58th Annual IEEE Symposium on Foundations of Computer Science

Fooling intersections of low-weight halfspaces

Rocco A. Servedio
Department of Computer Science
Columbia University
New York, USA
rocco@cs.columbia.edu

Li-Yang Tan
Toyota Technological Institute at Chicago
Chicago, USA
liyang@cs.columbia.edu

***Abstract*—A *weight-t halfspace* is a Boolean function $f(x) = \mathrm{sign}(w_1x_1 + \cdots + w_nx_n - \theta)$ where each w_i is an integer in $\{-t, \ldots, t\}$. We give an explicit pseudorandom generator that δ-fools any intersection of k weight-t halfspaces with seed length $\mathrm{poly}(\log n, \log k, t, 1/\delta)$. In particular, our result gives an explicit PRG that fools any intersection of any quasi$\mathrm{poly}(n)$ number of halfspaces of any $\mathrm{polylog}(n)$ weight to any $1/\mathrm{polylog}(n)$ accuracy using seed length $\mathrm{polylog}(n)$. Prior to this work no explicit PRG with non-trivial seed length was known even for fooling intersections of n weight-1 halfspaces to constant accuracy.**

The analysis of our PRG fuses techniques from two different lines of work on unconditional pseudorandomness for different kinds of Boolean functions. We extend the approach of Harsha, Klivans and Meka [HKM12] for fooling intersections of regular halfspaces, and combine this approach with results of Bazzi [Baz07] and Razborov [Raz09] on bounded independence fooling CNF formulas. Our analysis introduces new coupling-based ingredients into the standard Lindeberg method for establishing quantitative central limit theorems and associated pseudorandomness results.

***Keywords*-Unconditional derandomization; intersection of halfspaces; Lindeberg method**

I. Introduction

A *halfspace*, or *linear threshold function* (henceforth abbreviated LTF), over $\{-1,1\}^n$ is a Boolean function f that can be expressed as $f(x) = \mathrm{sign}(w_1x_1 + \cdots + w_nx_n - \theta)$ for some real values $w_1, \ldots, w_n, \theta$. LTFs are a natural class of Boolean functions which play a central role in many areas such as machine learning and voting theory, and have been intensively studied in complexity theory from many perspectives such as circuit complexity [GHR92], [Raz92], [Hås94], [SO03], communication complexity [Nis93], [Vio15], Boolean function analysis [Cho61], [GL94], [Per04], [Ser07], [O'D14], property testing [MORS09], [MORS10], pseudorandomness [DGJ$^+$10], [MZ13], [GKM15] and more.

Because of the limited expressiveness of a single LTF (even a parity function over two variables cannot be expressed as an LTF), it is natural to consider Boolean functions that are obtained by combining LTFs in various ways. Perhaps the simplest and most natural functions of this sort are *intersections of LTFs*, i.e. Boolean functions of the form $F_1 \wedge \cdots \wedge F_k$ where each F_j is an LTF. Intersections of LTFs have been studied in many contexts including Boolean function analysis [Kan14], [She13a], [She13b], computational learning (both algorithms [BK97], [KOS04], [KOS08], [Vem10] and hardness results [KS06], [KS11]), and pseudorandomness [GOWZ10], [HKM12]. We further note that the set of feasible solutions to an $\{0,1\}$-integer program with k constraints corresponds precisely to the set of satisfying assignments of an intersection of k LTFs; understanding the structure of these sets has been the subject of intensive study in computer science, optimization, and combinatorics.

This paper continues the study of intersections of LTFs from the perspective of unconditional pseudorandomness; in particular, we are interested in constructing explicit *pseudorandom generators* (PRGs) for intersections of LTFs. Recall the following standard definitions:

Definition 1 (Pseudorandom generator). *A function* $\mathsf{Gen} : \{-1,1\}^r \to \{-1,1\}^n$ *is said to* δ-fool a function $F : \{-1,1\}^n \to \{-1,1\}$ with seed length r *if*

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{U}' \leftarrow \{-1,1\}^r} \left[F(\mathsf{Gen}(\boldsymbol{U}')) \right] - \mathop{\mathbf{E}}_{\boldsymbol{U} \leftarrow \{-1,1\}^n} \left[F(\boldsymbol{U}) \right] \right| \leq \delta.$$

Such a function Gen *is said to be a* explicit pseudorandom generator that δ-fools a class $\mathcal{F}$ of n-variable functions *if* Gen *is computable by a deterministic uniform* $\mathrm{poly}(n)$*-time algorithm and* Gen δ*-fools every function* $F \in \mathcal{F}$.

A. Prior work

Before describing our results, we recall relevant prior work on fooling LTFs and intersections of LTFs.

Fooling a single LTF: In [DGJ$^+$10] Diakonikolas et al. showed that any $\tilde{O}(1/\delta^2)$-wise independent distribution over $\{-1,1\}^n$ suffices to δ-fool any LTF, and thereby gave a PRG for single LTFs with seed length $\tilde{O}(1/\delta^2) \cdot \log n$. Soon after, [MZ13] gave a more efficient PRG for LTFs with seed length $O(\log n + \log^2(1/\delta))$. They did this by first developing an alternative $\tilde{O}(1/\delta^2) \cdot \log n$ seed length PRG for *regular* LTFs; these are LTFs in which no individual weight is large compared to the total size of all the weights (we give precise definitions later). [MZ13] built on this PRG for regular LTFs using structural results for LTFs and PRGs for read-once branching programs to

0272-5428/17 $31.00 © 2017 IEEE
DOI 10.1109/FOCS.2017.81

obtain their improved $O(\log n + \log^2(1/\delta))$ seed length for fooling arbitrary LTFs. More recently, [GKM15] gave a PRG which δ-fools any LTF over $\{-1,1\}^n$ using seed length $O(\log(n/\delta)(\log\log(n/\delta))^2)$; this is the current state-of-the-art for fooling a single LTF.

Since the approach of [MZ13] for fooling regular LTFs is important for our discussion in later sections, we describe it briefly here. The [MZ13] PRG for regular LTFs employs hashing and other techniques; its analysis crucially relies on the *Berry–Esséen theorem* [Ber41], [Ess42]. Recall that the Berry–Esséen theorem is an "invariance principle" for the distribution of linear forms; it (or rather, a special case of it) says that for w a regular vector, the two random variables $w \cdot \boldsymbol{U}$ and $w \cdot \boldsymbol{G}$, where $\boldsymbol{U}$ is uniform over $\{-1,1\}^n$ and $\boldsymbol{G}$ is drawn from the standard n-dimensional Gaussian distribution $\mathcal{N}(0,1)^n$, are close in CDF distance. Roughly speaking, the [MZ13] PRG analysis for τ-regular LTFs proceeds by showing that the limited randomness provided by their generator is sufficient to apply the Berry–Esséen theorem (over a certain set of roughly $1/\tau^2$ independent random variables). We give a more detailed description of the structure of the [MZ13] PRG in Section II.

Fooling intersections of regular LTFs: Now we turn to results on fooling intersections of LTFs. Essentially simultaneously with [MZ13] (in terms of conference publication), [HKM12] gave a PRG for intersections of *regular* LTFs. Their PRG $\tilde{O}((\log k)^{8/5}\tau^{1/5})$-fools any intersection of k many τ-regular LTFs with seed length $O((\log n \log k)/\tau)$. As we discuss in in Section II, the [HKM12] generator has the same structure as the [MZ13] PRG for regular LTFs, but with different (larger) parameter settings and a significantly more involved analysis. At the heart of the correctness proof of the [HKM12] PRG is a new invariance principle that [HKM12] prove for *k-tuples* $(w^{(1)} \cdot \boldsymbol{U}, \ldots, w^{(k)} \cdot \boldsymbol{U})$ of regular linear forms, generalizing the Berry–Esséen theorem which as described above applies to a single regular linear form. With this new invariance principle in hand, to prove their PRG theorem [HKM12] argue (similar in spirit to [MZ13]) that the limited randomness provided by their generator is sufficient for their new k-dimensional invariance principle.

Note that even the $k = 1$ case of the invariance principle (the Berry–Esséen theorem) does not give a meaningful bound for non-regular linear forms. As a simple example, consider the trivial linear form x_1, which is highly non-regular: the two one-dimensional random variables $\boldsymbol{U}_1$ and $\boldsymbol{G}_1$, where $\boldsymbol{U}_1$ is uniform over $\{-1,1\}$ and $\boldsymbol{G}_1$ is distributed according to $\mathcal{N}(0,1)$, have CDF distance ≈ 0.341. And indeed the analysis of the [HKM12] PRG only goes through for intersections of LTFs in which all the LTFs are regular. So while the [HKM12] PRG has an extremely good (polylogarithmic) dependence on the number of LTFs in the intersection, the regularity requirement means that the [HKM12] PRG theorem cannot be applied, for example, to fool the class of intersections of LTFs in which each weight is either 0 or 1.

The PRG of Gopalan, O'Donnell, Wu, and Zuckerman: Around the same time, [GOWZ10] gave a PRG that δ-fools intersections of k arbitrary LTFs with seed length $O((k\log(k/\delta) + \log n) \cdot \log(k/\delta))$, and indeed δ-fools any depth-k size-s decision tree that queries LTFs at its internal nodes with seed length $O((k\log(ks/\delta) + \log n) \cdot \log(ks/\delta))$. Their approach builds on the PRG of [MZ13] for general LTFs; one central ingredient is a generalization of structural results for single LTFs used in [MZ13] to k-tuples of LTFs. Both this generalization, and the read-once branching program based techniques from [MZ13] (which are extended in [GOWZ10] to the context of k-tuples of LTFs), necessitate a seed length which is at least linear in k. So while the [GOWZ10] PRG is is notable for being able to handle intersections of general LTFs, their seed length's linear dependence in k means that their seed length is $n^{\Omega(1)}$ whenever $k = n^{\Omega(1)}$, and furthermore their result does not give a non-trivial PRG for intersections of $k \geq n$ many LTFs.

