec{\theta} \in \mathcal{O}] \right| \le \varepsilon. \tag{2}$$

Here is a high-level sketch of why our PRG ensures this.

(1) A (pseudorandom) JL projection of the k unit vectors $W^1, \dots, W^k \in \mathbb{R}^n$ results in much lower-dimensional vectors $V^1, \dots, V^k \in \mathbb{R}^d$, where $d = \Theta(\log(k)/\varepsilon'^2)$, which approximately preserve pairwise distances. Let us write Σ^W (Σ^V respectively) to denote the $k \times k$ covariance matrix of the k-dimensional Gaussian random variable $W\mathbf{G}^{(n)} - \vec{\theta}$ ($V\mathbf{G}^{(d)} - \vec{\theta}$ respectively, where $\mathbf{G}^{(d)}$ is distributed according to $\mathcal{N}(0,1)^d$). As we will see in Section 4.1, we have that Σ^W and Σ^V are entrywise close to each other (see Observation 6 for details).

(2) The entrywise closeness of Σ^W and Σ^V implies that the quadratic Wasserstein distance $\mathcal{W}_2(W\mathbf{G}^{(n)} - \vec{\theta}, V\mathbf{G}^{(d)} - \vec{\theta})$ is small; more precisely, we get that

$$\mathcal{W}_2(W\mathbf{G}^{(n)} - \vec{\theta}, V\mathbf{G}^{(d)} - \vec{\theta}) \le \tau, \quad \text{where } \tau = O(k^{\frac{7}{8}} \cdot (\varepsilon')^{1/4}). \tag{3}$$

(See Proposition 8 in Section 5.2 for details.)

(3) As the main step of our analysis, using an adaptation of an argument from [5], in Section 5.3 we use (3) to infer that for every union of orthants $\mathcal{O}$, we have

$$\left| \mathop{\mathbf{Pr}}_{\mathbf{G}^{(n)} \leftarrow \mathcal{N}(0,1)^n}[W\mathbf{G}^{(n)} - \vec{\theta} \in \mathcal{O}] - \mathop{\mathbf{Pr}}_{\mathbf{G}^{(d)} \leftarrow \mathcal{N}(0,1)^d}[V\mathbf{G}^{(d)} - \vec{\theta} \in \mathcal{O}] \right| \le O(k^{2/3}\tau^{2/3}) = \varepsilon. \tag{4}$$

This concludes the analysis since the inequality (4) is exactly the same as (2). This is because for each j we have $V^j = W^j \mathbf{A}^\mathsf{T}$ where $\mathbf{A}$ is the (pseudorandom) projection matrix.

[2] By the Kantorovich-Rubinstein duality theorem, there is an equivalent formulation $\mathcal{W}_2(\mathbf{X}, \mathbf{Y})$ in terms of Lipschitz test functions, but we will not need this alternative formulation.

2.2 An outline of our analysis for fooling intersections of LTFs

At a high level, our analysis for fooling intersections of LTFs exploits the rich and influential line of work on analyzing supremum (maximum) of Gaussian processes [24, 11, 31]. We recall that a Gaussian process is a set of jointly normal random variables (the set may be infinite, though we will only concerned with the finite case where it has cardinality k). To see the relationship between the maximum of a Gaussian process and an intersection of LTFs, let $W^1, \ldots, W^k \in \mathbb{R}^n$ be unit vectors and $\vec{\theta} \in \mathbb{R}^k$. Define the LTF $h_i(z) = \mathrm{sign}(W^i z - \theta_i)$ and consider the k-face polytope $h_1(z) \wedge \ldots \wedge h_k(z)$. Showing that our PRG ε-fools this k-face polytope (i.e., the function $h_1 \wedge \ldots \wedge h_k$) relative to the standard Gaussian distribution is equivalent to showing the following: Taking W to be the $k \times n$ matrix whose rows are $W^1, \ldots, W^k$,

$$\left| \Pr_{\mathbf{Z} \leftarrow \mathsf{Gen}}[W\mathbf{Z} \leq \vec{\theta}] - \Pr_{\mathbf{G} \leftarrow \mathcal{N}(0,1)^n}[W\mathbf{G} \leq \vec{\theta}] \right| \leq \varepsilon. \tag{5}$$

Note that $W\mathbf{Z} \leq \vec{\theta}$ if and only if $\max_{j \in [k]}((W\mathbf{Z})_j - \theta_j) \leq 0$. Likewise, $W\mathbf{G} \leq \vec{\theta}$ if and only if $\max_{j \in [k]}((W\mathbf{G})_j - \theta_j) \leq 0$.

Both $\{(W\mathbf{Z})_j - \theta_j\}_{1 \leq j \leq k}$ and $\{(W\mathbf{G})_j - \theta_j\}_{1 \leq j \leq k}$ are Gaussian processes, and we are interested in comparing the maxima of these two processes. If we were interested in comparing just the expectations of the maxima, i.e., $\mathbf{E}[\max_{j \in [k]}((W\mathbf{Z})_j - \theta_j)]$ versus $\mathbf{E}[\max_{j \in [k]}((W\mathbf{G})_j - \theta_j)]$, then the classical Sudakov-Fernique inequality [11, 30] provides a tool to compare (and prove the closeness of) these two quantities. Indeed, Meka [25] used this as a starting point in his work on a deterministic algorithm for estimating the supremum of a Gaussian process. We are interested in a somewhat more delicate quantity, and so we will use a generalization of a recent result of Chernozhukov *et al.* [6] which itself extends the Sudakov-Fernique inequality.

Now we turn from the above conceptual overview to a more detailed sketch of our analysis. Let the vectors $V^1, \ldots, V^k$ and the covariance matrix Σ^V be defined in the previous subsection.

(1') The first step of the argument is identical to Step 1 in the previous subsection: the covariance matrices Σ^W and Σ^V are entrywise close to each other.

(2') Next, we use the entrywise closeness of Σ^W and Σ^V to show that for any sufficiently smooth function g, we have that

$$\left| \mathbf{E}[g(\max_{j \in [k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j))] - \mathbf{E}[g(\max_{j \in [k]} (V^j \cdot \mathbf{G}^{(d)} - \theta_j))] \right| \quad \text{is small.} \tag{6}$$

is small. This is via an extension (to non-centered Gaussians) of Theorem 1 of [6], which in turn is a generalization of Chatterjee's quantitative Fernique-Sudakov bound [4].[3] We carry out this step in Section 6.2.

(3') Using a result of [16] (which follows almost directly from an influential work of Nazarov [27]), we have that the real-valued random variable

$$\max_{j \in [k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j),$$

which is a max of non-centered Gaussians, has good *anticoncentration*, meaning that it does not put very much mass in any small interval. See Section 6.3 for more details.

[3] Chatterjee's original argument in [4] bounds the difference in the expectations of $\max_{j \in [k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j)$ and $\max_{j \in [k]} (V^j \cdot \mathbf{G}^{(n)} - \theta_j)$, corresponding to the identity function $g(x) = x$.

(4') We specialize (6) to the case where g is a smooth approximator of the sign function. For a particular such g, combining (6) with the anticoncentration of $\max_{j\in[k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j)$ mentioned above, we can pass from g, which is a smooth approximator of $\mathrm{sign}(\cdot)$, to the actual $\mathrm{sign}(\cdot)$ function, and thereby show that

$$\left|\mathbf{Pr}[\mathrm{sign}(\max_{j\in[k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j)) = 1] - \mathbf{Pr}[\mathrm{sign}(\max_{j\in[k]} (V^j \cdot \mathbf{G}^{(d)} - \theta_j)) = 1]\right| \tag{7}$$

is small. We give this argument in Section 6.4.

(5') Having (7) be small is exactly the same as having the LHS of (5) is small, since for each j we have $V^j = W^j \mathbf{A}^\mathsf{T}$ where $\mathbf{A}$ is the (pseudorandom) projection matrix from Step 1 of our PRG. See Section 6.5 for more details.

3 The idea of our deterministic approximate counting results

In this section, we give an overview of our approximate counting algorithms for intersections and arbitrary functions of LTFs. We begin with the description for arbitrary functions as it relies on (extensions of) relatively well known tools from the literature such as regularity lemmas and invariance principles. In particular, we follow the (by now standard) paradigm of *reducing* the counting problem over the discrete cube to the Gaussian case by applying an appropriate regularity lemma; the proof of correctness relies on an *invariance principle for arbitrary functions of LTFs.* Once in the Gaussian case, we apply Theorem 3 which allows us to do counting over Gaussian space. This is explained in more detail in Section 3.1.

We then move on to the case of intersections of LTFs, which is somewhat more subtle. Similar to the first case, we also use a regularity lemma to *reduce* the Boolean case to the Gaussian case. However, instead of a naive approach of traversing all the root-to-leaf paths in the decision tree (constructed by the regularity lemma), we use the PRG construction of [28] to traverse only a small subset of the leaves. More details are given in Section 3.2.

3.1 Deterministic approximate counting for arbitrary functions of k LTFs via an invariance principle and a multiregularity lemma

A *regular* LTF is an LTF $\mathrm{sign}(\sum_{i=1}^n w_i x_i - \theta)$ in which, intuitively, no individual weight w_i has large magnitude compared to the overall magnitude of the weights (see Section 7.1 for a precise definition). The main structural result of [16] is an *invariance principle for intersections of LTFs*: roughly speaking, this states that if $F_0 = h_1 \wedge \cdots \wedge h_k$ is an intersection of k LTFs all of which are sufficiently regular, then the expected values of $F_0(\mathbf{U}^{(n)})$ (where the input is uniform over $\{-1,1\}^n$) and of $F_0(\mathbf{G}^{(n)})$ (where the input is a standard $\mathcal{N}(0,1)^n$ Gaussian) are close. A notable aspect of the [16] invariance principle is that its error bound has only a *poly-logarithmic* dependence on k (see Theorem 28 in Section 7.3 for a precise statement).

Now, consider any $F = g(h_1, \cdots, h_k)$ (where $g : \{-1,1\}^k \to \{-1,1\}$ is arbitrary). A naive approach based on just using the [16] invariance principle 2^k times together with a union bound would give an invariance principle for arbitrary functions of k LTFs with an error bound that depends exponentially on k. Instead, we use an analogue of the [16] invariance principle which goes beyond intersections of LTFs and works for arbitrary functions of k LTFs. The work of Gopalan et al. [14] gives an invariance principle for arbitrary functions of k LTFs that has a polynomial dependence on k in the error bound. We provide an alternate proof of this invariance principle for arbitrary functions of k LTFs. This polynomial dependence on k is crucial for obtaining a final overall running time for counting satisfying assignments with a singly exponential dependence on k, rather than a doubly exponential dependence which would follow from the naive approach.

As we explain in Section 7.2, the proof of our invariance principle is completely different from the proofs of of [16], [14]; we feel that our new proof of the invariance principle, Theorem 24, may be of independent interest. The [16] and [14] invariance principles are proved using a Lindeberg-type "replacement" argument; key ingredients are an analysis of hashing n coordinates into buckets and bounds on the derivatives of particular "smooth mollifiers" for functions of LTFs. Our proof of Theorem 24 uses none of these ingredients; instead, its main components are (a) a CLT for Wasserstein distance due to Valiant and Valiant [32], and (b) a conversion from Wasserstein distance to "union-of-orthants" distance. (Indeed, the ideas underlying the proof of Theorem 24 are very similar to the ideas underlying our PRG for arbitrary functions of k LTFs; this is analogous, at a high level, to how the proof of the [16] invariance principle is closely related to the analysis of the [16] PRG for intersections of regular LTFs.)

Using the invariance principle for deterministic approximate counting. By combining the invariance principle for arbitrary functions of LTFs with our PRG, which shows that a random variable $\mathbf{Z} \leftarrow \mathsf{Gen}$ is such that the expectation of $F(\mathbf{Z})$ is close to that of $F(\mathbf{G}^{(n)})$, it is straightforward to obtain a deterministic approximate counting algorithm for arbitrary functions of k regular LTFs over $\{-1,1\}^n$ simply by enumerating over all the seeds of our PRG. This algorithm has running time $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(k,1/\varepsilon)}$.) To obtain a deterministic approximate counting algorithm for arbitrary functions of k *general* LTFs over $\{-1,1\}^n$, we combine the above algorithm with the deterministic algorithmic version of the *multi-regularity lemma* of [14]. Briefly, this is a deterministic algorithm which builds a decision tree of depth roughly k, with the property that at almost every leaf ρ of the decision tree, either the restriction of $g(h_1, \cdots, h_k)$ according to ρ is very close to a constant function -1 or 1, or else each restricted LTF $h_1 \upharpoonright \rho, \ldots h_k \upharpoonright \rho$ is regular (and hence the deterministic approximate counting algorithm for arbitrary functions of regular LTFs can be used). We note that the total number of leaves in this decision tree is exponential in k. By running the approximate counting algorithm for functions of k-regular LTFs at each of the leaves, it is possible to approximate the overall number of satisfying assignments. We give the details of this (fairly standard) approach in Section 7.2.

3.2 Deterministic approximate counting for intersections of k LTFs

Let $F = h_1 \wedge \cdots \wedge h_k$. Recall that the invariance principle of [16] shows that if all the LTFs are sufficiently regular, then the expected values of $F_0(\mathbf{U}^{(n)})$ and of $F_0(\mathbf{G}^{(n)})$ are close, where crucially the error bound only has a polylogarithmic dependence on k. By combining this with our PRG, it is straightforward to obtain a deterministic approximate counting algorithm for intersections of k regular LTFs over $\{-1,1\}^n$ simply by enumerating over all the seeds of our PRG – the resulting running time is $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(\log k, 1/\varepsilon)}$. For intersections of general halfspaces, one can apply the multi-regularity lemma of [15] to reduce to the case of intersection of regular halfspaces. A naive application of this (similar to the previous subsection) will result in a running time exponential in k – this is because there are 2^k leaves in the resulting decision tree and running the algorithm for each of the leaves separately will result in an exponential in k overhead.

To instead get a $2^{\mathrm{poly}(\log k)}$ overhead, we crucially rely on two facts: (i) the decision tree constructed by the regularity lemma is non-adaptive, i.e., all nodes at the same level are labeled by the same variable. Further, if the set of internal variables is denoted by S, then this set can be enumerated in time $\mathrm{poly}(S)$. (ii) For any fixing of the set of variables in $\overline{S}$, the computation of the decision tree can be represented as an intersection of k halfspaces.

