Fooling Polytopes

RYAN O'DONNELL, Carnegie Mellon University, USA
ROCCO A. SERVEDIO, Columbia University, USA
LI-YANG TAN, Stanford University, USA

We give a pseudorandom generator that fools m-facet polytopes over $\{0,1\}^n$ with seed length $\mathrm{polylog}(m) \cdot \log n$. The previous best seed length had superlinear dependence on m.

CCS Concepts: • **Theory of computation**;

Additional Key Words and Phrases: Polytopes, pseudorandom generators, Littlewood-Offord, central limit theorems

ACM Reference format:
Ryan O'Donnell, Rocco A. Servedio, and Li-Yang Tan. 2022. Fooling Polytopes. *J. ACM* 69, 2, Article 9 (January 2022), 37 pages.
https://doi.org/10.1145/3460532

1 INTRODUCTION

Unconditional derandomization has been a major focus of research in computational complexity theory for more than 30 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, 43, 45] and branching programs [22, 44, 46], 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 [9, 15, 27, 29, 35, 39], low-degree polynomial threshold functions [10, 24, 25, 27, 28, 39], and, most relevant to this article, *intersections of halfspaces* [8, 18, 19, 56].

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

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.
Authors' addresses: R. O'Donnell, School of Computer Science, 5000 Forbes Avenue, Pittsburgh, PA 15213; email: odonnell@cs.cmu.edu; R. A. Servedio, Computer Science Department, 500 West 120 Street, Room 450, MC0401, New York, New York 10027; email: rocco@cs.columbia.edu; L.-Y. Tan, Gates Computer Science Building, 353 Jane Stanford Way, Stanford, CA 94305; email: liyang@cs.stanford.edu.

0004-5411/2022/01-ART9 $15.00
https://doi.org/10.1145/3460532

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 that 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 [4, 26, 40, 48, 58, 59] and computational learning theory [6, 16, 30, 32–34, 57, 64].

The main result of this article 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 $F(x) = \mathbb{1}[w \cdot x \le \theta]$ is said to be τ-*regular* if $|w_j| \le \tau \|w\|_2$ 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 [19] gave a PRG that δ-fools any intersection of m many τ-regular halfspaces with seed length $\mathrm{poly}(\log m, 1/\delta) \cdot \log n$, where τ has to be sufficiently small relative to m and δ (specifically, $\tau \le$ some $\mathrm{poly}(\frac{\delta}{\log m})$ 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 $F(x) = \mathbb{1}[x_1 \le 1/2]$, that are highly irregular.

Recently, Reference [56] built on the work of Reference [19] 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 $\mathrm{poly}(\log m, W, 1/\delta) \cdot \mathrm{polylog}\, n$; a halfspace has weight W if it can be expressed as $\mathbb{1}[w \cdot x \le \theta]$ 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 Reference [56] result also cannot be used to fool even two general halfspaces.

In Reference [18], 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 that is at least linear in m, and indeed their overall seed length is $O((m \log(m/\delta) + \log n) \cdot \log(m/\delta))$.[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 $m \ge n$ many halfspaces. (Indeed, this PRG of Reference [18] 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

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

Table 1. PRGs for Intersections of Halfspaces Over $\{0,1\}^n$

Reference	Function class	Seed length of PRG
[18]	Monotone functions of m halfspaces	$O((m\log(m/\delta)+\log n)\cdot\log(m/\delta))$ $O(m\log(m/\delta)+\log n)$, if $m/\delta \le$ any polylog(n)
[19]	Intersections of m τ-regular halfspaces	$\mathrm{poly}(\log m, 1/\delta)\cdot\log n$, if $\tau \le$ some $\mathrm{poly}(\frac{\delta}{\log m})$
[56]	Intersections of m weight-W halfspaces	$\mathrm{poly}(\log m, W, 1/\delta)\cdot\mathrm{polylog}\, n$
This work	Intersections of m halfspaces	$\mathrm{poly}(\log m, 1/\delta)\cdot\log n$

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 [39] 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 $\tau \in [\frac{1}{\sqrt{n}}, 1]$. Reference [19] 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, Reference [19] obtained a Gaussian PRG for intersections of m halfspaces with seed length $\mathrm{poly}(\log m, 1/\delta)\cdot\log n$. Recent work of Reference [8] gives a different Gaussian PRG with seed length $\mathrm{poly}(\log m, 1/\delta)+O(\log n)$.

The focus of the current work is on the standard setting of PRGs over Boolean space, and the rest of the article 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 $m \le n$ general halfspaces had a superlinear seed length dependence on m. The PRGs that could fool intersections of $m \ge n$ halfspaces imposed technical restrictions on the halfspaces: either regularity (hence excluding simple halfspaces such as $\mathbb{1}[x_1 \le 1/2]$) or small weights (hence excluding almost all halfspaces). Please refer to Table 1.

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

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* $\mathrm{poly}(\log m, 1/\delta)\cdot\log n$ *that δ-fools the class of intersections of m halfspaces over $\{0,1\}^n$.*

In particular, this PRG fools intersections of quasipoly(n) many halfspaces with seed length 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 $n^{\mathrm{polylog}(m)}$ 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 $\{0, 1, \ldots, N\}^n$ and the solid cube $[0, 1]^n$ (details are left to the interested reader).

1.3 Discussion

Since the initial conference publication of a preliminary version of this article [50], several works have appeared that are relevant to the topic of this article. One line of work has been on obtaining pseudorandom generators for functions of halfspaces (including intersections of halfspaces, i.e., polytopes) with a better dependence on the error parameter but a worse dependence on other parameters. Kabanets et al. [23] gave a PRG that δ-fools the class of n-variable size-s de Morgan formulas with halfspace gates at the bottom and has seed length $O(n^{1/2}s^{1/4}\log(n)\log(n/\delta))$, and Hatami et al. [20] gave a PRG that δ-fools the class of arbitrary functions of m halfspaces over n variables with seed length $\tilde{O}(\sqrt{n(m+\log(1/\delta))})$. The techniques employed in these works are very different from the methods of our article; even for the special case of intersections of m halfspaces, each of these results has a seed length that is polynomial in n and m (rather than logarithmic or polylogarithmic as in our work), but these results have a much better seed length dependence on the error parameter δ. Achieving a polylogarithmic dependence on all three parameters n, m, and $1/\delta$ is an interesting challenge for future work.

In a different related line of work, Arunachalam and Yao [2] have considered the problem of constructing explicit PRGs for positive spectrahedrons. A positive spectrahedron is a Boolean function $\mathbb{1}[x_1A^1 + \cdots + x_nA^n \preceq B]$, where the A^i's are $k \times k$ positive semidefinite matrices. Building on some of the ideas and ingredients in this article and in prior work [19], they establish invariance principles and give a PRG that δ-fools "sufficiently regular" positive spectrahedrons over $\{0, 1\}^n$ with seed length $\mathrm{poly}(\log k, 1/\delta) \cdot \log n$.

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 that 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 [12, 37] that we establish. We now describe our proof and the new ideas underlying it in detail.

2.1 Background: The Reference [19] PRG for Regular Polytopes

We begin by recalling the arguments of Harsha, Klivans, and Meka [19] for fooling regular polytopes. At a high level, Reference [19] builds on the work of Meka and Zuckerman [39], which gave a versatile and powerful framework for constructing pseudorandom generators from probabilistic *invariance principles*; the main technical ingredient underlying the Reference [19] PRG for regular polytopes is a new invariance principle for such polytopes, which we now describe.

Reference [19]'s invariance principle and the Lindeberg method. At a high level, the Reference [19] invariance principle for regular polytopes is as follows: Given an m-tuple of regular linear forms over n input variables $x = (x_1, \ldots, x_n)$ (denoted by Ax, where A is an m-by-n matrix), the distribution (over $\mathbb{R}^m$) of $A\boldsymbol{u}$, where $\boldsymbol{u} \sim \{-1, 1\}^n$ is uniform random, is very close to the distribution of $A\boldsymbol{g}$, where $\boldsymbol{g} \sim \mathcal{N}(0, 1)^n$ 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 $\mathbb{1}[Ax \le b]$ where $b \in \mathbb{R}^m$, which syncs up precisely with an intersection of m halfspaces $\mathbb{1}[A_1 x \le b_1] \wedge \cdots \wedge \mathbb{1}[A_m x \le b_m]$. To prove this invariance principle, Reference [19] employs the well-known Lindeberg method (see, e.g., Chapter §11 of References [47] and [61]) 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 $\mathbb{1}[Ax \le b]$, 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 $A\boldsymbol{u}$ and $A\boldsymbol{g}$ using test functions that are orthant indicators $O_b(v_1, \ldots, v_m) = \mathbb{1}[v \le b]$ (corresponding to multidimensional CDF distance), distance is measured using a sufficiently smooth *mollifier* $\widetilde{O}_b : \mathbb{R}^m \to [0,1]$ of O_b. Such mollifiers, with useful properties that we now discuss, were proposed and analyzed by Bentkus [5]. In more detail, Reference [19] proves that the difference between the expectations of $\widetilde{O}_b(A\boldsymbol{u})$ and $\widetilde{O}_b(A\boldsymbol{g})$ is bounded by a certain function involving $\widetilde{O}_b$'s derivatives. In fact, as in standard in Lindeberg-style proofs of invariance principles, Reference [19] actually bounds this difference with respect to *any* smooth test function $\Upsilon : \mathbb{R}^m \to \mathbb{R}$ in terms of Υ's derivatives; the only specific property of Bentkus's mollifier $\widetilde{O}_b$ 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 $\Upsilon(A\boldsymbol{u})$ to $\Upsilon(A\boldsymbol{g})$, using the multidimensional Taylor expansion of Υ 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 aforedescribed smooth invariance principle applied to Bentkus's mollifier. This step relies on a second key property of Bentkus's mollifier: $\widetilde{O}_b$ 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 $A\boldsymbol{g}$; intuitively, such anticoncentration establishes that $A\boldsymbol{g}$ does not place too much probability weight on the error region where $\widetilde{O}_b$ disagrees with O_b. In Reference [19], the required anticoncentration for $A\boldsymbol{g}$ follows immediately from a result of Nazarov [33, 42] on the Gaussian surface area of m-facet polytopes.