1) A conceptual challenge: We elaborate briefly on an issue related to the linear-in-k dependence of the [GOWZ10] generator discussed above. A standard approach to analyze non-regular LTFs, both in pseudorandomness and in other subfields of complexity theory such as analysis of Boolean functions and learning theory [DS13], [DRST14], [DSTW14], [DDS16], [FGRW09], [CSS16], is to reduce the analysis of non-regular LTFs to that of regular LTFs via a "critical index" argument (see [Ser07]). Indeed, most previous pseudorandomness results for classes involving non-regular LTFs and PTFs—general LTFs [DGJ+10], [MZ13], functions of LTFs [GOWZ10], degree-d PTFs and functions of such PTFs [DKN10], [MZ13], [DDS14], [DS14]—make use of such a reduction to the regular case. In working with functions that involve k LTFs (or PTFs), this analysis (see [DDS14], [GOWZ10]) involves "multi-critical-index" arguments, originating in [GOWZ10], which necessitate an $\Omega(k)$ seed length dependence; indeed, this linear-in-k dependence was highlighed in [HKM12] as a conceptual challenge to overcome in extending their results to intersections of k non-regular LTFs.

In this work we give the first analysis that is able to handle an interesting class of functions involving k non-regular LTFs while avoiding this linear-in-k cost that is inherent to multi-critical-index based arguments, and in fact achieving a polylogarithmic dependence on k.

B. Our main result: fooling intersections of low-weight LTFs

It is easy to see that every LTF $f : \{-1,1\}^n \to \{-1,1\}$ has some representation as $f(x) = \text{sign}(w \cdot x - \theta)$ where the coefficients $w_1, \ldots, w_n$ are all integers; a standard way of measuring the "complexity" of an LTF is by the size of its integer weights. It has been known at least since the 1960s [MTT61], [Hon87], [Rag88] that every n-variable LTF has an integer representation with $\max |w_i| \le n^{O(n)}$, and Håstad has shown [Hås94] that there are LTFs that in fact require $\max w_i = n^{\Omega(n)}$ for any integer representation. However, in many settings, LTFs with *small integer weights* are of special interest. Such LTFs are often the relevant ones in contexts such as voting systems or contexts where, e.g., biological or physical constraints may limit the size of the weights. From a more theoretical perspective, it is well known that sample complexity bounds for many commonly used LTF learning methods, such as the Perceptron and Winnow algorithms, are essentially determined by the size of the integer weights.

We say that f is a *weight-t LTF* if it can be represented as $f(x) = \text{sign}(w \cdot x - \theta)$ where each w_i is an integer satisfying $|w_i| \le t$. Note that arguably the simplest and most natural LTFs — unweighted threshold functions, with the majority function as a special case — have weight 1.

Our main result is an efficient PRG for fooling intersections of low-weight LTFs:

Theorem 1 (PRG for intersections of low-weight LTFs)**.** *For all values of $k, t \in \mathbb{N}$ and $\delta \in (0, 1)$, there is an explicit pseudorandom generator that δ-fools any intersection of k weight-t LTFs over $\{-1, 1\}^n$ with seed length* $\text{poly}(\log n, \log k, t, 1/\delta)$.

Recalling the results of [HKM12], [GOWZ10] described in Section I-A, prior to this work no explicit PRG with non-trivial seed length was known even for fooling intersections of n weight-1 LTFs to constant accuracy. (In fact, no $2^{0.99n}$-time algorithm was known for deterministic approximate counting of satisfying assignments of such an intersection; since such an algorithm is allowed to inspect the intersection of halfspaces which is its input, while a PRG is "input-oblivious", giving such an algorithm is an easier problem than constructing a PRG.) In contrast, our result gives an explicit PRG that fools any intersection of any $\text{quasipoly}(n)$ number of LTFs of any $\text{polylog}(n)$ weight to any $1/\text{polylog}(n)$ accuracy using seed length $\text{polylog}(n)$. For any $c > 0$ our result also gives an explicit PRG with seed length n^c that fools intersections of $\exp(n^{\Omega(1)})$ many LTFs of weight $n^{\Omega(1)}$ to accuracy $1/n^{\Omega(1)}$. Recalling the correspondence between intersections of LTFs and $\{0, 1\}$-integer programs, our PRG immediately yields new deterministic algorithms for approximately counting the number of feasible solutions to broad classes of $\{0, 1\}$-integer programs.

Our most general PRG result: We obtain Theorem 1 as an easy consequence of a PRG that fools a more general class of intersections of LTFs. To describe this class we require some terminology. We say that a vector w over $\mathbb{R}^n$ is *s-sparse* if at most s coordinates among $w_1, \ldots, w_n$ are nonzero. We similarly say that a linear threshold function $\text{sign}(w \cdot x - \theta)$ is s-sparse if w is s-sparse. Following [HKM12], we say that a linear form $w = (w_1, \ldots, w_n)$ with norm $\|w\| := \left(\sum_{i=1}^n w_i^2\right)^{1/2}$ is *τ-regular* if $\sum_{i=1}^n w_i^4 \le \tau^2 \|w\|^2$, and we say that a linear threshold function $\text{sign}(w \cdot x - \theta)$ is τ-regular if the linear form w is τ-regular. Finally, we say that $F : \{-1, 1\}^n \to \{-1, 1\}$ is a *(k, s, τ)-intersection of LTFs* if $F = F_1 \wedge \cdots \wedge F_k$ where each F_j is an LTF which is either s-sparse or τ-regular.

Our most general PRG result is the following:

Theorem 2 (Our most general PRG, informal statement)**.** *For all values of $k, s \in \mathbb{N}$ and $\tau \in (0, 1)$, there is an explicit pseudorandom generator with seed length* $\text{poly}(\log n, \log k, s, 1/\tau)$ *that fools any (k, s, τ)-intersection of LTFs to accuracy* $\delta = \text{poly}(\log k, \tau)$.

In Section IV-A we give the formal statement of Theorem 2 and show how Theorem 1 follows from Theorem 2.

II. Our approach

As explained in Section I-A, invariance-based arguments are not directly useful for our task of fooling intersections of low-weight LTFs, since the invariance principle does not give a non-trivial bound even for a single low-weight LTF. Nevertheless, we are able to show that a generator with the same structure as the [MZ13], [HKM12] generators (but now with slightly larger parameter settings than were used in the [HKM12] generator) indeed fools any (k, s, τ)-intersection of LTFs. We do this via an analysis that brings in ingredients that are novel in the context of fooling intersections of LTFs; in particular, we use results of Bazzi [Baz07] and Razborov [Raz09] on bounded independence fooling depth-2 circuits.

How are depth-2 circuits relevant to intersections of LTFs? A starting point for our work is to re-express a (k, s, τ)-intersection of LTFs using a different representation, in which we replace each s-sparse LTF by a CNF formula computing the same function over $\{-1, 1\}^n$. The following is an immediate consequence of the fact that any s-sparse LTF depends on at most s variables:

Fact II.1. *Let F be a (k, s, τ)-intersection of LTFs. Then $F \equiv H \wedge G$, where*

- *H is the intersection of at most k many τ-regular LTFs.*
- *G is a width-s CNF formula with at most $k \cdot 2^s$ clauses;*

We refer to a function of the form $H \wedge G$ as above as a (k, s, τ)-CnfLtf. We can thus restate our goal as that of designing a PRG to fool any (k, s, τ)-CnfLtf: with this perspective it is not surprising that pseudorandomness tools for fooling CNF formulas can be of use.

A. The structure of our PRG

To describe our approach we need to explain the general structure of the PRG which is used in [MZ13] for regular LTFs, in [HKM12] for intersections of regular LTFs, and in our work for (k, s, τ)-intersections of LTFs. The construction uses an r_{hash}-wise independent family

$\mathcal{H}$ of hash functions $h : [n] \to [\ell]$, and an r_{bucket}-wise independent generator outputting strings in $\{-1,1\}^n$, which we denote $\mathscr{G}$. The overall generator, which we denote Gen, on input $(h, X^{(1)}, \dots, X^{(\ell)})$ outputs the string $\mathsf{Gen}(h, X^{(1)}, \dots, X^{(\ell)}) := Y \in \{-1,1\}^n$, where $Y_{h^{-1}(b)} = \mathscr{G}(X^{(b)})_{h^{-1}(b)}$ for all $b \in [\ell]$. (Here and elsewhere, for Y an n-bit string and $S \subseteq [n]$ we write Y_S to denote the $|S|$-bit string obtained by restricting Y to the coordinates in S.)

The [MZ13] PRG for τ-regular LTFs instantiates this construction with

$$\ell = 1/\tau^2, \quad r_{\text{hash}} = 2, \quad \text{and} \quad r_{\text{bucket}} = 4,$$

while the [HKM12] PRG for intersections of k many τ-regular LTFs takes

$$\ell = 1/\tau, \quad r_{\text{hash}} = 2\log k, \quad \text{and} \quad r_{\text{bucket}} = 4\log k.$$

We state the exact parameter settings which we use to fool (k, s, τ)-intersections of LTFs in Section IV (the specific values are not important for our discussion in this section).

B. Sketch of the [MZ13], [HKM12] analysis

As our analysis (sketched in Section II-C) builds on [MZ13], [HKM12], in this subsection we sketch the [MZ13], [HKM12] arguments establishing correctness of the PRG Gen for regular LTFs and intersections of regular LTFs.

A high-level sketch of the [MZ13] analysis showing that Gen fools any regular LTF $F(x) = \operatorname{sign}(w \cdot x - \theta)$ is as follows: the hash function $h : [n] \to [\ell]$ partitions the n coefficients $w_1, \dots, w_n$ into ℓ buckets. The pairwise independence of $\boldsymbol{h} \leftarrow \mathcal{H}$ and the regularity of w are together used to show that each of the ℓ buckets receives essentially the same amount of "coefficient weight." The idea then is to view the sum $w \cdot \boldsymbol{Y}$, where $\boldsymbol{Y}$ is the output of the generator, as a sum of ℓ *independent* random variables (note that the inputs $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(\ell)} \in \{-1,1\}^r$ to Gen are indeed mutually independent), one for each bucket, and use the Berry–Esséen theorem on that sum.[1] The four-wise independence of $\mathscr{G}$ is used to ensure that each of the ℓ summands—the b-th summand corresponding to $w_{\boldsymbol{h}^{-1}(b)} \cdot \boldsymbol{Y}_{\boldsymbol{h}^{-1}(b)}$, the contribution from the b-th bucket—has the moment properties that are required to apply the Berry–Esséen theorem. Note that in this analysis the Berry–Esséen theorem is used as a "black box."

