Fooling Polytopes

Ryan O'Donnell
Carnegie Mellon University
USA

Rocco A. Servedio
Columbia University
USA

Li-Yang Tan
Stanford University
USA

ABSTRACT

We give a pseudorandom generator that fools m -facet polytopes over $\{0, 1\}^n$ with seed length $\text{polylog}(m) \cdot \log n$. The previous best seed length had superlinear dependence on m . An immediate consequence is a deterministic quasipolynomial time algorithm for approximating the number of solutions to any $\{0, 1\}$ -integer program.

CCS CONCEPTS

• **Theory of computation** → **Pseudorandomness and derandomization**.

KEYWORDS

polytopes, intersections of halfspaces, pseudorandom generators, derandomization, central limit theorems

ACM Reference Format:

Ryan O'Donnell, Rocco A. Servedio, and Li-Yang Tan. 2019. Fooling Polytopes. In *Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing (STOC '19)*, June 23–26, 2019, Phoenix, AZ, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3313276.3316321>

1 INTRODUCTION

Unconditional derandomization has been a major focus of research in computational complexity theory for more than thirty years. A significant line of work in this area has been on developing unconditional pseudorandom generators (PRGs) for various types of Boolean functions. Early seminal results in this vein focused on Boolean circuits [1, 31, 33] and branching programs [15, 32, 34], but over the past decade or so a new strand of research has emerged in which the goal is to construct PRGs against *halfspaces* and various generalizations of halfspaces. This work has included a sequence of successively more efficient PRGs against halfspaces [7, 10, 19, 21, 26, 28], low-degree polynomial threshold functions [8, 16, 17, 19, 20, 28], and, most relevant to this paper, *intersections of halfspaces* [6, 13, 14, 39].

Since intersections of m halfspaces correspond to m -facet polytopes, and also to $\{0, 1\}$ -integer programs with m constraints, these objects are of fundamental interest in high-dimensional geometry, optimization, and a range of other areas. A pseudorandom generator that δ -fools intersections of m halfspaces can equivalently be viewed as an explicit *discrepancy set* for m -facet polytopes: a small

subset of $\{0, 1\}^n$ that δ -approximates the $\{0, 1\}^n$ -volume of every m -facet polytope. (Discrepancy sets are stricter versions of *hitting sets*, which are only required to intersect every polytope of volume at least δ .) The problem of constructing a PRG for intersections of m halfspaces is also a stricter version of the algorithmic problem of deterministically approximating the number of solutions of a $\{0, 1\}$ -integer program with m constraints. It is stricter because a PRG yields an input-oblivious algorithm: the range of a PRG is a single fixed set of points which gives approximately the right answer for *every* $\{0, 1\}$ -integer program. Beyond pseudorandomness, intersections of halfspaces also play a significant role in other fields such as concrete complexity theory [3, 18, 29, 36, 41, 42] and computational learning theory [5, 11, 22–25, 40, 45].

The main result of this paper is a new PRG for intersections of m halfspaces. Its seed length grows polylogarithmically with m , which is an exponential improvement of the previous best PRG for this class. Before giving the precise statement of our result, we briefly describe the prior state of the art for this problem.

1.1 Prior Work on PRGs for Intersections of Halfspaces

A halfspace $F(x) = \mathbb{1}[w \cdot x \leq \theta]$ is said to be τ -regular if $|w_j| \leq \tau \|w\|_2$ for all $j \in [n]$; intuitively, a τ -regular halfspace is one in which no coefficient w_j is too large relative to the overall scale of all the coefficients. Harsha, Klivans, and Meka [14] gave a PRG which δ -fools any intersection of m many τ -regular halfspaces with seed length $\text{poly}(\log m, 1/\delta) \cdot \log n$, where τ has to be sufficiently small relative to m and δ (specifically, $\tau \leq \text{some poly}(\frac{\delta}{\log m})$ is required). While this seed length has the desirable property of being polylogarithmic in m , due to the regularity requirement this result cannot be used to fool intersections of even two general halfspaces. We note that there are very basic halfspaces, such as $F(x) = \mathbb{1}[x_1 \leq 1/2]$, that are highly irregular.

Recently, [39] built on the work of [14] to give a PRG that fools a different subclass of intersections of halfspaces. They give a PRG that δ -fools any intersection of m many *weight- W* halfspaces with seed length $\text{poly}(\log m, W, 1/\delta) \cdot \text{polylog } n$; a halfspace has weight W if it can be expressed as $\mathbb{1}[w \cdot x \leq \theta]$ where each coefficient w_j is an integer of magnitude at most W . Unfortunately, many n -variable halfspaces require weight polynomially or even exponentially large in n ; in fact, a counting argument shows that almost all halfspaces require exponentially large weight. Therefore, the [39] result also cannot be used to fool even two general halfspaces.

In [13], Gopalan, O'Donnell, Wu, and Zuckerman gave a PRG that *can* fool intersections of m general halfspaces. However, various aspects of their approach each necessitate a seed length which is at least linear in m , and indeed their overall seed length is

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

STOC '19, June 23–26, 2019, Phoenix, AZ, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6705-9/19/06...\$15.00

<https://doi.org/10.1145/3313276.3316321>

$O((m \log(m/\delta) + \log n) \cdot \log(m/\delta))$.¹ So while this PRG is notable for being able to handle intersections of general halfspaces, its seed length becomes trivial (greater than n) for intersections of $m \geq n$ many halfspaces. (Indeed, this PRG of [13] fools *arbitrary monotone functions* of m general halfspaces, with intersections (i.e. ANDs) being a special case. Due to the generality of this class—which of course includes every monotone function over $\{0, 1\}^m$ —it can be shown that any PRG for it has to have at least linear seed length dependence on m .)

1.1.1 PRGs over Gaussian Space. There has also been work on PRGs for functions over $\mathbb{R}^n$ endowed with the n -dimensional Gaussian distribution. Analyses in this setting are often facilitated by the continuous nature of $\mathbb{R}^n$ and rotational invariance of the Gaussian distribution, useful technical properties not afforded by the standard setting of Boolean space. For halfspaces and polytopes, PRGs over Gaussian space can be viewed as a first step towards PRGs over Boolean space; as we describe below, Boolean PRGs even for restricted subclasses of halfspaces and polytopes yield Gaussian PRGs for general halfspaces and polytopes, but the converse does not hold. We also note that the correspondence between polytopes and $\{0, 1\}$ -integer programs is specific to Boolean space, and in particular, Gaussian PRGs do not yield algorithms for counting solutions to these programs.

For halfspaces, Meka and Zuckerman [28] showed that any PRG for the subclass of $O(\frac{1}{\sqrt{n}})$ -regular halfspaces over Boolean space yields a PRG for all halfspaces over Gaussian space. Note that $O(\frac{1}{\sqrt{n}})$ -regular halfspaces are “the most regular” ones; every halfspace is τ -regular for some $\tau \in [\frac{1}{\sqrt{n}}, 1]$. [14] generalized this connection to polytopes: they showed that any PRG for intersections of m many $O((\log m)/\sqrt{n})$ -regular halfspaces over Boolean space yields a PRG for intersections of m many arbitrary halfspaces over Gaussian space. Combining this with their Boolean PRG for intersections of regular halfspaces discussed above, [14] obtained a Gaussian PRG for intersections of m halfspaces with seed length $\text{poly}(\log m, 1/\delta) \cdot \log n$. Recent work of [6] gives a different Gaussian PRG with seed length $\text{poly}(\log m, 1/\delta) + O(\log n)$.

The focus of the current work is on the standard setting of PRGs over Boolean space, and the rest of the paper addresses this (more challenging) setting.

1.2 This Work: A PRG for Intersections of General Halfspaces

Summarizing the prior state of the art on PRGs over Boolean space, there were no PRGs that could fool intersections of $m = n$ many general halfspaces, and relatedly, the best PRG for intersections of $m \leq n$ general halfspaces had a superlinear seed length dependence on m . The PRGs that could fool intersections of $m \geq n$ halfspaces imposed technical restrictions on the halfspaces: either regularity (hence excluding simple halfspaces such as $\mathbb{1}[x_1 \leq 1/2]$), or small weights (hence excluding almost all halfspaces).

The main result of this paper is a PRG which fools intersections of m general halfspaces with a polylogarithmic seed length dependence on m :

¹Their seed length improves to $O(m \log(m/\delta) + \log n)$ if m/δ is bounded by any $\text{polylog}(n)$.

THEOREM 1.1 (PRG FOR POLYTOPES). *For all $n, m \in \mathbb{N}$ and $\delta \in (0, 1)$, there is an explicit pseudorandom generator with seed length $\text{poly}(\log m, 1/\delta) \cdot \log n$ that δ -fools the class of intersections of m halfspaces over $\{0, 1\}^n$.*

In particular, this PRG fools intersections of $\text{quasipoly}(n)$ many halfspaces with seed length $\text{polylog}(n)$, and its seed length remains non-trivial for intersections of exponentially many halfspaces ($\exp(n^c)$ where $c > 0$ is an absolute constant).

An immediate consequence of Theorem 1.1 is a deterministic algorithm that runs in time $n^{\text{polylog}(m)}$ and additively approximates the number of solutions to any n -variable $\{0, 1\}$ -integer program with m constraints. Prior to our result, no non-trivial deterministic algorithm (running in time $< 2^n$) was known even for general $\{0, 1\}$ -integer programs with $m = n$ constraints. Theorem 1.1 also yields PRGs with comparable seed lengths for intersections of halfspaces over a range of other domains, such as the n -dimensional hypergrid $\{0, 1, \dots, N\}^n$ and the solid cube $[0, 1]^n$ (details are left to the interested reader).

2 OVERVIEW OF OUR PROOF

Our proof of Theorem 1.1 involves several novel extensions of the central technique driving this line of work, namely Lindeberg-style proofs of probabilistic invariance principles and derandomizations thereof. We develop these extensions to overcome challenges which arise due to the generality of our setting; specifically, the fact that we are dealing with intersections of arbitrary halfspaces, with no restrictions whatsoever on their structure. One of the key new ingredients in our analysis, which we believe is of independent interest, is a sharp high-dimensional generalization of the classic Littlewood–Offord anticoncentration inequality [9, 27] that we establish. We now describe our proof and the new ideas underlying it in detail.

2.1 Background: The [14] PRG for Regular Polytopes

We begin by recalling the arguments of Harsha, Klivans, and Meka [14] for fooling regular polytopes. At a high level, [14] builds on the work of Meka and Zuckerman [28], which gave a versatile and powerful framework for constructing pseudorandom generators from probabilistic *invariance principles*; the main technical ingredient underlying the [14] PRG for regular polytopes is a new invariance principle for such polytopes, which we now describe.

[14]’s Invariance Principle and the Lindeberg Method. At a high level, the [14] invariance principle for regular polytopes is as follows: given an m -tuple of regular linear forms over n input variables $x = (x_1, \dots, x_n)$ (denoted by Ax , where A is an m -by- n matrix), the distribution (over $\mathbb{R}^m$) of Au , where $u \sim \{-1, 1\}^n$ is uniform random, is very close to the distribution of Ag , where $g \sim \mathcal{N}(0, 1)^n$ is distributed according to a standard n -dimensional Gaussian. Here closeness is measured by *multidimensional CDF distance*; we observe that multidimensional CDF distance corresponds to test functions of the form $\mathbb{1}[Ax \leq b]$ where $b \in \mathbb{R}^m$, which synchs up precisely with an intersection of m halfspaces $\mathbb{1}[A_1x \leq b_1] \wedge \dots \wedge \mathbb{1}[A_mx \leq b_m]$. To prove this invariance principle, [14] employs the well-known Lindeberg method (see

e.g. Chapter §11 of [35] and [43]) and proceeds in two main conceptual steps. The first step establishes a version of the result for *smooth test functions*, proxies for the actual “hard threshold” test functions $\mathbb{1}[Ax \leq b]$, and the second step relates distance with respect to these smooth test functions to multidimensional CDF distance via *Gaussian anticoncentration*. We outline each of these two steps below.