The Reference [19] PRG via a derandomized invariance principle. Having proved this invariance principle for regular polytopes, Reference [19] then establishes a *pseudorandom* version by derandomizing its proof. That is, they argue that their proof in fact establishes multidimensional-CDF-closeness between Az and $A\boldsymbol{g}$, where $\boldsymbol{g} \sim \mathcal{N}(0,1)^n$ is distributed according to a standard Gaussian as before, but $z \sim \{-1,1\}^n$ is the output of a suitable pseudorandom suitable generator $\mathscr{G} : \{-1,1\}^r \to \{-1,1\}^n$ (rather than uniform random). Combining the "full-randomness" invariance principle (establishing closeness between $A\boldsymbol{u}$ and $A\boldsymbol{g}$) with this pseudorandom version (establishing closeness between Az and $A\boldsymbol{g}$), it follows from the triangle inequality that Az and $A\boldsymbol{u}$ are close. Recalling that multidimensional CDF distance corresponds to test functions of the form $\mathbb{1}[Ax \le b] = \mathbb{1}[A_1 x \le b_1] \wedge \cdots \wedge \mathbb{1}[A_m x \le b_m]$, this is precisely equivalent to the claim that $\mathscr{G}$ 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 Reference [19] generator (for fooling intersections of m many τ-regular halfspaces):

(1) Pseudorandomly hash the n variables into $L := \mathrm{poly}(1/\tau)$ buckets using an $(r_{\mathrm{hash}} := 2\log m)$-wise uniform hash function $\boldsymbol{h} : [n] \to [L]$.

(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 [39] for fooling a single regular halfspace, the only difference being that the relevant parameters L, r_{hash}, and r_{bucket} are larger in Reference [19] than in Reference [39] (naturally so, given that the Reference [19] generator fools intersections of m regular halfspaces instead of a single one).

Our analysis in this article can be used to show that the Reference [39] 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 Reference [39] 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 [17]. 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 Reference [19] PRG result from regular to general polytopes stems from the fact that an invariance principle *simply does not hold* for general polytopes $Ax \le b$. Without the regularity requirement on A, it is not true that $A\boldsymbol{u}$ and $A\boldsymbol{g}$ are close in CDF distance; indeed, even a single non-regular linear form such as x_1 is distributed very differently under $\boldsymbol{u} \sim \{-1, 1\}^n$ versus $\boldsymbol{g} \sim \mathcal{N}(0, 1)^n$. This therefore necessitates a significant conceptual departure from the Meka–Zuckerman framework for constructing pseudorandom generators from invariance principles: Rather than establishing closeness between $A\boldsymbol{u}$ and $A\boldsymbol{z}$ (where $\boldsymbol{z} \sim \{-1, 1\}^n$ is the output of a suitable pseudorandom generator) through $A\boldsymbol{g}$ by means of an invariance principle, one has to establish closeness between $A\boldsymbol{u}$ and $A\boldsymbol{z}$ "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 $A\boldsymbol{u}$ and $A\boldsymbol{z}$ are close with respect to Bentkus's smooth mollifiers $\widetilde{O}_b$ 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 Reference [19], 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 Reference [19] first proves a general invariance principle establishing closeness in expectation (with a quantitative bound that depends on Υ's derivatives) between $\Upsilon(A\boldsymbol{u})$ and $\Upsilon(A\boldsymbol{g})$ for *any* smooth test function Υ. They then apply this general invariance principle with Bentkus's orthant mollifier $\widetilde{O}_b$ being the test function, using the bounds on $\widetilde{O}_b$'s derivatives established in Reference [5] but no other properties of $\widetilde{O}_b$.

In contrast, we do not prove closeness between $A\boldsymbol{u}$ and $A\boldsymbol{z}$ for all smooth test functions; our argument is carefully tailored to Bentkus's specific mollifier. In addition to bounds on $\widetilde{O}_b$'s derivatives, we crucially rely on the specific structure of $\widetilde{O}_b$, in particular, the fact that it is the

product of m univariate functions, one for each coordinate (i.e., $\widetilde{O}_b(v) = \prod_{i=1}^{m} \psi_{b_i}(v_i)$, 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 that 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 $\widetilde{O}_b$ allows us to handle these head portions using pseudorandom generators for *small-width CNF formulas* [17]. (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 Reference [19]'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 $|\mathbf{E}[\Upsilon(\boldsymbol{v}+\boldsymbol{\Delta})] - \mathbf{E}[\Upsilon(\boldsymbol{v}+\boldsymbol{\Delta}')]|$ are bounded by considering two Taylor expansions of Υ, both taken around the "common point" $\boldsymbol{v}$. Lindeberg method arguments analyze the difference of these Taylor expansions using moment-matching properties of $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}'$ 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 $|\mathbf{E}[\Upsilon(\boldsymbol{v}+\boldsymbol{\Delta})] - \mathbf{E}[\Upsilon(\boldsymbol{v}'+\boldsymbol{\Delta}')]|$. Note that this involves considering the Taylor expansions of Υ around two *distinct* points $\boldsymbol{v}$ and $\boldsymbol{v}'$, which may be far from each other—indeed, *a priori* it is not even clear that $|\mathbf{E}[\Upsilon(\boldsymbol{v})] - \mathbf{E}[\Upsilon(\boldsymbol{v}')]|$ will be small. Because of these differences from the standard Lindeberg scenario, moment-matching properties of $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}'$ 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 $|\mathbf{E}[\Upsilon(\boldsymbol{v}+\boldsymbol{\Delta})] - \mathbf{E}[\Upsilon(\boldsymbol{v}'+\boldsymbol{\Delta}')]|$ (see Section 8).

Step 2: Anticoncentration. The next step is to pass from closeness of $\widetilde{O}_b(A\boldsymbol{u})$ and $\widetilde{O}_b(A\boldsymbol{z})$ in expectation, to closeness of $A\boldsymbol{u}$ and $A\boldsymbol{z}$ in multidimensional CDF distance. We recall that in the analogous step in Reference [19]'s proof, the starting point was closeness in expectation of $\widetilde{O}_b(A\boldsymbol{u})$ and $\widetilde{O}_b(A\boldsymbol{g})$, where $\boldsymbol{g} \sim \mathcal{N}(0,1)^n$ is a standard Gaussian (instead of $\widetilde{O}_b(A\boldsymbol{z})$ where $\boldsymbol{z} \sim \{-1,1\}^n$ is pseudorandom). For this reason, it sufficed for Reference [19] to bound the Gaussian anticoncentration of $A\boldsymbol{g}$ 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 $A\boldsymbol{u}$ where $\boldsymbol{u} \sim \{-1,1\}^n$ 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 [12, 37]. 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 $j \in [n]$,*

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

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

Littlewood and Offord [37] first proved a bound of $O((\log n)/\sqrt{n})$; Erdös [12] 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 $w = (1, 0, \ldots, 0)$ 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 $\mathbb{1}[w \cdot x \le \theta]$ where $|w_j| \ge 1$ 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 $Ax \le b$ where $|A_{ij}| \ge 1$ for all i and j?*

In more detail, we say that u lies within the "width-2 boundary" of the polytope $Ax \le b$ provided $Au \le b$ and $A_i \cdot u > b_i - 2$ for some $i \in [m]$; equivalently, u lies in the difference of the two polytopes $Ax \le b$ and $Ax \le b - 2 \cdot \mathbb{1}_m$, 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). *For all $m \ge 2$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$ with $|A_{ij}| \ge 1$ for all $i \in [m]$ and $j \in [n]$,*

$$\mathbf{Pr}[A\boldsymbol{u} \le b \ \&\ A_i \cdot \boldsymbol{u} > b_i - 2 \text{ for some } i \in [m]] \le \frac{5\sqrt{2 \ln m}}{\sqrt{n}},$$

where $\boldsymbol{u} \sim \{-1, 1\}^n$ 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 [26]. 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 $\widetilde{O}_b$ 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.