Since [HKM12] have to prove the k-dimensional invariance principle that they use in place of the Berry–Esséen theorem, their analysis is necessarily more involved, but at a high level it follows a similar approach to the [MZ13] analysis sketched above. A sketch of their argument that Gen fools any intersection $F = F_1 \wedge \cdots \wedge F_k$ of regular LTFs is as follows:

1) [HKM12] first argue that for *any* smooth test function $\psi : \mathbb{R}^k \to [0,1]$—replacing the "hard threshold" function $\mathbf{1}(v_1 \le \theta_1) \cdot \mathbf{1}(v_2 \le \theta_2) \cdots \mathbf{1}(v_k \le \theta_k)$, which corresponds to k-dimensional CDF distance—the pseudorandom distribution output by the generator fools the test function ψ relative to an $\mathcal{N}(0,1)^n$ Gaussian input to ψ. This is done by
   a) first arguing (similar to [MZ13]) that the $(2\log k)$-wise independent hash function $\boldsymbol{h} \leftarrow \mathcal{H}$ and the regularity of each LTF F_k together "spread the coefficient weight" of the k LTFs roughly evenly among the ℓ buckets (we note that this part of the argument has nothing to do with the function ψ);
   b) then a hybrid argument across the ℓ buckets, using the smoothness of ψ and moment properties of the random variables corresponding to the ℓ buckets (which now follow from the $(4\log k)$-wise independence of $\mathscr{G}$), is used to bound

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{Y} \leftarrow \mathsf{Gen}} \left[\psi(w^{(1)} \cdot \boldsymbol{Y}, \dots, w^{(k)} \cdot \boldsymbol{Y})\right] - \mathop{\mathbf{E}}_{\boldsymbol{G} \leftarrow \mathcal{N}(0,1)^n} \left[\psi(w^{(1)} \cdot \boldsymbol{G}, \dots, w^{(k)} \cdot \boldsymbol{G})\right] \right|. \tag{1}$$

   (Such a hybrid argument is a central ingredient in the Lindeberg-style "replacement method" proof of the Berry–Esséen theorem, and is also used in [HKM12]'s proof of their invariance principle for intersections of k regular LTFs.) We note that multidimensional Taylor's theorem plays a crucial role in bounding the difference in expectation between ψ applied to two random variables, which is done to "bound the distance" at each step of the hybrid.
2) Next [HKM12] use a particular smooth function ψ^* based on a result of Bentkus [Ben90] and a Gaussian surface area bound for intersections of k halfspaces due to Nazarov [Naz03] to pass from fooling the smooth test function ψ^* to fooling the "hard threshold" function corresponding to CDF distance. This essentially amounts to using the fact that (1) is small to show that $\left|\mathbf{E}_{\boldsymbol{Y} \leftarrow \mathsf{Gen}}[F(\boldsymbol{Y})] - \mathbf{E}_{\boldsymbol{G} \leftarrow \mathcal{N}(0,1)^n}[F(\boldsymbol{G})]\right|$ is also small. Given this, the fact that the generator fools F, i.e. that $\left|\mathbf{E}_{\boldsymbol{Y} \leftarrow \mathsf{Gen}}[F(\boldsymbol{Y})] - \mathbf{E}_{\boldsymbol{X} \leftarrow \{-1,1\}^n}[F(\boldsymbol{X})]\right|$ is small, follows from [HKM12]'s invariance principle, which bounds $\left|\mathbf{E}_{\boldsymbol{G} \leftarrow \mathcal{N}(0,1)^n}[F(\boldsymbol{G})] - \mathbf{E}_{\boldsymbol{X} \leftarrow \{-1,1\}^n}[F(\boldsymbol{X})]\right|$. We note that this second step of [HKM12]'s analysis does not use regularity of the F_j's at all (but their invariance principle does require that each F_j is regular).

[1] Note that if the weight vector w is non-regular, then it is in general impossible for any hash function, even a fully independent one, to spread the coefficient weight out evenly among the ℓ buckets, and consequently the Berry–Esséen theorem cannot be applied (as, intuitively, it requires that no individual random variable summand is "too heavy" compared to the "total weight" of the sum). This is why the overall approach requires regularity.

C. Sketch of our analysis

Here we give an overview of our proof that Gen, with suitable parameters, fools any (k, s, τ)-CNFLTF $F = H \wedge G$. Recall that H is an intersection of k many τ-regular LTFs and G is a $(k \cdot 2^s)$-clause CNF, and that the difference between our task and that of [HKM12] is that we must handle the CNF G in addition to the intersection of regular LTFs H. While it is not difficult to see, as a consequence of [Baz07], [Raz09], that the [HKM12] generator with suitable parameters (i) fools H, and (ii) fools G, it is far from clear *a priori* that it fools $H \wedge G$. We show this via a rather delicate argument, which involves a novel extension of the Lindeberg method that is at the heart of all PRGs in this line of work [GOWZ10], [MZ13], [HKM12]. To surmount the technical challenges that arise in our setting (which we described next), our analysis features several new ingredients which are not present in the analyses of [GOWZ10], [MZ13], [HKM12], or indeed in other Lindeberg-type proofs of quantitative central limit theorems that we are aware of. The ideas in this new style of coupling-based analysis, which we outline in Section II-C1 below, may be of use elsewhere.

The standard Lindeberg setup, and a new challenge in our setting: As is standard in Lindeberg-style proofs, our analysis focuses on a particular smooth test function, which for us takes $k+1$ arguments and which we denote ψ^*_{k+1}. This should be thought of as the $(k+1)$-variable version of the smooth function of Bentkus [Ben90], which was used by [HKM12] as mentioned in the preceding subsection. Crucially, while ψ^*_{k+1} maps all of $\mathbb{R}^{k+1}$ to $[-1, 1]$, in our arguments this test function will only ever receive inputs from $\mathbb{R}^k \times \{\pm 1\}$; indeed, its last ($(k+1)$-st) coordinate will always be a Boolean value which is the output of the CNF G.

The heart of our proof lies in showing that for this specific smooth test function ψ^*_{k+1} (which should be thought of as a proxy for $\text{AND}(\text{sign}(v_1 - \theta), \ldots, \text{sign}(v_k - \theta_k), v_{k+1}))$), the pseudorandom distribution output by the generator fools the test function ψ^*_{k+1} relative to a uniform random input drawn from $\{-1, 1\}^n$. This is done by means of a hybrid argument, the analysis of which (like that of [GOWZ10], [HKM12]) employs a multidimensional version of Taylor's theorem. However, the fact that the distinguished last coordinate of ψ^*_{k+1} always receives a $\{\pm 1\}$-valued input—in particular, an input whose magnitude changes by a large amount (namely 2) when it does change—introduces significant challenges in using the multidimensional Taylor's theorem. Recall that Taylor's theorem quantifies the following intuition: roughly speaking, if the input to a smooth function ψ is only changed by a small amount Δ, then the resulting change in its output value, $\psi(v + \Delta) - \psi(v)$, is correspondingly small as well. Naturally, if Δ is large then Taylor's theorem does not give useful bounds.

1) New ingredients in our approach: Taylor's theorem is the core ingredient in Lindeberg-style proofs of invariance principles (see e.g. [Tao10] and Chapter 11 of [O'D14]) and associated pseudorandomness results (see e.g. [GOWZ10], [MZ13], [HKM12]), where it is used to bound the distance incurred by a single step of the hybrid argument. As mentioned above, in order for Taylor's theorem to give a useful bound when it is applied to re-express $\psi^*_{k+1}(v + \Delta)$ (in terms of $\psi^*_{k+1}(v)$, various derivatives of ψ^*_{k+1} at v, Δ, and an error term), the quantity Δ must be "small." This is a problem in our context since the distinguished last coordinate of ψ^*_{k+1}'s argument (the output of the CNF G) is $\{\pm 1\}$-valued, so the last coordinate of Δ alone may already be as large as 2. We get around this difficulty by utilizing a carefully chosen coupling between two adjacent hybrid random variables and decomposing each of the two relevant arguments to which ψ^*_{k+1} is applied (each of which is a random variable) in a very careful way. One of these random variables is expressed as $\boldsymbol{v} + \boldsymbol{\Delta}^{\text{unif}}$ (corresponding to "filling in the current bucket uniformly at random") and the other is $\boldsymbol{v} + \boldsymbol{\Delta}^{\text{pseudo}}$ (corresponding to "filling in the current bucket pseudorandomly"); roughly speaking, in order to succeed our analysis must show that the magnitude of $\mathbf{E}[\psi^*_{k+1}(\boldsymbol{v} + \boldsymbol{\Delta}^{\text{unif}})] - \mathbf{E}[\psi^*_{k+1}(\boldsymbol{v} + \boldsymbol{\Delta}^{\text{pseudo}})]$ is suitably small. The key property of the coupling we employ is that it ensures that the last coordinates of both random variables $\boldsymbol{\Delta}^{\text{unif}}$ and $\boldsymbol{\Delta}^{\text{pseudo}}$ are almost always zero; in fact, one of them will actually be always zero, see Equation (7). (We note that if no coupling is used then the last coordinate of $\boldsymbol{\Delta}^{\text{pseudo}}$ can be as large as 2 with constant probability.) The existence of such a favorable coupling follows from the fact that each bucket of Gen is, by virtue of its bounded independence and the results of Bazzi [Baz07] and Razborov [Raz09], "sufficiently pseudorandom" to fool CNF formulas.

However, the way that we structure the random variables $\boldsymbol{v}, \boldsymbol{\Delta}^{\text{unif}}$, and $\boldsymbol{\Delta}^{\text{pseudo}}$ to ensure that the last coordinate of each $\boldsymbol{\Delta}$ is almost always small (as discussed above), introduces a new complication, which is that now the random variables $\boldsymbol{v}$ and $\boldsymbol{\Delta}^{\text{unif}}$ are not independent (and neither are $\boldsymbol{v}$ and $\boldsymbol{\Delta}^{\text{pseudo}}$). This situation does not arise in standard uses of the Lindeberg method, either in proving invariance principles or in applications to pseudorandom generators. In all of these previous proofs, independence is used to show that various first derivative, second derivative, etc. terms in the Taylor expansions for the two adjacent random variables cancel out perfectly upon subtraction (using matching moments). To surmount this lack of independence, we exploit the fact that our coupling lets us re-express the coupled joint distribution (over a pair of vectors in $\mathbb{R}^k \times \{\pm 1\}$) as a mixture of three joint distributions over pairs of $(k+1)$-dimensional vectors in such a way that one component of the mixture is entirely supported on $(\mathbb{R}^k \times \{1\}) \times (\mathbb{R}^k \times \{1\})$, one is entirely supported on $(\mathbb{R}^k \times \{-1\}) \times (\mathbb{R}^k \times \{-1\})$, and the third has a very small mixing weight. Under each of the first two joint distributions (supported entirely on pairs that agree in the last coordinate), $\boldsymbol{v}$ and $\boldsymbol{\Delta}^{\text{unif}}$ will indeed be independent, and so will $\boldsymbol{v}$ and $\boldsymbol{\Delta}^{\text{pseudo}}$.