The first step is to prove an invariance principle for *smooth test functions*. Here instead of measuring the distance between Au and Ag using test functions that are orthant indicators $O_b(v_1, \dots, v_m) = \mathbb{1}[v \leq b]$ (corresponding to multidimensional CDF distance), distance is measured using a sufficiently smooth *mollifier* $\tilde{O}_b : \mathbb{R}^m \rightarrow [0, 1]$ of O_b . Such mollifiers, with useful properties that we now discuss, were proposed and analyzed by Bentkus [4]. In more detail, [14] prove that the difference between the expectations of $\tilde{O}_b(Au)$ and $\tilde{O}_b(Ag)$ is bounded by a certain function involving $\tilde{O}_b$ ’s derivatives. In fact, as in standard in Lindeberg-style proofs of invariance principles, [14] actually bounds this difference with respect to *any* smooth test function $Y : \mathbb{R}^m \rightarrow \mathbb{R}$ in terms of Y ’s derivatives; the only specific property of Bentkus’s mollifier $\tilde{O}_b$ that is used is that its derivatives are appropriately small. At a high level, the proof of this smooth invariance principle proceeds by hybridizing from $Y(Au)$ to $Y(Ag)$, using the multidimensional Taylor expansion of Y to bound the error incurred in each step. (The regularity of the linear forms is used in a crucial way to control the approximation error that results from truncating the Taylor expansion at a certain fixed degree.)

The second step is to establish the desired bound on multidimensional CDF distance using the aforescribed smooth invariance principle applied to Bentkus’s mollifier. This step relies on a second key property of Bentkus’s mollifier: $\tilde{O}_b$ agrees with the orthant indicator O_b except on a small error region near the orthant boundary. With this property in hand, a fairly simple and standard argument shows that it suffices to bound the *anticoncentration* of the *Gaussian* random variable Ag ; intuitively, such anticoncentration establishes that Ag does not place too much probability weight on the error region where $\tilde{O}_b$ disagrees with O_b . In [14], the required anticoncentration for Ag follows immediately from a result of Nazarov [24, 30] on the Gaussian surface area of m -facet polytopes.

The [14] PRG via a Derandomized Invariance Principle. Having proved this invariance principle for regular polytopes, [14] then establish a *pseudorandom* version by derandomizing its proof. That is, they argue that their proof in fact establishes multidimensional CDF-closeness between Az and Ag , where $g \sim \mathcal{N}(0, 1)^n$ is distributed according to a standard Gaussian as before, but $z \sim \{-1, 1\}^n$ is the output of a suitable pseudorandom suitable generator $\mathcal{G} : \{-1, 1\}^r \rightarrow \{-1, 1\}^n$ (rather than uniform random). Combining the “full-randomness” invariance principle (establishing closeness between Au and Ag) with this pseudorandom version (establishing closeness between Az and Ag), it follows from the triangle inequality that Az and Au are close. Recalling that multidimensional CDF distance corresponds to test functions of the form $\mathbb{1}[Ax \leq b] = \mathbb{1}[A_1x \leq b_1] \wedge \dots \wedge \mathbb{1}[A_mx \leq b_m]$, this is precisely equivalent to the claim that $\mathcal{G}$ fools the intersection of m halfspaces with weight matrix $A \in \mathbb{R}^{m \times n}$ (and an arbitrary vector of thresholds $b \in \mathbb{R}^m$).

For later reference, we close this section with an informal description of the [14] generator (for fooling intersections of m many τ -regular halfspaces):

- (1) Pseudorandomly hash the n variables into $L := \text{poly}(1/\tau)$ buckets using an $(r_{\text{hash}} := 2 \log m)$ -wise uniform hash function $h : [n] \rightarrow [L]$.
- (2) Independently across buckets, assign values to the variables within each bucket using an $(r_{\text{bucket}} := 4 \log m)$ -wise uniform distribution.

We remark that this is the structure of the Meka–Zuckerman generator [28] for fooling a single regular halfspace, the only difference being that the relevant parameters L , r_{hash} , and r_{bucket} are larger in [14] than in [28] (naturally so, given that the [14] generator fools intersections of m regular halfspaces instead of a single one).

Our analysis in this paper can be used to show that the [28] generator, instantiated with suitable choices of L , r_{hash} , and r_{bucket} , fools intersections of m general halfspaces. However, for technical reasons (that are not essential for this high-level discussion), this results in a seed length that is $\text{poly}(\log m, 1/\delta, \log n)$. To achieve our seed length of $\text{poly}(\log m, 1/\delta) \cdot \log n$, we slightly extend the [28] generator in two ways. First, within each bucket the variables are assigned using an r_{bucket} -wise uniform distribution XOR-ed with an independent draw from a generator that fools small-width CNF formulas [12]. Second, we XOR the entire resulting n -bit string with an independent draw from a k -wise independent generator. (See Section 4 for a detailed description of our PRG.)

2.2 Some Key New Ingredients in Our Analysis

A fundamental challenge in extending the [14] PRG result from regular to general polytopes stems from the fact that an invariance principle *simply does not hold* for general polytopes $Ax \leq b$. Without the regularity requirement on A , it is not true that Au and Ag are close in CDF distance; indeed, even a single non-regular linear form such as x_1 is distributed very differently under $u \sim \{-1, 1\}^n$ versus $g \sim \mathcal{N}(0, 1)^n$. This therefore necessitates a significant conceptual departure from the Meka–Zuckerman framework for constructing pseudorandom generators from invariance principles: rather than establishing closeness between Au and Az (where $z \sim \{-1, 1\}^n$ is the output of a suitable pseudorandom generator) through Ag by means of an invariance principle, one has to establish closeness between Au and Az “directly” without using invariance.

Somewhat surprisingly, even though an invariance principle does not hold in our setting of general polytopes, our proof nonetheless proceeds via the Lindeberg method for proving invariance principles. Following the two main conceptual steps of the method (as outlined in the previous section), we first prove that Au and Az are close with respect to Bentkus’s smooth mollifiers $\tilde{O}_b$ for the orthant indicators O_b , and then use this to establish closeness in multidimensional CDF distance. However, the fact that we are dealing with matrices $A \in \mathbb{R}^{m \times n}$ whose rows are *arbitrary* linear forms (corresponding to the facets of general m -facet polytopes) instead of regular linear forms poses significant challenges in both steps of the Lindeberg method. We discuss some of these challenges, and the new ideas that we employ to overcome them, next. For concreteness we will discuss these challenges and new ingredients by contrasting our proof with that of [14], but we remark here that

these are in fact qualitative differences between our approach and the Lindeberg method in general.

Step 1: Fooling Bentkus's Mollifier. Recall that [14] first proves a general invariance principle establishing closeness in expectation (with a quantitative bound that depends on Y 's derivatives) between $Y(Au)$ and $Y(Ag)$ for *any* smooth test function Y . They then apply this general invariance principle with Bentkus's orthonant mollifier $\tilde{O}_b$ being the test function, using the bounds on $\tilde{O}_b$'s derivatives established in [4] but no other properties of $\tilde{O}_b$.

In contrast, we do not prove closeness between Au and Az for all smooth test functions; our argument is carefully tailored to Bentkus's specific mollifier. In addition to bounds on $\tilde{O}_b$'s derivatives, we crucially rely on the specific structure of $\tilde{O}_b$, in particular, the fact that it is the product of m univariate functions, one for each coordinate (i.e. $\tilde{O}_b(v) = \prod_{i=1}^m \psi_{b_i}(v_i)$, where each ψ_{b_i} maps $\mathbb{R}$ to $[0, 1]$). A high-level intuition for why such product structure is useful is as follows. By doing some structural analysis of halfspaces (see Section 5), we can decompose each of our m halfspaces into a small "head" portion, consisting of at most k variables, and a remaining "tail" portion which is regular. From this point of view, the difference between regular and general polytopes is therefore the presence of these size-at-most- k head portions in each of the m halfspaces. Very roughly speaking, the product structure of $\tilde{O}_b$ allows us to handle these head portions using pseudorandom generators for *small-width CNF formulas* [12]. (To see the relevance of CNF formulas in this context, at least at a conceptual level, observe that a product of $\{0, 1\}$ -valued k -juntas is a width- k CNF formula.)

Our proof incorporates these PRGs for CNFs into [14]'s analysis of the regular tail portions. We highlight one interesting aspect of our analysis: In all previous instantiations of the Lindeberg method that we are aware of, expressions like $|E[Y(v + \Delta)] - E[Y(v + \Delta')]|$ are bounded by considering two Taylor expansions of Y , both taken around the "common point" v . Lindeberg method arguments analyze the difference of these Taylor expansions using moment-matching properties of Δ and Δ' and the fact that they are "small" in a certain technical sense, which is directly related to the regularity assumptions that underlie these invariance principles. In contrast, in our setting, since we are dealing with arbitrary linear forms rather than regular ones, we end up having to bound expressions like $|E[Y(v + \Delta)] - E[Y(v' + \Delta')]|$. Note that this involves considering the Taylor expansions of Y around two *distinct* points v and v' , which may be far from each other — indeed, *a priori* it is not even clear that $|E[Y(v)] - E[Y(v')]|$ will be small. Because of these differences from the standard Lindeberg scenario, moment-matching properties of Δ and Δ' and their "smallness" no longer suffice to ensure that the overall expected difference is small. Instead, as alluded to above, our analysis additionally exploits the product structure of Bentkus's mollifier via PRGs for CNFs to bound $|E[Y(v + \Delta)] - E[Y(v' + \Delta')]|$ (see Section 8).

Step 2: Anticoncentration. The next step is to pass from closeness of $\tilde{O}_b(Au)$ and $\tilde{O}_b(Az)$ in expectation, to closeness of Au and Az in multidimensional CDF distance. We recall that in the analogous step in [14]'s proof, the starting point was closeness in expectation of $\tilde{O}_b(Au)$ and $\tilde{O}_b(Ag)$, where $g \sim \mathcal{N}(0, 1)^n$ is a standard Gaussian (instead of $\tilde{O}_b(Az)$ where $z \sim \{-1, 1\}^n$ is pseudorandom). For

this reason, it sufficed for [14] to bound the Gaussian anticoncentration of Ag , and as mentioned, such a bound is an immediate consequence of Nazarov's bound on the Gaussian surface area of m -facet polytopes.

In contrast, since the Gaussian distribution does not enter into our arguments at all (by necessity, as explained above), we instead have to bound the *Boolean* anticoncentration of Au where $u \sim \{-1, 1\}^n$ is uniform random. This task, which is carried out in Section 7, requires significantly more work; indeed, Boolean anticoncentration formally contains Gaussian anticoncentration as a special case. At the heart of our arguments for this step is a new Littlewood–Offord-type anticoncentration inequality for m -facet polytopes, a high-dimensional generalization of the classic Littlewood–Offord theorem [9, 27]. We discuss this new theorem, which we believe is of independent interest, next.

2.2.1 A Littlewood–Offord Theorem for Polytopes. We first recall the classic Littlewood–Offord anticoncentration inequality.

THEOREM 2.1 (LITTLEWOOD–OFFORD). *For all $\theta \in \mathbb{R}$ and $w \in \mathbb{R}^n$ such that $|w_j| \geq 1$ for all $j \in [n]$,*

$$\Pr[w \cdot u \in (\theta - 2, \theta)] = O\left(\frac{1}{\sqrt{n}}\right),$$

where $u \sim \{-1, 1\}^n$ is uniformly random.

Littlewood and Offord [27] first proved a bound of $O((\log n)/\sqrt{n})$; Erdős [9] subsequently sharpened this to $O(1/\sqrt{n})$, which is optimal by considering $w = 1^n$ and $\theta = 0$. (We observe that the question trivializes without the assumption on the magnitudes of w 's coordinates; for instance, the relevant probability is $1/2$ for $w = (1, 0, \dots, 0)$ and $\theta = 1$.)