Remark 2.3. Our generalization of the Littlewood–Offord theorem (Theorem 2.2) is, to our knowledge, incomparable to other high-dimensional generalizations that have been studied in the literature. In particular, References [14, 31, 62] (see also the references therein) study the probability that $A\boldsymbol{u}$ falls within a ball of fixed radius in $\mathbb{R}^m$, where $A \in \mathbb{R}^{m \times n}$ is a matrix whose columns have 2-norm at least 1 (i.e., $A\boldsymbol{u}$ is the random ± 1 sum of n many m-dimensional vectors of length at least 1).

2.3 Relation to Reference [56]

We close this section with a discussion of the connection between our techniques and those of the recent work cited in Reference [56]. Recall that the main result of Reference [56] is a PRG for

δ-fooling intersections of m weight-W halfspaces using seed length $\mathrm{poly}(\log m, W, 1/\delta) \cdot \mathrm{polylog}\, n$ (whereas our main result, which is strictly stronger, is a PRG for δ-fooling intersections of m general halfspaces using seed length $\mathrm{poly}(\log m, 1/\delta) \cdot \log n$, with no dependence on the weights of the halfspaces).

A key structural observation driving Reference [56] 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 Reference [56] PRG.) From this starting point, it suffices for Reference [56] to bound the multidimensional CDF distance between the $(\mathbb{R}^m \times \{\pm 1\})$-valued random variables $(A\boldsymbol{u}, G(\boldsymbol{u}))$ and $(A\boldsymbol{z}, G(\boldsymbol{z}))$, where $A \in \mathbb{R}^{m \times n}$ is the weight matrix of H, $\boldsymbol{u}$ is uniform random, and $\boldsymbol{z}$ is the output of the Reference [56] PRG (which is a slight variant of Reference [19]'s pseudorandom generator). Since H is an intersection of regular halfspaces, the fact that $A\boldsymbol{u}$ and $A\boldsymbol{z}$ are close in multidimensional CDF distance is precisely the main result of Reference [19]; the crux of the work in Reference [56], therefore, lies in dealing with the additional distinguished $(m+1)^{\mathrm{st}}$ coordinate corresponding to the CNF G. Very roughly speaking, Reference [56] employs a careful coupling $(\widehat{\boldsymbol{u}}, \widehat{\boldsymbol{z}})$ (whose existence is a consequence of the fact that bounded independence fools CNFs [3, 52]) to ensure that $G(\widehat{\boldsymbol{u}})$ and $G(\widehat{\boldsymbol{z}})$ almost always agree, and hence these $(m+1)^{\mathrm{st}}$ coordinates "have a negligible effect" throughout Reference [19]'s Lindeberg-based proof of the regular case establishing closeness between $A\boldsymbol{u}$ and $A\boldsymbol{z}$.

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 Reference [56] did not require as significant a departure from Reference [19]'s approach, and from the Lindeberg method as a whole, as the general case that is the subject of this article. In particular, the new ideas discussed in Section 2.2 that are central to our proof were not present in Reference [56]'s analysis for the low-weight case. To elaborate on this,

- Reference [56] did not have to exploit the product structure of Bentkus's orthant mollifier $\widetilde{O}_b$ to fool it. Like Reference [19], the arguments of Reference [56] establish closeness in expectation between $\Upsilon(A\boldsymbol{u}, G(\boldsymbol{u}))$ and $\Upsilon(A\boldsymbol{z}, G(\boldsymbol{z}))$ for *all* smooth test functions Υ, and the only properties of Bentkus's mollifier that are used are the bounds on its derivatives given in Reference [5] (which are used in a black box way). The simpler setting of Reference [56] also did not necessitate comparing the Taylor expansions of Υ around distinct points, as discussed in Section 2.2.
- Reference [56] 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 Reference [19], Reference [56] was 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 article, we view halfspaces as having the domain $\{-1,1\}^n$ rather than $\{0,1\}^n$. We remind the reader that a halfspace $F : \{-1,1\}^n \to \{0,1\}$ is a function of the form $F(x) = \mathbb{1}[w \cdot x \le \theta]$ for some $w \in \mathbb{R}^n$, $\theta \in \mathbb{R}$.

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 ith 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 $w \in \mathbb{R}^n$ is *τ-regular* if $|w_j| \le \tau \|w\|_2$ for all $j \in [n]$. More generally, we say that w is *(k,τ)-regular* if there is a partition $[n] = \textsc{Head} \sqcup \textsc{Tail}$ where $|\textsc{Head}| \le k$ and the subvector $w_{\textsc{Tail}}$ is τ-regular. We say that w is *(k,τ)-standardized* if w is (k,τ)-regular and $\sum_{j \in \textsc{Tail}} w_j^2 = 1$. 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 $Ax \le b$.

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

$$O_b = \{v \in \mathbb{R}^m : v_i \le b_i \text{ for all } i \in [m]\}.$$

We will overload notation and also write "O_b" to denote the indicator $\mathbb{R}^m \to \{0,1\}$ of the orthant O_b (i.e., $O_b(v) = \mathbb{1}[v \le b]$). We write $\partial O_b \subset O_b$ to denote O_b's *surface*,

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

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

$$\partial_{-\Lambda} O_b = O_b \setminus O_{b-(\Lambda,\ldots,\Lambda)}, \qquad \partial_{+\Lambda} O_b = O_{b+(\Lambda,\ldots,\Lambda)} \setminus O_b, \tag{1}$$

and $\partial_{\pm\Lambda} O_b$ to denote the disjoint union $\partial_{\pm\Lambda} O_b = \partial_{+\Lambda} O_b \sqcup \partial_{-\Lambda} O_b$.

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

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

where

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

3.2 Pseudorandomness Preliminaries

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

We recall the definition of a pseudorandom generator:

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

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

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

Bounded independence and hash families. A sequence of random variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is said to be *r-wise independent* if any collection of r of them is independent. In case the $\boldsymbol{x}_i$'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 $\mathscr{G}$ is that they 0-fool the class of degree-r polynomials $\{-1,1\}^n \to \mathbb{R}$.

A distribution $\mathcal{H}$ on functions $[n] \to [L]$ is said to be an *r-wise uniform hash family* if, for $\boldsymbol{h} \sim \mathcal{H}$, the sequence $(\boldsymbol{h}(1), \ldots, \boldsymbol{h}(n))$ is r-wise uniform. Such a distribution also has the property that for any $\ell \in [L]$, the sequence $(\mathbb{1}_{\boldsymbol{h}(1)=\ell}, \ldots, \mathbb{1}_{\boldsymbol{h}(n)=\ell})$ 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 Reference [63]) give that for every n, L and r, there is an r-wise uniform hash family $\mathcal{H}$ of functions $[n] \to [L]$ 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 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 [17] 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 $\mathscr{G}_{\mathrm{GMR}} = \mathscr{G}_{\mathrm{GMR}}(w, \delta_{\mathrm{CNF}})$ that δ_{CNF}-fools the class of all width-w CNF formulas over $\{-1,1\}^n$ and has seed length*

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

4 OUR PRG

The Meka–Zuckerman generator. As stated earlier, the PRG that we will analyze is a slight variant of a PRG first proposed by Meka and Zuckerman for fooling a single halfspace [39]. We begin by recalling the Meka–Zuckerman PRG.

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

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

In words, an r_{hash}-wise uniform hash $\boldsymbol{h}$ is used to partition the variables $x_1, \ldots, x_n$ 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 References [19, 56] also have this structure (though the choice of parameters L, r_{bucket}, and r_{hash} are different than those in Reference [39]).

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 $\mathscr{G}$, 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:

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

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

$$\breve{\boldsymbol{y}}_{\boldsymbol{h}^{-1}(\ell)} = \left(\boldsymbol{y}^\ell \oplus \tilde{\boldsymbol{y}}^\ell\right)_{\boldsymbol{h}^{-1}(\ell)} \qquad \text{for all } \ell \in [L],$$

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

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 $\mathscr{G}$ with parameters $L, r_{\text{hash}}, r_{\text{bucket}}, k, w, \delta_{\text{CNF}}$ is*

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

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:

$$\begin{aligned}
\lambda &= \frac{\delta}{\sqrt{\log(m/\delta)\log m}} && \text{(Dictated by Equation (31))}\\
\tau &= \frac{\delta^{1+\varepsilon}}{(\log m)^{2.5+2\varepsilon}} && \text{(Dictated by Equation (30))}\\
d &= \text{constant depending only on } \varepsilon. && \text{(Dictated by Equation (30))}
\end{aligned}$$

The precise value of $d = d(\varepsilon)$ will be specified in the proof of Theorem 8.1. We will instantiate our generator $\mathscr{G} = \mathscr{G}(L, r_{\text{hash}}, r_{\text{bucket}}, k, w, \delta_{\text{CNF}})$ with parameters:

$$L = \frac{(\log m)^5}{\delta^{2+\varepsilon}} \qquad \text{(Constrained by Equation (30), chosen to optimize seed length)}$$

$$r_{\text{hash}} = C_1 \log(Lm/\delta) \qquad \text{(Dictated by Proposition 8.11)}$$
$$r_{\text{bucket}} = \log(m/\delta) \qquad \text{(Dictated by Lemma 8.10)}$$
$$k = \frac{C_2 \log(m/\delta) \log\log(m/\delta)}{\tau^2} \qquad \text{(Dictated by Theorem 5.1)}$$
$$w = \frac{2k}{L} \qquad \text{(Dictated by Proposition 8.11)}$$
$$\delta_{\text{CNF}} = \frac{\delta}{L} \cdot \left(\frac{\lambda}{m\sqrt{n}}\right)^{d-1}, \qquad \text{(Dictated by Equation (30))}$$

where C_1 and C_2 are absolute constants specified in the proofs of Proposition 8.11 and Theorem 5.1, respectively.