However, performing the hybrid method using these conditional distributions presents another challenge: while now $\boldsymbol{v}$ and $\boldsymbol{\Delta}^{\text{unif}}$ are independent (and likewise for $\boldsymbol{v}$ and $\boldsymbol{\Delta}^{\text{pseudo}}$), the moments of these conditional random variables may not match perfectly. We deal with this by exploiting the fact that each pseudorandom distribution that we consider "filling in a single bucket" can in fact fool, to very high accuracy, any of $\text{poly}(n)$ many new circuits which arise in our analysis of the multidimensional Taylor expansion (intuitively, these are "slightly augmented" CNFs or DNFs). This allows us to show that while we do not get perfect cancellation, the relevant moments under the conditional distributions are adequately close to each other. Finally, our coupling-based perspective also allows us to bound the (crucial) final error term resulting from Taylor's theorem by reducing its analysis to that of the corresponding error term in [HKM12].

The above is a sketch of how we show that Gen fools the smooth test function ψ^*_{k+1}. To pass from fooling ψ^*_{k+1} to fooling the "hard threshold" AND function, we combine the [HKM12] invariance principle with a simple relationship, Claim V.2, which we establish between the anti-concentration of the $(k+1)$-dimensional input to the ψ^*_{k+1} function (with its distinguished last coordinate corresponding to outputs of the CNF) and its k-dimensional marginal which excludes the last coordinate (all coordinates of which correspond to outputs of regular linear forms, i.e. the setting of [HKM12]).

III. Notation and preliminaries

LTFs and regularity: We recall that a linear threshold function (LTF) is a function of the form $\text{sign}(w \cdot x - \theta)$, where $\text{sign}(z)$ is 1 if $z > 0$ and is -1 otherwise. We view -1 as TRUE and 1 as FALSE throughout the paper.

We write $W \in \mathbb{R}^{n \times k}$ to denote the matrix whose j-th column is the weight vector of the j-th LTF in an intersection 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 to denote the j-th column of W (so $\|W^j\| = 1$ for all j), and for $B \subseteq [n]$ (a subset of variables) we write W_B to denote the matrix formed by the rows of W with indices in B. Combining these notations, W^j_B denotes the $|B|$-element column vector which is obtained from W^j by taking those entries given by the indices in B. Throughout the paper we will write $\vec{\theta}$ to denote the k-tuple $\vec{\theta} = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$.

We say that a vector $w \in \mathbb{R}^n$ is *τ-regular* if $\sum_{i=1}^n w_i^4 \leq \tau^2 \|w\|^2$, and that it is *$s$-sparse* if it has at most s non-zero entries. We use the same terminology to refer to an LTF $\text{sign}(w \cdot x - \theta)$. We say that a matrix $W \in \mathbb{R}^{n \times k}$ is τ-regular if each of its columns is τ-regular.

A *restriction* ρ fixing a subset $S \subseteq [n]$ of n input variables is an element of $\{0,1\}^S$; it corresponds to setting the variables in S in the obvious way and leaving the variables outside S free. Given an n-variable function f and a restriction ρ we write $f \upharpoonright \rho$ to denote the function obtained by setting some of the input variables as dictated by ρ.

Probability background: We recall some standard definitions of bounded-independence distributions and hash families. A distribution $\mathcal{D}$ over $\{-1,1\}^n$ is *r-wise independent* if for every $1 \leq i_1 < \cdots < i_r \leq n$ and every $(b_1, \ldots, b_r) \in \{-1,1\}^r$, we have

$$\Pr_{\boldsymbol{X} \leftarrow \mathcal{D}} \left[\boldsymbol{X}_{i_1} = b_1 \text{ and } \cdots \text{ and } \boldsymbol{X}_{i_r} = b_r \right] = 2^{-r}.$$

We recall the results of [Baz07], [Raz09] which state that bounded-independence distributions fool CNF formulas:

Theorem 3 (Bounded independence fools depth-2 circuits)**.** *Let f be any M-clause CNF formula or M-term DNF formula. Then f is δ-fooled by any $O((\log(M/\delta))^2)$-wise independent distribution.*

A family $\mathcal{H}$ of functions from $[n]$ to $[\ell]$ is said to be an *r-wise independent hash family* if for every $1 \leq i_1 < \cdots < i_r \leq n$ and $(j_1, \ldots, j_r) \in [\ell]^r$, we have

$$\Pr_{\boldsymbol{h} \leftarrow \mathcal{H}} \left[\boldsymbol{h}(i_1) = j_1 \text{ and } \cdots \text{ and } \boldsymbol{h}(i_r) = j_r \right] = \ell^{-r}.$$

When S is a set the notations $\mathbf{Pr}_{\boldsymbol{X} \leftarrow S}[\cdot], \mathbf{E}_{\boldsymbol{X} \leftarrow S}[\cdot]$ indicate that the relevant probability or expectation is over a uniform draw of $\boldsymbol{X}$ from set S. Throughout the paper we use bold fonts such as $\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{h}$, etc. to indicate random variables. We write $\mathcal{N}(0,1)$ to denote the standard normal distribution with mean 0 and variance 1.

Calculus: We say that a function $\psi : \mathbb{R}^k \to \mathbb{R}$ is *smooth* if its first through fourth derivatives are uniformly bounded. For smooth $\psi : \mathbb{R}^k \to \mathbb{R}$, $v \in \mathbb{R}^k$, and $j_1, \ldots, j_r \in [k]$, we write $(\partial_{j_1,\ldots,j_r}\psi)(x)$ to denote $\partial_{j_1}\partial_{j_2}\cdots\partial_{j_r}\psi(x)$, and for $s = 1, 2, \ldots$ we write $\|\psi^{(s)}\|_1$ to denote

$$\sup_{v \in \mathbb{R}^k} \left\{ \sum_{j_1,\ldots,j_s \in [k]} |(\partial_{j_1,\ldots,j_s}\psi)(v)| \right\}.$$

Given indices $j_1, \ldots, j_r \in [k]$, we write $(j_1, \ldots, j_r)!$ to denote $s_1! s_2! \cdots s_k!$, where for each $\ell \in [k]$, s_ℓ denotes the number of occurrences of ℓ in $(j_1, \ldots, j_r)$. We will use the following form of multidimensional Taylor's theorem (see e.g. Fact 4.3 of [HKM12]):

Fact III.1 (Multidimensional Taylor's theorem)**.** *Let ψ :*

$\mathbb{R}^k \to \mathbb{R}$ *be smooth and let* $v, \Delta \in \mathbb{R}^k$. *Then*

$$\psi(v+\Delta) = \psi(v) + \sum_{j\in[k]} (\partial_j \psi)(v)\Delta_j + \sum_{j,j'\in[k]} \frac{1}{(j,j')!}(\partial_{j,j'}\psi)(v)\Delta_j\Delta_{j'} + \sum_{j,j',j''\in[k]} \frac{1}{(j,j',j'')!}(\partial_{j,j',j''}\psi)(v)\Delta_j\Delta_{j'}\Delta_{j''} + \mathrm{err}(v,\Delta),$$

where $|\mathrm{err}(v,\Delta)| \le \|\psi^{(4)}\|_1 \cdot \max_{j\in[k]} |\Delta_j|^4$.

Useful results from [HKM12]: The following notation will be useful: for $0 < \lambda < 1$, $k \ge 1$, and $\vec{\theta} = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$, we define

$$\mathrm{Inner}_{k,\vec{\theta}} = \{v \in \mathbb{R}^k \colon v_j \le \theta_j \text{ for all } j \in [k]\},$$

$$\mathrm{Outer}_{\lambda,k,\vec{\theta}} = \{v \in \mathbb{R}^k \colon v_j \ge \theta_j + \lambda \text{ for some } j \in [k]\},$$

$$\mathrm{Strip}_{\lambda,k,\vec{\theta}} = \mathbb{R}^k \setminus (\mathrm{Inner}_{k,\vec{\theta}} \cup \mathrm{Outer}_{\lambda,k,\vec{\theta}}).$$

We recall the main result of [HKM12]:

Theorem 4 (Invariance principle for polytopes, Theorem 3.1 of [HKM12])**.** *Let* $W \in \mathbb{R}^{n\times k}$ *be* τ*-regular with each column* W^j *satisfying* $\|W^j\| = 1$. *Then for all* $\vec{\theta} \in \mathbb{R}^k$, *we have*

$$\left| \Pr_{\boldsymbol{U}\leftarrow\{-1,1\}^n}\left[W^T\boldsymbol{U} \in \mathrm{Inner}_{k,\vec{\theta}}\right] - \Pr_{\boldsymbol{G}\leftarrow\mathcal{N}(0,1)^n}\left[W^T\boldsymbol{G} \in \mathrm{Inner}_{k,\vec{\theta}}\right] \right| = O\big((\log k)^{8/5}(\tau\log(1/\tau))^{1/5}\big).$$

We will also use the following anti-concentration bound for Gaussian random variables (which is an easy consequence of the $O(\sqrt{\log k})$ Gaussian surface area upper bound of Nazarov [Naz03] for intersections of k LTFs):

Theorem 5 (Anti-concentration bound for Gaussian random variables landing in a strip, Lemma 3.4 of [HKM12])**.** *For all* $\vec{\theta} \in \mathbb{R}^k$ *and all* $0 < \lambda < 1$, *we have*

$$\Pr_{\boldsymbol{G}\leftarrow\mathcal{N}(0,1)^n}\left[W^T\boldsymbol{G} \in \mathrm{Strip}_{\lambda,k,\vec{\theta}}\right] = O(\lambda\sqrt{\log k}).$$