Theorem 2.1 has the following natural geometric interpretation: the maximum fraction of hypercube points that can fall within the "width-2 boundary" of a halfspace $\mathbb{1}[w \cdot x \leq \theta]$ where $|w_j| \geq 1$ for all j is $O(1/\sqrt{n})$. Given this geometric interpretation, it is natural to seek a generalization from single halfspaces (i.e. 1-facet polytopes) to m -facet polytopes:

What is the maximum fraction of hypercube points $u \in \{-1, 1\}^n$ that can lie within the "width-2 boundary" of an m -facet polytope $Ax \leq b$ where $|A_{ij}| \geq 1$ for all i and j ?

In more detail, we say that u lies within the "width-2 boundary" of the polytope $Ax \leq b$ provided $Au \leq b$ and $A_i \cdot u > b_i - 2$ for some $i \in [m]$; equivalently, u lies in the difference of the two polytopes $Ax \leq b$ and $Ax \leq b - 2 \cdot \mathbb{1}_m$, where $\mathbb{1}_m$ denotes the all-1's vector in $\mathbb{R}^m$. The Littlewood–Offord theorem (Theorem 2.1), along with a naive union bound, implies a bound of $O(m/\sqrt{n})$; we are not aware of any improvement of this naive bound prior to our work.

We give an essentially complete answer to this question, with upper and lower bounds that match up to constant factors. In Section 7 we prove the following "Littlewood–Offord theorem for polytopes":

THEOREM 2.2 (LITTLEWOOD–OFFORD THEOREM FOR POLYTOPES). *There is a universal constant C ($C = 5\sqrt{2}$ suffices) such that the following holds. For all $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ with $|A_{ij}| \geq 1$ for all $i \in [m]$ and $j \in [n]$,*

$$\Pr[Au \leq b \text{ \& } A_i \cdot u > b_i - 2 \text{ for some } i \in [m]] \leq \frac{C\sqrt{\ln m}}{\sqrt{n}},$$

where $\mathbf{u} \sim \{-1, 1\}^n$ is uniformly random.

Our proof of [Theorem 2.2](#) draws on and extends techniques from Kane’s bound on the Boolean average sensitivity of m -facet polytopes [18]. We complement [Theorem 2.2](#) with a matching lower bound, which establishes the existence of an m -facet polytope with an $\Omega(\sqrt{\ln m}/\sqrt{n})$ -fraction of hypercube points lying within its width-2 boundary. (In fact, our lower bound is slightly stronger: it establishes the existence of a polytope with an $\Omega(\sqrt{\ln m}/\sqrt{n})$ -fraction of hypercube points lying on its *surface*, corresponding to its width-0 boundary.)

[Theorem 2.2](#) does not suffice for the purpose of passing from closeness with respect to Bentkus’s orthant mollifier $\tilde{O}_b$ to closeness in multidimensional CDF distance (i.e. Step 2 in [Section 2.2](#)): while the assumption on the magnitudes of A ’s entries is essential to [Theorem 2.2](#) (just as the analogous assumption on w ’s coordinates is essential to the Littlewood–Offord theorem), the weight matrix of a general m -facet polytope need not have this property. In [Section 7](#) we establish various technical extensions of [Theorem 2.2](#) that are required to handle this issue.

2.3 Relation to [39]

We close this section with a discussion of the connection between our techniques and those of the recent work [39]. Recall that the main result of [39] is a PRG for δ -fooling intersections of m weight- W halfspaces using seed length $\text{poly}(\log m, W, 1/\delta) \cdot \text{polylog } n$. (Our main result, which is strictly stronger, is a PRG for δ -fooling intersections of m general halfspaces using seed length $\text{poly}(\log m, 1/\delta) \cdot \log n$, with no dependence on the weights of the halfspaces.)

A key structural observation driving [39] is that every intersection of m low-weight halfspaces can be expressed as $H \wedge G$, where H is an intersection of m regular halfspaces and G is a small-width CNF. (The width of G grows polynomially with the weights of the halfspaces, and this polynomial growth is responsible for the polynomial dependence on W in the seed length of the [39] PRG.) From this starting point, it suffices for [39] to bound the multidimensional CDF distance between the $(\mathbb{R}^m \times \{\pm 1\})$ -valued random variables $(A\mathbf{u}, G(\mathbf{u}))$ and $(A\mathbf{z}, G(\mathbf{z}))$, where $A \in \mathbb{R}^{m \times n}$ is the weight matrix of H , $\mathbf{u}$ is uniform random, and $\mathbf{z}$ is the output of the [39] PRG (which is a slight variant of [14]’s pseudorandom generator). Since H is an intersection of regular halfspaces, the fact that $A\mathbf{u}$ and $A\mathbf{z}$ are close in multidimensional CDF distance is precisely the main result of [14]; the crux of the work in [39] therefore lies in dealing with the additional distinguished $(m+1)$ st coordinate corresponding to the CNF G . Very roughly speaking, [39] employs a careful coupling $(\hat{\mathbf{u}}, \hat{\mathbf{z}})$ (whose existence is a consequence of the fact that bounded independence fools CNFs [2, 37]) to ensure that $G(\hat{\mathbf{u}})$ and $G(\hat{\mathbf{z}})$ almost always agree, and hence these $(m+1)$ st coordinates “have a negligible effect” throughout [14]’s Lindeberg-based proof of the regular case establishing closeness between $A\mathbf{u}$ and $A\mathbf{z}$.

Because of the aforementioned structural fact (that an m -tuple of low-weight halfspaces is equivalent to “an m -tuple of regular halfspaces plus a CNF”), the low-weight case analyzed in [39] did not require as significant a departure from [14]’s approach, and from the Lindeberg method as a whole, as the general case which is the subject of this paper. In particular, the new ideas discussed

in [Section 2.2](#) that are central to our proof were not present in [39]’s analysis for the low-weight case. To elaborate on this,

- [39] did not have to exploit the product structure of Bentkus’s orthant mollifier $\tilde{O}_b$ in order to fool it. Like [14], the arguments of [39] establish closeness in expectation between $Y(A\mathbf{u}, G(\mathbf{u}))$ and $Y(A\mathbf{z}, G(\mathbf{z}))$ for *all* smooth test functions Y , and the only properties of Bentkus’s mollifier that are used are the bounds on its derivatives given in [4] (which are used in a black box way). The simpler setting of [39] also did not necessitate comparing the Taylor expansions of Y around distinct points, as discussed in [Section 2.2](#).
- [39] did not have to reason about Boolean anticoncentration, which as discussed above requires significant novel conceptual and technical work, including our new Littlewood–Offord theorem for polytopes. Like [14], [39] were able to apply Nazarov’s Gaussian anticoncentration bounds as a black box to pass from fooling Bentkus’s mollifier to closeness in multidimensional CDF distance.

3 PRELIMINARIES

For convenience, in the rest of the paper we view halfspaces as having the domain $\{-1, 1\}^n$ rather than $\{0, 1\}^n$. We remind the reader that a halfspace $F : \{-1, 1\}^n \rightarrow \{0, 1\}$ is a function of the form $F(\mathbf{x}) = \mathbb{1}[\mathbf{w} \cdot \mathbf{x} \leq \theta]$ for some $\mathbf{w} \in \mathbb{R}^n$, $\theta \in \mathbb{R}$.

For an n -dimensional vector y and subset $B \subseteq [n]$, we write y_B to denote the $|B|$ -dimensional vector obtained by restricting y to the coordinates in B . For an $m \times n$ matrix A and subset $B \subseteq [n]$, we write A^B to denote the $m \times |B|$ matrix obtained by restricting A to the columns in B . For indices $i \in [m]$ and $j \in [n]$, we write A_i to denote the n -dimensional vector corresponding to the i -th row of A , and A^j to denote the m -dimensional vector corresponding to the j -column of A .

3.1 Regularity, Orthants, and Taylor’s Theorem

Definition 3.1 ((k, τ) -regular vectors and matrices). We say that a vector $\mathbf{w} \in \mathbb{R}^n$ is τ -regular if $|w_j| \leq \tau \|\mathbf{w}\|_2$ for all $j \in [n]$. More generally, we say that $\mathbf{w}$ is (k, τ) -regular if there is a partition $[n] = \text{HEAD} \sqcup \text{TAIL}$ where $|\text{HEAD}| \leq k$ and the subvector $\mathbf{w}_{\text{TAIL}}$ is τ -regular. We say that $\mathbf{w}$ is (k, τ) -standardized if $\mathbf{w}$ is (k, τ) -regular and $\sum_{j \in \text{TAIL}} w_j^2 = 1$. We say that a matrix $A \in \mathbb{R}^{m \times n}$ is τ -regular (respectively: (k, τ) -regular, (k, τ) -standardized) if all its rows are τ -regular (respectively: (k, τ) -regular, (k, τ) -standardized). We also use this terminology to refer to polytopes $A\mathbf{x} \leq \mathbf{b}$.

Translated orthants and their boundaries. For $\mathbf{b} \in \mathbb{R}^m$, we write $O_{\mathbf{b}} \subset \mathbb{R}^m$ to denote the *translated orthant*

$$O_{\mathbf{b}} = \{\mathbf{v} \in \mathbb{R}^m : v_i \leq b_i \text{ for all } i \in [m]\}.$$

We will overload notation and also write “ $O_{\mathbf{b}}$ ” to denote the indicator $\mathbb{R}^m \rightarrow \{0, 1\}$ of the orthant $O_{\mathbf{b}}$ (i.e., $O_{\mathbf{b}}(\mathbf{v}) = \mathbb{1}[\mathbf{v} \leq \mathbf{b}]$). We write $\partial O_{\mathbf{b}} \subset O_{\mathbf{b}}$ to denote $O_{\mathbf{b}}$ ’s *surface*,

$$\partial O_{\mathbf{b}} = \{\mathbf{v} \in O_{\mathbf{b}} : v_i = b_i \text{ for some } i \in [m]\}.$$

For $\Lambda > 0$, we write $\partial_{-\Lambda} O_{\mathbf{b}}$ and $\partial_{+\Lambda} O_{\mathbf{b}}$ to denote the *inner and outer Λ -boundaries* of $O_{\mathbf{b}}$,

$$\partial_{-\Lambda} O_{\mathbf{b}} = O_{\mathbf{b}} \setminus O_{\mathbf{b} - (\Lambda, \dots, \Lambda)}, \quad \partial_{+\Lambda} O_{\mathbf{b}} = O_{\mathbf{b} + (\Lambda, \dots, \Lambda)} \setminus O_{\mathbf{b}}, \quad (1)$$

and $\mathcal{O}_{\pm\Lambda}$ to denote the disjoint union $\mathcal{O}_{\pm\Lambda} = \mathcal{O}_{+\Lambda} \sqcup \mathcal{O}_{-\Lambda}$.

Derivatives and multidimensional Taylor expansion. We write $\psi^{(d)}$ to denote the d -th derivative of a C^d function $\psi : \mathbb{R} \rightarrow \mathbb{R}$. For an m -dimensional multi-index $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^m$, we write $|\alpha|$ to denote $\alpha_1 + \dots + \alpha_m$, and $\alpha!$ to denote $\alpha_1! \alpha_2! \dots \alpha_m!$. Given a vector $\Delta \in \mathbb{R}^m$, the expression Δ^α denotes $\prod_{i=1}^m \Delta_i^{\alpha_i}$. Given a function $\Upsilon : \mathbb{R}^m \rightarrow \mathbb{R}$, the expression $\partial_\alpha \Upsilon$ denotes the mixed partial derivative taken α_i times in the i -th coordinate.