Glossing over some subtleties, this suggests that instead of doing approximate counting for all the leaves in the decision tree, one can just perform this computation on a subset of the leaves given by the output of a PRG. In particular, we use the PRG due to [28] to select the subset. While the PRG in [28] has a $(\log n) \cdot \text{poly}(\log k)$ seed length (where n is the ambient dimension), in this application 'n' is set to $|S|$ which has polynomial dependence on k (for constant error $\varepsilon > 0$). Putting this together, we obtain a deterministic algorithm for counting intersection of k arbitrary halfspaces with running time $\text{poly}(n) \cdot 2^{\text{poly}(\log k, 1/\varepsilon)}$. The full details are given in Section 7.3.

4 Notation and setup

We write $W \in \mathbb{R}^{k \times n}$ to denote the matrix whose j-th row is the weight vector of the j-th LTF in a function of k LTFs. We assume that each such LTF has been normalized so that its weight vector has norm 1. For $j \in [k]$ (indexing one of the LTFs) we write $W^j = (W^j_1, \ldots, W^j_n)$ to denote the j-th row of W, so $\|W^j\| = 1$ for all j. Thus an arbitrary function of k LTFs is $g(h_1, \ldots, h_k)$, where $g : \{-1,1\}^k \to \{-1,1\}$ and

$$h_j(x) = \text{sign}(W^j \cdot x - \theta_j) \qquad \text{where } W^j = (W^j_1, \ldots, W^j_n) \in \mathbb{R}^n \text{ has } \|W^j\| = 1$$

(we take -1 to represent True and 1 to represent False throughout), and an intersection of k LTFs is a function $h_1(x) \wedge \cdots \wedge h_k(x)$.

Throughout this paper we will use notation like $\vec{\theta}$ to denote vectors in $\mathbb{R}^k$, i.e. $\vec{\theta} = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$. We write $\mathbf{G}$ or simply $\mathbf{G}^{(n)}$ to denote $(\mathbf{G}_1, \ldots, \mathbf{G}_n)$, a random variable distributed according to $\mathcal{N}(0,1)^n$ (so each of $\mathbf{G}_1, \ldots, \mathbf{G}_n$ is an i.i.d. $\mathcal{N}(0,1)$ Gaussian).

4.1 Entrywise closeness of the original covariance matrix and the pseudorandomly-projected covariance matrix

As above let $W \in \mathbb{R}^{k \times n}$ have j-th row W^j with $\|W^j\| = 1$ for all $j \in [k]$. For convenience we also define $W^0 \in \mathbb{R}^n$ to be the all-0 vector.

Let $d = O(\log(k/\delta')/\varepsilon'^2)$ (where ε' will be taken to be at most 1) and let $V \in \mathbb{R}^{k \times d}$ satisfy the following:

$$\text{For all } 0 \le i,j \le k \text{ we have } \|W^i - W^j\| \le \|V^i - V^j\| \le (1+\varepsilon')\|W^i - W^j\| \tag{8}$$

where we take $V^0 = (0, \ldots, 0) \in \mathbb{R}^d$. (As we will see in the next subsection, $V^1, \ldots, V^k$ should be thought of as the vectors we get by doing a pseudorandom JL-projection of $W^1, \ldots, W^k$ to d dimensions.)

We will consider the two k-dimensional Gaussian random vectors $W\mathbf{G}^{(n)}$ and $V\mathbf{G}^{(d)}$. The covariance matrix of $W\mathbf{G}^{(n)}$, which we denote Σ^W, is the $k \times k$ matrix $W^{\mathsf{T}}W$ which has $\sigma^W_{ij} := W^i \cdot W^j$ as its (i,j) entry, and similarly the covariance matrix Σ^V of $V\mathbf{G}^{(d)}$ has $\sigma^V_{ij} := V^i \cdot V^j$ as its (i,j) entry. We define

$$\Delta := \max_{1 \le i,j \le k} |\sigma^W_{ij} - \sigma^V_{ij}| = \max_{1 \le i,j \le k} |W^i \cdot W^j - V^i \cdot V^j|, \tag{9}$$

the maximum entry-wise difference between the two covariance matrices. The following simple observation upper bounds Δ:

▶ **Observation 6.** *If $W^0, \ldots, W^k \in \mathbb{R}^n$, $V^0, \ldots, V^k \in \mathbb{R}^d$ satisfy (8), then $\Delta \le 9\varepsilon'$.*

Proof. Taking $i = 0$, (8) implies that each V^j, $j \in [k]$, has $\|V^j\| \in [1, 1+\varepsilon']$. Now fix any $i, j \in [k]$. We have

$$\|W^i - W^j\|^2 = W^i \cdot W^i - 2W^i \cdot W^j + W^j \cdot W^j = 2 - 2W^i \cdot W^j$$

and similarly (using the fact that each $\|V^\ell\|^2 \le (1+\varepsilon')^2$)

$$\|V^i - V^j\|^2 = V^i \cdot V^i - 2V^i \cdot V^j + V^j \cdot V^j = 2 + 2\gamma - 2V^i \cdot V^j$$

for some $0 \le \gamma \in 2\varepsilon' + \varepsilon'^2 \le 3\varepsilon'$. Hence

$$2\gamma + 2W^i \cdot W^j - 2V^i \cdot V^j = \|V^i - V^j\|^2 - \|W^i - W^j\|^2,$$

which implies

$$\begin{aligned} |W^i \cdot W^j - V^i \cdot V^j| &\le \gamma + \frac{1}{2}\left(\|V^i - V^j\|^2 - \|W^i - W^j\|^2\right) \\ &\le 3\varepsilon' + \frac{1}{2}\left(\left((1+\varepsilon')\|W^i - W^j\|\right)^2 - \|W^i - W^j\|^2\right) \\ &= 3\varepsilon' + \frac{1}{2}\left((2\varepsilon' + \varepsilon'^2)\|W^i - W^j\|^2\right) \\ &\le 3\varepsilon' + 2(2\varepsilon' + \varepsilon'^2) \le 9\varepsilon', \end{aligned}$$

where for the penultimate inequality we used $\|W^i - W^j\|^2 \le 4$ and $\varepsilon'^2 \le \varepsilon'$ which holds since $0 < \varepsilon' < 1$. ◀

4.2 Formalizing step (1) of the intuitive sketch: Getting d-dimensional vectors $V^1, \ldots, V^k$ via pseudorandom projection

Recall that Steps 1 and 1′ of the analysis are identical for arbitrary functions of LTFs (in Section 2.1) and for intersections of LTFs (in Section 2.2). We give the details of this step here.

We use the following derandomized JL lemma given by Kane, Meka, and Nelson [20]:

▶ **Theorem 7** (Derandomized Johnson-Lindenstrauss [20]). *Let $0 \le \varepsilon', \delta' < 1/2$ and let $\delta'' = \delta'/k^2$. There is a distribution $\mathcal{D}$ over random matrices $\mathbf{A} \in \mathbb{R}^{d \times n}$, $d = O(\log(k/\delta')/\varepsilon'^2)$, such that (i) a draw of $\mathbf{A} \leftarrow \mathcal{D}$ can be generated using $O(\log n + \log(1/\delta'') \cdot \log((\log(1/\delta''))/\varepsilon'))$ bits, and (ii) the following holds: Fix unit vectors $W^1, \ldots, W^k \in \mathbb{R}^n$. Then*

$$\mathop{\boldsymbol{Pr}}_{\mathbf{A} \leftarrow \mathcal{D}}\left[\|W^i - W^j\| \le \|W^i\mathbf{A}^\mathsf{T} - W^j\mathbf{A}^\mathsf{T}\| \le (1+\varepsilon')\|W^i - W^j\| \text{ for all } i,j \in [k]\right] \ge 1 - \delta'. \tag{10}$$

Let $\mathbf{V}^j = W^j\mathbf{A}^\mathsf{T}$ where $\mathbf{A} \leftarrow \mathcal{D}$. By Theorem 7, except with failure probability at most δ', (8) is satisfied. We will always take $\delta' = \varepsilon$, and so this δ' failure probability just gets absorbed into the overall $O(\varepsilon)$ error bound of the PRG. Fix $V^1, \ldots, V^k$ to be any such outcome of $\mathbf{V}^1, \ldots, \mathbf{V}^k$; in the rest of the argument we will work with this $V^1, \ldots, V^k$. Note that by Observation 6 we have that Δ, which is defined in terms of this $V^1, \ldots, V^k$, satisfies $\Delta \le 9\varepsilon'$.

5 Fooling arbitrary functions of LTFs: Proof of Theorem 2

5.1 Parameter settings

As will be seen in the analysis below, in order for the overall PRG to $O(\varepsilon)$-fool arbitrary functions of k LTFs, we take $\varepsilon' = \frac{\varepsilon^6}{k^{15/2}}$. Recalling that $\delta' = \varepsilon$, by (1) the overall seed length (as a function of n, k and ε) is $O(\log n) + \tilde{O}(\frac{k^{15}}{\varepsilon^{12}})$, as claimed in Theorem 2. In the rest of this section we establish correctness of the PRG.

5.2 Formalizing step (2) of the intuitive sketch: Upper bounding the quadratic Wasserstein distance

Recall that the *quadratic Wasserstein distance* between random variables $\mathbf{X}, \mathbf{Y}$ in $\mathbb{R}^k$ is defined to be

$$\mathcal{W}_2(\mathbf{X}, \mathbf{Y}) = \inf_{(\widehat{\mathbf{X}}, \widehat{\mathbf{Y}})} (\mathbf{E}[\|\widehat{\mathbf{X}} - \widehat{\mathbf{Y}}\|^2])^{1/2}, \tag{11}$$

where the infimum is taken over all couplings $(\widehat{\mathbf{X}}, \widehat{\mathbf{Y}})$ of $\mathbf{X}$ and $\mathbf{Y}$.

▶ **Proposition 8.** *Let $W^1, \ldots, W^k$ be unit vectors in $\mathbb{R}^n$, $V^1, \ldots, V^k$ be vectors in $\mathbb{R}^d$ satisfying (8) and let $\vec{\theta} \in \mathbb{R}^k$. Then we have*

$$\mathcal{W}_2(W\mathbf{G}^{(n)} - \vec{\theta}, V\mathbf{G}^{(d)} - \vec{\theta}) \leq \tau, \qquad \textit{where } \tau = O(k^{\frac{7}{8}} \cdot (\varepsilon')^{1/4}). \tag{12}$$

Proof. Observe that $W\mathbf{G}^{(n)} - \vec{\theta}$ and $V\mathbf{G}^{(d)} - \vec{\theta}$ have the same mean. For this case, Proposition 7 of Givens and Shortt [12] shows that

$$\mathcal{W}_2^2(W\mathbf{G}^{(n)} - \vec{\theta}, V\mathbf{G}^{(d)} - \vec{\theta}) = \mathrm{Tr}(\Sigma^W + \Sigma^V - 2((\Sigma^W)^{1/2}\Sigma^V(\Sigma^W)^{1/2})^{1/2}). \tag{13}$$

Here Σ^W and Σ^V are the covariance matrices of the distribution $W\mathbf{G}^{(n)} - \vec{\theta}$ and $V\mathbf{G}^{(d)} - \vec{\theta}$ respectively[4]. To bound the expression on the right hand side, first observe that

$$\left|\mathrm{Tr}(\Sigma^W + \Sigma^V) - 2\mathrm{Tr}(\Sigma^W)\right| \leq \left|\mathrm{Tr}(\Sigma^W - \Sigma^V)\right| \leq 9k \cdot \varepsilon'. \tag{14}$$

The last inequality uses Observation 6. To proceed further, we recall the following very useful fact from Bhatia [2] (Theorem X.1.3)

▶ **Fact 9.** *Let $\|\cdot\|$ be any unitarily invariant matrix norm. For psd matrices A and B, we have the following*

$$\|\,|A^{\frac{1}{2}} - B^{\frac{1}{2}}|\,\| \leq \|\sqrt{|A - B|}\|,$$

where $|X|$ denotes the psd matrix $\sqrt{X^*X}$.

For any symmetric matrix X, let $\|X\|_{\mathsf{tr}}$ denotes its trace norm, i.e., the sum of the singular values of X. Note that the trace-norm is unitarily invariant. With this, we now have

$$\begin{aligned}
\left|2\mathrm{Tr}(\Sigma^W - ((\Sigma^W)^{1/2}\Sigma^V(\Sigma^W)^{1/2})^{1/2})\right| &\leq 2\|\Sigma^W - ((\Sigma^W)^{1/2}\Sigma^V(\Sigma^W)^{1/2})^{1/2}\|_{\mathsf{tr}} \\
&\leq 2\|\sqrt{|(\Sigma^W)^2 - (\Sigma^W)^{1/2}\Sigma^V(\Sigma^W)^{1/2}|}\|_{\mathsf{tr}} \\
&= 2\|\sqrt{|(\Sigma^W)^{1/2}(\Sigma^W - \Sigma^V)(\Sigma^W)^{1/2}|}\|_{\mathsf{tr}} \qquad (15)
\end{aligned}$$

In the above, the first inequality uses the fact that for any symmetric matrix X, $|\mathrm{Tr}(X)| \leq \|X\|_{\mathsf{tr}}$ and the second inequality follows from Fact 9. We now recall the following fact:

▶ **Fact 10.** *For any symmetric $X \in \mathbb{R}^{k \times k}$,*

$$\|\sqrt{|X|}\|_{\mathsf{tr}} \leq \sqrt{k} \cdot \sqrt{\|X\|_{\mathsf{tr}}}.$$

[4] [12] states their theorem for non-singular Σ^V and Σ^W. However, we can always perturb our Gaussians infinitesimally, apply (13) and then take a limit.

Proof. If $\sigma_1, \ldots, \sigma_k$ denotes the singular values of X, then the left hand side is $\sum_{j=1}^{k} \sqrt{\sigma_j}$ and the right hand side is $\sqrt{k} \cdot \sqrt{\sigma_1 + \ldots + \sigma_k}$, so the inequality is a consequence of the AM-GM inequality. ◀

Applying Fact 10 to (15), we have that

$$\begin{aligned} \left|2\mathrm{Tr}(\Sigma^W - ((\Sigma^W)^{1/2}\Sigma^V(\Sigma^W)^{1/2})^{1/2})\right| &\leq 2\sqrt{k}\sqrt{\| |(\Sigma^W)^{1/2}(\Sigma^W - \Sigma^V)(\Sigma^W)^{1/2}| \|_{\mathsf{tr}}}. \\ &= 2\sqrt{k}\sqrt{\|(\Sigma^W)^{1/2}(\Sigma^W - \Sigma^V)(\Sigma^W)^{1/2}\|_{\mathsf{tr}}}. \qquad (16) \end{aligned}$$

The second equality simply uses that for symmetric X, $\| |X| \|_{\mathsf{tr}} = \|X\|_{\mathsf{tr}}$. Next, we recall the following useful inequality for unitarily invariant norms (see [2], p.94).