Our seed length: By Fact 4.3, our overall seed length is

$$\text{polylog}(m) \cdot \delta^{-(2+\varepsilon)} \cdot \log n \tag{2}$$

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

Remark 4.4. As alluded to in the introduction, our techniques can also be used to show that the Meka–Zuckerman generator itself fools the class of intersections of m halfspaces over $\{-1, 1\}^n$. However, this would require setting the parameters L, r_{hash}, and r_{bucket} to be somewhat larger than the values used above and would result in a slightly worse seed length of $\text{poly}(\log m, 1/\delta, \log n)$ than our $\text{poly}(\log m, 1/\delta) \cdot \log n$. Briefly, such an analysis would use the fact that bounded-uniformity distributions fool CNF formulas [3, 52]; our analysis instead uses the (more efficient) Reference [17] generator for this purpose.

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 (Approximating Arbitrary Polytopes 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*

$$k := \frac{C_2 \log(m/\delta) \log\log(m/\delta)}{\tau^2} \leq \frac{n}{2}. \tag{3}$$

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 $\boldsymbol{y} \sim \{-1, 1\}^n$ is $2k$-wise uniform, then

$$\mathbf{Pr}[\mathbb{1}[A\boldsymbol{y} \leq b] \neq \mathbb{1}[A'\boldsymbol{y} \leq b']] \leq \delta. \tag{4}$$

Remark 5.2. Had we been content in this theorem with the worse value of $k = O\left(\log^2(m/\delta)/\tau^2\right)$, then the result would essentially be implicit in Reference [9, Theorem 5.4] (and Reference [18, Theorem 7.4]), using only $(k + 2)$-wise uniformity. To save essentially a $\log(m/\delta)$ factor, we give a modified proof in Appendix A.

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 $\boldsymbol{z} = \check{\boldsymbol{y}} \oplus \boldsymbol{y}^\star$ from our generator is indeed $2k$-wise uniform, since $\boldsymbol{y}^\star$ is; indeed,Lemma 5.1 is the motivation for why our

construction includes a bitwise-Xor with $\boldsymbol{y}^\star$.) 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., $\mathbf{Pr}[F(\boldsymbol{u}) \neq F'(\boldsymbol{u})]$ is small), fooling F' does not suffice to fool F itself.

Given Lemma 5.1, to prove Theorem 1.1 it is sufficient to prove the following:

Theorem 5.3 (Fooling (k,τ)-standardized Polytopes). *Let $\mathscr{G}$ be our generator with parameters as set in Section 4.1. For all m-facet (k,τ)-standardized polytopes $A'x \le b'$,*

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

Proof of Theorem 1.1 Assuming Theorem 5.3 and Lemma 5.1. Let $Ax \le b$ be any m-facet polytope in $\mathbb{R}^n$. Given $\delta > 0$, we recall that $\tau = \delta^{1+\varepsilon}/(\log m)^{2.5+\varepsilon}$. If the quantity (3) is greater than $n/2$, then the claimed seed length from Fact 4.3 is greater than n and the conclusion of Theorem 1.1 trivially holds, so we suppose that (3) is less than $n/2$. Let $A'x \le b'$ be the m-facet (k,τ)-standardized polytope given by Lemma 5.1. We have

$$\begin{aligned}
\Pr_{\boldsymbol{u}\sim\{-1,1\}^n}[A\boldsymbol{u} \in O_b] &= \Pr_{\boldsymbol{u}\sim\{-1,1\}^n}[A'\boldsymbol{u} \in O_{b'}] \pm \delta && \text{(Lemma 5.1 applied to } \boldsymbol{u}\text{)}\\
&= \Pr_{\boldsymbol{z}\sim\mathcal{G}}[A'\boldsymbol{z} \in O_{b'}] \pm \delta \pm \delta && \text{(Theorem 5.3)}\\
&= \Pr_{\boldsymbol{z}\sim\mathcal{G}}[A\boldsymbol{z} \in O_b] \pm \delta \pm \delta \pm \delta && \text{(Lemma 5.1 applied to } \boldsymbol{z}\text{)}
\end{aligned}$$

and Theorem 1.1 follows by rescaling δ appropriately. □

The rest of the article is devoted to proving Theorem 5.3.

6 BENTKUS'S MOLLIFIER AND ITS PROPERTIES

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

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

$$\widetilde{\mathbb{1}}_{\theta,\lambda}(t) = \mathop{\mathbf{E}}_{\boldsymbol{g}\sim N(0,1)}\left[\mathbb{1}[t + \lambda\boldsymbol{g} \le \theta]\right].$$

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 $\widetilde{O}_{b,\lambda} : \mathbb{R}^m \to (0,1)$,

$$\widetilde{O}_{b,\lambda}(v) = \mathop{\mathbf{E}}_{\boldsymbol{g}\sim N(0,1)^m}\left[O_b(v + \lambda\boldsymbol{g})\right].$$

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

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

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 $\widetilde{O}_{b,\lambda}(v + \Delta) = \widetilde{O}_{b-v,\lambda}(\Delta)$.*

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 Reference [47]). *For all $\theta \in \mathbb{R}$, $\lambda > 0$, and integer $d \geq 1$,*

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

The following result, from Bentkus [5, Theorem 3(ii)], can be viewed as a multidimensional generalization of Fact 6.4. (Strictly speaking, Reference [5] only considers b's of the form $(\theta, \theta, \ldots, \theta)$, 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 $d \geq 1$,*

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

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

Definition 6.6 (Inner and Outer Approximators for Orthants). We say that $\Upsilon : \mathbb{R}^m \to [0,1]$ is a (Λ, δ)-*inner approximator for* O_b if

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

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

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

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 $\widetilde{O}_{b^{\text{in}},\lambda}, \widetilde{O}_{b^{\text{out}},\lambda}$ are (Λ, δ)-inner and -outer approximators for O_b, respectively, where $\Lambda = \Theta(\lambda\sqrt{\log(m/\delta)})$.*

Proof. Let $b^{\text{in}} = b - \beta \mathbb{1}_m$ where $\beta = \Theta(\lambda\sqrt{\log(m/\delta)}) < \Lambda$ will be specified in more detail later. We show below that $\widetilde{O}_{b^{\text{in}},\lambda}$ is an (Λ, δ)-inner approximator for O_b; an analogous argument in which the $v \in O_b$ and $v \notin O_b$ cases switch roles shows that $\widetilde{O}_{b^{\text{out}},\lambda}$ is a (Λ, δ)-outer approximator for O_b, where $b^{\text{out}} = b + \beta \mathbb{1}_m$.

Fix $v \notin \supset_{-\Lambda} O_b$. There are two possibilities: either $v \in O_b$ or $v \notin O_b$. We first consider the case in which v lies in O_b. Since $v \notin \supset_{-\Lambda} O_b$, we have $v_i \leq b_i - \Lambda$ for all $i \in [m]$. Since $O_b(v) = 1$, we must show that $\widetilde{O}_{b^{\text{in}},\lambda}(v) \geq 1 - \delta$. Recalling Equation (5) and the fact that the function $\widetilde{\mathbb{1}}_{\theta,\lambda} : \mathbb{R} \to (0,1)$ is monotone decreasing for all $\theta \in \mathbb{R}$ and $\lambda > 0$, it suffices to show that $\widetilde{O}_{b^{\text{in}},\lambda}(b - \Lambda \mathbb{1}_m) \geq 1 - \delta$. Again by Equation (5) this holds if and only if

$$\prod_{i=1}^m \widetilde{\mathbb{1}}_{b_i - \beta,\lambda}(b_i - \Lambda) \geq 1 - \delta,$$

which is equivalent to

$$\left(\Pr_{g \sim \mathcal{N}(0,1)}[g \leq (\Lambda - \beta)/\lambda]\right)^m \geq 1 - \delta,$$

which holds if

$$\Pr_{g \sim \mathcal{N}(0,1)}[g \leq (\Lambda - \beta)/\lambda] \geq 1 - \delta/m. \tag{6}$$

By the well-known Gaussian tail bound $\mathbf{Pr}[\boldsymbol{g} \geq t] \leq 1 - \frac{1}{t\sqrt{2\pi}} e^{-t^2/2}$ for $t > 0$ (see, e.g., Reference [13], Section 7.1), we see that to achieve Equation (6) it suffices to have $\Lambda - \beta \geq C\lambda\sqrt{\ln(m/\delta)}$ for an absolute constant $C > 0$, and hence $\Lambda = \Theta(\lambda\sqrt{\log(m/\delta)})$ suffices.