IV. Our PRG and the statements of our main results

Our PRG for (k,s,τ)-intersections of LTFs is the generator Gen described in Section II-A, instantiated with the following parameters:

$$\ell = 1/\tau,$$
$$r_{\text{hash}} = 2\log k,$$
$$r_{\text{bucket}} = 4\log k + O((\log(M/\delta_{\text{CNF}}))^2$$

where

$$M = k \cdot 2^s \quad \text{and} \quad \delta_{\text{CNF}} = 1/\mathrm{poly}(n)$$

(the exact value for δ_{CNF} will be specified later). By standard constructions of r_{hash}-wise independent hash families and r_{bucket}-wise independent random variables, the total seed length of our generator is

$$O(\log(n\log\ell)\cdot r_{\text{hash}} + \ell\cdot(\log n)\cdot r_{\text{bucket}}) = O\left(\frac{1}{\tau}\cdot\log n\cdot(\log k + s + \log n)^2\right) = \mathrm{poly}(\log n, \log k, s, 1/\tau).$$

A. Formal statements of our main results

We begin with our most general PRG result:

Theorem 6 (Formal statement of Theorem 2)**.** *For all values of* $k, s \in \mathbb{N}$ *and* $\tau \in (0,1)$, *the pseudorandom generator* Gen *instantiated with the parameters above fools the class of* (k,s,τ)*-intersections of LTFs to accuracy*

$$\delta := O((\log k)^{8/5}(\tau\log(1/\tau))^{1/5})) \tag{2}$$

with seed length $\mathrm{poly}(\log n, \log k, s, 1/\tau)$.

Our PRG for the intersections of low-weight LTFs (Theorem 1) follows as a consequence of Theorem 6 via the following observation:

Observation 7 (Sparse-or-regular dichotomy)**.** Let $F(x) = \mathrm{sign}(w\cdot x - \theta)$ be a weight-t LTF. Then for any s, either F is s-sparse or F is $(t/\sqrt{s+1})$-regular.

Proof of Theorem 1 assuming Theorem 6: We fix

$$\tau := \tilde{\Theta}\left(\frac{\delta^5}{(\log k)^8}\right)$$

so as to satisfy (2). By Observation 7, we have that every weight-t LTF is either τ-regular or $(s := (t/\tau)^2)$-sparse. By our choice of τ, the parameters ℓ, r_{hash}, and r_{bucket} of the pseudorandom generator Gen instantiated with our parameters are all bounded by $\mathrm{poly}(\log n, \log k, t, 1/\delta)$, and hence the overall seed length is

$$O\left(\log(n\log\ell)\cdot r_{\text{hash}} + \ell\cdot(\log n)\cdot r_{\text{bucket}}\right)$$

which is indeed $\mathrm{poly}(\log n, \log k, t, 1/\delta)$ as claimed. ■

The remainder of this paper will be devoted to proving Theorem 6.

V. Fooling the smooth test function ψ^*_{k+1}

An intermediate goal, which in fact takes us most of the way to establishing Theorem 6, is to show that Gen fools a particular smooth test function $\psi^*_{\lambda,k+1,(\vec{\theta},0)}$. In this section we define this smooth test function, establish some of its basic properties, and formally state our intermediate goal (Theorem 8 below).

A. The smooth test function $\psi^*_{\lambda,k+1,(\vec{\theta},0)}$ and its basic properties

As discussed in Section II-C, our analysis crucially features a particular smooth function $\psi^*_{\lambda,k+1,(\vec{\theta},0)} : \mathbb{R}^{k+1} \to [-1,1]$, which is essentially the $(k+1)$-dimensional version of a function due to Bentkus [Ben90]. Fact V.1 below states the key properties of this function.

Fact V.1 (Main result of [Ben90], see Theorem 3.5 of [HKM12])**.** *For all positive integers k, $0 < \lambda < 1$, and $\vec{\theta} \in \mathbb{R}^k$, there exists a smooth function $\psi^*_{\lambda,k,\vec{\theta}} : \mathbb{R}^k \to [-1,1]$ such that the following holds: for every $s = 1,2,3,4$, we have $\|(\psi^*_{\lambda,k,\vec{\theta}})^{(s)}\|_1 \leq C \log^{s-1}(k+1)/\lambda^s$, and for all $v \in \mathbb{R}^k$, we have*

$$\psi^*_{\lambda,k,\vec{\theta}}(v) = \begin{cases} -1 & \text{if } v \in \mathrm{Inner}_{k,\vec{\theta}} \\ 1 & \text{if } v \in \mathrm{Outer}_{\lambda,k,\vec{\theta}} \\ \in [-1,1] & \text{otherwise (i.e. if } v \in \mathrm{Strip}_{\lambda,k,\vec{\theta}}). \end{cases} \tag{3}$$

For intuition, the test function $\psi^*_{\lambda,k,\vec{\theta}} : \mathbb{R}^k \to [-1,1]$ may loosely be thought of as a smooth approximation to the k-variable AND function; recall that on input $(b_1,\ldots,b_k) \in \{-1,1\}^k$, the AND function outputs -1 iff $(b_1,\ldots,b_k) = (-1,\ldots,-1)$. (We note that [HKM12] only require the $s = 4$ case of the above theorem (this is their Theorem 3.5), since in their framework they can obtain perfect cancellation of the first, second and third derivative terms in the relevant difference of Taylor expansions. In contrast we need to use all of the $s = 1,2,3,4$ cases.)

As mentioned earlier, in our analysis of $\psi^*_{\lambda,k+1,(\vec{\theta},0)}$ the last argument will always receive a Boolean value from $\{-1,1\}$ (corresponding to the output of the CNF G). We will use the following simple claim to control the behavior of $\psi^*_{\lambda,k+1,(\vec{\theta},0)}$ on inputs of this sort:

Claim V.2. *Given $0 < \lambda < 1, k \geq 1$, and $\vec{\theta} \in \mathbb{R}^k$, let $v \in \mathbb{R}^k$ be such that $v \notin \mathrm{Strip}_{\lambda,k,\vec{\theta}}$. Then both vectors $(v,-1) \in \mathbb{R}^{k+1}$ and $(v,1) \in \mathbb{R}^{k+1}$ lie outside of $\mathrm{Strip}_{\lambda,k+1,(\vec{\theta},0)}$.*

B. Towards Theorem 6: fooling the test function $\psi^*_{\lambda,k+1,(\vec{\theta},0)}$

As an intermediate step towards Theorem 6 we will first establish the following "pseudorandom generator" for the smooth function $\psi^*_{\lambda,k+1,(\vec{\theta},0)}$:

Theorem 8 (Gen fools the smooth test function $\psi^*_{\lambda,k+1,(\vec{\theta},0)}$)**.** *Let $H \wedge G$ be a (k,s,τ)-CNFLTF, and let $W \in \mathbb{R}^{n\times k}$ be the matrix of weight vectors (each of norm 1) of the τ-regular LTFs that comprise H, and $\vec{\theta} \in \mathbb{R}^k$ be the vector of their thresholds (so $\mathrm{sign}(W^j \cdot x - \theta_j)$ is the j-th LTF). For $0 < \lambda < 1$, let $\psi^*_{\lambda,k+1,(\vec{\theta},0)} : \mathbb{R}^{k+1} \to [-1,1]$ be as described in Fact V.1. Then when* Gen *is instantiated with the parameters from Section IV,*

$$\begin{aligned} &\Bigg| \mathop{\mathbf{E}}_{\boldsymbol{Y}\leftarrow \mathsf{Gen}} \left[\psi^*_{\lambda,k+1,(\vec{\theta},0)}(W^T\boldsymbol{Y}, G(\boldsymbol{Y}))\right] \\ &- \mathop{\mathbf{E}}_{\boldsymbol{U}\leftarrow\{-1,1\}^n} \left[\psi^*_{\lambda,k+1,(\vec{\theta},0)}(W^T\boldsymbol{U}, G(\boldsymbol{U}))\right] \Bigg| \\ &= O\Bigg(\frac{(\log k)^3}{\lambda^4} \left((\log k)^3 \cdot \tau \log(1/\tau) + \frac{1}{\tau} \cdot \delta_{\mathrm{CNF}} \cdot n^2 \right) \\ &\quad + \frac{1}{\tau} \left(\sqrt{\delta_{\mathrm{CNF}}} + \sum_{a=1}^{3} n^a \sqrt{\delta_{\mathrm{CNF}}} \cdot \frac{(\log k)^{a-1}}{\lambda^a} \right) \Bigg). \end{aligned} \tag{4}$$

VI. Setup for our coupling-based hybrid argument

We begin by defining the sequence of random variables that we will use to hybridize between $\boldsymbol{Y} \leftarrow \mathsf{Gen}$, the n-bit pseudorandom input, and $\boldsymbol{U}$, the n-bit uniform random input.

Definition 2 (Hybrid random variables)**.** *For any index $b \in \{0,1,\ldots,\ell\}$ and any hash $h\colon [n] \to [\ell]$, we define the hybrid random variable $\boldsymbol{X}^{h,b}$ over $\{-1,1\}^n$ as follows: Independently across each $c \in [\ell]$,*

- *If $c > b$, then the coordinates $\boldsymbol{X}^{h,b}_{h^{-1}(c)}$ of $\boldsymbol{X}^{h,b}$ are distributed according to a uniform random draw from $\{-1,1\}^n$;*
- *If $c \leq b$, then the coordinates $\boldsymbol{X}^{h,b}_{h^{-1}(c)}$ of $\boldsymbol{X}^{h,b}$ are distributed according to a draw from an r_{bucket}-wise independent random variable over $\{-1,1\}^n$.*

Let $\mathcal{H}$ be a $(2\log k)$-wise independent family of hashes $h : [n] \to [\ell]$. For each $b \in \{0,1,\ldots,\ell\}$, the hybrid random variable $\boldsymbol{X}^{\boldsymbol{h},b}$ is defined by drawing $\boldsymbol{h} \leftarrow \mathcal{H}$ and then taking $\boldsymbol{X}^{\boldsymbol{h},b}$ as above.

Remark 9. Note that $\boldsymbol{X}^{\boldsymbol{h},0}$ is a uniform random variable over $\{-1,1\}^n$ (indeed $\boldsymbol{X}^{h,0}$ is uniform for every fixed hash h), while $\boldsymbol{X}^{\boldsymbol{h},\ell}$ is distributed according to Gen.