▶ **Fact 11.** *Let A, B, C be symmetric matrices and let $\| \cdot \|$ be any unitarily invariant norm. Then, $\|ABC\| \leq \|A\|_2 \cdot \|B\| \cdot \|C\|_2$.*

Applying Fact 11 to the right hand side of (16), we obtain

$$\begin{aligned} \left|2\mathrm{Tr}(\Sigma^W - ((\Sigma^W)^{1/2}\Sigma^V(\Sigma^W)^{1/2})^{1/2})\right| &\leq 2\sqrt{k}\sqrt{\|(\Sigma^W)^{1/2}\|_2 \|\Sigma^W - \Sigma^V\|_{\mathsf{tr}} \|(\Sigma^W)^{1/2}\|_2}. \\ &= 2\sqrt{k}\|(\Sigma^W)^{1/2}\|_2 \cdot \sqrt{\|\Sigma^W - \Sigma^V\|_{\mathsf{tr}}}. \qquad (17) \end{aligned}$$

Now, Σ^W is a matrix in which each entry $W^i \cdot W^j$ is upper bounded by 1 in absolute value. Thus, $\|\Sigma^W\|_2 \leq k$. This immediately implies that $\|(\Sigma^W)^{1/2}\|_2 \leq \sqrt{k}$. Similarly,

$$\|\Sigma^W - \Sigma^V\|_{\mathsf{tr}} \leq \sqrt{k} \cdot \|\Sigma^W - \Sigma^V\|_F \leq 9\sqrt{k} \cdot k \cdot \varepsilon' = 9\varepsilon' \cdot k^{3/2}.$$

Here the last inequality is again using Observation 6. Combining this with (17), we have

$$\left|2\mathrm{Tr}(\Sigma^W - ((\Sigma^W)^{1/2}\Sigma^V(\Sigma^W)^{1/2})^{1/2})\right| \leq 6k^{\frac{7}{4}} \cdot \sqrt{\varepsilon'}.$$

Combining the above equation with (14) and (13) (and using triangle inequality), we get that

$$\mathcal{W}_2^2(W\mathbf{G}^{(n)} - \vec{\theta}, V\mathbf{G}^{(d)} - \vec{\theta}) \leq 9k\varepsilon' + 2k^{\frac{7}{4}} \cdot \sqrt{\varepsilon'}.$$

This immediately yields the proposition. ◀

5.3 Formalizing step (3) of the intuitive sketch: Upper bounding the "union-of-orthants distance"

The following definition will be convenient: Given two random variables $\mathbf{X}, \mathbf{Y}$ over $\mathbb{R}^k$, the *union-of-orthants distance between* $\mathbf{X}$ *and* $\mathbf{Y}$ is defined to be

$$d_{\mathrm{UO}}(\mathbf{X}, \mathbf{Y}) := \max_{\mathcal{O}} \left|\mathbf{Pr}[\mathbf{X} \in \mathcal{O}] - \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}]\right|, \qquad (18)$$

where the max is taken over all 2^{2^k} possible unions of orthants $\mathcal{O}$ in $\mathbb{R}^k$. This definition aligns well with arbitrary functions of k LTFs $g(h_1, \ldots, h_k)$ because of the following easy observation:

▶ **Observation 12.** *For any $g : \{-1,1\}^k \to \{-1,1\}$ and any random variables $\mathbf{X}, \mathbf{Y}$ over $\mathbb{R}^k$, we have*

$$|\boldsymbol{Pr}[g(\mathrm{sign}(\mathbf{X}_1), \ldots, \mathrm{sign}(\mathbf{X}_k)) = 1] - \boldsymbol{Pr}[g(\mathrm{sign}(\mathbf{Y}_1), \ldots, \mathrm{sign}(\mathbf{Y}_k)) = 1]| \leq d_{\mathrm{UO}}(\mathbf{X}, \mathbf{Y}).$$

▶ **Lemma 13.** *Let $W^1, \dots, W^k$ be unit vectors in $\mathbb{R}^n$, $V^1, \dots, V^k$ be vectors in $\mathbb{R}^d$ satisfying (8) and let $\vec{\theta} \in \mathbb{R}^k$. Then we have*

$$d_{\mathrm{UO}}(W\mathbf{G}^{(n)} - \vec{\theta}, V\mathbf{G}^{(d)} - \vec{\theta}) \le O(k^{2/3}\tau^{2/3}), \tag{19}$$

where τ is as defined in Proposition 8.

The argument here is similar to the proof of Theorem 5 in [5]. That result used a CLT due to Valiant and Valiant (which gave an upper bound on the L^1 (as opposed to quadratic, i.e. $\mathcal{W}_2$) transportation distance between a certain sum of vector-valued random variables and a Gaussian distribution) to obtain an upper bound on union-of-orthants distance between those two distributions. We briefly explain the main idea (which is quite simple) behind the argument in our setting.

We consider an optimal coupling of the random variables $\mathbf{X} = W\mathbf{G}^{(n)} - \vec{\theta}$ and $\mathbf{Y} = V\mathbf{G}^{(d)} - \vec{\theta}$ which achieves the minimal quadratic transportation distance as in (11). Since by Proposition 8 the quadratic transportation cost $\mathcal{W}_2(\mathbf{X}, \mathbf{Y})$ of transforming $\mathbf{X}$ to $\mathbf{Y}$ is "small", the optimal coupling cannot move a "non-small" amount of mass by a distance that is not "small." Assume (contrary to our desired conclusion) that the union-of-orthants distance between $\mathbf{X}$ and $\mathbf{Y}$ is not small, and fix a union of orthants $\mathcal{O}$ that achieves the max in (18). Without loss of generality we may suppose that $\mathbf{X}$ puts more mass on $\mathcal{O}$ than $\mathbf{Y}$ (and this difference is large by the above assumption). Gaussian anticoncentration tells us that $\mathbf{X}$ can only have a small amount of mass overall that is close to orthant boundaries, and hence $\mathbf{X}$ can have only a small amount of such mass in $\mathcal{O}$. This means that a non-small amount of mass from $\mathbf{X}$ must be moved a non-small distance (since it must go from being within $\mathcal{O}$ and not close to any orthant boundary, to being outside of $\mathcal{O}$) in order to transform $\mathbf{X}$ to $\mathbf{Y}$; but this contradicts the premise that $\mathcal{W}_2(\mathbf{X}, \mathbf{Y})$ is small.

We now proceed to the formal argument.

Proof of Lemma 13. As above let $\mathbf{X} = W\mathbf{G}^{(n)} - \vec{\theta}$ and $\mathbf{Y} = V\mathbf{G}^{(d)} - \vec{\theta}$. By Proposition 8 we have that $\mathcal{W}_2(\mathbf{X}, \mathbf{Y}) \le \tau$. We define

$$B_r := \left\{x \in \mathbb{R}^k \colon |x_i| \le r \text{ for some } i \in [k]\right\}$$

to be the region of all points in $\mathbb{R}^k$ whose L^∞-distance from any orthant boundary point is at most r. With foresight we choose $r = \tau^{2/3}/k^{1/3}$ (the rationale for this choice will be evident toward the end of the proof). We partition $\mathcal{O}$ into $\mathcal{O}_{\mathrm{bd}} := \mathcal{O} \cap B_r$ (the points in $\mathcal{O}$ that lie close to the orthant boundaries) and $\mathcal{O}_{\mathrm{in}} := \mathcal{O} \setminus B_r$ (the points in $\mathcal{O}$ that lie far away from the orthant boundaries). We have

$$\begin{aligned}\left|\mathbf{Pr}[\mathbf{X} \in \mathcal{O}] - \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}]\right| &= \left|(\mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{\mathrm{in}}] + \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{\mathrm{bd}}]) - (\mathbf{Pr}[\mathbf{Y} \in \mathcal{O}_{\mathrm{in}}] + \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}_{\mathrm{bd}}])\right| \\ &\le \underbrace{\left|\mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{\mathrm{in}}] - \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}_{\mathrm{in}}]\right|}_{=\Xi} + \underbrace{\mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{\mathrm{bd}}] + \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}_{\mathrm{bd}}]}_{=\Gamma}.\end{aligned}$$

We bound the quantities Ξ and Γ separately.

For Γ, we have that

$$\Gamma \le \sum_{i=1}^{k} \mathbf{Pr}\left[\mathbf{X}_i \in [-r, r]\right] + \mathbf{Pr}\left[\mathbf{Y}_i \in [-r, r]\right] \le O(kr), \tag{20}$$

where we used the fact that each coordinate $\mathbf{X}_i$ of $\mathbf{X}$ is a one-dimensional Gaussian with variance $\|W^i\|^2 = 1$ and each coordinate $\mathbf{Y}_i$ of $\mathbf{Y}$ is a one-dimensional Gaussian with variance $1 \le \|V^i\|^2 \le (1 + \varepsilon')^2 = O(1)$.

For Ξ, let us assume without loss of generality (a symmetrical argument works in the other case) that $\mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{\text{in}}] \geq \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}_{\text{in}}]$, so $\Xi = \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{\text{in}}] - \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}_{\text{in}}]$. Let $\mathcal{D}$ be any coupling of $\mathbf{X}$ and $\mathbf{Y}$ that achieves

$$\mathop{\mathbf{E}}_{(\widehat{\mathbf{X}},\widehat{\mathbf{Y}})\sim\mathcal{D}}[\|\widehat{\mathbf{X}} - \widehat{\mathbf{Y}}\|^2]^{1/2} = 2\tau,$$

so $\mathcal{D}$ is the joint distribution of a pair $(\mathbf{U}, \mathbf{V})$ of $\mathbb{R}^k$-valued random variables with marginals distributed according to $\mathbf{X}$ and $\mathbf{Y}$ respectively. Since

$$\int_{\mathcal{O}_{\text{in}}} \int_{\mathbb{R}^k} \mathcal{D}(u,v)\, dv\, du = \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{\text{in}}]$$

and

$$\int_{\mathcal{O}_{\text{in}}} \int_{\mathcal{O}_{\text{in}}} \mathcal{D}(u,v)\, dv\, du \leq \int_{\mathbb{R}^k} \int_{\mathcal{O}_{\text{in}}} \mathcal{D}(u,v)\, dv\, du = \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}_{\text{in}}],$$

it follows that

$$\int_{\mathcal{O}_{\text{in}}} \int_{\mathbb{R}^k \setminus \mathcal{O}_{\text{in}}} \mathcal{D}(u,v)\, dv\, du = \int_{\mathcal{O}_{\text{in}}} \int_{\mathbb{R}^k} \mathcal{D}(u,v)\, dv\, du - \int_{\mathcal{O}_{\text{in}}} \int_{\mathcal{O}_{\text{in}}} \mathcal{D}(u,v)\, dv\, du \geq \Xi. \tag{21}$$

Next we define the quantities

$$\Xi_{\text{near}}(\mathcal{D}) \quad := \quad \int_{\mathcal{O}_{\text{in}}} \int_{\mathcal{O}_{\text{bd}}} \mathcal{D}(u,v)\, dv\, du$$

(in words, this is the probability that $\mathbf{U}$ lies “well inside” $\mathcal{O}$ and $\mathbf{V}$ lies “close to the boundary” in $\mathcal{O}$), and

$$\Xi_{\text{far}}(\mathcal{D}) \quad := \quad \int_{\mathcal{O}_{\text{in}}} \int_{\mathbb{R}^k \setminus \mathcal{O}} \mathcal{D}(u,v)\, dv\, du$$

(in words, this is the probability that $\mathbf{U}$ lies “well inside” $\mathcal{O}$ and $\mathbf{V}$ lies outside $\mathcal{O}$). Note that $\Xi_{\text{near}}(\mathcal{D})$ and $\Xi_{\text{far}}(\mathcal{D})$ sum to the quantity on the left-hand side of (21), and so $\Xi_{\text{near}}(\mathcal{D}) + \Xi_{\text{far}}(\mathcal{D}) \geq \Xi$. (In words, since $\mathbf{X}$ places Ξ more mass on $\mathcal{O}_{\text{in}}$ than $\mathbf{Y}$ does, any scheme $\mathcal{D}$ of moving the mass of $\mathbf{X}$ to obtain $\mathbf{Y}$ must move at least Ξ amount from within $\mathcal{O}_{\text{in}}$ to outside it. $\Xi_{\text{near}}(\mathcal{D})$ is the amount moved from within $\mathcal{O}_{\text{in}}$ to $\mathcal{O}$'s boundary $\mathcal{O}_{\text{bd}}$, and $\Xi_{\text{far}}(\mathcal{D})$ is the rest, moved from within $\mathcal{O}_{\text{in}}$ to locations entirely out of $\mathcal{O}$.) Since $\|u - v\|^2 \geq r^2$ for any pair of points $u \in \mathcal{O}_{\text{in}}$ and $y \notin \mathcal{O}$, it follows that

$$(2\tau)^2 = \mathop{\mathbf{E}}_{(\mathbf{U},\mathbf{V})\sim\mathcal{D}}[\|\mathbf{U} - \mathbf{V}\|^2] \geq r^2 \cdot \Xi_{\text{far}}(\mathcal{D}).$$

We consider two cases, depending on the relative magnitudes of $\Xi_{\text{near}}(\mathcal{D})$ and $\Xi_{\text{far}}(\mathcal{D})$. If $\Xi_{\text{far}}(\mathcal{D}) \geq \Xi_{\text{near}}(\mathcal{D})$, then we have

$$r^2 \cdot \frac{\Xi}{2} \leq r^2 \cdot \Xi_{\text{far}}(\mathcal{D}) \leq 4\tau^2,$$

and hence $\Xi \leq 8\tau^2/r^2$, which along with our upper bound on Γ given by (20) completes the proof. If on the other hand $\Xi_{\text{near}}(\mathcal{D}) > \Xi_{\text{far}}(\mathcal{D})$, then

$$\frac{\Xi}{2} \leq \Xi_{\text{near}}(\mathcal{D}) \leq \int_{\mathbb{R}^k} \int_{\mathcal{O}_{\text{bd}}} \mathcal{D}(u,v)\, dv\, du = \mathbf{Pr}[\mathbf{Y} \in \mathcal{O}_{\text{bd}}] \leq \Gamma,$$

and again our upper bound on Γ completes the proof. ◀

Observing that by our setting of parameters we have that $k^{2/3}\tau^{2/3} = O(\varepsilon)$, we get that

$$d_{\mathrm{UO}}(W\mathbf{G}^{(n)} - \vec{\theta}, V\mathbf{G}^{(d)} - \vec{\theta}) \leq O(\varepsilon)$$

provided that $W^1, \dots, W^k$, $V^1, \dots, V^k$ satisfy (8). Recalling from Section 4.2 that all but a $\delta' = \varepsilon$ fraction of outcomes $V^1, \dots, V^k$ of $\mathbf{V}^j = W^j\mathbf{A}^\intercal$ satisfy (8), we have

$$d_{\mathrm{UO}}(W\mathbf{G}^{(n)} - \vec{\theta}, W\mathbf{A}^\intercal\mathbf{G}^{(d)} - \vec{\theta}) \leq O(\varepsilon),$$

and recalling that a draw $\mathbf{Z}$ from our generator Gen is $\mathbf{Z} = \mathbf{A}^\intercal\mathbf{G}^{(d)}$, this is equivalent to

$$d_{\mathrm{UO}}(W\mathbf{G}^{(n)} - \vec{\theta}, W\mathbf{Z} - \vec{\theta}) \leq O(\varepsilon),$$

and the proof of Theorem 2 is complete.