Now, we turn to the case in which $v \notin O_b$, and hence for some $i \in [m]$, we have $v_i > b_i$; without loss of generality, we suppose that $v_1 > b_1$. Since $O_b(v) = 0$ in this case, we must show that $\widetilde{O}_{b^{\text{in}},\lambda}(v) \leq \delta$. By Equation (5) this holds if and only if

$$\prod_{i=1}^{m} \widetilde{\mathbb{1}}_{b_i-\beta,\lambda}(v_i) \leq \delta,$$

which holds if

$$\widetilde{\mathbb{1}}_{b_1-\beta,\lambda}(v_1) \leq \delta,$$

which is equivalent to

$$\Pr_{\boldsymbol{g}\sim\mathcal{N}(0,1)}[v_1 + \lambda\boldsymbol{g} \leq b_1 - \beta] \leq \delta.$$

Recalling that $v_1 > b_1$, it suffices to have

$$\Pr_{\boldsymbol{g}\sim\mathcal{N}(0,1)}[\boldsymbol{g} \leq -\beta/\lambda] \leq \delta,$$

which holds (with room to spare) for our choice of β by the standard Gaussian tail bound. □

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 for an $\mathbb{R}^m$-valued random variable $\tilde{\boldsymbol{v}}$ to fool a translated orthant O_b relative to another $\mathbb{R}^m$-valued random variable $\boldsymbol{v}$, it suffices to (i) have $\tilde{\boldsymbol{v}}$ fool both inner and outer approximators for O_b, and (ii) establish anticoncentration of the original random variable $\boldsymbol{v}$ 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 $\Upsilon^{\text{in}}, \Upsilon^{\text{out}} : \mathbb{R}^m \to [0,1]$ be (Λ, δ)-inner and -outer approximators for O_b. Let $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$ be $\mathbb{R}^m$-valued random variables satisfying:*

$$\left|\mathbf{E}\left[\Upsilon(\boldsymbol{v})\right] - \mathbf{E}\left[\Upsilon(\tilde{\boldsymbol{v}})\right]\right| \leq \gamma \tag{7}$$

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

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

Proof. The proof follows similar lines to the arguments used to prove Lemma 3.3 in Reference [19]. We first note that

$$\begin{aligned}
\mathbf{Pr}\left[\tilde{\boldsymbol{v}} \in O_b\right] &\leq \mathbf{E}\left[\Upsilon^{\text{out}}(\tilde{\boldsymbol{v}})\right] + \delta && \text{(Item 1 of Fact 6.8)}\\
&\leq \left(\mathbf{E}\left[\Upsilon^{\text{out}}(\boldsymbol{v})\right] + \gamma\right) + \delta && \text{(Equation (7) with } \Upsilon = \Upsilon^{\text{out}})\\
&\leq \mathbf{Pr}\left[\boldsymbol{v} \in O_{b+\Lambda\mathbb{1}_m}\right] + \gamma + 2\delta. && \text{(Item 2 of Fact 6.8)}
\end{aligned}$$

Combining this with a symmetric argument for the lower bound, we have:

$$\mathbf{Pr}\left[\boldsymbol{v} \in O_{b-\Lambda\mathbb{1}_m}\right] - \gamma - 2\delta \le \mathbf{Pr}\left[\tilde{\boldsymbol{v}} \in O_b\right] \le \mathbf{Pr}\left[\boldsymbol{v} \in O_{b+\Lambda\mathbb{1}_m}\right] + \gamma + 2\delta. \tag{8}$$

To convert this type of closeness into CDF closeness, we observe that

$$\mathbf{Pr}\left[\boldsymbol{v} \in O_{b+\Lambda\mathbb{1}_m}\right] = \mathbf{Pr}\left[\boldsymbol{v} \in O_b\right] + \mathbf{Pr}\left[\boldsymbol{v} \in \eth_{+\Lambda} O_b\right]$$
$$\mathbf{Pr}\left[\boldsymbol{v} \in O_{b-\Lambda\mathbb{1}_m}\right] = \mathbf{Pr}\left[\boldsymbol{v} \in O_b\right] - \mathbf{Pr}\left[\boldsymbol{v} \in \eth_{-\Lambda} O_b\right].$$

Plugging these identities into Equation (8), we conclude that

$$\begin{aligned}\mathbf{Pr}\left[\tilde{\boldsymbol{v}} \in O_b\right] &= \mathbf{Pr}\left[\boldsymbol{v} \in O_b\right] \pm \left(\gamma + 2\delta + \mathbf{Pr}\left[\boldsymbol{v} \in \eth_{+\Lambda} O_b\right] + \mathbf{Pr}\left[\boldsymbol{v} \in \eth_{-\Lambda} O_b\right]\right) \\ &= \mathbf{Pr}\left[\boldsymbol{v} \in O_b\right] \pm \left(\gamma + 2\delta + \mathbf{Pr}\left[\boldsymbol{v} \in \eth_{\pm\Lambda} O_b\right]\right),\end{aligned}$$

thus completing the proof of Lemma 6.9. □

6.1.1 Applying Lemma 6.9 in the context of Theorem 5.3, and the organization of the rest of this article. Applying Lemma 6.9 with $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$ being $A\boldsymbol{u}$ and Az, respectively, the task of bounding

$$\left| \mathop{\mathbf{Pr}}_{\boldsymbol{u}\sim\{-1,1\}^n}\left[A\boldsymbol{u} \in O_b\right] - \mathop{\mathbf{Pr}}_{z\sim\mathscr{G}_{\mathrm{MZ}}}\left[Az \in O_b\right] \right|$$

reduces to the following two-step program:

(1) Establishing anticoncentration within orthant boundaries: bounding $\mathbf{Pr}[A\boldsymbol{u} \in \eth_{\pm\Lambda} O_b]$; and,
(2) Fooling Bentkus's mollifier: bounding $|\mathbf{E}[\widetilde{O}(A\boldsymbol{u})] - \mathbf{E}[\widetilde{O}(Az)]|$ for $\widetilde{O} \in \{\widetilde{O}_{b^{\mathrm{out}},\lambda}, \widetilde{O}_{b^{\mathrm{in}},\lambda}\}$, 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 5.3.

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.*[2] *Then for all $b \in \mathbb{R}^m$ and $\Lambda \ge \tau$, we have*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{u}\sim\{-1,1\}^n}[A\boldsymbol{u} \in \eth_{\pm\Lambda} O_b] = O\left(\Lambda\sqrt{\log m}\right).$$

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. Theorem 2.2 will in fact be obtained as a special case of a more general result about intersections of m arbitrary *unate* functions (namely, Lemma 7.13).

Definition 7.2 (Unateness). A function $F : \{-1,1\}^n \to \{0,1\}$ is *unate in direction* $\sigma \in \{-1,1\}^n$ if the function $G(x_1,\ldots,x_n) = F(\sigma_1 x_1,\ldots,\sigma_n x_n)$ is a monotone Boolean function, meaning that $G(x) \le G(y)$ whenever $x_j \le y_j$ for all $j \in [n]$. We refer to σ as the *orientation* of F.

Our analysis, dealing as it does with intersections of unate functions, is somewhat reminiscent of that of Reference [26], and indeed we will establish the main result of Reference [26]—an upper bound of $O(\sqrt{n\log m})$ on the average sensitivity of any intersection of m unate functions—in the course of our analysis.

[2]Equivalently, A is (n, τ)-standardized.

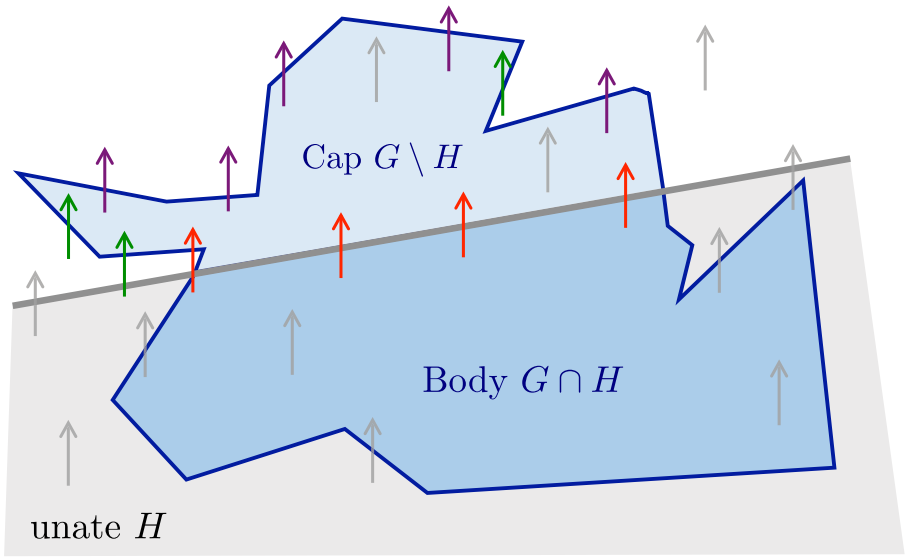

Fig. 1. Illustration of a cap and body.

7.1 Caps and Their Boundary Edges

Let G and H be subsets of $\{-1,1\}^n$. We typically think of G as a *G*eneral/arbitrary set and H as a *H*alfspace, though formally H will only need to be unate. Throughout this section, we write $\sigma \in \{-1,1\}^n$ to denote the orientation of H.