The following is a straightforward consequence of the multidimensional Taylor theorem, upper-bounding the error term by the L_1 -norm of the derivatives times the L_∞ -norm of the offset-powers:

FACT 3.2 (MULTIDIMENSIONAL TAYLOR APPROXIMATION). *Let $d \in \mathbb{N}$ and let $\Upsilon : \mathbb{R}^m \rightarrow \mathbb{R}$ be a C^d function. Then for all $v, \Delta \in \mathbb{R}^m$,*

$$\Upsilon(v + \Delta) = \sum_{0 \leq |\alpha| \leq d-1} \frac{\partial_\alpha \Upsilon(v)}{\alpha!} \Delta^\alpha + \text{err}(v, \Delta),$$

where

$$|\text{err}(v, \Delta)| \leq \sup_{v^* \in \mathbb{R}^m} \left\{ \sum_{|\alpha|=d} |\partial_\alpha \Upsilon(v^*)| \right\} \cdot \|\Delta\|_d^d.$$

3.2 Pseudorandomness Preliminaries

Throughout this work we use **boldface** for random variables and random vectors. If $\mathcal{D}$ is a probability distribution, we write $\mathbf{x} \sim \mathcal{D}$ to denote that $\mathbf{x}$ is drawn from that distribution. For example, $\mathcal{N}(0, 1)$ will denote the standard normal distribution, so $\mathbf{g} \sim \mathcal{N}(0, 1)$ means $\mathbf{g}$ is a standard Gaussian random variable. In case S is a finite set, the notation $\mathbf{x} \sim S$ will mean that $\mathbf{x}$ is chosen uniformly at random from S . The most common case for this will be $\mathbf{u} \sim \{-1, 1\}^n$, meaning that $\mathbf{u}$ is chosen uniformly from $\{-1, 1\}^n$. We will reserve $\mathbf{u}$ for this specific random vector.

We recall the definition of a pseudorandom generator:

Definition 3.3 (PRG). A function $\mathcal{G} : \{-1, 1\}^r \rightarrow \{-1, 1\}^n$ is said to δ -fool a function $F : \{-1, 1\}^n \rightarrow \mathbb{R}$ with seed length r if

$$\left| \mathbb{E}_{s \sim \{-1, 1\}^r} [F(\mathcal{G}(s))] - \mathbb{E}_{\mathbf{u} \sim \{-1, 1\}^n} [F(\mathbf{u})] \right| \leq \delta.$$

Such a function $\mathcal{G}$ is said to be a *explicit pseudorandom generator (PRG)* that δ -fools a class $\mathcal{F}$ of n -variable functions if $\mathcal{G}$ is computable by a deterministic uniform poly(n)-time algorithm and $\mathcal{G}$ δ -fools every function $F \in \mathcal{F}$. We will also use the notation $\mathbf{z} \sim \mathcal{G}$ to mean that $\mathbf{z} = \mathcal{G}(s)$ for $s \sim \{-1, 1\}^r$.

Bounded Independence and Hash Families. A sequence of random variables $\mathbf{x}_1, \dots, \mathbf{x}_n$ is said to be *r -wise independent* if any collection of r of them is independent. In case the $\mathbf{x}_i$'s are uniformly distributed on their range, we say the sequence is *r -wise uniform*. We will also use this terminology for distributions $\mathcal{D}$ on $\{-1, 1\}^n$. An obvious but useful fact about r -wise uniform PRGs $\mathcal{G}$ is that they 0-fool the class of degree- r polynomials $\{-1, 1\}^n \rightarrow \mathbb{R}$.

A distribution $\mathcal{H}$ on functions $[n] \rightarrow [L]$ is said to be an *r -wise uniform hash family* if, for $\mathbf{h} \sim \mathcal{H}$, the sequence $(\mathbf{h}(1), \dots, \mathbf{h}(n))$ is r -wise uniform. Such a distribution also has the property that for any $\ell \in [L]$, the sequence $(\mathbb{1}_{\mathbf{h}(1)=\ell}, \dots, \mathbb{1}_{\mathbf{h}(n)=\ell})$ is r -wise independent on $\{0, 1\}^n$, with each individual random variable being

Bernoulli($1/L$). Well-known constructions (see e.g. Section 3.5.5 of [44]) give that for every n, L and r , there is an r -wise uniform hash family $\mathcal{H}$ of functions $[n] \rightarrow [L]$ such that choosing a random function from $\mathcal{H}$ takes $O(r \log(nL))$ random bits (and evaluating a function from $\mathcal{H}$ takes time $\text{poly}(r, \log n, \log L)$), and consequently there are known efficient constructions of r -wise uniform distributions over $\{0, 1\}^n$ with seed length $O(r \log n)$.

Fooling CNFs. Gopalan, Meka, and Reingold [12] have given an efficient explicit PRG that fools the class of small-width CNFs:

THEOREM 3.4 (PRG FOR SMALL-WIDTH CNFS). *There is an explicit PRG $\mathcal{G}_{\text{GMR}} = \mathcal{G}_{\text{GMR}}(w, \delta_{\text{CNF}})$ that δ_{CNF} -fools the class of all width- w CNF formulas over $\{-1, 1\}^n$ and has seed length*

$$O(w^2 \log^2(w \log(1/\delta_{\text{CNF}})) + w \log(w) \log(1/\delta_{\text{CNF}}) + \log \log n).$$

4 OUR PRG

The Meka–Zuckerman Generator. As stated earlier the PRG which we will analyze is a slight variant of a PRG first proposed by Meka and Zuckerman for fooling a single halfspace [28]. We begin by recalling the Meka–Zuckerman PRG.

Definition 4.1 (Meka–Zuckerman generator). The *Meka–Zuckerman generator* with parameters $L, r_{\text{hash}}, r_{\text{bucket}} \in [n]$, denoted $\mathcal{G}_{\text{MZ}}$, is defined as follows. Let $\mathbf{h} : [n] \rightarrow [L]$ be an r_{hash} -wise uniform hash function. Let $\mathbf{y}^1, \dots, \mathbf{y}^L \sim \{-1, 1\}^n$ be independent random variables, each r_{bucket} -wise uniform. A draw from $\mathcal{G}_{\text{MZ}}(L, r_{\text{hash}}, r_{\text{bucket}})$ is $\mathbf{z} \sim \{-1, 1\}^n$ where

$$z_{\mathbf{h}^{-1}(\ell)} = \mathbf{y}_{\mathbf{h}^{-1}(\ell)}^\ell \quad \text{for all } \ell \in [L].$$

In words, an r_{hash} -wise uniform hash $\mathbf{h}$ is used to partition the variables $x_1, \dots, x_n$ into L “buckets,” and then independently across buckets, the variables in each bucket are assigned according to an r_{bucket} -wise uniform distribution.

We note in passing that the generators of [14, 39] also have this structure (though the choice of parameters L, r_{bucket} , and r_{hash} are different than those in [28]).

Our Generator. Now we are ready to describe our generator and bound its seed length. Roughly speaking, our generator extends the Meka–Zuckerman generator by (i) additionally XOR-ing each bucket with an independent pseudorandom variable that fools CNF formulas; and (ii) globally XOR-ing the entire resulting n -bit string with an independent draw from a $2k$ -wise uniform distribution.

Definition 4.2 (Our generator). Our generator, denoted $\mathcal{G}$, is parameterized by values $L, r_{\text{hash}}, r_{\text{bucket}}, k, w \in [n]$, $\delta_{\text{CNF}} \in (0, 1)$ and is defined as follows. Let:

- $\mathbf{h}, \mathbf{y}^1, \dots, \mathbf{y}^L$ be defined as in the Meka–Zuckerman generator with parameters L, r_{hash} , and r_{bucket} .
- $\tilde{\mathbf{y}}^1, \dots, \tilde{\mathbf{y}}^L \sim \{-1, 1\}^n$ be independent draws from $\mathcal{G}_{\text{GMR}}(w, \delta_{\text{CNF}})$.
- $\mathbf{y}^* \sim \{-1, 1\}^n$ be $2k$ -wise uniform.

Define the random variable $\tilde{\mathbf{y}} \sim \{-1, 1\}^n$ by

$$\tilde{y}_{\mathbf{h}^{-1}(\ell)} = (\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_{\mathbf{h}^{-1}(\ell)} \quad \text{for all } \ell \in [L],$$

where $\oplus$ denotes bitwise XOR. A draw from our generator $\mathcal{G} = \mathcal{G}(L, r_{\text{hash}}, r_{\text{bucket}}, k, w, \delta_{\text{CNF}})$ is $\mathbf{z} \sim \{-1, 1\}^n$ where $\mathbf{z} = \tilde{\mathbf{y}} \oplus \mathbf{y}^*$.

Recalling the standard constructions of r -wise uniform hash functions and random variables described at the end of [Section 3](#), we have the following:

FACT 4.3 (SEED LENGTH). *The seed length of our PRG $\mathcal{G}$ with parameters $L, r_{\text{hash}}, r_{\text{bucket}}, k, w, \delta_{\text{CNF}}$ is*

$$\begin{aligned} &\lesssim r_{\text{hash}} \cdot \log(nL) + L \cdot r_{\text{bucket}} \cdot \log n && \text{(Seed length for } \mathcal{G}_{\text{MZ}}) \\ &+ L(w^2 \log^2(w \log(1/\delta_{\text{CNF}})) + w \log(w) \log(1/\delta_{\text{CNF}}) + \log \log n) && (L \text{ copies of } \mathcal{G}_{\text{GMR}}) \\ &+ k \log n. && (2k\text{-wise uniform string}) \end{aligned}$$

4.1 Setting of Parameters

We close this section with the parameter settings for fooling intersections of m halfspaces over $\{-1, 1\}^n$. Fix $\varepsilon \in (0, 1)$ to be an arbitrarily small absolute constant; the parameters we now specify will be for fooling to accuracy $O_\varepsilon(\delta) = O(\delta)$. We first define a few auxiliary parameters:

$$\begin{aligned} \lambda &= \frac{\delta}{\sqrt{\log(m/\delta) \log m}} && \text{(Dictated by Equation (17))} \\ \tau &= \frac{\delta^{1+\varepsilon}}{(\log m)^{2.5+\varepsilon}} && \text{(Dictated by Equation (16))} \\ d &= \text{constant depending only on } \varepsilon. && \text{(Dictated by Equation (16))} \end{aligned}$$

The precise value of $d = d(\varepsilon)$ will be specified in the proof of [Theorem 8.1](#). We will instantiate our generator $\mathcal{G}$ with parameters:

$$\begin{aligned} L &= \frac{(\log m)^5}{\delta^{2+\varepsilon}} && \text{(Constrained by Equation (16),} \\ &&& \text{chosen to optimize seed length)} \\ r_{\text{hash}} &= C_1 \log(Lm/\delta) && \text{(Dictated by Proposition 8.11)} \\ r_{\text{bucket}} &= \log(m/\delta) && \text{(Dictated by Lemma 8.10)} \\ k &= \frac{C_2 \log(m/\delta) \log \log(m/\delta)}{\tau^2} && \text{(Dictated by Lemma 5.1)} \\ w &= \frac{2k}{L} && \text{(Dictated by Proposition 8.11)} \\ \delta_{\text{CNF}} &= \frac{\delta}{L} \cdot \left(\frac{\lambda}{m\sqrt{n}} \right)^{d-1}, && \text{(Dictated by Equation (16))} \end{aligned}$$

where C_1 and C_2 are absolute constants specified in the proofs of [Proposition 8.11](#) and [Lemma 5.1](#) respectively.

Our seed length: By [Fact 4.3](#), our overall seed length is

$$\tilde{O}\left(\frac{(\log m)^{6+\varepsilon}}{\delta^{2+\varepsilon}}\right) \cdot \log n + \tilde{O}\left(\frac{(\log m)^{7+\varepsilon}}{\delta^{2+\varepsilon}}\right) = \text{poly}(\log m, 1/\delta) \cdot \log n \quad (2)$$

for any absolute constant $\varepsilon \in (0, 1)$.