6 Fooling intersections of LTFs: Proof of Theorem 3

6.1 Parameter settings, notation and terminology

As we will see in the analysis given below, in order for the overall PRG to ε-fool k-facet Gaussian polytopes it suffices to take $\varepsilon' = O(\varepsilon^3/\log^2 k)$ and $\delta' = \varepsilon'/k^2$, so by (1) the overall seed length (as a function of n, k and ε) is $O(\log n) + \tilde{O}(\frac{\log^6 k}{\varepsilon^6})$ as claimed in Theorem 3.

The following notation will be useful: For $0 < \lambda$, $k \geq 1$, and $\vec{\theta} = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$, we define

$$\mathrm{Strip}_{\lambda,k,\vec{\theta}} = \{x \in \mathbb{R}^k : \text{ some } j \in [k] \text{ has } x_j \in (\theta_j, \theta_j + \lambda) \text{ and every } j \in [k] \text{ has } x_j < \theta_j + \lambda\}.$$

We recall that the *Kolmogorov distance* between two real-valued random variables $\mathbf{S}$ and $\mathbf{T}$ is defined to be

$$d_{\mathrm{K}}(\mathbf{S}, \mathbf{T}) = \sup_{\theta \in \mathbb{R}} \big| \mathbf{Pr}[\mathbf{S} \leq \theta] - \mathbf{Pr}[\mathbf{T} \leq \theta] \big|.$$

For $f : \mathbb{R}^k \to \mathbb{R}$ a smooth function we write $\partial_j f(z)$ to denote $\frac{\partial f}{\partial z_j}(z)$ and write $\partial_i\partial_j f(z)$ to denote $\frac{\partial^2 f}{\partial z_i \partial z_j}(z)$.

6.2 Formalizing step (2′) of the intuitive sketch: Fooling smooth test functions of max of non-centered Gaussians

A crucial ingredient in executing step (2′) of our analysis is the the following "soft-max" function which is used in [4, 6] and many other works. The soft-max function $F_\beta : \mathbb{R}^k \to \mathbb{R}$ is defined as

$$F_\beta(x_1, \dots, x_k) = \frac{1}{\beta} \cdot \ln \left(\sum_{i=1}^{k} e^{\beta x_i} \right).$$

For conciseness let us write e_β to denote $\beta^{-1} \ln k$. We record some useful facts about the soft-max function:

▶ **Fact 14.** *For any vector $v \in \mathbb{R}^k$, and any parameter $\beta > 0$,*

$$0 \leq F_\beta(v) - \max_{i \in [k]} v_i \leq e_\beta.$$

▶ **Fact 15** (Lemma 3 of [6]). *For every $1 \le i, j \le k$, we have*

$$\partial_i F_\beta(z) = \pi_i(z), \qquad \partial_i \partial_j F_\beta(z) = \beta w_{ij}(z),$$

where

$$\pi_i(z) := \frac{e^{\beta z_i}}{\sum_{\ell=1}^k e^{\beta z_\ell}}, \qquad w_{ij}(z) := \mathbb{1}[i = j]\pi_i(z) - \pi_i(z)\pi_j(z).$$

Furthermore, we have

$$\pi_j(z) \ge 0, \qquad \sum_{j=1}^k \pi_j(z) = 1, \qquad \sum_{i=1}^k \sum_{j=1}^k |w_{ij}(z)| \le 2.$$

▶ **Fact 16** (Lemma 4 of [6]). *Let $m(z) = g(F_\beta(z))$ where $g \in C^2(\mathbb{R})$. Then for every $1 \le i, j \le k$, we have*

$$\partial_i \partial_j m(z) = (g''(F_\beta(z))\pi_i(z)\pi_j(z) + \beta g'(F_\beta(z)) w_{ij}(z),$$

where π_i and w_{ij} are defined as in Fact 15 above.

Fact 14 follows almost directly from the definition of F_β. Facts 15 and 16 can be routinely verified by calculus.

The following is the main result of this section (cf. (6)):

▶ **Theorem 17** (Fooling smooth test functions of max of non-centered Gaussians). *Let $W^1, \dots, W^k$ be unit vectors in $\mathbb{R}^n$, $V^1, \dots, V^k$ be vectors in $\mathbb{R}^d$ satisfying (8) and let $\vec{\theta} \in \mathbb{R}^k$. Fix any function $g \in C^2(\mathbb{R})$, $g : \mathbb{R} \to [-1, 1]$ such that $\|g'\|_\infty := \sup_{x \in \mathbb{R}} |g'(x)| < \infty$ and $\|g''\|_\infty := \sup_{x \in \mathbb{R}} |g''(x)| < \infty$. Then for any $\beta > 0$, we have*

$$\Big| \boldsymbol{E}[g(F_\beta(W^1 \cdot \mathbf{G}^{(n)} - \theta_1, \dots, W^k \cdot \mathbf{G}^{(n)} - \theta_k))] - \\ \boldsymbol{E}[g(F_\beta(V^1 \cdot \mathbf{G}^{(d)} - \theta_1, \dots, V^k \cdot \mathbf{G}^{(d)} - \theta_k))] \Big| \le O(\|g''\|_\infty \varepsilon' + \|g'\|_\infty \varepsilon' \beta).$$

Further,

$$\left| \boldsymbol{E}[g(\max_{j \in [k]}(W^j \cdot \mathbf{G}^{(n)} - \theta_j))] - \boldsymbol{E}[g(\max_{j \in [k]}(V^j \cdot \mathbf{G}^{(d)} - \theta_j))] \right| \le O(\|g''\|_\infty \varepsilon' + \|g'\|_\infty \sqrt{\varepsilon' \ln k}).$$

We use the rest of this subsection to prove Theorem 17. The proof extends the proofs of similar results in [4, 6] to the case of non-centered Gaussians.

For ease of presentation, for $i \in [k]$ define the non-centered Gaussian random variables $\mathbf{X}_i := W^i \cdot \mathbf{G}^{(n)} - \theta_i$ and $\mathbf{Y}_i := V^i \cdot \mathbf{G}^{(d)} - \theta_i$. We may suppose, without loss of generality, that $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_k)$ and $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_k)$ are defined over the same probability space and that $\mathbf{X}$ and $\mathbf{Y}$ are independent of each other. Our goal is to bound the magnitude of the difference

$$\mathbf{E}[g(F_\beta(\mathbf{X}_1, \dots, \mathbf{X}_k))] - \mathbf{E}[g(F_\beta(\mathbf{Y}_1, \dots, \mathbf{Y}_k))]. \tag{22}$$

Let μ_i denote $\mathbf{E}[\mathbf{X}_i] = \mathbf{E}[\mathbf{Y}_i]$, and let $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \mu_i$ be the centered version of $\mathbf{X}_i$ and similarly let $\widetilde{\mathbf{Y}_i} = \mathbf{Y}_i - \mu_i$. Observe that by independence we have $\mathbf{E}[\mathbf{X}_i \mathbf{Y}_j] = 0$ for all $i, j \in [k]$. Now, as is standard, we do a Slepian interpolation; so for $t \in [0, 1]$, we define

$\mathbf{Z}_{t,i} := \sqrt{t}\widetilde{\mathbf{X}}_i + \sqrt{1-t}\widetilde{\mathbf{Y}}_i + \mu_i$, and we write $\mathbf{Z}_t$ to denote $(\mathbf{Z}_{t,1}, \dots, \mathbf{Z}_{t,k})$. We define the function

$$\Psi(t) = \mathbf{E}[g(F_\beta(\mathbf{Z}_{t,1}, \dots, \mathbf{Z}_{t,k}))],$$

and we observe that

$$(22) = \Psi(1) - \Psi(0) = \int_0^1 \Psi'(t)dt. \tag{23}$$

Thus to upper bound the magnitude of (22) it suffices to upper bound $\int_0^1 |\Psi'(t)|dt$.

For $x \in \mathbb{R}^k$ let us write $m(x)$ to denote $g(F_\beta(x))$. By applying the chain rule, we have

$$\Psi'(t) = \frac{1}{2}\sum_{i=1}^{k} \mathbf{E}\left[\partial_i m(\mathbf{Z}_t) \cdot \left(\frac{\widetilde{\mathbf{X}_i}}{\sqrt{t}} - \frac{\widetilde{\mathbf{Y}_i}}{\sqrt{1-t}}\right)\right].$$

Now we recall the following "integration by parts" lemma, which is sometimes referred to as "Stein's identity:"

▶ **Lemma 18** (Lemma 2 of [6], see also Lemma 2.1 of [4]). *Let $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_p)$ be a p-dimensional Gaussian random vector with mean zero and let $f : \mathbb{R}^p \to \mathbb{R}$ be a C^1 function with $\boldsymbol{E}[|\partial_\ell f(\mathbf{A})|] < \infty$ for all $\ell \in [p]$. Then for each $\ell \in [p]$, we have*

$$\boldsymbol{E}[\mathbf{A}_\ell f(\mathbf{A})] = \sum_{j=1}^{p} \boldsymbol{E}[\mathbf{A}_\ell \mathbf{A}_j]\, \boldsymbol{E}[\partial_j f(\mathbf{A})].$$

We now set (i) $p = k+1$, (ii) $\mathbf{A}_j = \mathbf{Z}_{t,j}$ (for $1 \le j \le k$), (iii) $\mathbf{A}_{k+1} = \frac{\tilde{\mathbf{X}}_i}{\sqrt{t}} - \frac{\tilde{\mathbf{Y}}_i}{\sqrt{1-t}}$ and (iv) $f(\mathbf{A}) = \partial_i m(\mathbf{Z}_t)$. Observe that with this setting, $\partial_{k+1} f(\mathbf{A}) = 0$. Applying Lemma 18 with $\ell = k+1$, we get that

$$\begin{aligned}\Psi'(t) &= \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \mathbf{E}\left[\left(\frac{\tilde{\mathbf{X}}_i}{\sqrt{t}} - \frac{\tilde{\mathbf{Y}}_i}{\sqrt{1-t}}\right)\left(\sqrt{t}\tilde{\mathbf{X}}_j - \sqrt{1-t}\tilde{\mathbf{Y}}_j\right)\right] \mathbf{E}[\partial_{i,j} m(\mathbf{Z}_t)] \\ &= \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} (\sigma^W_{i,j} - \sigma^V_{i,j}) \cdot \mathbf{E}[\partial_{i,j} m(\mathbf{Z}_t)],\end{aligned}$$

where the second equality uses the independence between $\mathbf{X}$ and $\mathbf{Y}$. We get that

$$\begin{aligned}\int_{t=0}^{1} |\Psi'(t)|dt &\le \frac{1}{2}\int_{t=0}^{1} \sum_{i,j=1}^{k} |\sigma^W_{i,j} - \sigma^V_{i,j}| \cdot |\mathbf{E}[\partial_{i,j} m(\mathbf{Z}_t)]|\, dt \\ &\le \frac{\Delta}{2} \cdot \int_{t=0}^{1} \sum_{i,j=1}^{k} |\mathbf{E}[\partial_{i,j} m(\mathbf{Z}_t)]|\, dt, \end{aligned} \tag{24}$$

where $\Delta = \max_{i,j \in [k]} |\sigma^W_{i,j} - \sigma^V_{i,j}|$ is the quantity defined in (9). Thus, we are left with the task of upper bounding the double derivatives. We have

$$\partial_i m(x) = \partial_i (g(F_\beta(x_1, \dots, x_k))) = g'(F_\beta(x_1, \dots, x_k)) \cdot \frac{\partial F_\beta}{\partial x_i}$$

and hence

$$\partial_{i,j} m(x) = \partial_{i,j}(g(F_\beta(x_1, \dots, x_k))) = g''(F_\beta(x_1, \dots, x_k)) \cdot \frac{\partial F_\beta}{\partial x_i}\frac{\partial F_\beta}{\partial x_j} + g'(F_\beta(x_1, \dots, x_k)) \cdot \frac{\partial^2 F_\beta}{\partial x_i \partial x_j}.$$

Applying Facts 15 and 16, it follows that

$$\sum_{i,j=1}^{k} |\mathbf{E}[\partial_{i,j} m(\mathbf{Z}_t)]| = O(\|g''\|_\infty + \|g'\|_\infty \cdot \beta).$$

Hence combining (23), (24), and the above, and recalling that $\Delta \leq 9\varepsilon'$ (see Observation 6), we get that

$$|\mathbf{E}[g(F_\beta(\mathbf{X}_1, \ldots, \mathbf{X}_k))] - \mathbf{E}[g(F_\beta(\mathbf{Y}_1, \ldots, \mathbf{Y}_k))]| \leq O(||g''||_\infty \cdot \varepsilon' + ||g'||_\infty \cdot \varepsilon' \cdot \beta),$$

giving the first claim of the theorem. For the second claim, using Fact 14, it follows that

$$\begin{aligned} |\,\mathbf{E}[g(\max_{j\in[k]}(\mathbf{X}_j))] - \mathbf{E}[g(\max_{j\in[k]}(\mathbf{X}_j))]| &\leq O(||g''||_\infty \cdot \varepsilon' + ||g'||_\infty \cdot \varepsilon' \cdot \beta) + ||g'||_\infty \cdot \frac{\ln k}{\beta} \\ &\leq O\Big(||g'||_\infty \cdot \big(\varepsilon' \cdot \beta + (\ln k)/\beta\big) + ||g''||_\infty \cdot \varepsilon'\Big). \end{aligned}$$

The second claim of the theorem now follows by setting $\beta = \sqrt{(\ln k)/\varepsilon'}$.