A. Coupling adjacent random variables in the hybrid argument

Fix a hash $h\colon [n] \to [\ell]$, a bucket $b \in [\ell]$, and a restriction $\rho \in \{-1,1\}^{[n]\setminus h^{-1}(b)}$ fixing the variables outside bucket $h^{-1}(b)$. Recall that $\boldsymbol{X}^{h,b-1}$ is distributed according to the uniform distribution within $h^{-1}(b)$, and $\boldsymbol{X}^{h,b}$ is distributed according to a r_{bucket}-wise independent distribution within this same bucket $h^{-1}(b)$. For the remainder of this paper, for notational clarity unless otherwise indicated $\boldsymbol{U}$ denotes a uniformly distributed random variable over $\{-1,1\}^{h^{-1}(b)}$ and $\boldsymbol{Z}$ denotes a r_{bucket}-wise independent random variable over $\{-1,1\}^{h^{-1}(b)}$.

Our CNF-fooling-based coupling: By the results of Bazzi and Razborov (Theorem 3) and the choice of r_{bucket} from Section IV, the random variable $\boldsymbol{Z}$ δ_{CNF}-fools $G \restriction \rho$ (which, like G, is an M-clause CNF). Consequently there exists a coupling $(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})$ between $\boldsymbol{U}$ and $\boldsymbol{Z}$ such that

$$\Pr_{(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})}\left[(G \restriction \rho)(\widehat{\boldsymbol{U}}) \neq (G \restriction \rho)(\widehat{\boldsymbol{Z}})\right] \leq \delta_{\mathrm{CNF}}. \tag{5}$$

(Note that this coupling depends on $G \restriction \rho$.)

Consider the following joint distribution over a pair of random variables $(\widehat{\boldsymbol{X}}^{h,b-1}(\rho), \widehat{\boldsymbol{X}}^{h,b}(\rho))$, both supported on $\{-1,1\}^n$: First make a draw $(\widehat{U}, \widehat{Y}) \leftarrow (\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})$, and output $(\widehat{X}^{h,b-1}(\rho), \widehat{X}^{h,b}(\rho))$ where

- $\widehat{X}^{h,b-1}(\rho)$ assigns variables according to $\widehat{U}$ within $h^{-1}(b)$, and according to ρ outside $h^{-1}(b)$.
- $\widehat{X}^{h,b}(\rho)$ assigns variables according to $\widehat{Y}$ within $h^{-1}(b)$, and according to ρ outside $h^{-1}(b)$.

Remark 10. Note that for $\boldsymbol{\rho} \leftarrow \boldsymbol{X}^{h,b}_{[n]\setminus h^{-1}(b)}$, we have that $\widehat{\boldsymbol{X}}^{h,b-1}(\boldsymbol{\rho})$ is distributed identically as $\boldsymbol{X}^{h,b-1}$ and likewise $\widehat{\boldsymbol{X}}^{h,b}(\boldsymbol{\rho})$ is distributed identically as $\boldsymbol{X}^{h,b}$.

VII. The hybrid argument: Proof of Theorem 8

Throughout this section for notational clarity we simply write ψ instead of $\psi^*_{\lambda,k+1,(\vec{\theta},0)}$. We also write $F_\psi : \{-1,1\}^n \to [-1,1]$ to denote the function

$$F_\psi(x) = \psi(W^T x, G(x)).$$

Our core technical result, which we prove in Section VIII, is the following:

Lemma VII.1 (Error incurred in one step of hybrid)**.** *For all hashes $h : [n] \to [\ell]$, buckets $b \in [\ell]$, and restrictions $\rho \in \{-1,1\}^{[n]\setminus h^{-1}(b)}$, we have that*

$$\begin{aligned}&\left|\mathbf{E}\left[F_\psi(\widehat{\boldsymbol{X}}^{h,b-1}(\rho))\right] - \mathbf{E}\left[F_\psi(\widehat{\boldsymbol{X}}^{h,b}(\rho))\right]\right| \\ &= O\Bigg(\frac{(\log k)^3}{\lambda^4}\left((\log k)^2 \cdot h(W,b) + \delta_{\mathrm{CNF}} \cdot n^2\right) \\ &\qquad + \sqrt{\delta_{\mathrm{CNF}}} + \sum_{a=1}^{3} n^a \sqrt{\delta_{\mathrm{CNF}}} \cdot \frac{(\log k)^{a-1}}{\lambda^a}\Bigg),\end{aligned} \tag{6}$$

where

$$h(W,b) := \left(\sum_{j=1}^{k} \|W^j_{h^{-1}(b)}\|^{4\log k}\right)^{1/\log k}.$$

The following corollary follows as an immediate consequence of Lemma VII.1, Remark 10, and the triangle inequality:

Corollary VII.2 (Averaging Lemma VII.1 over $\boldsymbol{\rho}$ and summing over $b \in [\ell]$)**.** *For all hashes $h : [n] \to [\ell]$, we have that*

$$\begin{aligned}&\left|\mathbf{E}\left[F_\psi(\boldsymbol{X}^{h,0})\right] - \mathbf{E}\left[F_\psi(\boldsymbol{X}^{h,\ell})\right]\right| \\ &= \frac{O((\log k)^3)}{\lambda^4} \cdot (\log k)^2 \cdot \sum_{b=1}^{\ell} h(W,b) \\ &\quad + \ell \cdot O\Bigg(\frac{(\log k)^3}{\lambda^4} \cdot \delta_{\mathrm{CNF}} \cdot n^2 + \sqrt{\delta_{\mathrm{CNF}}} \\ &\qquad + \sum_{a=1}^{3} n^a \sqrt{\delta_{\mathrm{CNF}}} \cdot \frac{(\log k)^{a-1}}{\lambda^a}\Bigg).\end{aligned}$$

We do not have a good bound on the quantity $h(W,b)$ for an arbitrary hash $h : [n] \to [\ell]$ and bucket $b \in [\ell]$. Instead, we shall use the following:

Lemma VII.3 (Lemma 4.1 of [HKM12])**.** *For $\ell = 1/\tau$ and $\boldsymbol{h}$ drawn from a $(2\log k)$-wise independent hash family $\mathcal{H}$,*

$$\begin{aligned}\mathbf{E}\left[\sum_{b=1}^{\ell} \boldsymbol{h}(W,b)\right] &\leq \sum_{b=1}^{\ell}\left(\mathbf{E}\left[\sum_{j=1}^{k} \|W^j_{h^{-1}(b)}\|^{4\log k}\right]\right)^{1/\log k} \\ &\leq 4\log k \cdot \tau \log(1/\tau).\end{aligned}$$

(The middle quantity is what [HKM12] denotes by $\mathcal{H}(W)$ and is the quantity they bound; the left inequality is by the power-mean inequality.)

We are now ready to prove Theorem 8:

Proof of Theorem 8 assuming Lemma VII.1:

$$\begin{aligned}&\Bigg|\mathop{\mathbf{E}}_{\boldsymbol{Y}\leftarrow\mathsf{Gen}}\left[\psi^*_{\lambda,k+1,(\vec{\theta},0)}(W^T\boldsymbol{Y}, G(\boldsymbol{Y}))\right] \\ &\qquad - \mathop{\mathbf{E}}_{\boldsymbol{U}\leftarrow\{-1,1\}^n}\left[\psi^*_{\lambda,k+1,(\vec{\theta},0)}(W^T\boldsymbol{U}, G(\boldsymbol{U}))\right]\Bigg| \\ &= \left|\mathbf{E}\left[F_\psi(\boldsymbol{X}^{\boldsymbol{h},0})\right] - \mathbf{E}\left[F_\psi(\boldsymbol{X}^{\boldsymbol{h},\ell})\right]\right| \\ &\qquad\qquad \text{(Remark 9 and definition of } F_\psi\text{)} \\ &\leq \mathop{\mathbf{E}}_{\boldsymbol{h}\leftarrow\mathcal{H}}\left[\left|\mathbf{E}\left[F_\psi(\boldsymbol{X}^{\boldsymbol{h},0})\right] - \mathbf{E}\left[F_\psi(\boldsymbol{X}^{\boldsymbol{h},\ell})\right]\right|\right] \\ &= O\Bigg(\frac{(\log k)^3}{\lambda^4}\left((\log k)^3 \cdot \tau\log(1/\tau) + \frac{1}{\tau} \cdot \delta_{\mathrm{CNF}} \cdot n^2\right) \\ &\qquad + \frac{1}{\tau}\left(\sqrt{\delta_{\mathrm{CNF}}} + \sum_{a=1}^{3} n^a \sqrt{\delta_{\mathrm{CNF}}} \cdot \frac{(\log k)^{a-1}}{\lambda^a}\right)\Bigg),\end{aligned}$$

where the final equality is by Corollary VII.2, Lemma VII.3, and recalling that $\ell = 1/\tau$. ■

VIII. A single step of the hybrid argument: Proof of Lemma VII.1

Fix a hash $h : [n] \to [\ell]$, a bucket $b \in [\ell]$, and a restriction $\rho \in \{-1,1\}^{[n]\setminus h^{-1}(b)}$. As is standard in applications of the Lindeberg method, we will express $F_\psi(\widehat{\boldsymbol{X}}^{h,b-1}(\rho))$ and $F_\psi(\widehat{\boldsymbol{X}}^{h,b}(\rho))$ as $\psi(\boldsymbol{v} + \boldsymbol{\Delta}^{\mathrm{unif}})$ and $\psi(\boldsymbol{v} + \boldsymbol{\Delta}^{\mathrm{pseudo}})$ respectively, where $\boldsymbol{v}$ is common to both random variables. (Very roughly speaking, the Lindeberg method employs Taylor's theorem to show that quantities such as (6) are small if $\boldsymbol{\Delta}^{\text{unif}}$ and $\boldsymbol{\Delta}^{\text{pseudo}}$ are sufficiently "small" and ψ is sufficiently "nice."). We now describe the choice of random variables $\boldsymbol{v}, \boldsymbol{\Delta}^{\text{unif}}, \boldsymbol{\Delta}^{\text{pseudo}} \in \mathbb{R}^{k+1}$ to accomplish this.