5 REDUCTION TO STANDARDIZED POLYTOPES

5.1 A Reduction from Fooling Polytopes to Fooling Standardized Polytopes

In this section we reduce from the problem of fooling general m -facet polytopes to the problem of fooling m -facet (k, τ) -standardized polytopes ([Definition 3.1](#)). The main technical result we prove in this section is the following:

LEMMA 5.1 (APPROXIMATION BY (k, τ) -STANDARDIZED POLYTOPES UNDER BOUNDED-UNIFORMITY DISTRIBUTIONS). *There is a universal constant $C_2 > 0$ such that the following holds. Fix $m \geq 1$ and $0 < \delta, \tau < 1/2$ such that the right-hand side of [Equation \(3\)](#) below is at most $n/2$. Let*

$$k = \frac{C_2 \log(m/\delta) \log \log(m/\delta)}{\tau^2}. \quad (3)$$

For every m -facet polytope $Ax \leq b$ in $\mathbb{R}^n$, there is an m -facet (k, τ) -standardized polytope $A'x \leq b'$ in $\mathbb{R}^n$ such that if $\mathbf{y} \sim \{-1, 1\}^n$ is $2k$ -wise uniform, then

$$\Pr[\mathbb{1}[A\mathbf{y} \leq b] \neq \mathbb{1}[A'\mathbf{y} \leq b']] \leq \delta. \quad (4)$$

We stress that [Lemma 5.1](#) establishes that $\mathbb{1}[Ax \leq b]$ is well-approximated by $\mathbb{1}[A'x \leq b']$ under *both* the uniform distribution and the pseudorandom distribution constructed by our generator, since both of these distributions are $2k$ -wise uniform. (Note that a draw $\mathbf{z} = \tilde{\mathbf{y}} \oplus \mathbf{y}^*$ from our generator is indeed $2k$ -wise uniform, since $\mathbf{y}^*$ is; indeed, [Lemma 5.1](#) is the motivation for why our construction includes a bitwise-XOR with $\mathbf{y}^*$.) This is crucial: in general, given a function F and an approximator F' that is close to F only under the uniform distribution (i.e. $\Pr[F(\mathbf{u}) \neq F'(\mathbf{u})]$ is small), fooling F' does not suffice to fool F itself.

Given [Lemma 5.1](#), in order to prove [Theorem 1.1](#) it is sufficient to prove the following:

THEOREM 5.2 (FOOLING (k, τ) -STANDARDIZED POLYTOPES). *Let $\mathcal{G}$ be our generator with parameters as set in [Section 4.1](#). For all m -facet (k, τ) -standardized polytopes $A'x \leq b'$,*

$$\left| \Pr_{\mathbf{u} \sim \{-1, 1\}^n} [A'\mathbf{u} \in O_{b'}] - \Pr_{\mathbf{z} \sim \mathcal{G}} [A'\mathbf{z} \in O_{b'}] \right| = O(\delta).$$

The rest of the paper following this section is devoted to proving [Theorem 5.2](#).

6 BENTKUS'S MOLLIFIER AND ITS PROPERTIES

In this section we introduce and analyze Bentkus's orthant mollifier $\tilde{O}_b : \mathbb{R}^m \rightarrow (0, 1)$, which is a smoothed version of the translated orthant indicator function $O_b : \mathbb{R}^m \rightarrow \{0, 1\}$ from [Section 3.1](#).

Definition 6.1 (Gaussian-mollified halfline). For $\theta \in \mathbb{R}$ and $\lambda > 0$, we define the C^∞ function $\tilde{\mathbb{1}}_{\theta, \lambda} : \mathbb{R} \rightarrow (0, 1)$,

$$\tilde{\mathbb{1}}_{\theta, \lambda}(t) = \mathbb{E}_{\mathbf{g} \sim N(0, 1)} [\mathbb{1}[t + \lambda \mathbf{g} \leq \theta]].$$

Definition 6.2 (Bentkus's orthant mollifier). For $b \in \mathbb{R}^m$ and $\lambda > 0$, the *Bentkus λ -mollifier* for O_b is defined to be the C^∞ function $\tilde{O}_{b, \lambda} : \mathbb{R}^m \rightarrow (0, 1)$,

$$\tilde{O}_{b, \lambda}(v) = \mathbb{E}_{\mathbf{g} \sim N(0, 1)^m} [O_b(v + \lambda \mathbf{g})].$$

Since $O_b(v) = \prod_{i=1}^m \mathbb{1}[v_i \leq b_i]$ and $N(0, 1)^m$ is a product distribution, the mollifier $\tilde{O}_{b, \lambda}$ can be equivalently defined as follows:

$$\tilde{O}_{b, \lambda}(v) = \prod_{i=1}^m \tilde{\mathbb{1}}_{b_i, \lambda}(v_i). \quad (5)$$

This product structure of Bentkus's mollifier will be crucially important for us in the analysis that we carry out in [Section 8.1](#). We note the following translation property of Bentkus's mollifier:

FACT 6.3. For all $b, v, \Delta \in \mathbb{R}^m$ and $\lambda > 0$, we have $\tilde{O}_{b,\lambda}(v + \Delta) = \tilde{O}_{b-v,\lambda}(\Delta)$.

In Section 8.1 we will also use the following global bound on the magnitude of the derivatives of the Gaussian-mollified halfline:

FACT 6.4 (STANDARD; SEE EXERCISE 11.41 IN [35]). For all $\theta \in \mathbb{R}$, $\lambda > 0$, and integer $d \geq 1$,

$$\left\| \tilde{\mathbb{I}}_{\theta,\lambda}^{(d)} \right\|_{\infty} = O_d \left(\frac{1}{\lambda} \right)^d.$$

The following result, from Bentkus [4, Theorem 3(ii)], can be viewed as a multidimensional generalization of Fact 6.4. (Strictly speaking [4] only considers b 's of the form $(\theta, \theta, \dots, \theta)$, but by translation-invariance the bound holds for all $b \in \mathbb{R}^m$.)

THEOREM 6.5 (BOUNDED SUM OF DERIVATIVES). For all $m \geq 2$, $b \in \mathbb{R}^m$, $\lambda > 0$, and integer $d \geq 1$,

$$\sup_{v \in \mathbb{R}^m} \left\{ \sum_{|\alpha|=d} |\partial_{\alpha} \tilde{O}_{b,\lambda}(v)| \right\} = O_d \left(\frac{\sqrt{\log m}}{\lambda} \right)^d.$$

Recall from (1) that $\partial_{-\Lambda} O_b = O_b \setminus O_{b-(\Lambda, \dots, \Lambda)}$ and $\partial_{+\Lambda} O_b = O_{b+(\Lambda, \dots, \Lambda)} \setminus O_b$. We will use the following notions of approximation for translated orthants:

Definition 6.6 (Inner and outer approximators for orthants). We say that $\Upsilon : \mathbb{R}^m \rightarrow [0, 1]$ is a (Λ, δ) -inner approximator for O_b if

$$|\Upsilon(v) - O_b(v)| \leq \delta \quad \text{for all } v \notin \partial_{-\Lambda} O_b.$$

Similarly, we say that Υ is a (Λ, δ) -outer approximator for O_b if

$$|\Upsilon(v) - O_b(v)| \leq \delta \quad \text{for all } v \notin \partial_{+\Lambda} O_b.$$

The connection between Bentkus's mollifier and these notions of approximation is established in the following claim.

LEMMA 6.7 (BENTKUS'S MOLLIFIER, APPROPRIATELY TRANSLATED, YIELDS INNER AND OUTER APPROXIMATORS FOR TRANSLATED ORTHANTS). For all $b \in \mathbb{R}^m$ and $\lambda, \delta \in (0, 1)$, there are $b^{\text{in}}, b^{\text{out}} \in \mathbb{R}^m$ such that $\tilde{O}_{b^{\text{in}},\lambda}, \tilde{O}_{b^{\text{out}},\lambda}$ are (Λ, δ) -inner and -outer approximators for O_b respectively, where $\Lambda = \Theta(\lambda \sqrt{\log(m/\delta)})$.

PROOF. $\square$

6.1 The Connection between Inner/Outer Approximators and CDF Distance

The following elementary properties of inner/outer approximators will be useful for us:

FACT 6.8. Fix $b \in \mathbb{R}^m$ and let $\Upsilon^{\text{in}}, \Upsilon^{\text{out}}$ be (Λ, δ) -inner and -outer approximators for O_b . Then

- (1) $\Upsilon^{\text{in}}(v) - \delta \leq O_b(v) \leq \Upsilon^{\text{out}}(v) + \delta$ for all $v \in \mathbb{R}^m$.
- (2) Υ^{in} is a (Λ, δ) -outer approximator for $O_{b-\Lambda \mathbb{1}_m}$, and similarly Υ^{out} is a (Λ, δ) -inner approximator for $O_{b+\Lambda \mathbb{1}_m}$.

The next lemma is straightforward but very useful for us. Intuitively, it says that in order for an $\mathbb{R}^m$ -valued random variable $\tilde{v}$ to fool a translated orthant O_b relative to another $\mathbb{R}^m$ -valued random variable v , it suffices to (i) have $\tilde{v}$ fool both inner and outer approximators for O_b , and (ii) establish anticoncentration of the

original random variable v at the inner and outer boundaries of O_b . We explain in detail how we will use this lemma after giving its proof below.

LEMMA 6.9. Let $\Upsilon^{\text{in}}, \Upsilon^{\text{out}} : \mathbb{R}^m \rightarrow [0, 1]$ be (Λ, δ) -inner and -outer approximators for O_b . Let v and $\tilde{v}$ be $\mathbb{R}^m$ -valued random variables satisfying:

$$|\mathbb{E}[\Upsilon(v)] - \mathbb{E}[\Upsilon(\tilde{v})]| \leq \gamma \quad (6)$$

for both $\Upsilon \in \{\Upsilon^{\text{out}}, \Upsilon^{\text{in}}\}$. Then

$$|\Pr[v \in O_b] - \Pr[\tilde{v} \in O_b]| \leq \gamma + 2\delta + \Pr[v \in \partial_{\pm\Lambda} O_b].$$

PROOF. $\square$

6.1.1 Applying Lemma 6.9 in the context of Theorem 5.2, and the organization of the rest of this paper. Applying Lemma 6.9 with v and $\tilde{v}$ being Au and Az respectively, the task of bounding

$$\left| \Pr_{u \sim \{-1, 1\}^n} [Au \in O_b] - \Pr_{z \sim \mathcal{G}_{\text{MZ}}} [Az \in O_b] \right|$$

reduces to the following two-step program:

- (1) Establishing anticoncentration within orthant boundaries: bounding $\Pr[Au \in \partial_{\pm\Lambda} O_b]$; and,
- (2) Fooling Bentkus's mollifier: bounding $|\mathbb{E}[\tilde{O}(Au)] - \mathbb{E}[\tilde{O}(Az)]|$ for $\tilde{O} \in \{\tilde{O}_{b^{\text{out}},\lambda}, \tilde{O}_{b^{\text{in}},\lambda}\}$, the inner and outer approximators for O_b given by Lemma 6.7.

Section 7 is devoted to the former, and Section 8 the latter. In Section 9 we put these pieces together to prove Theorem 5.2.

7 BOOLEAN ANTICONCENTRATION WITHIN ORTHANT BOUNDARIES

The main result of this section is Theorem 7.1, which provides the first step of the two-step program described at the end of Section 6:

THEOREM 7.1 (BOOLEAN ANTICONCENTRATION WITHIN ORTHANT BOUNDARIES). Assume $A \in \mathbb{R}^{m \times n}$ satisfies the following property: each of its row vectors has a τ -regular subvector of 2-norm 1, where τ is as set in Section 4.1. Then for all $b \in \mathbb{R}^m$ and $\Lambda \geq \tau$, we have

$$\Pr_{u \sim \{-1, 1\}^n} [Au \in \partial_{\pm\Lambda} O_b] = O(\Lambda \sqrt{\log m}).$$

En route to proving Theorem 7.1 we will establish a ‘‘Littlewood–Offord theorem for polytopes,’’ Theorem 2.2, that was stated in Section 2.2.1.

8 FOOLING BENTKUS'S MOLLIFIER

The main result of this section is the following theorem, which provides the second step of the two-step program described at the end of Section 6:

THEOREM 8.1 ($\mathcal{G}$ FOOLS BENTKUS'S MOLLIFIER). Let $\mathcal{G}$ be our generator with parameters as given in Section 4.1, and likewise let $\lambda > 0$ be as set in Section 4.1. For all (k, τ) -standardized matrices $A \in \mathbb{R}^{m \times n}$ and all $b \in \mathbb{R}^m$,

$$\left| \mathbb{E}_{u \sim \{-1, 1\}^n} [\tilde{O}_{b,\lambda}(Au)] - \mathbb{E}_{z \sim \mathcal{G}_{\text{MZ}}} [\tilde{O}_{b,\lambda}(Az)] \right| = O(\delta).$$

At a very high level, in line with the usual Lindeberg approach, [Theorem 8.1](#) is proved by hybridizing between $\mathbf{u}$ and $\mathbf{z}$ via a sequence of intermediate distributions. In our setting there are $L + 1$ such distributions, the first of which is $\mathbf{u}$ and the last of which is $\mathbf{z}$, and the ℓ -th of which may be viewed as “filling in buckets $\ell, \dots, L$ according to $\mathbf{u}$ and filling in buckets $1, \dots, \ell - 1$ according to $\mathbf{z}$,” where the L buckets correspond to the partition of $[n]$ induced by the choice of the random hash function in the Meka–Zuckerman generator.