6.3 Formalizing step (3') of the intuitive sketch: anticoncentration of max of non-centered Gaussians

We recall the following useful anticoncentration result from [16], which follows almost directly from a result of Nazarov [27]:

▶ **Lemma 19** (Lemma 3.4 of [16]: anticoncentration of multidimensional Gaussian). *Let $W^1, \ldots, W^k$ be unit vectors in $\mathbb{R}^n$. For all $\vec{\theta} \in \mathbb{R}^k$ and all $\lambda > 0$, we have*

$$\mathop{\boldsymbol{Pr}}_{\mathbf{G}\leftarrow\mathcal{N}(0,1)^n}\big[W\mathbf{G} \in \mathrm{Strip}_{\lambda,k,\vec{\theta}}\big] = O(\lambda\sqrt{\log k}). \tag{25}$$

This can be viewed as a k-dimensional analogue of Theorem 3 from [6], which gives an anticoncentration bound on $\max\{W^1 \cdot \mathbf{G}, \cdots W^k \cdot \mathbf{G}\}$ (and also the above lemma is for non-centered Gaussians, whereas Theorem 3 of [6] is about centered Gaussians). As an immediate consequence of Lemma 19 we obtain the following:

▶ **Theorem 20** (anticoncentration of max of non-centered Gaussians). *Fix any $\vec{\theta} \in \mathbb{R}^k$. For all $\lambda > 0$ and all $t \in \mathbb{R}$ it holds that*

$$\boldsymbol{Pr}[\max_{j\in[k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j) \in [t-\lambda, t]] = O(\lambda\sqrt{\log k}).$$

6.4 Formalizing step (4') of the intuitive sketch: Passing from a smooth approximator of sign(·) to sign(·)

In this section we prove the following theorem, which upper bounds the Kolmogorov distance between the random variables $\max_{j\in[k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j)$ and $\max_{j\in[k]} (V^j \cdot \mathbf{G}^{(d)} - \theta_j)$:

▶ **Theorem 21.** *Let $W^1, \ldots, W^k$ be unit vectors in $\mathbb{R}^n$, $V^1, \ldots, V^k$ be vectors in $\mathbb{R}^d$ satisfying (8). For all $\vec{\theta} \in \mathbb{R}^k$, the following bound holds:*

$$d_{\mathrm{K}}\left(\max_{j\in[k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j), \max_{j\in[k]} (V^j \cdot \mathbf{G}^{(d)} - \theta_j)\right) \leq O(\varepsilon' \log^2 k)^{1/3}.$$

This is equivalent to showing that for all $\vec{\theta} \in \mathbb{R}^k$ and all $t \in \mathbb{R}$, we have

$$|\,\mathbf{Pr}[\max_{j\in[k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j) \le t] - \mathbf{Pr}[\max_{j\in[k]} (V^j \cdot \mathbf{G}^{(d)} - \theta_j) \le t]| \le O(\varepsilon' \log^2 k)^{1/3}. \quad (26)$$

Our argument follows the proof of Theorem 2 in [6]; the main idea is to combine Theorem 17, where g is a smooth approximation of the sign function, with Theorem 20, which establishes anticoncentration of the max of non-centered Gaussians. The particular $g \in C^2(\mathbb{R}), g : \mathbb{R} \to [-1, 1]$ which we use is the following smooth approximator of the sign function:

$$g(z) = \begin{cases} -1 & z \le -1 \\ -60 \int_{(z+1)/2}^{1} s^2(1-s)^2 ds + 1 & -1 < z < 1 \\ 1 & z \ge 1. \end{cases}$$

Given parameters $x \in \mathbb{R}, \beta > 0$, and $\delta > 0$, define the function $g_{x,\beta,\delta}(z) = g((z - x - e_\beta)/\delta)$. We record a simple claim that can be verified by direct calculation:

▷ Claim 22. For any $x \in \mathbb{R}$, $\beta > 0$ and $\delta > 0$, the following hold:

1. $||g'_{x,\beta,\delta}||_\infty = ||g'||_\infty/\delta \le O(1/\delta)$,
2. $||g''_{x,\beta,\delta}||_\infty = ||g'||_\infty/\delta^2 \le O(1/\delta^2)$,
3. $\mathbb{1}(z \le x + e_\beta) \le g_{x,\beta,\delta}(z) \le \mathbb{1}(z \le x + e_\beta + \delta)$, for all $z \in \mathbb{R}$.

We now proceed to prove (26). As before, for ease of presentation define the random variables $\mathbf{X}_i = W^i \cdot \mathbf{G}^{(n)} - \theta_i$ and $\mathbf{Y}_i = V^i \cdot \mathbf{G}^{(d)} - \theta_i$, $i \in [k]$.

For arbitrary $x \in \mathbb{R}, \beta > 0$, and $\delta > 0$, we have

$$\begin{aligned}
\mathbf{Pr}\left[\max_{j\in[k]} \mathbf{X}_j \le x\right] &\le \mathbf{Pr}[F_\beta(\mathbf{X}) \le x + e_\beta] && \text{(Claim 14)}\\
&\le \mathbf{E}[g_{x,\beta,\delta}(F_\beta(\mathbf{X}))] && \text{(Claim 22)}\\
&\le \mathbf{E}[g_{x,\beta,\delta}(F_\beta(\mathbf{Y}))] + O\Big(||g''||_\infty \cdot \frac{\varepsilon'}{\delta^2} + ||g'||_\infty \cdot \frac{\varepsilon'\beta}{\delta}\Big) && \text{(Theorem 17, Claim 22)}\\
&\le \mathbf{Pr}[F_\beta(\mathbf{Y}) \le x + e_\beta + \delta] + O\Big(\frac{\varepsilon'}{\delta^2} + \frac{\varepsilon'\beta}{\delta}\Big) && \text{(Claim 22)}\\
&\le \mathbf{Pr}[\max_{j\in[k]} \mathbf{Y}_j \le x + e_\beta + \delta] + e_\beta + O\Big(\frac{\varepsilon'}{\delta^2} + \frac{\varepsilon'\beta}{\delta}\Big) && \text{(Claim 14)}\\
&= \mathbf{Pr}[\max_{j\in[k]} \mathbf{Y}_j \le x] + (\mathbf{Pr}[\max_{j\in[k]} \mathbf{Y}_j \le x + e_\beta + \delta] - \mathbf{Pr}[\max_{j\in[k]} \mathbf{Y}_j \le x])+\\
&\quad e_\beta + O\Big(\frac{\varepsilon'}{\delta^2} + \frac{\varepsilon'\beta}{\delta}\Big)\\
&\le \mathbf{Pr}[\max_{j\in[k]} \mathbf{Y}_j \le x] + O((e_\beta + \delta)\sqrt{\log k}) + e_\beta + O\Big(\frac{\varepsilon'}{\delta^2} + \frac{\varepsilon'\beta}{\delta}\Big) && \text{(Theorem 20)}
\end{aligned}$$

Setting $\beta = (\log k)/\delta$ and $\delta = O(\varepsilon'\sqrt{\log k})^{1/3}$ completes the proof of (26).

6.5 Formalizing step (5') of the intuitive sketch: Re-interpreting the Kolmogorov distance bound as a PRG

We conclude the proof of our PRG construction from the bound proved in Theorem 21; recall that this gives CDF-closeness at every point in $\mathbb{R}$, specifically

$$d_{\mathrm{K}}(\max_{j\in[k]} (W^j \cdot \mathbf{G}^{(n)} - \theta_j), \max_{j\in[k]} (V^j \cdot \mathbf{G}^{(d)} - \theta_j)) \le O(\varepsilon' \log^2 k)^{1/3}$$

Specializing this to CDF-closeness at the point 0, we get that

$$\left|\mathbf{Pr}[W^j \cdot \mathbf{G}^{(n)} \le \theta_j \text{ for all } j \in [m]] - \mathbf{Pr}[V^j \cdot \mathbf{G}^{(d)} \le \theta_j \text{ for all } j \in [m]]\right| \le O(\varepsilon' \log^2 k)^{1/3}.$$

Now we recall that, from Section 4.2, all but a $\delta' = \varepsilon$ fraction of outcomes $V^1, \dots, V^k$ of $\mathbf{V}^j = W^j \mathbf{A}^\mathsf{T}$ satisfy (8). Hence we have

$$\left|\mathbf{Pr}[W^j \cdot \mathbf{G}^{(n)} \leq \theta_j \text{ for all } j \in [m]] - \mathbf{Pr}[W^j \mathbf{A}^\mathsf{T} \cdot \mathbf{G}^{(d)} \leq \theta_j \text{ for all } j \in [m]]\right| \leq O(\varepsilon' \log^2 k)^{1/3} + \varepsilon,$$

and recalling that a draw $\mathbf{Z}$ from our generator Gen is $\mathbf{Z} = \mathbf{A}^\mathsf{T}\mathbf{G}^{(d)}$, we get that this is equivalent to

$$\left|\mathbf{Pr}[W^j \cdot \mathbf{G}^{(n)} \leq \theta_j \text{ for all } j \in [m]] - \mathbf{Pr}[W^j \cdot \mathbf{Z} \leq \theta_j \text{ for all } j \in [m]]\right| \leq O(\varepsilon' \log^2 k)^{1/3} + \varepsilon.$$

Setting $\varepsilon' = \varepsilon^3 / \log^2 k$ completes the proof of correctness of our PRG construction.

7 Application of our PRG: Deterministic approximate counting for functions of LTFs over $\{-1,1\}^n$

In this section we prove Theorems 4 and 5, which we state with precise bounds as two parts of the following theorem.

▶ **Theorem 23** (Restatements of Theorem 4 and 5).

1. (Arbitrary functions of LTFs). *There is a deterministic algorithm which, given as input k LTFs $h_1, \dots, h_k$ over $\{-1,1\}^n$, an explicit function $g : \{-1,1\}^k \to \{-1,1\}$, and an error parameter $\varepsilon > 0$, runs in* $\mathrm{poly}(n) \cdot 2^{\tilde{O}(\frac{k^{15}}{\varepsilon^{12}})}$ *time and outputs a value $\tilde{v} \in [0,1]$ such that $|\tilde{v} - v| \leq \varepsilon$, where v is the fraction of points in $\{-1,1\}^n$ that satisfy $g(h_1, \dots, h_k)$.*
2. (Intersections of LTFs). *There is a deterministic algorithm which, given as input k LTFs $h_1, \dots, h_k$ over $\{-1,1\}^n$ and an error parameter $\varepsilon > 0$, runs in* $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(\log k, 1/\varepsilon)}$ *time and outputs a value $\tilde{v} \in [0,1]$ such that $|\tilde{v} - v| \leq \varepsilon$, where v is the fraction of points in $\{-1,1\}^n$ that satisfy $h_1(x) \wedge \cdots \wedge h_k(x)$.*

We prove Part 1 first since it is simpler and relies on (extensions of) known tools such as regularity lemmas and invariance principles. In particular, Part 1 requires an invariance principle for arbitrary functions of LTFs. Such an invariance principle was proved in [14]; we provide an alternate proof of the invariance principle that we require in Appendix C, which we believe could be of independent interest. For Part 2, the main ingredients are an invariance principle of [16] for intersections of LTFs and a "multi-regularity lemma" for k-tuples of LTFs due to [14] along with a subtle application of the PRG for intersections of LTFs due to [28].

7.1 A useful notion: Regularity

Given an LTF $h(x) = \mathrm{sign}(w_1 x_1 + \cdots + w_n x_n - \theta)$ and a parameter $0 < \tau < 1$, we say that h is *τ-regular* if

$$\sum_{j=1}^{n} w_j^4 \leq \tau^2 \cdot (\sum_{j=1}^{n} w_j^2)^2.$$

Intuitively, τ-regularity (when τ is small) captures the property that no weight in $w_1, \dots, w_n$ has magnitude which is large relative to "the overall scale of the weights." Regularity is a useful condition because if w is a τ-regular weight vector with two-norm 1, then by the Berry-Esseen theorem [1, 10] the CDF of the real random variable $w \cdot \mathbf{X}$ (where $\mathbf{X}$ is uniform over $\{-1,1\}^n$) is τ-close to the CDF of an $\mathcal{N}(0,1)$ Gaussian. Thus the Berry-Esseen theorem implies that regular LTFs will "behave similarly" whether they are given uniform inputs $\mathbf{X} \leftarrow \{-1,1\}^n$ or Gaussian inputs $\mathbf{G} \leftarrow \mathcal{N}(0,1)^n$; in this sense, it can be viewed as an invariance principle for a single LTF.

7.2 Proof of Part 1 of Theorem 23: Arbitrary functions of k LTFs

The first principal ingredient that we use is an *invariance principle for arbitrary functions of LTFs.* As mentioned earlier, such a result was established in [14] via a "Lindeberg-method" type proof. In Appendix C we give an alternate proof (which is very different from the proofs of [14, 16]) of the version that we require, which is stated below:

▶ **Theorem 24** (Invariance principle for arbitrary functions of k LTFs). *Let $h_1, h_2, \ldots, h_k$ be τ-regular LTFs and let $F(x) = g(h_1(x), \cdots, h_k(x))$ where $g : \{-1,1\}^k \to \{-1,1\}$ may be any function. Then*

$$\left| \underset{\mathbf{X}\leftarrow\{-1,1\}^n}{\boldsymbol{Pr}}[F(\mathbf{X}) = -1] - \underset{\mathbf{Z}\leftarrow\mathcal{N}(0,1)^n}{\boldsymbol{Pr}}[F(\mathbf{Z}) = -1] \right| \leq O(k^{3/2}\tau\sqrt{\log(k/\tau)}). \tag{27}$$

Combining Theorem 2 (our PRG for arbitrary functions of LTFs over Gaussian space) and Theorem 24, an algorithm that simply enumerates over all the seeds of our PRG yields the following deterministic approximate counting algorithm for intersections of sufficiently regular LTFs:

▶ **Corollary 25** (Deterministic approximate counting for arbitrary functions of regular LTFs). *There is a deterministic algorithm with the following performance guarantee: Given $\varepsilon > 0$, a collection $h_1, \ldots, h_k$ of LTFs over $\{-1,1\}^n$ each of which is τ-regular where $\tau = O(\frac{\varepsilon}{k^{3/2}\sqrt{(\log k)(\log \frac{k}{\varepsilon})}})$, and a function $g : \{-1,1\}^k \to \{-1,1\}$, the algorithm runs in time $\mathrm{poly}(n) \cdot 2^{\tilde{O}(\frac{k^{15}}{\varepsilon^{12}})}$ and outputs a value $\tilde{v} \in [0,1]$ such that $|\tilde{v} - v| \leq \varepsilon$, where v is the fraction of points in $\{-1,1\}^n$ that satisfy $g(h_1, \ldots, h_k)$.*

We next extend Corollary 25 to obtain a deterministic approximate counting algorithm for arbitrary functions of k *general* LTFs using a slight extension of the "multi-regularity lemma" established in [14] (see Theorem 5.4 of the ArXiV version, available at [15]) for k-tuples of general LTFs.