We define $v : \{-1,1\}^{h^{-1}(b)} \to \mathbb{R}^{k+1}$ as follows:

$$\begin{aligned} v(x)_j &= \sum_{i \in [n] \setminus h^{-1}(b)} W_i^j \rho_i \qquad \text{for } j \in [k], \\ v(x)_{k+1} &= (G \upharpoonright \rho)(x). \end{aligned}$$

Recalling that ρ is a fixed restriction, we observe that only the final coordinate of v depends on its input x. We further define $\Delta^{\text{unif}} : \{-1,1\}^{h^{-1}(b)} \to \mathbb{R}^{k+1}$ and $\Delta^{\text{pseudo}} : \{-1,1\}^{h^{-1}(b)} \times \{-1,1\}^{h^{-1}(b)} \to \mathbb{R}^{k+1}$ as follows:

$$\begin{aligned} \Delta^{\text{unif}}(x)_j &= \sum_{i \in h^{-1}(b)} W_i^j x_i \qquad \text{for } j \in [k], \\ \Delta^{\text{unif}}(x)_{k+1} &= 0, \end{aligned} \tag{7}$$

and

$$\begin{aligned} \Delta^{\text{pseudo}}(x,z)_j &= \sum_{i \in h^{-1}(b)} W_i^j z_i \qquad \text{for } j \in [k], \\ \Delta^{\text{pseudo}}(x,z)_{k+1} &= (G \upharpoonright \rho)(z) - (G \upharpoonright \rho)(x). \end{aligned}$$

We observe that

$$\begin{aligned} F_\psi(\widehat{\boldsymbol{X}}^{h,b-1}(\rho)) &\equiv \psi(v(\boldsymbol{U}) + \Delta^{\text{unif}}(\boldsymbol{U})) \\ F_\psi(\widehat{\boldsymbol{X}}^{h,b}(\rho)) &\equiv \psi(v(\widehat{\boldsymbol{U}}) + \Delta^{\text{pseudo}}(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})), \end{aligned}$$

and so the desired quantity (6) of Lemma VII.1 that we wish to upper bound may be re-expressed as

$$\begin{aligned} (6) &= \left| \mathbf{E}[F_\psi(\widehat{\boldsymbol{X}}^{h,b-1}(\rho))] - \mathbf{E}[F_\psi(\widehat{\boldsymbol{X}}^{h,b}(\rho))] \right| \\ &= \Bigg| \mathop{\mathbf{E}}_{\boldsymbol{U}} \left[\psi(v(\boldsymbol{U}) + \Delta^{\text{unif}}(\boldsymbol{U})) \right] \\ &\qquad - \mathop{\mathbf{E}}_{(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})} \left[\psi(v(\widehat{\boldsymbol{U}}) + \Delta^{\text{pseudo}}(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})) \right] \Bigg|. \end{aligned} \tag{8}$$

We observe that unlike standard Lindeberg-style proofs of invariance principles and associated pseudorandomness results, in our setup $v(\boldsymbol{U})$ and $\Delta^{\text{unif}}(\boldsymbol{U})$ are not independent, and likewise neither are $v(\widehat{\boldsymbol{U}})$ and $\Delta^{\text{pseudo}}(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})$. This motivates the definitions of the following subsection.

A. Mixtures of conditional distributions

Let $\boldsymbol{U}^1$ denote the distribution $\boldsymbol{U}$ conditioned on outcomes $x \in \{-1,1\}^{h^{-1}(b)}$ such that $(G \upharpoonright \rho)(x) = 1$, and similarly $\boldsymbol{U}^{-1}$. Equivalently, $\boldsymbol{U}^1$ and $\boldsymbol{U}^{-1}$ are uniform distributions over $(G \upharpoonright \rho)^{-1}(1)$ and $(G \upharpoonright \rho)^{-1}(-1)$ respectively. We note that $\boldsymbol{U}$ can be expressed as the mixture of $\boldsymbol{U}^1$ and $\boldsymbol{U}^{-1}$ with mixing weights

$$\begin{aligned} \pi_1 &:= \mathop{\mathbf{Pr}}_{\boldsymbol{U}} \left[(G \upharpoonright \rho)(\boldsymbol{U}) = 1 \right] \\ \pi_{-1} &:= \mathop{\mathbf{Pr}}_{\boldsymbol{U}} \left[(G \upharpoonright \rho)(\boldsymbol{U}) = -1 \right]. \end{aligned}$$

We may suppose without loss of generality that $\mathbf{Pr}_{\boldsymbol{U}}[(G \upharpoonright \rho)(\boldsymbol{U}) = -1] \geq \mathbf{Pr}_{\boldsymbol{Z}}[(G \upharpoonright \rho)(\boldsymbol{Z}) = -1]$ (the other case is entirely similar).

Next, we similarly express the joint distribution $(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})$ as the mixture of conditional distributions $(\widehat{\boldsymbol{U}}^1, \widehat{\boldsymbol{Z}}^1)$, $(\widehat{\boldsymbol{U}}^{-1}, \widehat{\boldsymbol{Z}}^{-1})$, $(\widehat{\boldsymbol{U}}^{\text{err}}, \widehat{\boldsymbol{Z}}^{\text{err}})$, where

- $(\widehat{\boldsymbol{U}}^1, \widehat{\boldsymbol{Z}}^1)$ is supported on pairs (x,z) such that $(G \upharpoonright \rho)(x) = (G \upharpoonright \rho)(z) = 1$
- $(\widehat{\boldsymbol{U}}^{-1}, \widehat{\boldsymbol{Z}}^{-1})$ is supported on pairs (x,z) such that $(G \upharpoonright \rho)(x) = (G \upharpoonright \rho)(z) = -1$
- $(\widehat{\boldsymbol{U}}^{\text{err}}, \widehat{\boldsymbol{Z}}^{\text{err}})$ is supported on pairs (x,z) such that $(G \upharpoonright \rho)(x) = -1, (G \upharpoonright \rho)(z) = 1$.

The mixing weights are $\tilde{\pi}_1, \tilde{\pi}_{-1}$, and $\tilde{\pi}_{\text{err}}$ respectively, where

$$\tilde{\pi}_1 = \pi_1, \qquad \tilde{\pi}_{-1} = \pi_{-1} - \tilde{\pi}_{\text{err}}, \qquad \tilde{\pi}_{\text{err}} \leq \delta_{\text{CNF}}$$

and the bound $\tilde{\pi}_{\text{err}} \leq \delta_{\text{CNF}}$ follows from (5). We stress that while $\widehat{\boldsymbol{U}}^1$ is distributed identically as $\boldsymbol{U}^1$, this is not the case for $\widehat{\boldsymbol{U}}^{-1}$ and $\boldsymbol{U}^{-1}$, because of the small fraction of pairs that do not align perfectly under the coupling $(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})$ and are captured by $(\widehat{\boldsymbol{U}}^{\text{err}}, \widehat{\boldsymbol{Z}}^{\text{err}})$.

Proposition VIII.1 (Expressing $\boldsymbol{U}$ and $(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})$ as mixtures of conditional distributions)**.** *For any function $f : \{-1,1\}^{h^{-1}(b)} \to \mathbb{R}$,*

$$\mathop{\mathbf{E}}_{\boldsymbol{U}} \left[f(\boldsymbol{U}) \right] = \pi_1 \mathop{\mathbf{E}}_{\boldsymbol{U}^1} \left[f(\boldsymbol{U}^1) \right] + \pi_{-1} \mathop{\mathbf{E}}_{\boldsymbol{U}^{-1}} \left[f(\boldsymbol{U}^{-1}) \right].$$

Similarly, for any function $f : \{-1,1\}^{h^{-1}(b)} \times \{-1,1\}^{h^{-1}(b)} \to \mathbb{R}$,

$$\begin{aligned} &\mathop{\mathbf{E}}_{(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}})} \left[f(\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{Z}}) \right] \\ &= \tilde{\pi}_1 \mathop{\mathbf{E}}_{(\widehat{\boldsymbol{U}}^1, \widehat{\boldsymbol{Z}}^1)} \left[f(\widehat{\boldsymbol{U}}^1, \widehat{\boldsymbol{Z}}^1) \right] + \tilde{\pi}_{-1} \mathop{\mathbf{E}}_{(\widehat{\boldsymbol{U}}^{-1}, \widehat{\boldsymbol{Z}}^{-1})} \left[f(\widehat{\boldsymbol{U}}^{-1}, \widehat{\boldsymbol{Z}}^{-1}) \right] \\ &\qquad + \tilde{\pi}_{\text{err}} \mathop{\mathbf{E}}_{(\widehat{\boldsymbol{U}}^{\text{err}}, \widehat{\boldsymbol{Z}}^{\text{err}})} \left[f(\widehat{\boldsymbol{U}}^{\text{err}}, \widehat{\boldsymbol{Z}}^{\text{err}}) \right] \\ &= \pi_1 \mathop{\mathbf{E}}_{(\widehat{\boldsymbol{U}}^1, \widehat{\boldsymbol{Z}}^1)} \left[f(\widehat{\boldsymbol{U}}^1, \widehat{\boldsymbol{Z}}^1) \right] + \pi_{-1} \mathop{\mathbf{E}}_{(\widehat{\boldsymbol{U}}^{-1}, \widehat{\boldsymbol{Z}}^{-1})} \left[f(\widehat{\boldsymbol{U}}^{-1}, \widehat{\boldsymbol{Z}}^{-1}) \right] \\ &\qquad \pm 2\, \delta_{\text{CNF}} \cdot \|f\|_\infty. \end{aligned}$$

These conditional distributions are useful because of the following two simple but crucial observations:

Observation 11 ($\boldsymbol{v}$ becomes constant)**.** Fix $c \in \{-1,1\}$. For all $x \in \text{supp}(\boldsymbol{U}^c)$ we have that $v(x)$ is the same fixed vector $v^* \in \mathbb{R}^{k+1}$ given by

$$\begin{aligned} v_j^* &= \sum_{i \in [n] \setminus h^{-1}(b)} W_i^j \rho_i \qquad \text{for } j \in [k], \\ v_{k+1}^* &= (G \upharpoonright \rho)(x) = c. \end{aligned}$$

The same is true for $\widehat{\boldsymbol{U}}^c$: for all $x \in \text{supp}(\widehat{\boldsymbol{U}}^c)$ we have $v(x) = v^*$.