In [Section 8.1](#) we upper bound the error incurred by taking a single step through this sequence of hybrid distributions. The upper bound given there (see [Lemma 8.3](#)) has a first component corresponding to the terms of order $0, \dots, d - 1$ in a $(d - 1)$ -st order Taylor expansion, and a second component corresponding to the error term in Taylor’s theorem. The first component is upper bounded in [Section 8.1](#), and the second component is upper bounded in [Section 8.2](#). [Section 8.3](#) formalizes the hybrid argument and uses the results of these earlier subsections to establish [Theorem 8.1](#).

Remark 8.2 (Head and tail matrices). Recalling the definition of a (k, τ) -standardized matrix A ([Definition 3.1](#)), for every $i \in [m]$ there is a partition $[n] = \text{HEAD}_i \sqcup \text{TAIL}_i$ such that $|\text{HEAD}_i| \leq k$ and $(A_i)_{\text{TAIL}_i}$ is τ -regular with 2-norm $\|(A_i)_{\text{TAIL}_i}\|_2$ equal to 1. Therefore, we may write A as $H + T$ where

$$H_{ij} = A_{ij} \cdot \mathbb{1}[j \in \text{HEAD}_i] \quad \text{and} \quad T_{ij} = A_{ij} \cdot \mathbb{1}[j \in \text{TAIL}_i]$$

for all $j \in [n]$ and $i \in [m]$. Note that every row of H is k -sparse, and every row of T is τ -regular with 2-norm 1.

8.1 Single Swap in the Hybrid Argument

LEMMA 8.3 (ERROR INCURRED BY A SINGLE SWAP). Fix $B \subseteq [n]$. Let $H^B, T^B \in \mathbb{R}^{m \times B}$, where every row of H^B is w -sparse and every row of T^B has 2-norm at most 1. Let $\mathbf{u}, \mathbf{y}$ be random variables over $\{-1, 1\}^B$, where $\mathbf{u}$ is uniform and $\mathbf{y}$ δ_{CNF} -fools the class of width- w CNFs. For all $b \in \mathbb{R}^m$, $\lambda > 0$, and all integers $d \geq 2$

$$|\mathbb{E}[\tilde{O}_{b,\lambda}(H^B \mathbf{u} + T^B \mathbf{u})] - \mathbb{E}[\tilde{O}_{b,\lambda}(H^B \mathbf{y} + T^B \mathbf{y})]| \quad (7)$$

$$= \delta_{\text{CNF}} \cdot m^{d-1} \cdot O_d \left(\frac{\sqrt{n}}{\lambda} \right)^{d-1} + O_d \left(\frac{\sqrt{\log m}}{\lambda} \right)^d \left(\mathbb{E}[\|T^B \mathbf{u}\|_\infty^d] + \mathbb{E}[\|T^B \mathbf{y}\|_\infty^d] \right). \quad (8)$$

As we will see later, [Equation \(7\)](#) is a useful bound because we can (and will) take δ_{CNF} to be very small, and when we apply [Lemma 8.3](#) we will be able to ensure that both expectations on the right-hand side of [Equation \(7\)](#) are small as well.

The main ingredient in the proof of [Lemma 8.3](#) is the following claim:

CLAIM 8.4. For all integers $c \geq 1$ and $\alpha \in \mathbb{N}^m$ such that $|\alpha| = c$,

$$\left| \mathbb{E}[\partial_\alpha \tilde{O}_{b,\lambda}(H^B \mathbf{u}) \cdot (T^B \mathbf{u})^\alpha] - \mathbb{E}[\partial_\alpha \tilde{O}_{b,\lambda}(H^B \mathbf{y}) \cdot (T^B \mathbf{y})^\alpha] \right| = \delta_{\text{CNF}} \cdot O_c \left(\frac{\sqrt{n}}{\lambda} \right)^c. \quad (9)$$

Remark 8.5. Recalling the discussion of Step 1 in [Section 2.2](#), we remark that [Claim 8.4](#) provides the key ingredient of the arguments sketched there. This claim plays an essential role in enabling us to

get a strong bound on the magnitude of the difference of two expectations (which was denoted “ $|\mathbb{E}[Y(\mathbf{v} + \Delta)] - \mathbb{E}[Y(\mathbf{v}' + \Delta')]|$ ” in [Section 2.2](#) and corresponds precisely to the left-hand side of [Lemma 8.3](#) above) through an application of Taylor’s theorem around two different points. As will be seen in [Section 8.1.1](#), the proof of [Claim 8.4](#) exploits the product structure of $\tilde{O}_b$ by using pseudorandom generators for small-width CNF formulas.

8.1.1 Proof of [Claim 8.4](#).

Definition 8.6. We say that a function $\xi : \{-1, 1\}^B \rightarrow \mathbb{R}$ is Boolean if its range is contained in $\{0, 1\}$. For $\xi_1, \dots, \xi_m : \{-1, 1\}^B \rightarrow \mathbb{R}$, we say that the associated product function $\Xi = \prod_{i \in [m]} \xi_i$ is a Boolean product function in case all the ξ_i ’s are Boolean.

Definition 8.7. We say that ξ is a weight- W combination of Boolean functions if it is expressible as a linear combination $\xi = \sum_\ell c_\ell \xi_\ell$ where each ξ_ℓ is a Boolean function and where $\sum_\ell |c_\ell| \leq W$. Likewise, Ξ is a weight- W combination of Boolean product functions if it is expressible as a linear combination $\Xi = \sum_\ell c_\ell \Xi_\ell$ where each Ξ_ℓ is a Boolean product function and where $\sum_\ell |c_\ell| \leq W$.

The following facts are easy to establish:

- FACT 8.8.** (1) A function $\xi : \{-1, 1\}^B \rightarrow [0, 1]$ is a weight-1 combination of Boolean functions.
 (2) A function $\xi : \{-1, 1\}^B \rightarrow [-W, W]$ is a weight- $(2W)$ combination of Boolean functions.
 (3) A weight- W_1 combination of weight- W_2 combinations of Boolean functions is a weight- $(W_1 W_2)$ combination of Boolean functions.
 (4) If ξ_1 and ξ_2 are weight- W_1 and weight- W_2 combinations of Boolean product functions respectively, then $\xi_1 \cdot \xi_2$ is a weight- $(W_1 W_2)$ combination of Boolean product functions.

We are now ready to prove [Claim 8.4](#).

PROOF OF CLAIM 8.4. We define the function $G_\alpha : \{-1, 1\}^B \rightarrow \mathbb{R}$,

$$G_\alpha(x) := \partial_\alpha \tilde{O}_{b,\lambda}(H^B x) \cdot (T^B x)^\alpha = \left(\prod_{i \notin S} \tilde{\mathbb{I}}_{b_i,\lambda}(H_i^B x) \prod_{i \in S} \tilde{\mathbb{I}}_{b_i,\lambda}^{(\alpha_i)}(H_i^B x) \right) \cdot \prod_{i \in S} (T_i^B x)^{\alpha_i}, \quad (10)$$

where S denotes $\text{supp}(\alpha) = \{i \in [m] : \alpha_i > 0\}$. ([Equation \(10\)](#) crucially relies on the product structure of $\tilde{O}_{b,\lambda} : \mathbb{R}^m \rightarrow (0, 1)$; recall [Equation \(5\)](#).)

Note that [Claim 8.4](#) is equivalent to the claim that $\mathbf{y}$ δ -fools G_α for $\delta = \delta_{\text{CNF}} \cdot O_c(\sqrt{n}/\lambda)^c$. We analyze the three types of functions in [Equation \(10\)](#) in turn:

- Recalling the assumptions of [Lemma 8.3](#), by Item 1 of [Fact 8.8](#), the function $x \mapsto \tilde{\mathbb{I}}_{b_i,\lambda}(H_i^B x)$ is a weight-1 combination of Boolean functions. Furthermore, since $|\text{supp}(H_i^B)| \leq w$, it is in fact a weight-1 combination of Boolean w -juntas.
- Similarly, by Item 2 of [Fact 8.8](#), the function $x \mapsto \tilde{\mathbb{I}}_{b_i,\lambda}^{(\alpha_i)}(H_i^B x)$ is a weight- $(2\|\tilde{\mathbb{I}}_{b_i,\lambda}^{(\alpha_i)}\|_\infty)$ combination of Boolean w -juntas.
- Since $\|T_i^B\|_1 \leq \sqrt{B} \cdot \|T_i^B\|_2 \leq \sqrt{B} \leq \sqrt{n}$ and $x_j \in \{-1, 1\}$ for all $j \in B$, by Items 2 and 3 of [Fact 8.8](#) the function $x \mapsto T_i^B x$ is a weight- $(2\sqrt{n})$ combination of Boolean functions. Furthermore, it is a weight- $(2\sqrt{n})$ combination of Boolean 1-juntas.

Combining the above with Item 4 of [Fact 8.8](#), it follows that $G_\alpha : \{-1, 1\}^B \rightarrow \mathbb{R}$ is a weight- W combination of Boolean product functions $\Xi : \{-1, 1\}^B \rightarrow \{0, 1\}$, where

$$\begin{aligned} W &= \left(\prod_{i \in S} 2 \|\tilde{\mathbf{I}}_{b_i, \lambda}^{(\alpha_i)}\|_\infty \right) \cdot \left(\prod_{i \in S} (2\sqrt{n})^{\alpha_i} \right) \\ &= \left(\prod_{i \in S} O_{\alpha_i} \left(\frac{1}{\lambda^{\alpha_i}} \right) \right) \cdot \left(\prod_{i \in S} (2\sqrt{n})^{\alpha_i} \right) \quad (\text{Fact 6.4}) \\ &= O_c \left(\frac{\sqrt{n}}{\lambda} \right)^c. \quad (|\alpha| = \alpha_1 + \dots + \alpha_m = c) \end{aligned}$$

Furthermore, every Ξ in this combination is the product of m Boolean w -juntas and $|\alpha|$ Boolean 1-junta(s). Since each such Ξ is computable by a width- w CNF, and $\mathbf{y}$ δ_{CNF} -fools the class of width- w CNFs, we conclude that $\mathbf{y}$ δ -fools G_α where $\delta = \delta_{\text{CNF}} \cdot W$. This completes the proof of [Claim 8.4](#). $\square$

8.2 Bounding the Error Terms

We will use the following technical result, which is almost an immediate consequence of Rosenthal's inequality [\[38\]](#). (

CLAIM 8.9. *Let $\beta \in [0, 1]$ and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be independent $\{0, \pm 1\}$ -valued random variables, each being 0 with probability $1 - \beta$ and ± 1 with probability $\beta/2$ each. Let $\mathbf{w} \in \mathbb{R}^n$ be a τ -regular vector of 2-norm 1. Then for any $q \geq 2$,*

$$\mathbb{E}[|\mathbf{w} \cdot \mathbf{x}|^q] = O(q\tau \cdot (\beta/\tau^2)^{1/q} + \sqrt{q}\sqrt{\beta})^q.$$

Of course, if q is an even integer, then the above continues to hold even if $\mathbf{x}_1, \dots, \mathbf{x}_n$ are merely q -wise independent.