While not precisely stated in these terms, we recall that this multi-regularity lemma, roughly speaking, asserts the following: Given a k-tuple of LTFs $h_1, \ldots, h_k$, there is a relatively shallow non-adaptive decision tree on the variables such that for all $i \in [k]$, one of the two following two possibilities hold:

1. For every leaf ρ of the decision tree (corresponding to a restriction), the restricted LTF $h_i \upharpoonright \rho$ is regular.
2. With high probability, the restricted LTF $h_i \upharpoonright \rho$ is close to a constant.

Similar to Lemma 18 of [7], the multi-regularity lemma of [14] can be implemented as a deterministic algorithm. In fact, because the decision tree is *non-adaptive*, the set of variables appearing in the internal nodes can be computed in time polynomial in depth of the decision tree (as opposed to exponential in the depth which is the size of the tree). This is because at each node, in order to choose which variable from $x_1, \ldots, x_n$ should be placed at that node it suffices to compute the influence of each variable in each of the k restricted linear forms, and this is a straightforward deterministic computation. We remark that the ability to *compute the tree* in polynomial time (in terms of its depth) is not crucial for this subsection. However, it is vital for the application in the next subsection – deterministic counting for intersections of general LTFs. Viewed as an algorithmic procedure from this perspective, Theorem 5.4 of [15] yields the following in our setting:

▶ **Lemma 26** (Algorithmic regularity lemma for LTFs, general k, based on Theorem 5.4 of [15]). *There is an algorithm* `ConstructTree` *with the following properties: Let $h_1, \ldots, h_k$ be LTFs over $\{-1,1\}^n$. Algorithm* `ConstructTree` *(which is deterministic) receives $h_1, \ldots, h_k$ and*

$0 < \tau, \gamma < 1/4$ as input, runs in time $\mathrm{poly}(n, D_k(\tau, \gamma))$ *and outputs a set of variables $S \subseteq [n]$ and a k-tuple of labels* $(\mathrm{label}_1, \dots, \mathrm{label}_k) \in \{R, J\}^k$ *such that the following holds:*

1. *$|S| \leq D_k(\tau, \gamma)$ where*

$$D_k(\tau, \gamma) := k \cdot \frac{1}{\tau} \cdot \mathrm{poly}\left(\log \frac{1}{\gamma}\right).$$

2. *For each leaf ρ and $i \in [k]$, if* $\mathrm{label}_i = R$, *then the LTF $h_i \restriction \rho$ is τ-regular.*
3. *For each $i \in [k]$, if* $\mathrm{label}_i = J$, *then the LTF h'_i obtained by zeroing the coordinates outside S satisfies $\boldsymbol{Pr}_{x \in \{-1,1\}^n}[h_i(x) \neq h'_i(x)] \leq \gamma$. In particular, observe that for any leaf ρ, $h'_i \restriction \rho$ is fixed at either $+1$ or -1.*

▶ Remark 27. The theorem above can be obtained by essentially observing the proof of Theorem 5.4 in [15]. In particular, the S in the above theorem corresponds to the H_0 in their theorem. Similarly, the coordinates $i \in [k]$ which are labeled 'R' (resp. labeled 'J') in our theorem correspond exactly to the coordinates $i \in [d]$ which fall in the first case (resp. second case) of Theorem 5.4 in [15]. To get the guarantee for the third case, we define h'_i as follows. Let $h_i(x) = \mathrm{sign}(\sum_j w_{i,j} x_j - \theta_j)$. We then define $h'_i(x) = \mathrm{sign}(\sum_{j \in S} w_{i,j} x_j - \theta_j)$, i.e., simply erase the coordinates outside of S. The upper bound on the quantity $\mathbf{Pr}_{x \in \{-1,1\}^n}[h_i(x) \neq h'_i(x)]$ can essentially be derived from the event whose probability is upper bound in the centered equation in item (2) of Theorem 5.4 of [15].

We now extend the algorithm in Corollary 29 to handle arbitrary functions of k general LTFs using the algorithmic regularity lemma for multiple LTFs given in Lemma 26. The parameter"δ" in Lemma 26 is set to ε and the parameter "γ" is set to ε/k, and the parameter "τ" is set to $O(\frac{\varepsilon}{k^{3/2}\sqrt{(\log k)(\log \frac{k}{\varepsilon})}})$ so that Corollary 25 can be applied. Constructing the decision tree in the first step of the algorithm for general LTFs takes time $\mathrm{poly}(n, D_k(\tau, \varepsilon, \delta)) = \mathrm{poly}(n, k, 1/\varepsilon)$. In the second step of the algorithm for general LTFs, for each leaf ρ in the decision tree,

- If any of the k labels are "fail" the contribution from that leaf is 0;
- If all k labels are bits $b_1, \dots, b_k \in \{-1, 1\}$, then the contribution from that leaf is $2^{-D_k} \cdot \mathbb{1}[g(b_1, \dots, b_k) = -1]$;
- If $k - t$ of the labels (for notational convenience, say these are the ones corresponding to $h_{t+1}, \dots, h_k$) are bits $b_{t+1}, \dots, b_k$ and the remaining t labels (say the ones corresponding to $h_1 \restriction \rho, \dots, h_t \restriction \rho$) are "regular," we run the approximate counting algorithm for the regular case from Corollary 25 to compute an $\pm\varepsilon$-accurate estimate (call it v_ρ) of the fraction of satisfying assignments of $g((h_1 \restriction \rho) \wedge \cdots \wedge (h_t \restriction \rho), b_{t+1}, \dots, b_k)$, and the contribution from that leaf is $2^{-D_k} \cdot v_\rho$.

The overall running time for the algorithm is at most $\mathrm{poly}(n)\cdot$ (number of leaves) $\cdot$ (running time of Corollary 25), which is $\mathrm{poly}(n) \cdot 2^{\tilde{O}(k^{3/2}/\varepsilon) + \tilde{O}(k^{15}/\varepsilon^{12})}$. To establish correctness, we observe that the final value $\tilde{v}$ may be viewed as a sum of contributions across all the leaves. Property 3 of Lemma 26 and the setting of $\delta = \varepsilon$ in Step 1 ensures that leaves that have any "fail" label contribute a total of $O(\varepsilon)$ to the error $|v - \tilde{v}|$. The setting of the γ parameter to be ε/k ensures that leaves containing any $+1$ label, or having all -1's as their labels, collectively contribute a total of at most $O(\varepsilon)$ to $|v - \tilde{v}|$. Finally, Theorem 2 ensures that leaves as in the last bullet above contribute a total of $O(\varepsilon)$ to $|v - \tilde{v}|$. This concludes the proof of Theorem 23.

7.3 Proof of Part 2 of Theorem 23: Intersections of k LTFs

We begin by recalling the main structural result of [16], which extends the Berry-Esseen theorem to *intersections* of LTFs (also known as polytopes). (Recall that we view -1 as "true" and $+1$ as "false.")

▶ **Theorem 28** (Theorem 3.1 of [16]: invariance principle for polytopes)**.** *Let $h_1, h_2, \ldots, h_k$ be τ-regular LTFs and let $F(x) = h_1(x) \wedge \cdots \wedge h_k(x)$. Then*

$$\left| \Pr_{\mathbf{U}^{(n)} \leftarrow \{-1,1\}^n}[F(\mathbf{U}^{(n)}) = -1] - \Pr_{\mathbf{G}^{(n)} \leftarrow \mathcal{N}(0,1)^n}[F(\mathbf{G}^{(n)}) = -1] \right| \leq C(\log k)^{8/5}(\tau \log(1/\tau))^{1/5}$$

where C is an absolute constant.

Combining Theorem 3 (our PRG for intersections of LTFs over Gaussian space) and Theorem 28, an algorithm that simply enumerates over all the seeds of our PRG yields the following deterministic approximate counting algorithm for intersections of sufficiently regular LTFs:

▶ **Corollary 29** (Deterministic approximate counting for intersections of regular LTFs)**.** *There is a deterministic algorithm with the following performance guarantee: Given $\varepsilon > 0$ and a collection $h_1, \ldots, h_k$ of LTFs over $\{-1,1\}^n$, each of which is τ-regular where $\tau = O(\frac{\varepsilon^5}{\log^8(k) \cdot \log(\frac{\log k}{\varepsilon})})$, the algorithm runs in time $\mathrm{poly}(n) \cdot 2^{\tilde{O}(\frac{\log^6 k}{\varepsilon^6})}$ and outputs a value $\tilde{v} \in [0,1]$ such that $|\tilde{v} - v| \leq \varepsilon$, where v is the fraction of points in $\{-1,1\}^n$ that satisfy $h_1 \wedge \cdots \wedge h_k$.*

The above algorithm works only for intersections of sufficiently regular LTFs. We will now extend Corollary 29 to obtain a deterministic approximate counting algorithm for intersections of k *general* LTFs using two tools. The first is the multi-regularity lemma (Lemma 26) from the previous subsection. The second ingredient we require is the recent construction of a PRG for intersection of LTFs by O'Donnell, Servedio and Tan [28] where they construct a PRG for the uniform distribution on $\{-1,1\}^n$ which fools intersections of k LTFs with seed length $(\log n) \cdot \mathrm{poly}(\log k, 1/\varepsilon)$. More precisely, we have the following theorem from [28].

▶ **Theorem 30.** *There is an efficiently computable ε-PRG $\mathcal{G}_{\mathsf{OST}} : \{-1,1\}^s \to \{-1,1\}^n$ for intersections of k LTFs over $\{-1,1\}^n$ with $s = (\log n) \cdot \mathrm{poly}(\log k, 1/\varepsilon)$.*

Observe that while $\mathcal{G}_{\mathsf{OST}}$ simultaneously achieves polylogarithmic dependence on both n and k, to get a deterministic approximate counting algorithm with the kind of guarantee we want, we would need a seed length of the form $\log n + \mathrm{poly}(\log k, 1/\varepsilon)$. While we will crucially use $\mathcal{G}_{\mathsf{OST}}$, we will essentially bootstrap it with the algorithms from Corollary 29 and Lemma 26 as follows.

Given as input $h_1, \ldots, h_k$ and a desired accuracy parameter ε, the algorithm proceeds as follows:

1. Run the algorithm `ConstructTree` on the LTFs $h_1, \ldots, h_k$ with its "γ" parameter as $\varepsilon/4k$ and "τ" parameter as $O(\frac{\varepsilon^5}{\log^8(k) \cdot \log(\frac{\log k}{\varepsilon})})$.
2. Let S be the set of variables returned by the algorithm `ConstructTree`. We set $n_{\mathsf{OST}} = |S|$, $\varepsilon_{\mathsf{OST}} = \varepsilon/4$ and $k_{\mathsf{OST}} = k$.
3. Let us run $\mathcal{G}_{\mathsf{OST}}$ with parameters n_{OST}, $\varepsilon_{\mathsf{OST}}$ and k_{OST}. Let s_{OST} be the seed length. Let $\mathcal{O}_{\mathsf{OST}} \subseteq \{-1,1\}^S$ denote the range of $\mathcal{G}_{\mathsf{OST}}$. We treat each $\rho \in \mathcal{O}_{\mathsf{OST}}$ as an assignment for the coordinates in S.

4. For each $\rho \in \mathcal{O}_{\mathsf{OST}}$, compute v_ρ as follows: If there is any $i \in [k]$ such that $\text{label}_i = J$ and $h'_i \upharpoonright \rho = -1$, then set $v_\rho = 0$. Otherwise, observe that for all $i \in [k]$ such that $\text{label}_i = R$, $h_i \upharpoonright \rho$ is τ-regular (for τ specified earlier). Run the algorithm from Corollary 29 to compute $\mathbf{Pr}[\wedge_{i:\text{label}_i=R}(h_i \upharpoonright \rho)]$. Let the output be v_ρ.
5. Output the value $\mathbf{E}_{\rho\in\mathcal{O}_{\mathsf{OST}}}[v_\rho]$.

The analysis of the running time of the above routine is straightforward: Observe that for our choice of τ and γ, the value $D_k(\tau,\gamma)$ (from Lemma 26) is $\tilde{O}(k\cdot\varepsilon^{-5})$. The running time of the first step, i.e., `ConstructTree` is bounded by $\text{poly}(n, D_k(\tau,\gamma))$. Now, observing that $|S| \le D_k(\tau,\gamma)$, from Theorem 30, we get that $s_{\mathsf{OST}} = \mathsf{poly}(\log k, \varepsilon^{-1})$ and thus $|\mathcal{O}_{\mathsf{OST}}| = 2^{\mathsf{poly}(\log k,\varepsilon^{-1})}$. For each $\rho \in \mathcal{O}_{\mathsf{OST}}$, the running time of the algorithm from Lemma 26 is bounded by $\text{poly}(n)\cdot 2^{\tilde{O}(\log^6 k/\varepsilon^6)}$. Thus, the total running time is $|\mathcal{O}_{\mathsf{OST}}|\cdot\text{poly}(n)\cdot 2^{\tilde{O}(\log^6 k/\varepsilon^6)}$ which is $\text{poly}(n)\cdot 2^{\text{poly}(\log k, 1/\varepsilon)}$.