We call the set $G \setminus H$ the *cap*, the set $G \cap H$ the *body*, and the complement of G the *exterior*. Please refer to Figure 1, where G is the union of the two regions with blue shading and H is the gray-shaded region (depicted as a halfspace in the figure). The upward arrows in the diagram illustrate some edges of the hypercube. We have oriented these edges according to σ: For an edge $\{x, y\}$ in the jth direction in which $x_j = -1$ and $y_j = 1$, the tail of the corresponding arrow represents x if $\sigma_j = -1$ and y if $\sigma_j = 1$. Note in particular that the edges are oriented "away" from H (i.e., so that H is antimonotone with respect to the edge orientations).

We will be concerned with the *boundary* edges for the cap $G \setminus H$; these are edges that have one endpoint inside $G \setminus H$ and one endpoint outside it.

Definition 7.3 (Edge Boundary). For a general set $F \subseteq \{-1,1\}^n$, let $\mathcal{E}(F)$ denote the fraction of all $n2^{n-1}$ hypercube edges that are boundary edges for F.

We distinguish the three possible types of boundary edges of the cap $G \setminus H$:

- **Body→Cap (BC) edges:** the red edges in the diagram. Formally, these are edges where the tail is in the body $G \cap H$ and the head is in the cap $G \setminus H$.
- **Exterior→Cap (EC) edges:** the green edges in the diagram. Formally, these are edges where the tail is not in G, and the head is in the cap $G \setminus H$.
- **Cap→Exterior (CE) edges:** the purple edges in the diagram. Formally, these are edges where the tail is in the cap $G \setminus H$ and the head is not in G.

Remark 7.4. Note that **there are no Cap→Body (CB) edges**. Formally, these would be the last possibility for $G \setminus H$ boundary edges, namely, ones with tail in the cap $G \setminus H$ and head in the body $G \cap H$. But these cannot exist due to the antimonotonicity of H vis-à-vis the edges; if the tail is already not in H, then the head cannot be in H.

Given a cap $C = G \setminus H$, we write $\mathrm{BC}(G,H)$, $\mathrm{EC}(G,H)$, $\mathrm{CE}(G,H)$ for the fraction of hypercube edges of each of the three above types. Therefore, $\mathcal{E}(C) = \mathrm{BC}(G,H) + \mathrm{EC}(G,H) + \mathrm{CE}(G,H)$.

We will also be interested in the *directed* edge boundary of caps:

Definition 7.5 (Directed Edge Boundary). For a cap $G \setminus H$, define

$$\vec{\mathcal{E}}(G,H) = \mathrm{BC}(G,H) + \mathrm{EC}(G,H) - \mathrm{CE}(G,H), \tag{9}$$

the fraction of inward boundary edges minus the fraction of outward boundary edges.

It will be very useful for us to have an upper bound on $\mathcal{E}(G \cap H) - \mathcal{E}(G)$, the change in $\mathcal{E}(G)$ when we intersect G with H (note that this quantity can be either positive or negative). The following fact is immediate from the definitions:

Fact 7.6 (Change in Boundary Size). *If $G \setminus H$ is a cap, then*

$$\mathcal{E}(G \cap H) - \mathcal{E}(G) = \mathrm{BC}(G,H) - \mathrm{EC}(G,H) - \mathrm{CE}(G,H). \tag{10}$$

Comparing Equations (10) and (9), we plainly have:

Fact 7.7. $\mathcal{E}(G \cap H) - \mathcal{E}(G) \le \vec{\mathcal{E}}(G,H)$.

To get a quantitative bound, we have the following lemma:

Lemma 7.8. *For any cap $C = G \setminus H$,*

$$\vec{\mathcal{E}}(G,H) \le \frac{U(\mathrm{vol}(C))}{\sqrt{n}},$$

where $\mathrm{vol}(C) = |C|/2^n$ and U denotes the function $U(p) = 2p\sqrt{2\ln(1/p)}$.

Proof. This is a basic fact in analysis of Boolean functions. Identifying C with its indicator function $C : \{-1,1\}^n \to \{0,1\}$, we have $\mathrm{vol}(C) = \mathbf{E}[C(\boldsymbol{u})]$ and

$$\vec{\mathcal{E}}(G,H) = 2 \mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \{-1,1\}^n \\ j \sim [n]}} [C(\boldsymbol{u})\sigma_j \boldsymbol{u}_j] = \frac{2}{n} \sum_{j=1}^{n} \sigma_j \widehat{C}(\{j\}),$$

where $\widehat{C}(\{j\})$ denotes the degree-1 Fourier coefficient of C corresponding to coordinate j. It is well known and elementary that for $F : \{-1,1\}^n \to \{0,1\}$ with $\mathbf{E}[F] = p$, one has $\sum_{j=1}^{n} |\widehat{F}(\{j\})| \le O(p\sqrt{\ln(1/p)})\sqrt{n}$; see, e.g., Kane's paper [26, Lemma 6] for the short proof. For the sake of an asymptotically tight constant, we can use the Cauchy–Schwarz inequality and the Fourier "Level-1 Inequality" [7, 21, 60] to get

$$\sum_{j=1}^{n} \sigma_j \widehat{C}(\{j\}) \le \sqrt{n} \cdot \sqrt{\sum_{j=1}^{n} \widehat{C}(\{j\})^2} \le \sqrt{n} \cdot p\sqrt{2\ln(1/p)}. \qquad \square$$

7.1.1 Reproving the Main Result of Reference [26]. We can now reprove the main result of Reference [26] (which we will use later):

Theorem 7.9 ([26]). *Let F be the intersection of $m \ge 2$ unate functions over $\{-1,1\}^n$. Then*

$$\mathcal{E}(F) \le \frac{2\sqrt{2\ln m}}{\sqrt{n}}. \tag{11}$$

(Equivalently, an intersection of $m \ge 2$ unate functions has average sensitivity at most $2\sqrt{2\ln m}\sqrt{n}$.)

Proof. Let $H_1, \dots, H_m$ be unate functions and define associated caps

$$C_i = (H_1 \cap \cdots \cap H_{i-1}) \setminus H_i, \tag{12}$$

with $C_1 = \{-1,1\}^n \setminus H_1$ (i.e., $H_0 = \{-1,1\}^n$). Letting $F = H_1 \cap \cdots \cap H_m$, we have that the complement $F^c = \{-1,1\}^n \setminus F$ of F can be expressed as a disjoint union of caps:

$$F^c = C_1 \sqcup \cdots \sqcup C_m. \tag{13}$$

For intuition, we may think of the intersection of m unate sets F as being formed in m stages, starting with $\{-1,1\}^n$ and successively intersecting with each H_i; given this interpretation, C_i is the portion of $\{-1,1\}^n$ that is removed in the ith stage. With this notation in hand, we have that

$$\begin{aligned}
\mathcal{E}(F) &= \sum_{i=1}^m \mathcal{E}((H_1 \cap \cdots \cap H_{i-1}) \cap H_i) - \mathcal{E}(H_1 \cap \cdots \cap H_{i-1}) \\
&\le \sum_{i=1}^m \vec{\mathcal{E}}(H_1 \cap \cdots \cap H_{i-1}, H_i) && \text{(Fact 7.7 with } G = H_1 \cap \cdots \cap H_{i-1} \text{ and } H = H_i) \\
&\le \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^m U(\mathrm{vol}(C_i)). && \text{(Lemma 7.8)}
\end{aligned}$$

Finally,

$$\sum_{i=1}^m U(\mathrm{vol}(C_i)) \le m \cdot U\left(\frac{\sum_{i=1}^m \mathrm{vol}(C_i)}{m}\right) = m \cdot U\left(\frac{\mathrm{vol}(F^c)}{m}\right) \le m \cdot U(\tfrac{1}{m}) = 2\sqrt{2 \ln m},$$

where we used concavity of U, then Equation (13), then the fact that U is increasing on $[0, 1/2]$. This completes the proof of Theorem 7.9. □

7.2 A Littlewood–Offord Theorem for Polytopes (Theorem 2.2)

In this section we prove Theorem 2.2:

Theorem 7.10. *For all $m \ge 2$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ with $|A_{ij}| \ge 1$ for all $i \in [m]$ and $j \in [n]$,*

$$\Pr_{\boldsymbol{u} \sim \{-1,1\}^n}[A\boldsymbol{u} \in \partial_{-2} O_b] \le \frac{5\sqrt{2 \ln m}}{\sqrt{n}}.$$

We note in passing that the anticoncentration bound given by Theorem 2.2 is best possible up to constant factors. Indeed, our matching lower bound applies even to the stricter event of falling *on the surface* of O_b:

Claim 7.11 (Optimality of Theorem 2.2). *For $2 \le m \le 2^n$, there is a matrix $A \in \{-1,1\}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ such that*

$$\Pr_{\boldsymbol{u} \sim \{-1,1\}^n}[A\boldsymbol{u} \in \partial O_b] = \Omega\left(\frac{\sqrt{\ln m}}{\sqrt{n}}\right).$$

We prove Claim 7.11 in Appendix B.