The following lemma will be used to bound the expectations on the right-hand side of [Equation \(7\)](#):

LEMMA 8.10. *Let $L, r_{\text{hash}}, r_{\text{bucket}}$, and τ be as set in [Section 4.1](#). Let $\mathbf{h} : [n] \rightarrow [L]$ be an r_{hash} -wise uniform hash function, and fix a bucket $\ell \in [L]$. Let $\mathbf{y} \sim \{-1, 1\}^n$ be an r_{bucket} -wise uniform random variable. Let $T \in \mathbb{R}^{m \times n}$ be a τ -regular matrix in which each row has 2-norm 1. Then for all integers $d \geq 2$,*

$$\mathbb{E}_{\mathbf{h}, \mathbf{y}} \left[\|T^{\mathbf{h}^{-1}(\ell)} \mathbf{y}_{\mathbf{h}^{-1}(\ell)}\|_\infty^d \right] = O_d \left(\tau \log m + \sqrt{(\log m)/L} \right)^d.$$

PROOF. $\square$

8.3 Proof of [Theorem 8.1](#): The Hybrid Argument

In this subsection we put together the two main results of the two previous subsections ([Lemma 8.3](#) and [Lemma 8.10](#)) to prove [Theorem 8.1](#).

Recalling [Remark 8.2](#), we can write A as $H + T$, where every row of H is k -sparse and every row of T is τ -regular with 2-norm 1. Let us say that a hash $\mathbf{h} : [n] \rightarrow [L]$ is H -good if

$$|h^{-1}(\ell) \cap \text{supp}(H_i)| \leq w := \frac{2k}{L} \quad (11)$$

for all buckets $\ell \in [L]$ and rows $i \in [m]$. Equivalently, for all $\ell \in [L]$, every row of the submatrix $H^{\mathbf{h}^{-1}(\ell)}$ is w -sparse.

PROPOSITION 8.11 (EVEN DISTRIBUTION OF HEAD VARIABLES). *There is a universal constant $C_1 > 0$ such that the following holds. If $\mathbf{h} : [n] \rightarrow [L]$ is r_{hash} -wise uniform where $r_{\text{hash}} \geq C_1 \log(Lm/\delta)$, then*

$$\Pr[\mathbf{h} \text{ is not } H\text{-good}] \leq \delta.$$

PROOF. $\square$

We are now ready to prove [Theorem 8.1](#):

PROOF OF [THEOREM 8.1](#). Let $\mathbf{h}, \mathbf{y}^1, \dots, \mathbf{y}^L, \tilde{\mathbf{y}}^1, \dots, \tilde{\mathbf{y}}^L, \check{\mathbf{y}}$, and $\mathbf{y}^\star$ be the random hash function and random variables associated with our generator $\mathcal{G}$, as defined in [Definition 4.2](#). Recall that a draw from $\mathbf{z} \sim \mathcal{G}$ is $\mathbf{z} := \check{\mathbf{y}} \oplus \mathbf{y}^\star$. We will show that in fact $\check{\mathbf{y}}$ alone satisfies:

$$\left| \mathbb{E}_{\mathbf{u} \sim \{-1, 1\}^n} [\tilde{O}_{b, \lambda}(A\mathbf{u})] - \mathbb{E}[\tilde{O}_{b, \lambda}(A\check{\mathbf{y}})] \right| = O(\delta). \quad (12)$$

Since $\mathbf{y}^\star$ and $\check{\mathbf{y}}$ are independent, [Theorem 8.1](#) follows as a consequence of [Equation \(12\)](#).

We recall the definition of $\check{\mathbf{y}}$:

$$\check{\mathbf{y}}_{\mathbf{h}^{-1}(\ell)} = (\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_{\mathbf{h}^{-1}(\ell)} \quad \text{for all } \ell \in [L].$$

We observe first that for each $\ell \in [L]$, the random variable $\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell \sim \{-1, 1\}^n$

- (i) is r_{bucket} -wise uniform (since $\mathbf{y}^\ell$ is); and
- (ii) δ_{CNF} -fools the class of width- w CNF formulas (since $\tilde{\mathbf{y}}^\ell$ does).

We will use both properties in this proof. For each hash $\mathbf{h} : [n] \rightarrow [L]$ and index $\ell \in \{0, 1, \dots, L\}$, we define the hybrid random variable $\mathbf{x}^{\mathbf{h}, \ell} \sim \{-1, 1\}^n$,

$$\mathbf{x}_{\mathbf{h}^{-1}(c)}^{\mathbf{h}, \ell} = \begin{cases} \mathbf{u}_{\mathbf{h}^{-1}(c)} & \text{if } c > \ell \\ (\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_{\mathbf{h}^{-1}(c)} & \text{if } c \leq \ell. \end{cases}$$

Averaging over $\mathbf{h}$, we get that $\mathbf{x}^{\mathbf{h}, 0} \equiv \mathbf{u}$ and $\mathbf{x}^{\mathbf{h}, L} \equiv \check{\mathbf{y}}$, and so we may write

LHS of (12)

$$\begin{aligned} &= \left| \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{u})] - \mathbb{E}[\tilde{O}_{b, \lambda}(A\check{\mathbf{y}})] \right| \\ &= \left| \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{x}^{\mathbf{h}, 0})] - \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{x}^{\mathbf{h}, L})] \right| \\ &\leq \mathbb{E}_{\mathbf{h}} \left[\left| \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{x}^{\mathbf{h}, 0})] - \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{x}^{\mathbf{h}, L})] \right| \right] \\ &\leq \mathbb{E}_{\mathbf{h}} \left[\left| \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{x}^{\mathbf{h}, 0})] - \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{x}^{\mathbf{h}, L})] \right| \cdot \mathbb{1}[\mathbf{h} \text{ is } H\text{-good}] \right] \\ &\quad + \Pr[\mathbf{h} \text{ is not } H\text{-good}] \\ &\leq \underbrace{\mathbb{E}_{\mathbf{h}} \left[\left| \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{x}^{\mathbf{h}, 0})] - \mathbb{E}[\tilde{O}_{b, \lambda}(A\mathbf{x}^{\mathbf{h}, L})] \right| \cdot \mathbb{1}[\mathbf{h} \text{ is } H\text{-good}] \right]}_{\heartsuit} + \delta. \end{aligned}$$

The penultimate inequality uses the fact that $\tilde{O}_{b, \lambda}$ is $(0, 1)$ -valued (and hence the difference in its expectations under any two distributions is at most 1), and the final inequality is by [Proposition 8.11](#) (note that we indeed have $r_{\text{hash}} \geq C_1 \log(Lm/\delta)$).

It remains to bound $\heartsuit$ by $O(\delta)$. Fix a H -good hash h . By the triangle inequality,

$$\begin{aligned} & \left| \mathbb{E} [\tilde{O}_{b,\lambda}(A\mathbf{x}^{h,0})] - \mathbb{E} [\tilde{O}_{b,\lambda}(A\mathbf{x}^{h,L})] \right| \\ & \leq \sum_{\ell=1}^L \left| \mathbb{E} [\tilde{O}_{b,\lambda}(A\mathbf{x}^{h,\ell-1})] - \mathbb{E} [\tilde{O}_{b,\lambda}(A\mathbf{x}^{h,\ell})] \right|. \end{aligned} \quad (13)$$

Fix $\ell \in [L]$ and consider the corresponding summand

$$\left| \mathbb{E} [\tilde{O}_{b,\lambda}(A\mathbf{x}^{h,\ell-1})] - \mathbb{E} [\tilde{O}_{b,\lambda}(A\mathbf{x}^{h,\ell})] \right|. \quad (14)$$

For notational clarity, let us write B for $h^{-1}(\ell)$ and $\bar{B}$ to denote $[n] \setminus B$. Furthermore, since these “adjacent” hybrid random variables $\mathbf{x}^{h,\ell-1}$ and $\mathbf{x}^{h,\ell}$ agree on all coordinates outside B , we introduce the random variable $\mathbf{s} \sim \{-1, 1\}^{\bar{B}}$ where $\mathbf{s}_{h^{-1}(c)} \equiv \mathbf{x}_{h^{-1}(c)}^{h,\ell-1} \equiv \mathbf{x}_{h^{-1}(c)}^{h,\ell}$ for all $c \neq \ell$. Note that $\mathbf{s}$, $\mathbf{u}_B$, and $(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_B$ are mutually independent. We have that

$$\begin{aligned} & (14) \\ & = \left| \mathbb{E}_{\mathbf{s}} \left[\mathbb{E}_{\mathbf{u}} [\tilde{O}_{b,\lambda}(A\bar{\mathbf{B}}\mathbf{s} + A^B\mathbf{u}_B)] \right. \right. \\ & \quad \left. \left. - \mathbb{E}_{\mathbf{y}^\ell, \tilde{\mathbf{y}}^\ell} [\tilde{O}_{b,\lambda}(A\bar{\mathbf{B}}\mathbf{s} + A^B(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_B)] \right] \right| \\ & \leq \mathbb{E}_{\mathbf{s}} \left[\left| \mathbb{E}_{\mathbf{u}} [\tilde{O}_{b,\lambda}(A\bar{\mathbf{B}}\mathbf{s} + A^B\mathbf{u}_B)] \right. \right. \\ & \quad \left. \left. - \mathbb{E}_{\mathbf{y}^\ell, \tilde{\mathbf{y}}^\ell} [\tilde{O}_{b,\lambda}(A\bar{\mathbf{B}}\mathbf{s} + A^B(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_B)] \right| \right] \\ & = \mathbb{E}_{\mathbf{s}} \left[\left| \mathbb{E}_{\mathbf{u}} [\tilde{O}_{b-A\bar{\mathbf{B}}\mathbf{s},\lambda}(A^B\mathbf{u}_B)] - \mathbb{E}_{\mathbf{y}^\ell, \tilde{\mathbf{y}}^\ell} [\tilde{O}_{b-A\bar{\mathbf{B}}\mathbf{s},\lambda}(A^B(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_B)] \right| \right] \\ & \quad \quad \quad (\text{Fact 6.3}) \\ & = \mathbb{E}_{\mathbf{s}} \left[\left| \mathbb{E}_{\mathbf{u}} [\tilde{O}_{b-A\bar{\mathbf{B}}\mathbf{s},\lambda}(H^B\mathbf{u}_B + T^B\mathbf{u}_B)] \right. \right. \\ & \quad \left. \left. - \mathbb{E}_{\mathbf{y}^\ell, \tilde{\mathbf{y}}^\ell} [\tilde{O}_{b-A\bar{\mathbf{B}}\mathbf{s},\lambda}(H^B(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_B + T^B(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_B)] \right| \right]. \end{aligned}$$

Since h is H -good, every row of H^B is indeed w -sparse, and since every row of T has 2-norm 1, every row of T^B has 2-norm at most 1. Recalling (ii) from above, we may apply Lemma 8.3 to each outcome $\mathbf{s}$ of $\mathbf{s}$, and we get that this quantity is at most

$$\begin{aligned} & \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} \\ & + O\left(\frac{\sqrt{\log m}}{\lambda}\right)^d \left(\mathbb{E}_{\mathbf{u}} [\|T^B\mathbf{u}_B\|_\infty^d] + \mathbb{E}_{\mathbf{y}^\ell, \tilde{\mathbf{y}}^\ell} [\|T^B(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_B\|_\infty^d] \right), \end{aligned}$$

and therefore

RHS of (13)

$$\begin{aligned} & \leq L \cdot \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} + O\left(\frac{\sqrt{\log m}}{\lambda}\right)^d \\ & \sum_{\ell=1}^L \left(\mathbb{E}_{\mathbf{u}} [\|T^{h^{-1}(\ell)}\mathbf{u}_{h^{-1}(\ell)}\|_\infty^d] + \mathbb{E}_{\mathbf{y}^\ell, \tilde{\mathbf{y}}^\ell} [\|T^{h^{-1}(\ell)}(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_{h^{-1}(\ell)}\|_\infty^d] \right). \end{aligned} \quad (15)$$