We now move to the proof of correctness of the algorithm. Observe that if $\text{label}_i = J$ for any $i \in [k]$, then by guarantee of Lemma 26, $\mathbf{Pr}_{x\in\{-1,1\}^n}[h_i(x) \neq h'_i(x)] \le \gamma$. Thus, if we define $\mathcal{A}_J = \{i \in [k] : \text{label}_i = J\}$ and $\mathcal{A}_R = \{i \in [k] : \text{label}_i = R\}$,

$$\big|\mathbf{Pr}_{x\in\{-1,1\}^n}[h_1(x)\wedge\ldots\wedge h_k(x)] - \mathbf{Pr}_{x\in\{-1,1\}^n}[\wedge_{i\in\mathcal{A}_J} h'_i(x)\ \wedge_{i\in\mathcal{A}_R} h_i(x)]\big| \le k\gamma = \frac{\varepsilon}{4}. \quad (28)$$

Now, consider any assignment $z \in \{-1,1\}^{[n]\setminus S}$ of the variables in $[n]\setminus S$. Then, using the guarantee of $\mathcal{G}_{\mathsf{OST}}$, we get

$$\big|\mathbf{Pr}_{x\in\{-1,1\}^S}[\wedge_{i\in\mathcal{A}_J} h'_i \upharpoonright z(x)\ \wedge_{i\in\mathcal{A}_R} h_i \upharpoonright z(x)] - \mathbf{Pr}_{\rho\in\mathcal{O}_{\mathsf{OST}}}[\wedge_{i\in\mathcal{A}_J} h'_i \upharpoonright z(\rho)\ \wedge_{i\in\mathcal{A}_R} h_i \upharpoonright z(\rho)]\big| \le \frac{\varepsilon}{4}.$$

Averaging over all possible values of $z \in \{-1,1\}^{[n]\setminus S}$ and combining with (28), we get

$$\big|\mathbf{Pr}_{x\in\{-1,1\}^n}[h_1(x)\wedge\ldots\wedge h_k(x)] - \mathbf{Pr}_{z\in\{-1,1\}^{[n]\setminus S},\rho\in\mathcal{O}_{\mathsf{OST}}}[\wedge_{i\in\mathcal{A}_J} h'_i(z,\rho)\ \wedge_{i\in\mathcal{A}_R} h_i(z,\rho)]\big| \le \frac{\varepsilon}{2}. \quad (29)$$

Now, observe that for any $\rho \in \mathcal{O}_{\mathsf{OST}}$, $h'_i(z,\rho) = h'_i(\rho)$ (since h'_i does not depend on the variables outside S). Further, for each $i \in \mathcal{A}_R$, the LTF $h_i \upharpoonright \rho$ is τ-regular. Consequently, for each choice of ρ, Step 4 of our routine outputs v_ρ such that

$$\big|\mathbf{Pr}_{z\in\{-1,1\}^{[n]\setminus S}}[\wedge_{i\in\mathcal{A}_J} h'_i(z,\rho)\ \wedge_{i\in\mathcal{A}_R} h_i(z,\rho)] - v_\rho\big| \le \frac{\varepsilon}{2}.$$

Averaging it over all choices of ρ, we get that output in the final step $\mathbf{E}_{\rho\in\mathcal{O}_{\mathsf{OST}}}[v_\rho]$ satisfies

$$\big|\mathbf{E}_{\rho\in\mathcal{O}_{\mathsf{OST}}}[v_\rho] - \mathbf{Pr}_{z\in\{-1,1\}^{[n]\setminus S},\rho\in\mathcal{O}_{\mathsf{OST}}}[\wedge_{i\in\mathcal{A}_J} h'_i(z,\rho)\ \wedge_{i\in\mathcal{A}_R} h_i(z,\rho)]\big| \le \frac{\varepsilon}{2}.$$

Combining this with (29) finishes the proof.

References

1 Andrew C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.

2 R. Bhatia. *Matrix analysis*, volume 169. Springer, 2013.

3 Thomas Bonis. Rates in the Central Limit Theorem and diffusion approximation via Stein's Method. *arXiv preprint*, 2015. `arXiv:1506.06966`.

4 Sourav Chatterjee. An error bound in the Sudakov-Fernique inequality. *arXiv preprint*, 2005. `arXiv:math/0510424`.

5 Xi Chen, Rocco Servedio, and Li-Yang Tan. New algorithms and lower bounds for testing monotonicity. In *Proceedings of the 55th Annual Symposium on Foundations of Computer Science (FOCS 2014)*, pages 286–295, 2014.

6 Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields*, 162(1-2):47–70, 2015.

7 Anindya De, Ilias Diakonikolas, and Rocco A. Servedio. Deterministic approximate counting for juntas of degree-2 polynomial threshold functions. In *Proceedings of the 29th Annual Conference on Computational Complexity (CCC)*, pages 229–240. IEEE, 2014.

8 Ilias Diakonikolas, Parikshit Gopalan, Rajesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded Independence Fools Halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010.

9 Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 11–20. IEEE, 2010.

10 Carl-Gustav Esseen. On the Liapunoff limit of error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, A:1–19, 1942.

11 X. Fernique. Regularité des trajectoires des fonctions aléatoires gaussiennes. In *Ecole d'Eté de Probabilités de Saint-Flour IV—1974*, pages 1–96. Springer, 1975.

12 C. Givens and R. Shortt. A class of Wasserstein metrics for probability distributions. *The Michigan Mathematical Journal*, 31(2):231–240, 1984.

13 Parikshit Gopalan, Daniel Kane, and Raghu Meka. Pseudorandomness via the discrete fourier transform. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 903–922. IEEE, 2015.

14 Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *Computational Complexity (CCC), 2010 IEEE 25th Annual Conference on*, pages 223–234. IEEE, 2010.

15 Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling Functions of Halfspaces under Product Distributions, 2010. `arXiv:1001.1593`.

16 Prahladh Harsha, Adam Klivans, and Raghu Meka. An invariance principle for polytopes. *Journal of the ACM (JACM)*, 59(6):29, 2012.

17 Daniel Kane. A Small PRG for Polynomial Threshold Functions of Gaussians. In *FOCS*, pages 257–266, 2011.

18 Daniel Kane. k-independent Gaussians fool polynomial threshold functions. In *IEEE Conference on Computational Complexity*, pages 252–261, 2011.

19 Daniel Kane. A pseudorandom generator for polynomial threshold functions of Gaussian with subpolynomial seed length. In *Proceedings of the 29th Annual Conference on Computational Complexity (CCC)*, pages 217–228, 2014.

20 Daniel Kane, Raghu Meka, and Jelani Nelson. Almost optimal explicit Johnson-Lindenstrauss families. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 628–639. Springer, 2011.

21 Daniel M. Kane. A Polylogarithmic PRG for Degree 2 Threshold Functions in the Gaussian Setting. In *30th Conference on Computational Complexity, CCC 2015*, pages 567–581, 2015.

22 Daniel M. Kane. A polylogarithmic PRG for degree 2 threshold functions in the Gaussian setting. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 33. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

23 Pravesh K. Kothari and Raghu Meka. Almost Optimal Pseudorandom Generators for Spherical Caps. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 247–256, 2015.

24 H. Landau and L. Shepp. On the supremum of a Gaussian process. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 369–378, 1970.

25 R. Meka. A polynomial time approximation scheme for computing the supremum of Gaussian processes. *Ann. Appl. Probab.*, 25(2):465–476, April 2015.

26 Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM Journal on Computing*, 42(3):1275–1301, 2013.

27 Fedor Nazarov. On the Maximal Perimeter of a Convex Set in $\mathbb{R}^n$ with Respect to a Gaussian Measure. In *Geometric Aspects of Functional Analysis: Israel Seminar 2001-2002*, pages 169–187, 2003.

28 R. O'Donnell, R.A. Servedio, and L.-Y. Tan. Fooling Polytopes. Available at `arXiv:1808.04035`, to appear in STOC 2019, 2018.

29 Rocco A. Servedio and Li-Yang Tan. Deterministic Search for CNF Satisfying Assignments in Almost Polynomial Time. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 813–823. IEEE, 2017.

30 V. Sudakov. *Geometric problems in the theory of infinite-dimensional probability distributions*, volume 141. American Mathematical Soc., 1979.

31 M. Talagrand. Majorizing measures: the generic chaining. *The Annals of Probability*, pages 1049–1103, 1996.

32 Gregory Valiant and Paul Valiant. Estimating the unseen: an $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the 43rd Symposium on Theory of Computing (STOC)*, pages 685–694, 2011.

33 Alex Zhai. A high-dimensional CLT in W2 distance with near optimal convergence rate. *Probability Theory and Related Fields*, 170(3):821–845, April 2018.

A Lower bound on seed length for PRG fooling arbitrary functions of k LTFs

The following simple claim gives an $\Omega(\log n)$ lower bound even for $k = 1$:

▷ Claim 31. Let $\mathcal{G}$ be a 0.49-PRG for the class of all LTFs over Gaussian space $\mathcal{N}(0,1)^n$. Then the seed length of $\mathcal{G}$ is at least $\lfloor \log n \rfloor$.

Proof. Suppose that $\mathcal{G}$ is a generator with seed length $s \leq \lfloor \log n \rfloor - 1$. Let $S = \{v^1, \dots, v^m\} \subset \mathbb{R}^n, |S| \leq n/2$ be the set of all points $\mathcal{G}(\{-1,1\}^s)$. Since $m < n$ there is a unit vector $w \in \mathbb{R}^n$ which hs orthogonal to all of $v^1, \dots, v^m$; fix such a w. Fix any value $\kappa = o_n(1)$. It is easy to see that the LTF $f(x) = \text{sign}(w \cdot x - \kappa)$ has $\mathbf{Pr}_{\mathbf{G}^{(n)} \leftarrow \mathcal{N}(0,1)^n}[f(\mathbf{G}^{(n)}) = 1] = \frac{1}{2} - o_n(1)$, but each of $v^1, \dots, v^m$ has $\text{sign}(w \cdot x - \kappa) = \text{sign}(-\kappa) = -1$, so $\mathbf{Pr}[f(\mathcal{G}(\mathbf{U}^{(s)})) = 1] = 0$. Hence $\mathcal{G}$ cannot be a 0.49-PRG for the class of all LTFs over Gaussian space. ◁

▷ Claim 32. Let $k \leq n$ and let $\mathcal{G}$ be a 0.49-PRG for the class of all functions $g(h_1, \dots, h_k) : \mathbb{R}^n \to \{-1,1\}^n$ where $g : \{-1,1\}^k \to \{-1,1\}$ and each h_i is an LTF. Then the seed length of $\mathcal{G}$ is at least k.

Proof. Suppose that $\mathcal{G}$ is a generator with seed length $s \leq k-1$. Let $S = \{v^1, \dots, v^m\} \subset \mathbb{R}^n, |S| \leq 2^{k-1}$ be the set of all points $\mathcal{G}(\{-1,1\}^s)$. Say that $b \in \{-1,1\}^k$ is *good* if some $j \in [m]$ satisfies $\text{sign}(v_i^j) = b_i$ for all $i \in [k]$ (i.e. b is the sign-pattern of the first k coordinates of some string in S). Let $g : \{-1,1\}^k \to \{-1,1\}$ be any function which outputs -1 on each good string in $\{-1,1\}^k$ and outputs 1 on exactly 2^{k-1} strings in $\{-1,1\}^k$ (such a g must exist since $|S| \leq 2^{k-1}$ and hence there are at most 2^{k-1} good strings in $\{-1,1\}^k$). Let $h_i(x)$ be the LTF $\text{sign}(x_i)$ for each $i \in [k]$. Then for $f(x) = g(h_1(x), \dots, h_k(x))$, we have $\mathbf{Pr}[f(\mathcal{G}(\mathbf{U}^{(s)})) = 1] = 0$ but $\mathbf{Pr}[f(\mathbf{G}^{(n)}) = 1] = 1/2$. Hence $\mathcal{G}$ cannot be a 0.49-PRG for the class of all functions of k LTFs over Gaussian space. ◁

B Simulating draws from the Gaussian distribution

In the analysis of our PRGs for arbitrary functions of k LTFs and for intersections of k LTFs, we assumed that we can sample from d-dimensional Gaussians, but to do this with perfect fidelity clearly requires infinitely many random bits. In this section we show that $O(d\log(kd/\varepsilon))$ truly random bits suffice to produce d-dimensional "approximate Gaussian" distributions that suffice for our applications.

▶ **Definition 33.** *We say that a random variable $\mathbf{G}'$ on $\mathbb{R}$ is a δ-approximate Gaussian random variable if there is a standard (correlated) Gaussian $\hat{\mathbf{G}}$ such that $\mathbf{Pr}[|\mathbf{G}' - \hat{\mathbf{G}}| > \delta] < \delta$.*

We recall a lemma proved by Kane [22] which generates such approximate Gaussians in a randomness efficient way. It is based on the Box-Muller transform.

▶ **Lemma 34** ([22]). *There is an explicit construction of a δ-approximate Gaussian random variable using $O(\log(1/\delta))$ bits of randomness.*

Let $\mathbf{G}^d$ be a $\mathcal{N}(0,1)^d$ Gaussian. Let $\hat{\mathbf{G}}^{(d)}$ denote a coordinate-wise independent distribution in which the i-th coordinate $\hat{\mathbf{G}}_i^{(d)}$ is a δ-approximate Gaussian random variable with respect to $\mathbf{G}_i^{(d)}$ as given by Lemma 34. We set (with foresight) the parameter $\delta = \varepsilon/(k\sqrt{d})$. By Lemma 34, a draw of $\hat{\mathbf{G}}^{(d)}$ can be generated using $O(d\log(kd/\varepsilon))$ bits of randomness. Below we prove that $\hat{\mathbf{G}}^{(d)}$ can be used instead of $\mathbf{G}^d$ in our PRGs, at the cost of an additional additive ε error for our PRG.

Let $\mathbf{X} = V\mathbf{G}^{(d)} - \vec{\theta}$ and $\hat{\mathbf{X}} = V\hat{\mathbf{G}}^{(d)} - \theta$. We prove that the "union-of-orthants" distance $d_{\mathrm{UO}}(\mathbf{X}, \hat{\mathbf{X}})$ between $\mathbf{X}$ and $\hat{\mathbf{X}}$ (see (18)) is at most ε. This directly implies that the approximation works since, as observed in Section 5.3, for any function $g : \{-1,1\}^k \to \{-1,1\}$, we have

$$\left|\mathbf{Pr}[g(\mathrm{sign}(\mathbf{X}_1), \ldots, \mathrm{sign}(\mathbf{X}_k)) = 1] - \mathbf{Pr}[g(\mathrm{sign}(\hat{\mathbf{X}}_1), \ldots, \mathrm{sign}(\hat{\mathbf{X}}_k)) = 1]\right| \leq d_{\mathrm{UO}}(\mathbf{X}, \hat{\mathbf{X}}).$$

In order to prove that $d_{\mathrm{UO}}(\mathbf{X}, \hat{\mathbf{X}}) \leq \varepsilon$, we recall some definitions from Section 5.3. Recall that

$$B_r := \left\{x \in \mathbb{R}^k : |x_i| \leq r \text{ for some } i \in [k]\right\}$$

is the region of all points in $\mathbb{R}^k$ whose L^∞-distance from any orthant boundary point is at most r. Set $r = 2\delta\sqrt{d}$. For any union of orthants $\mathcal{O}$, we partition $\mathcal{O}$ into $\mathcal{O}_{\mathrm{bd}} := \mathcal{O} \cap B_r$ (the points in $\mathcal{O}$ that lie close to the orthant boundaries) and $\mathcal{O}_{\mathrm{in}} := \mathcal{O} \setminus B_r$ (the points in $\mathcal{O}$ that lie far away from the orthant boundaries).