Note that as a consequence of Observation 11, the random variables $v(\boldsymbol{U}^c)$ and $\Delta^{\mathrm{unif}}(\boldsymbol{U}^c)$ are independent for $c \in \{-1,1\}$, and likewise $v(\widehat{\boldsymbol{U}}^c)$ and $\Delta^{\mathrm{pseudo}}(\widehat{\boldsymbol{U}}^c, \widehat{\boldsymbol{Z}}^c)$ are independent as well; cf. our remark following Equation (8). The next observation further motivates our couplings $(\widehat{\boldsymbol{U}}^1, \widehat{\boldsymbol{Z}}^1)$ and $(\widehat{\boldsymbol{U}}^{-1}, \widehat{\boldsymbol{Z}}^{-1})$:

Observation 12 **($\Delta^{\mathrm{pseudo}}_{k+1} = 0$).** Fix $c \in \{-1,1\}$. For all $(\widehat{U}, \widehat{Z}) \in \mathrm{supp}(\widehat{\boldsymbol{U}}^c, \widehat{\boldsymbol{Z}}^c)$, we have

$$\Delta^{\mathrm{pseudo}}_{k+1}(\widehat{U}, \widehat{Z}) = (G \upharpoonright \rho)(\widehat{Z}) - (G \upharpoonright \rho)(\widehat{U}) = 0.$$

Due to space considerations, the remainder of the proof of Lemma VII.1 is deferred to the full version. In the full version, we apply Proposition VIII.1 to bound the RHS of (8) by:

$$\sum_{c\in\{-1,1\}} \pi_c \cdot \Bigg| \mathop{\mathbf{E}}_{\boldsymbol{U}^c}\left[\psi(v(\boldsymbol{U}^c) + \Delta^{\mathrm{unif}}(\boldsymbol{U}^c))\right] - \mathop{\mathbf{E}}_{(\widehat{\boldsymbol{U}}^c, \widehat{\boldsymbol{Z}}^c)}\left[\psi(v(\widehat{\boldsymbol{U}}^c) + \Delta^{\mathrm{pseudo}}(\widehat{\boldsymbol{U}}^c, \widehat{\boldsymbol{Z}}^c))\right]\Bigg|,$$

which we then proceed to analyze via the Taylor expansion of $\psi(v + \Delta)$ (Fact III.1). Having established Lemma VII.1, and hence Theorem 8, we then show how Theorem 8 yields Theorem 6.

Acknowledgment

RAS is supported by NSF grants CCF-1420349 and CCF-1563155. LYT is supported by NSF grant CCF-1563122.

References

[Baz07] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. In *Proc. 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 63–73, 2007. (document), II, II-C, II-C1, III

[Ben90] Vidmantas Bentkus. Smooth approximations of the norm and differentiable functions with bounded support in Banach space l^k_∞. *Lithuan. Math. J.*, 30(3):223–230, 1990. 2, II-C, V-A, V.1

[Ber41] 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. I-A

[BK97] Avrim Blum and Ravi Kannan. Learning an intersection of a constant number of halfspaces under a uniform distribution. *Journal of Computer and System Sciences*, 54(2):371–380, 1997. I

[Cho61] Chao-Kong Chow. On the characterization of threshold functions. In *Proceedings of the Symposium on Switching Circuit Theory and Logical Design (FOCS)*, pages 34–38, 1961. I

[CSS16] Ruiwen Chen, Rahul Santhanam, and Srikanth Srinivasan. Average-case lower bounds and satisfiability algorithms for small threshold circuits. In *Proceedings of the 31st Conference on Computational Complexity (CCC)*, 2016. I-A1

[DDS14] 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. I-A1

[DDS16] Anindya De, Ilias Diakonikolas, and Rocco A. Servedio. A robust Khintchine inequality, and algorithms for computing optimal constants in Fourier analysis and high-dimensional geometry. *SIAM J. Discrete Math.*, 30(2):1058–1094, 2016. I-A1

[DGJ+10] 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. I, I-A, I-A1

[DKN10] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *Proc. 51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–20, 2010. I-A1

[DRST14] Ilias Diakonikolas, Prasad Raghavendra, Rocco A. Servedio, and Li-Yang Tan. Average sensitivity and noise sensitivity of polynomial threshold functions. *SIAM Journal on Computing*, 43(1):231–253, 2014. I-A1

[DS13] Ilias Diakonikolas and Rocco A. Servedio. Improved approximation of linear threshold functions. *Computational Complexity*, 22(3):623–677, 2013. I-A1

[DS14] Anindya De and Rocco A. Servedio. Efficient deterministic approximate counting for low-degree polynomial threshold functions. In *Proceedings of the 46th Annual Symposium on Theory of Computing (STOC)*, pages 832–841, 2014. I-A1

[DSTW14] Ilias Diakonikolas, Rocco A. Servedio, Li-Yang Tan, and Andrew Wan. A regularity lemma and low-weight approximators for low-degree polynomial threshold functions. *Theory of Computing*, 10:27–53, 2014. I-A1

[Ess42] Carl-Gustav Esseen. On the Liapunoff limit of error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, A:1–19, 1942. I-A

[FGRW09] Vitaly Feldman, Venkatesan Guruswami, Prasad Raghavendra, and Yi Wu. Agnostic learning of monomials by halfspaces is hard. In *Proc. 50th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 385–394, 2009. I-A1

[GHR92] Mikhail Goldmann, Johan Håstad, and Alexander Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992. I

[GKM15] Parikshit Gopalan, Daniel M. Kane, and Raghu Meka. Pseudorandomness via the discrete fourier transform. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS*, pages 903–922, 2015. I, I-A

[GL94] Craig Gotsman and Nathan Linial. Spectral properties of threshold functions. *Combinatorica*, 14(1):35–50, 1994. I

[GOWZ10] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *Proceedings of the 25th Annual Conference on Computational Complexity (CCC)*, pages 223–234, 2010. I, I-A, I-A1, I-B, II-C, II-C, II-C1

[Hås94] Johan Håstad. On the size of weights for threshold gates. *SIAM Journal on Discrete Mathematics*, 7(3):484–492, 1994. I, I-B

[HKM12] Prahladh Harsha, Adam R. Klivans, and Raghu Meka. An invariance principle for polytopes. *J. ACM*, 59(6):29:1–29:25, 2012. (document), I, I-A, I-A1, I-B, I-B, II, II-A, II-B, 1, 1, 2, II-C, II-C, II-C1, III, III, 4, 5, V.1, V-A, VII.3, VII

[Hon87] Jiawei Hong. On connectionist models. Technical Report Technical Report 87-012, Dept. of Computer Science, University of Chicago, 1987. I-B

[Kan14] Daniel M. Kane. The average sensitivity of an intersection of half spaces. In *Symposium on Theory of Computing (STOC)*, pages 437–440, 2014. I

[KOS04] Adam Klivans, Ryan O'Donnell, and Rocco A. Servedio. Learning intersections and thresholds of halfspaces. *Journal of Computer & System Sciences*, 68(4):808–840, 2004. I

[KOS08] Adam Klivans, Ryan O'Donnell, and Rocco A. Servedio. Learning geometric concepts via Gaussian surface area. In *Proceedings of the 49th Symposium on Foundations of Computer Science (FOCS)*, pages 541–550, 2008. I

[KS06] Adam Klivans and Alexander Sherstov. Cryptographic hardness for learning intersections of halfspaces. In *Proc. 47th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 553–562, 2006. I

[KS11] Subhash Khot and Rishi Saket. On the hardness of learning intersections of two halfspaces. *J. Comput. Syst. Sci.*, 77(1):129–141, 2011. I

[MORS09] Kevin Matulef, Ryan O'Donnell, Ronitt Rubinfeld, and Rocco A. Servedio. Testing ±1-weight halfspaces. In *APPROX-RANDOM*, pages 646–657, 2009. I

[MORS10] Kevin Matulef, Ryan O'Donnell, Ronitt Rubinfeld, and Rocco A. Servedio. Testing halfspaces. *SIAM J. on Comput.*, 39(5):2004–2047, 2010. I

[MTT61] Saburo Muroga, Iwao Toda, and Satoru Takasu. Theory of majority switching elements. *J. Franklin Institute*, 271(5):376–418, 1961. I-B

[MZ13] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM J. Comput.*, 42(3):1275–1301, 2013. I, I-A, I-A, I-A, I-A1, II, II-A, II-B, 1a, II-C, II-C1

[Naz03] 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 (2001-2002)*, pages 169–187. Lecture Notes in Math., Vol. 1807, Springer, 2003. 2, III

[Nis93] Noam Nisan. The communication complexity of threshold gates. In *In Proceedings of Combinatorics, Paul Erdos is Eighty*, pages 301–315, 1993. I

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. Available at http://analysisofbooleanfunctions.org/. I, II-C1

[Per04] Yuval Peres. Noise stability of weighted majority, 2004. Available at http://arxiv.org/abs/math/0412377. I

[Rag88] Prabhakar Raghavan. Learning in threshold networks. In *First Workshop on Computational Learning Theory*, pages 19–27, 1988. I-B

[Raz92] Alexander Razborov. On small depth threshold circuits. In *Proceedings of the Third Scandinavian Workshop on Algorithm Theory (SWAT)*, pages 42–52, 1992. I

[Raz09] Alexander Razborov. A simple proof of Bazzi's theorem. *ACM Trans. Comput. Theory*, 1(1):3:1–3:5, February 2009. (document), II, II-C, II-C1, III

[Ser07] Rocco A. Servedio. Every linear threshold function has a low-weight approximator. *Comput. Complexity*, 16(2):180–209, 2007. I, I-A1

[She13a] Alexander A. Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013. I

[She13b] Alexander A. Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013. I

[SO03] Jirí Síma and Pekka Orponen. General-purpose computation with neural networks: A survey of complexity theoretic results. *Neural Computation*, 15(12):2727–2778, 2003. I

[Tao10] Terence Tao. 254A Notes: Topics in random matrix theory. https://terrytao.wordpress.com/tag/lindeberg-replacement-trick/, 2010. II-C1

[Vem10] Santosh Vempala. A random-sampling-based algorithm for learning intersections of halfspaces. *J. ACM*, 57(6:32), 2010. I

[Vio15] Emanuele Viola. The communication complexity of addition. *Combinatorica*, 35(6):703–747, 2015. I