Since Equation (15) holds for every H -good hash h , we have shown that $\heartsuit$ is at most

$$\begin{aligned} & \mathbb{E}_{\mathbf{h}} [(\text{RHS of (13)}) \cdot \mathbb{1}[\mathbf{h} \text{ is } H\text{-good}]] \\ & \leq \mathbb{E}_{\mathbf{h}} [(\text{RHS of (15)}) \cdot \mathbb{1}[\mathbf{h} \text{ is } H\text{-good}]] \\ & \leq \mathbb{E}_{\mathbf{h}} [(\text{RHS of (15)})] \\ & = L \cdot \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} + O\left(\frac{\sqrt{\log m}}{\lambda}\right)^d \\ & \quad \sum_{\ell=1}^L \left(\underbrace{\mathbb{E}_{\mathbf{h}, \mathbf{u}} [\|T^{h^{-1}(\ell)}\mathbf{u}_{h^{-1}(\ell)}\|_\infty^d] + \mathbb{E}_{\mathbf{h}} [\|T^{h^{-1}(\ell)}(\mathbf{y}^\ell \oplus \tilde{\mathbf{y}}^\ell)_{h^{-1}(\ell)}\|_\infty^d]}_{\diamond} \right). \end{aligned}$$

Applying Lemma 8.10 to bound each of the $2L$ many summands of $\diamond$, we have that

$$\begin{aligned} \heartsuit & \leq L \cdot \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} \\ & \quad + O\left(\frac{\sqrt{\log m}}{\lambda}\right)^d \cdot 2L \cdot O\left(\tau \log(m/\delta) + \sqrt{\frac{\log(m/\delta)}{L}}\right)^d \\ & = L \cdot \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} \\ & \quad + L \cdot O\left(\frac{\tau \sqrt{\log(m)} \log(m/\delta)}{\lambda} + \frac{\sqrt{\log(m) \log(m/\delta)}}{\lambda \sqrt{L}}\right)^d. \end{aligned} \quad (16)$$

By our choice of parameters as set in Section 4.1, and by taking d to be sufficiently large relative to ε , the above expression can be bounded by $O(\delta)$. This establishes Equation (12), and the proof of Theorem 8.1 is complete. $\square$

9 PROOF OF THEOREM 5.2

Having completed both steps of the two-step program described at the end of Section 6 we are finally ready to prove Theorem 5.2:

PROOF OF THEOREM 5.2. Let $\lambda \in (0, 1)$ be as set in Section 4.1. By Lemma 6.7, there are $b^{\text{in}}, b^{\text{out}} \in \mathbb{R}^m$ such that $\tilde{O}_{b^{\text{in}},\lambda}, \tilde{O}_{b^{\text{out}},\lambda}$ are (Λ, δ) -inner and -outer approximators for O_b respectively, where $\Lambda = \Theta(\lambda \sqrt{\log(m/\delta)})$. Next, we apply Lemma 6.9 with $\mathbf{v}$ and $\tilde{\mathbf{v}}$ being $A\mathbf{u}$ and $A\mathbf{z}$ respectively, using Theorem 8.1 to show that Equation (6) is satisfied for both $\tilde{O}_{b^{\text{in}},\lambda}$ and $\tilde{O}_{b^{\text{out}},\lambda}$ with $\gamma = O(\delta)$. We conclude that:

$$\begin{aligned} & \left| \mathbb{Pr}_{\mathbf{u} \sim \{-1, 1\}^n} [A\mathbf{u} \in O_b] - \mathbb{Pr}_{\mathbf{z} \sim \mathcal{G}} [A\mathbf{z} \in O_b] \right| \\ & = O(\delta) + \mathbb{Pr}[A\mathbf{u} \in \mathcal{D}_{\pm\Lambda} O_b] \quad (\text{Lemma 6.9 and Theorem 8.1}) \\ & = O(\delta) + O(\Lambda \sqrt{\log m}) \\ & \quad (\text{Theorem 7.1; note that } \Lambda \geq \tau \text{ is indeed satisfied}) \\ & = O(\delta) + O(\lambda \sqrt{\log(m/\delta) \log m}) \\ & = O(\delta). \end{aligned} \quad (17)$$

This completes the proof of Theorem 5.2. $\square$

ACKNOWLEDGMENTS

R.O. is supported by NSF grants CCF-1618679 and CCF-1717606. R.S. is supported by NSF grant CCF-1563155. L.-Y.T. is supported by NSF grant CCF-1563122; part of this work was performed while the author was at TTI-Chicago.

This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] Miklós Ajtai and Avi Wigderson. 1989. Deterministic Simulation of Probabilistic Constant Depth Circuits. *Advances in Computing Research* 5 (1989), 199–222.
- [2] Louay Bazzi. 2009. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.* 38, 6 (2009), 2220–2272.
- [3] Richard Beigel, Nick Reingold, and Daniel Spielman. 1995. PP is closed under intersection. *Journal of Computer & System Sciences* 50, 2 (1995), 191–202.
- [4] Vidmantas Bentkus. 1990. Smooth approximations of the norm and differentiable functions with bounded support in Banach space l_{∞}^k . *Lithuan. Math. J.* 30, 3 (1990), 223–230.
- [5] Avrim Blum and Ravi Kannan. 1997. Learning an intersection of a constant number of halfspaces under a uniform distribution. *J. Comput. System Sci.* 54, 2 (1997), 371–380.
- [6] Eshan Chattopadhyay, Anindya De, and Rocco Servedio. 2018. Simple and efficient pseudorandom generators from Gaussian processes. Available at <https://eccc.weizmann.ac.il/report/2018/100/>.
- [7] Ilias Diakonikolas, Parikshit Gopalan, Rajesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. 2010. Bounded Independence Fools Halfspaces. *SIAM J. Comput.* 39, 8 (2010), 3441–3462.
- [8] Ilias Diakonikolas, Daniel Kane, and Jelani Nelson. 2010. Bounded independence fools degree-2 threshold functions. In *Proceedings of the 51st Annual Symposium on Foundations of Computer Science (FOCS)*. 11–20.
- [9] Paul Erdős. 1945. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* 51 (1945), 898–902.
- [10] Parikshit Gopalan, Daniel Kane, and Raghu Meka. 2015. Pseudorandomness via the Discrete Fourier Transform. In *Proceedings of the 56th Annual Symposium on Foundations of Computer Science (FOCS)*. 903–922.
- [11] Parikshit Gopalan, Adam Klivans, and Raghu Meka. 2012. Learning Functions of Halfspaces using Prefix Covers. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*. 15.1–15.10.
- [12] Parikshit Gopalan, Raghu Meka, and Omer Reingold. 2013. DNF sparsification and a faster deterministic counting algorithm. *Comput. Complexity* 22, 2 (2013), 275–310. <https://doi.org/10.1007/s00037-013-0068-6>
- [13] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. 2010. Fooling Functions of Halfspaces under Product Distributions. In *Proceedings of the 25th Annual Conference on Computational Complexity (CCC)*. 223–234. Full version available at <https://arxiv.org/abs/1001.1593>.
- [14] Prahladh Harsha, Adam R. Klivans, and Raghu Meka. 2012. An invariance principle for polytopes. *J. ACM* 59, 6 (2012), 29:1–29:25. <https://doi.org/10.1145/2395116.2395118>
- [15] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. 1994. Pseudorandomness for network algorithms. In *Proceedings of the 26th Annual Symposium on Theory of Computing (STOC)*. 356–364.
- [16] Daniel Kane. 2011. k -Independent Gaussians Fool Polynomial Threshold Functions. In *Proceedings of the 26th IEEE Conference on Computational Complexity (CCC)*. 252–261.
- [17] Daniel Kane. 2011. A Small PRG for Polynomial Threshold Functions of Gaussians. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 257–266.
- [18] Daniel Kane. 2014. The average sensitivity of an intersection of halfspaces. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*. 437–440.
- [19] Daniel Kane. 2014. A pseudorandom generator for polynomial threshold functions of Gaussians with subpolynomial seed length. In *Proceedings of the 29th Annual Conference on Computational Complexity (CCC)*. 217–228.
- [20] Daniel Kane and Sankeerth Rao. 2018. A PRG for Boolean PTF of degree 2 with seed length subpolynomial in ϵ and logarithmic in n . In *Proceedings of the 33rd Computational Complexity Conference (CCC)*. 2:1–2:24.
- [21] Zohar Shay Karnin, Yuval Rabani, and Amir Shpilka. 2012. Explicit Dimension Reduction and Its Applications. *SIAM J. Comput.* 41, 1 (2012), 219–249.
- [22] Subhash Khot and Rishi Saket. 2011. On the hardness of learning intersections of two halfspaces. *J. Comput. Syst. Sci.* 77, 1 (2011), 129–141.
- [23] Adam Klivans, Ryan O'Donnell, and Rocco A. Servedio. 2004. Learning intersections and thresholds of halfspaces. *Journal of Computer & System Sciences* 68, 4 (2004), 808–840.
- [24] Adam Klivans, Ryan O'Donnell, and Rocco A. Servedio. 2008. Learning Geometric Concepts via Gaussian Surface Area. In *Proceedings of the 49th Symposium on Foundations of Computer Science (FOCS)*. 541–550.
- [25] Adam Klivans and Alexander Sherstov. 2006. Cryptographic Hardness for Learning Intersections of Halfspaces. In *Proceedings of the 47th Symposium on Foundations of Computer Science (FOCS)*. 553–562.
- [26] Praveesh K. Kothari and Raghu Meka. 2015. Almost Optimal Pseudorandom Generators for Spherical Caps. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing (STOC)*. 247–256.
- [27] John Littlewood and Albert Cyril Offord. 1943. On the number of real roots of a random algebraic equation. *III. Rec. Math. [Mat. Sbornik] N.S.* 12 (1943), 277–286.
- [28] Raghu Meka and David Zuckerman. 2013. Pseudorandom generators for polynomial threshold functions. *SIAM J. Comput.* 42, 3 (2013), 1275–1301.
- [29] Marvin Minsky and Seymour Papert. 1968. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, MA.
- [30] Fedor Nazarov. 2003. 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)*. Lecture Notes in Math., Vol. 1807, Springer, 169–187.
- [31] Noam Nisan. 1991. Pseudorandom bits for constant depth circuits. *Combinatorica* 11, 1 (1991), 63–70.
- [32] Noam Nisan. 1992. Pseudorandom generators for space-bounded computations. *Combinatorica* 12, 4 (1992), 449–461.
- [33] Noam Nisan and Avi Wigderson. 1994. Hardness vs Randomness. *J. Comput. Syst. Sci.* 49, 2 (1994), 149–167. [https://doi.org/10.1016/S0022-0000\(05\)80043-1](https://doi.org/10.1016/S0022-0000(05)80043-1)
- [34] Noam Nisan and David Zuckerman. 1996. Randomness is Linear in Space. *J. Comput. Syst. Sci.* 52, 1 (1996), 43–52. <https://doi.org/10.1006/jcss.1996.0004>
- [35] Ryan O'Donnell. 2014. *Analysis of Boolean Functions*. Cambridge University Press. Available at <http://analysisofbooleanfunctions.net/>.
- [36] Ryan O'Donnell and Rocco Servedio. 2010. New Degree Bounds for Polynomial Threshold Functions. *Combinatorica* 30, 3 (2010), 327–358.
- [37] Alexander Razborov. 2009. A simple proof of Bazzi's theorem. *ACM Transactions on Computation Theory* 1, 1 (2009), 3.
- [38] Haskell Rosenthal. 1970. On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables. *Israel J. Math* 8 (1970), 273–303.
- [39] Rocco A. Servedio and Li-Yang Tan. 2017. Fooling intersections of low-weight halfspaces. In *Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*. 824–835.
- [40] Rocco A. Servedio and Li-Yang Tan. 2017. What circuit classes can be learned with non-trivial savings?. In *Proceedings of the 8th Innovations in Theoretical Computer Science (ITCS)*. 30:1–30:21.
- [41] Alexander A. Sherstov. 2013. The Intersection of Two Halfspaces Has High Threshold Degree. *SIAM J. Comput.* 42, 6 (2013), 2329–2374.
- [42] Alexander A. Sherstov. 2013. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica* 33, 1 (2013), 73–96.
- [43] Terence Tao. 2010. 254A Notes: Topics in random matrix theory. (2010). <https://terrytao.wordpress.com/tag/lindeberg-replacement-trick/>.
- [44] Salil P. Vadhan. 2012. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science* 7, 1–3 (2012), 1–336.
- [45] Santosh Vempala. 2010. A random-sampling-based algorithm for learning intersections of halfspaces. *J. ACM* 57, 6:32 (2010).