7.2.1 Proof of Theorem 2.2. As mentioned at the beginning of this section, we will obtain Theorem 2.2 as a corollary of a more general result about intersections of unate functions. Let $H_1, \dots, H_m \subseteq \{-1,1\}^n$ be unate sets, $m \ge 2$, and further suppose that we have additional unate sets $\overline{H}_1, \dots, \overline{H}_m$ such that $H_i \subseteq \overline{H}_i$ for all i. (For intuition, it may be helpful to think of H_i as the

"interior" of $\overline{H}_i$; see the proof of Theorem 2.2 using Lemma 7.13 just below for a typical example of sets H_i and $\overline{H}_i$.) We define the following subsets of $\{-1,1\}^n$:

$$\begin{aligned} F &= \overline{H}_1 \cap \cdots \cap \overline{H}_m \\ F^\circ &= H_1 \cap \cdots \cap H_m && \text{(interior of } F\text{)} \\ \partial F &= F \setminus F^\circ && \text{(boundary of } F\text{)} \\ F^c &= \{-1,1\}^n \setminus F && \text{(exterior of } F\text{)} \\ \partial H_i &= \overline{H}_i \setminus H_i \text{ (for each } i \in [m]\text{).} && \text{(boundary of } \overline{H}_i\text{)} \end{aligned}$$

Definition 7.12 (Thin Sets). We say that ∂H_i is *thin* if it does not contain any induced edges of the hypercube.

Lemma 7.13. *If ∂H_i is thin for each $i \in [m]$, then* $\mathrm{vol}(\partial F) \le \frac{5\sqrt{2\ln m}}{\sqrt{n}}$.

Proof of Theorem 2.2 Assuming Lemma 7.13. Fix any $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ such that $|A_{ij}| \ge 1$ for all $i \in [m]$ and $j \in [n]$, and let

$$\overline{H}_i = \left\{x \in \{-1,1\}^n : A^i \cdot x \le b_i\right\}, \qquad H_i = \left\{x \in \{-1,1\}^n : A^i \cdot x \le b_i - 2\right\},$$

so

$$\begin{aligned} \partial H_i &= \left\{x \in \{-1,1\}^n : b_i - 2 < A^i \cdot x \le b_i\right\} \quad \text{and} \\ \partial F &= \left\{x \in \{-1,1\}^n : Ax \le b \;\&\; A^i \cdot x > b_i - 2 \text{ for some } i \in [m]\right\} \\ &= \left\{x \in \{-1,1\}^n : Ax \in \circlearrowleft_{-2} O_b\right\}. \end{aligned}$$

Since $|A_{ij}| \ge 1$ for all i, j, it follows that each ∂H_i is thin, and hence Lemma 7.13 directly gives Theorem 2.2. □

The rest of this section will be devoted to the proof of Lemma 7.13. Recalling that F° is called the *interior* of F and ∂F is called the *boundary* of F, we say that an edge in the hypercube is *boundary-to-interior* if it has one endpoint in ∂F and the other endpoint in F°, and we write ν_{BI} for the fraction of all edges that are of this type. We similarly define *boundary-to-exterior* edges and ν_{BE}, with F^c. Note that every boundary-to-interior edge is a boundary edge for $F^\circ = H_1 \cap \cdots \cap H_m$, which is an intersection of m unate sets. By applying Theorem 7.9 to F°, we get that

$$\nu_{BI} \le \frac{2\sqrt{2\ln m}}{\sqrt{n}}. \tag{14}$$

Similarly, every boundary-to-exterior edge is a boundary edge for $F = \overline{H}_1 \cap \cdots \cap \overline{H}_m$; applying Theorem 7.9 to this intersection yields

$$\nu_{BE} \le \frac{2\sqrt{2\ln m}}{\sqrt{n}}. \tag{15}$$

Next, we bound the fraction of edges that have both endpoints in ∂F and go between "two different parts of ∂F. More precisely, for $x \in \partial F$, define $i^\star(x)$ to be the least i for which $x \in \partial H_i$ (equivalently, the least i for which $x \notin H_i$). We say that an edge $\{x, y\}$ is *boundary-to-boundary′* if $x, y \in \partial F$ but $i^\star(x) \ne i^\star(y)$; we write $\nu_{BB'}$ for the fraction of such edges.

Observation 7.14. *If every ∂H_i is thin, then every edge with both endpoints in ∂F is boundary-to-boundary′. In this case, $\nu_{BI} + \nu_{BE} + \nu_{BB'}$ is exactly the fraction of edges in the cube that touch ∂F, which in turn is an upper bound on* $\mathrm{vol}(\partial F)$.

Thus, Lemma 7.13 follows from Equations (14) and (15) and the following claim:

Claim 7.15 (Boundary-to-boundary′ Edges). $\nu_{BB'} \le \frac{\sqrt{2\ln m}}{\sqrt{n}}$.

Proof. We define the caps $C_1, \dots, C_m$ with respect to the H_i's as in Equation (12) in the proof of Theorem 7.9. Subtracting Equation (10) from Equation (9) for each C_i and summing over $i \in [m]$,

$$\begin{aligned} 2\sum_{i=1}^{m} \mathrm{EC}(H_1 \cap \cdots \cap H_{i-1}, H_i) &= \sum_{i=1}^{m} \vec{\mathcal{E}}(H_1 \cap \cdots \cap H_{i-1}, H_i) \\ &\quad - \left(\sum_{i=1}^{m} \mathcal{E}((H_1 \cap \cdots \cap H_{i-1}) \cap H_i) - \mathcal{E}(H_1 \cap \cdots \cap H_{i-1}) \right) \\ &= \sum_{i=1}^{m} \vec{\mathcal{E}}(H_1 \cap \cdots \cap H_{i-1}, H_i) - \mathcal{E}(H_1 \cap \cdots \cap H_m) \\ &= \sum_{i=1}^{m} \vec{\mathcal{E}}(H_1 \cap \cdots \cap H_{i-1}, H_i) - \mathcal{E}(F^\circ). \end{aligned}$$

Since $\mathcal{E}(F^\circ) \ge 0$, it follows that

$$\sum_{i=1}^{m} \mathrm{EC}(H_1 \cap \cdots \cap H_{i-1}, H_i) \le \frac{1}{2} \sum_{i=1}^{m} \vec{\mathcal{E}}(H_1 \cap \cdots \cap H_{i-1}, H_i) \le \frac{\sqrt{2\ln m}}{\sqrt{n}}, \tag{16}$$

where the derivation of the second inequality is exactly as in the proof of Theorem 7.9. By Equation (16), it suffices to show

$$\nu_{BB'} \le \sum_{i=1}^{m} \mathrm{EC}(H_1 \cap \cdots \cap H_{i-1}, H_i). \tag{17}$$

Let $\{x, y\}$ be a boundary-to-boundary′ edge and assume without loss of generality that $i^\star(x) < i^\star(y)$. We now show that edge $\{x, y\}$ contributes to $\mathrm{EC}(H_1 \cap \cdots \cap H_{i^\star(y)-1}, H_{i^\star(y)})$. For brevity, write $G = H_1 \cap \cdots \cap H_{i^*(y)-1}$, $H = H_{i^*(y)}$, and $C = G \setminus H = C_{i^\star(y)}$. Since $x \in \partial H_{i^\star(x)} = \overline{H}_{i^\star(x)} \setminus H_{i^\star(x)}$ (in particular, $x \notin H_{i^\star(x)}$) and $i^\star(x) < i^\star(y)$, we have that $x \notin G$. However, $y \in G \setminus H = C$ by definition of $i^\star(y)$. Since $x \notin G$ and $y \in G \setminus H$, we conclude that indeed $\{x, y\} \in \mathrm{EC}(G, H)$, as claimed. □

This completes the proof of Lemma 7.13 and hence Theorem 2.2.

7.3 A Robust Generalization of the Littlewood–Offord Theorem for Polytopes

In the previous section, we proved Theorem 2.2, which establishes anticoncentration of $A\boldsymbol{u}$ under the assumption that all its entries have magnitude at least 1. The goal of this section is to prove the following robust generalization of Theorem 2.2:

Theorem 7.16. *Let $A \in \mathbb{R}^{m \times n}$ have the property that in every row, at least an α fraction of the entries have magnitude at least λ. Then for any $b \in \mathbb{R}^m$,*

$$\mathbf{Pr}[A\boldsymbol{u} \in \partial_{-2\lambda} O_b] \le \frac{5\sqrt{2\ln m}}{\alpha\sqrt{n}}.$$

Recall that Theorem 2.2 followed as an easy consequence of the fact that $\mathrm{vol}(\partial F) \le \frac{5\sqrt{2\log m}}{\sqrt{n}}$ when all ∂H_i's are "thin" (Lemma 7.13). We slightly generalize this notion here.

Definition 7.17 (Semi-thin). For $\alpha \in [0,1]$, say that ∂H_i is *α-semi-thin* if the following holds: For each $x \in \partial H_i$, at least an α fraction of its hypercube neighbors are outside ∂H_i. (Note that "1-semi-thin" is equivalent to "thin.")

Example 7.18. Suppose $H = \{x \in \{-1,1\}^n : a \cdot x \le b_1\}$ and $\overline{H} = \{x \in \{-1,1\}^n : a \cdot x \le b_2\}$ where $b_1 \le b_2$, so $\partial H = \{x \in \{-1,1\}^n : b_1 < a \cdot x \le b_2\}$. If $|a_j| \ge (b_2 - b_1)/2$ for at least an α fraction of the coordinates $j \in [n]$, then ∂H is α-semi-thin.