We have

$$|\,\mathbf{Pr}[\mathbf{X} \in \mathcal{O}] - \mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}]| \leq \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{bd}] + |\,\mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{in}] - \mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}]|.$$

By Lemma 34 and a union bound, it follows that with probability at least $1 - k\delta$, $|\mathbf{G}_i^{(d)} - \hat{\mathbf{G}}_i^{(d)}| \leq \delta$ for each $i \in [k]$. Thus, with probability at least $1 - k\delta$, for each $i \in [k]$, we have

$$|\mathbf{X}_i - \hat{\mathbf{X}}_i| = V^i \cdot (\mathbf{G}_i^{(d)} - \hat{\mathbf{G}}_i^{(d)}) \leq ||V^i||_2||\mathbf{G}_i^{(d)} - \hat{\mathbf{G}}_i^{(d)}||_2 \leq \delta\sqrt{d}.$$

As a direct consequence, we have that $\mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}|\mathbf{X} \notin \mathcal{O}_{in}]| \leq k\delta$ and $\mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}|\mathbf{X} \in \mathcal{O}_{in}] \geq 1 - k\delta$. Thus,

$$\begin{aligned}\mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}] &\leq \mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}|\mathbf{X} \in \mathcal{O}_{in}] \cdot \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{in}] + \mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}|\mathbf{X} \notin \mathcal{O}_{in}] \cdot \mathbf{Pr}[\mathbf{X} \notin \mathcal{O}_{in}] \\ &\leq \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{in}] + k\delta,\end{aligned}$$

and

$$\begin{aligned}\mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}] &\geq \mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O} | \mathbf{X} \in \mathcal{O}_{in}] \cdot \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{in}] \\ &\geq (1 - k\delta) \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{in}] \geq \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{in}] - k\delta.\end{aligned}$$

Hence, $|\mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{in}] - \mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}]| \leq k\delta$.
Finally note that, as estimated in Section 5.3, using anti-concentration of Gaussians,

$$\mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{bd}] \leq O(kr).$$

Combining the above estimates, we have

$$\begin{aligned}|\mathbf{Pr}[\mathbf{X} \in \mathcal{O}] - \mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}]| &\leq \mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{bd}] + |\mathbf{Pr}[\mathbf{X} \in \mathcal{O}_{in}] - \mathbf{Pr}[\hat{\mathbf{X}} \in \mathcal{O}]| \\ &\leq O(k\delta\sqrt{d}) = O(\varepsilon),\end{aligned}$$

which concludes our proof.

C Proof of Theorem 24: An invariance principle for arbitrary functions of LTFs

C.1 Our starting point: a Wasserstein distance bound

Our proof of Theorem 24 closely parallels the arguments underlying our PRG for arbitrary functions of k LTFs that were given in Section 5. However, for technical reasons we will now be using the (non-quadratic) Wasserstein distance. We recall the definition of this distance measure between distributions that we will use. (As was the case earlier for quadratic Wasserstein distance, there is an equivalent formulation in terms of Lipschitz test functions, but we will not need this alternative formulation.)

▶ **Definition 35.** *For any two distributions* $\mathbf{X}$ *and* $\mathbf{Y}$ *over* $\mathbb{R}^k$*, the Wasserstein distance between* $\mathbf{X}$ *and* $\mathbf{Y}$ *is defined to be*

$$d_W(\mathbf{X}, \mathbf{Y}) = \inf_{(\widehat{\mathbf{X}}, \widehat{\mathbf{Y}})} (\boldsymbol{E}[\|\widehat{\mathbf{X}} - \widehat{\mathbf{Y}}\|]),$$

where the infimum is taken over all couplings $(\widehat{\mathbf{X}}, \widehat{\mathbf{Y}})$ *of* $\mathbf{X}$ *and* $\mathbf{Y}$*.*

As in the analysis of our PRG for arbitrary functions of k LTFs, we need an upper bound on the Wasserstein distance between the two random variables of interest as a starting point. In Section 5 the two relevant random variables were both multi-dimensional Gaussians and the desired (quadratic) Wasserstein closeness was given by Proposition 8. In the context of Theorem 24, the two relevant random variables are (i) a sum of independent vector-valued random variables and (ii) the Gaussian with matching mean and covariance, so it is natural to turn to the literature on *central limit theorems* for sums of vector-valued random variables for the desired upper bound on Wasserstein distance.

A range of central limit theorems for sums of independent vector-valued random variables have been established in the literature, but we are not aware of one which can be used "out of the box" for our purposes. Valiant and Valiant [32] gave a central limit theorem which upper bounds the Wasserstein distance between a sum of n vector-valued random variables and the corresponding Gaussian, but their quantitative bound has a $\log n$ factor which would spoil our desired final result. Zhai [33] gave a variant of the [32] CLT, but only for the setting of i.i.d. vector-valued random variables, whereas our summands are not identically distributed. Bonis [3] gave a sharpening of Zhai's bound, but it assumes that each summand random variable has identity covariance, which need not hold for us. While we do not know of any CLTs in the literature which directly yield our desired starting point, below we show how a "bucketing" scheme can be applied to the Valiant-Valiant CLT to yield a CLT of exactly the type that we need (where there is no dependence on n in the upper bound).

We begin by recalling the Valiant-Valiant CLT:

▶ **Theorem 36** (Valiant-Valiant CLT for Wasserstein distance [32]). *Let $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ be independent distributions in $\mathbb{R}^k$ with mean 0 and $||\mathbf{Z}_i||_2 \leq \beta$. Then, writing Σ to denote the covariance matrix of $\mathbf{Z}_1 + \cdots + \mathbf{Z}_n$, we have*

$$d_W(\sum_{a=1}^{n} \mathbf{Z}_a, \mathcal{N}(0, \Sigma)) \leq \beta k(2.7 + 0.83 \log n).$$

We use this to prove the following:

▶ **Proposition 37.** *Let $h_1, h_2, \dots, h_k$ be τ-regular LTFs, $h_i(x) = \text{sign}(W^i_1 x_1 + \cdots + W^i_n x_n - \theta)$ where we have normalized so that each vector $W^i = (W^i_1, \dots, W^i_n)$ has two-norm 1. Let W be the $k \times n$ matrix with (i,j) entry W^i_j, and for $\ell \in [n]$ let W_ℓ denote the column vector with entries $W^1_\ell, \dots, W^k_\ell$. For $\ell \in [n]$ let $\mathbf{Z}_\ell$ denote the k-dimensional random variable $\mathbf{Z}_\ell = \boldsymbol{x}_\ell W_\ell$ where $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ is uniform over $\{-1,1\}^n$ and let $\mathbf{Z} = \mathbf{Z}_1 + \cdots + \mathbf{Z}_n$. Let $\mathbf{G}'$ be the k-dimensional random Gaussian vector $\mathbf{G}' = W\mathbf{G}$ where $\mathbf{G}$ is distributed as $\mathcal{N}(0,1)^n$. Then*

$$d_W(\mathbf{Z}, \mathbf{G}') \leq O(k^2 \log(k) \cdot \tau^2 + k). \tag{30}$$

Further, if $\tau < 10/\sqrt{k}$, then the following bound also holds:

$$d_W(\mathbf{Z}, \mathbf{G}') \leq O(k^2 \tau^2 \log(k/\tau)). \tag{31}$$

(We note that while (30) does not provide a very strong upper bound on Wasserstein distance, for suitably small values of τ the bound (31) does give a useful upper bound, and it is this bound that we will employ in the next subsection.)

Proof. We begin by observing that the random variables $\mathbf{Z}_1, \dots, \mathbf{Z}_\ell$ are independent, have mean zero (indeed each has support size two, on the two points W_ℓ and $-W_\ell$), and lie in $\mathbb{R}^k$. However, at this point, just having the condition that the *rows* of W are τ-regular and have two-norm 1 doesn't provide much useful information about the two-norms of the columns W_ℓ. Our approach is to bucket the columns according to the two-norms and use the Valiant-Valiant CLT (Theorem 36) separately on each of these buckets. We now proceed to give more details.

Let A_i be the subset of those $\ell \in [n]$ such that $2^{-i-1} \leq \|W_\ell\|_2 \leq 2^{-i}$, i.e.

$$2^{-2i-2} \leq (W^1_\ell)^2 + \cdots + (W^k_\ell)^2 \leq 2^{-2i}.$$

Fix an $\ell \in [n]$ and consider the column vector $W_\ell = (W^1_\ell, \dots, W^k_\ell)$. We have that each $|W^i_\ell| \leq \tau$ (using the τ-regularity of each row and the fact that each row is normalized to have 2-norm 1). Thus, we have $0 \leq (W^1_\ell)^2 + \cdots + (W^k_\ell)^2) \leq k\tau^2$. It follows that A_i is empty if $i < i_0 := (\log(1/k\tau^2))/2 - 1$. (Note that if k is large and τ is not very small then i_0 may be a negative value; this will come up below.)

The sum of squares of all $W_{i,j}$ is k, so each A_i can have at most $k \cdot 2^{2i+2} = 4k2^{2i}$ many elements. Fix an i such that A_i is nonempty (so $i \geq i_0$). Each $\ell \in A_i$ has $\|W_\ell\|_2 \leq 2^{-i}$, and hence applying the Valiant-Valiant CLT to $\sum_{\ell \in A_i} \mathbf{Z}_\ell$ (setting its parameter "β" to 2^{-i}) gives

$$\begin{aligned} d_W\left(\sum_{\ell \in A_i} \mathbf{Z}_\ell, \mathcal{N}(0, \Sigma_{(i)})\right) &\leq 2^{-i} \cdot k \cdot (2.7 + \log |A_i|) \leq 2^{-i} \cdot k \cdot (O(1) + \log k + 2i) \\ &= O(k \log(k) \cdot 2^{-i} + k \cdot i \cdot 2^{-i}). \end{aligned}$$

Now we use the fact that if $\mathbf{X}, \mathbf{Y}$ are two independent random variables and $\mathbf{U}, \mathbf{V}$ are two independent random variables, then

$$d_W(\mathbf{X} + \mathbf{Y}, \mathbf{U} + \mathbf{V}) \leq d_W(\mathbf{X}, \mathbf{U}) + d_W(\mathbf{Y}, \mathbf{V})$$

(this is easy to see from the coupling-based definition that we have given for d_W). Applying this, where the sum is over all $i \geq i_0$, since $\sum_i \sum_{\ell \in A_i} \mathbf{Z}_\ell = \mathbf{Z}$ and $\sum_i \mathcal{N}(0, \Sigma_{(i)}) = \mathbf{G}'$, we get that

$$d_W(\mathbf{Z}, \mathbf{G}') \leq \sum_{i \geq i_0} O(k \log(k) \cdot 2^{-i}) + \sum_{i \geq i_0} O(k \cdot i \cdot 2^{-i}).$$

Let us upper bound this sum, keeping in mind that $\log(1/k\tau^2)$ may be negative. The first sum is at most

$$\sum_{i \geq i_0} O(k \log(k) \cdot 2^{-i}) \leq O(k^2 \log(k) \cdot \tau^2).$$

The second sum is

$$\sum_{i \geq i_0} O(k \cdot i \cdot 2^{-i})$$

which needs to be considered with a bit of care since i_0 may be negative. Summing over any negative values of i obviously gives a negative contribution. Summing over positive values of i gives at most $O(k)$ (and we note that indeed the contribution when $i = 1$ is $\Theta(k)$). So the total sum is at most

$$O(k^2 \log(k) \cdot \tau^2 + k).$$

We note that either of the two summands may dominate depending on the relation between τ and k). However, if we assume that $\tau < 10/\sqrt{k}$ (so i_0 is a positive number), then the upper bound on the second sum above becomes $O(k^2\tau^2 \log(1/k\tau^2))$, which is at most $O(k^2\tau^2 \log(1/\tau))$, and we can bound the whole quantity by $O(k^2\tau^2 \log(k/\tau))$ as claimed. ◀

C.2 The invariance principle for arbitrary functions of LTFs

The CLT in Proposition 37 gives closeness in (non-quadratic) Wasserstein distance. As in Section 5, using arguments from [5] this can be translated into closeness in union-of-orthants distance. The details of the arguments are almost identical to the analysis from [5] since now (as in that work) one of the random variables is a sum of independent vector-valued random variables, the other is Gaussian, and the relevant Wasserstein distance under consideration is the non-quadratic Wasserstein distance. In a bit more detail, the analogue of (20) is now established, as in [5], using the Berry-Esseen theorem and the fact that each linear form is τ-regular, yielding $\Gamma \leq O(k(r + \tau))$. The upper bound on Wasserstein distance that was provided by Theorem 7 in the [5] analysis is now provided by our Proposition 37; to be more precise, the analogue to the next-to-last centered equation in the proof of Theorem 5 of [5] in our setting is that we have $r\Delta/2 \leq d_W(\mathbf{Z}, \mathbf{G}')$ which is $O(k^2\tau^2 \log(k/\tau))$ by Proposition 37. Optimizing the choice of r to make $\Gamma + \Delta$ as small as possible, we obtain the following (we refer the reader to the proof of Theorem 5 of [5] for more details):

▶ **Theorem 38.** *Let $h_1, h_2, \ldots, h_k$ be τ-regular LTFs, $h_i(x) = \mathrm{sign}(W_1^i x_1 + \cdots + W_n^i x_n - \theta)$ where we have normalized so that each vector $W^i = (W_1^i, \ldots, W_n^i)$ has two-norm 1. Let W be the $k \times n$ matrix with (i,j) entry W_j^i, and for $\ell \in [n]$ let W_ℓ denote the column vector with entries $W_\ell^1, \ldots, W_\ell^k$. For $\ell \in [n]$ let $\mathbf{Z}_\ell$ denote the k-dimensional random variable $\mathbf{Z}_\ell = \boldsymbol{x}_\ell W_\ell$ where $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is uniform over $\{-1,1\}^n$ and let $\mathbf{Z} = \mathbf{Z}_1 + \cdots + \mathbf{Z}_n$. Let $\mathbf{G}'$ be the k-dimensional random Gaussian vector $\mathbf{G}' = W\mathbf{G}$ where $\mathbf{G}$ is distributed as $\mathcal{N}(0,1)^n$. Then*

$$d_{\mathrm{UO}}(\mathbf{Z}, \mathbf{G}') \leq O(k^{3/2}\tau\sqrt{\log(k/\tau)}).$$

(The condition $\tau < 10/\sqrt{k}$ in Proposition 37 does not necessitate any condition on τ in Theorem 38, because if $\tau \geq 10/\sqrt{k}$ then the claimed bound of Theorem 38 holds trivially.) Finally, we note that the desired invariance principle, Theorem 24, is a restatement of Theorem 38, using the connection between union-of-orthants distance and any k-variable Boolean combining function g that was formalized in Observation 12.