Theorem 7.16 follows as a direct consequence of the following lemma (by the same reasoning that derives Theorem 2.2 as a corollary of Lemma 7.13):

Lemma 7.19 (Robust Version of Lemma 7.13). *In the setup of Section 7.2.1, suppose each ∂H_i is α-semi-thin. Then*

$$\mathrm{vol}(\partial F) \le \frac{5\sqrt{2\ln m}}{\alpha\sqrt{n}}.$$

Proof. Our proof of Lemma 7.13 (a combination of Equation (14), Equation (15), and Claim 7.15) shows that

$$\nu_{BI} + \nu_{BE} + \nu_{BB'} \le \frac{5\sqrt{2\ln m}}{\sqrt{n}}. \tag{18}$$

However, in our current setting the left-hand side of the above is *not* a bound on $\mathrm{vol}(\partial F)$; Observation 7.14 no longer holds and we now may have edges (x,y) where $i^\star(x) = i^\star(y)$. Given an $x \in \partial F$ and y a Hamming neighbor of x, we say that y is *x-bad* if $y \in \partial F$ and $i^\star(y) = i^\star(x)$; otherwise, we say that y is *x-good*. With this terminology, we can rewrite Equation (18) as

$$\mathbf{Pr}\Big[\boldsymbol{u} \in \partial F \;\&\; \boldsymbol{u}^{\oplus \boldsymbol{j}} \text{ is } \boldsymbol{u}\text{-good}\Big] \le \frac{5\sqrt{2\ln m}}{\sqrt{n}}, \tag{19}$$

where $\boldsymbol{u} \sim \{-1,1\}^n$ and $\boldsymbol{j} \sim [n]$ are uniformly random, and $\boldsymbol{u}^{\oplus j}$ denotes $\boldsymbol{u}$ with its $\boldsymbol{j}$th coordinate flipped. By the α-semi-thin property, for any $x \in \partial F$, the fraction of $\boldsymbol{j}$'s such that $x^{\oplus \boldsymbol{j}}$ is x-good is at least α. Therefore,

$$\mathbf{Pr}\Big[\boldsymbol{u} \in \partial F \;\&\; \boldsymbol{u}^{\oplus \boldsymbol{j}} \text{ is } \boldsymbol{u}\text{-good}\Big] \ge \mathbf{Pr}[\boldsymbol{u} \in \partial F] \cdot \alpha, \tag{20}$$

and the lemma follows by combining Equations (19) and (20). □

7.4 Proof of Theorem 7.1

In this section, we prove Theorem 7.1 using Lemma 7.19 established in the previous section. In more detail, we use a bound on the anticoncentration of $A\boldsymbol{u}$ under the assumption that at least an α fraction of entries of each row of A have magnitude at least τ (given by Lemma 7.19) to establish a bound on the anticoncentration of $A\boldsymbol{u}$ under the assumption that each of A's rows has a τ-regular subvector of 2-norm 1 (Theorem 7.1).

The following result regarding τ-regular linear forms is fairly standard:

Proposition 7.20. *Let $w \in \mathbb{R}^n$ be a τ-regular vector with $\|w\|_2 = 1$. Let $\boldsymbol{\pi} : [n] \to [B]$ be a random hash function that independently assigns each coordinate in $[n]$ to a uniformly random bucket in $[B]$. For $b \in [B]$, write $\boldsymbol{\sigma}_b^2 = \sum_{j \in \boldsymbol{\pi}^{-1}(b)} w_j^2$, and say that bucket b is* good *if $\boldsymbol{\sigma}_b^2 > \frac{1}{2B}$. Assume $B \le 1/\tau^2$. Then*

$$\mathbf{Pr}\Big[\textit{at most } \frac{B}{16} \textit{ buckets } b \in [B] \textit{ are good}\Big] \le \exp\Big(-\frac{B}{64}\Big).$$

Proof. Let $X_b = \mathbb{1}[\sigma_b^2 > \frac{1}{2B}]$ be the indicator that the bth bucket is good. Since $\mathbf{E}[\sigma_b^2] = \frac{1}{B}$ and

$$\mathbf{E}[\sigma_b^4] = \mathbf{E}\left[\left(\sum_{j=1}^n w_j^2 \mathbb{1}[\pi(j) = b]\right)^2\right] = \frac{1}{B}\sum_{j=1}^n w_j^4 + \frac{1}{B^2}\sum_{j \neq j'} w_j^2 w_{j'}^2 \leq \frac{\tau^2}{B} + \frac{1}{B^2} \leq \frac{2}{B^2},$$

the Paley–Zygmund inequality implies that $\mathbf{E}[X_b] = \mathbf{Pr}[\sigma_b^2 > \frac{1}{2}\mathbf{E}[\sigma_b^2]] \geq \frac{1}{8}$.

The joint random variables $\sigma_1^2, \ldots, \sigma_B^2$ are of "balls in bins" type (where the jth "ball" has "mass" w_j^2) and are therefore negatively associated (see, e.g., Reference [11, Example 3.1]; the fact that the balls have different "masses" does not change the argument). Since $\mathbb{1}_{(\frac{1}{2B}, \infty)}$ is a nondecreasing function, it follows that the random variables $X_1, \ldots, X_B$ are also negatively associated. Thus, we may apply the Chernoff bound to $\sum_{k=1}^B X_k$, which has mean at least $\frac{B}{8}$. The result follows. □

Recall the following fact, which can also be easily proven using Paley–Zygmund (see, e.g., Proposition 3.7 of the full version of Reference [18]):

Fact 7.21. *For all $w \in \mathbb{R}^n$ and $\boldsymbol{u} \sim \{-1, 1\}^n$, we have $\mathbf{Pr}[|w \cdot \boldsymbol{u}| \geq \frac{1}{2}\|w\|_2] \geq \frac{1}{16}$.*

We combine these as follows:

Proposition 7.22. *Let $w \in \mathbb{R}^n$ and assume that some subvector w' of w is τ-regular with $\|w'\|_2 = 1$. Let $\boldsymbol{\pi} : [n] \to [B]$ be as in Proposition 7.20, where $B \leq 1/\tau^2$. Let $\boldsymbol{u} \sim \{-1, 1\}^n$, and define $\overline{\boldsymbol{w}} \in \mathbb{R}^B$ by $\overline{\boldsymbol{w}}_b = \sum_{j \in \boldsymbol{\pi}^{-1}(b)} w_j \boldsymbol{u}_j$. Call a bucket $b \in [B]$ big if $|\overline{\boldsymbol{w}}_b| > \frac{1}{2\sqrt{2B}}$. Then*

$$\mathbf{Pr}\left[\textit{fewer than } \frac{B}{512} \textit{ buckets are big}\right] \leq \exp\left(-\frac{B}{2048}\right).$$

Proof. First apply Proposition 7.20 to w' and observe that the presence of additional coordinates from w cannot harm "goodness." Then apply Fact 7.21 to the good buckets. Each becomes "big" independently with probability at least $\frac{1}{16}$, and the proof follows from another Chernoff bound. □

We take $B = \lfloor 1/\tau^2 \rfloor$ in the above. This yields the following:

Corollary 7.23. *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. Fix $B = \lfloor 1/\tau^2 \rfloor$ and let $\overline{\boldsymbol{A}} \in \mathbb{R}^{m \times B}$ be the matrix obtained from A by randomly partitioning its columns into B buckets and adding them up with uniformly random ± 1 signs within each bucket. Say that a row of $\overline{\boldsymbol{A}}$ is* spread *if at least a $\frac{1}{512}$-fraction of its entries exceed $\frac{\tau}{2\sqrt{2}}$. Then except with probability at most $m \cdot \exp(-\Omega(1/\tau^2))$, all of $\overline{\boldsymbol{A}}$'s rows are spread.*

7.4.1 Proof of Theorem 7.1. We can now prove Theorem 7.1, which we restate here for convenience:

Theorem 7.24. *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*

$$\Pr_{\boldsymbol{u} \sim \{-1,1\}^n}[A\boldsymbol{u} \in \mathcal{O}_{\pm\Lambda} O_b] = O\left(\Lambda\sqrt{\log m}\right).$$

Proof. By union-bounding over $2\lceil \Lambda/\tau \rceil$ choices of b, it suffices to prove the following: Whenever $A \in \mathbb{R}^{m \times n}$ has a τ-regular subvector of 2-norm 1 in each row, it holds that $\mathbf{Pr}[A\boldsymbol{u} \in \mathcal{O}_{-\tau} O_b] \leq O(\tau\sqrt{\log m})$. Note that the distribution of $A\boldsymbol{u}$ is the same as that of $\overline{\boldsymbol{A}}\boldsymbol{u}'$, where $\overline{\boldsymbol{A}}$ is as in Corollary 7.23, and $\boldsymbol{u}' \sim \{-1, 1\}^B$ is uniform. Thus, applying Corollary 7.23 and then Theorem 7.16 (with $\alpha = \frac{1}{512}$ and $\lambda = \frac{\tau}{2} \geq \frac{\tau}{2\sqrt{2}}$), we conclude that

$$\mathbf{Pr}[A\boldsymbol{u} \in \mathcal{O}_{-\tau} O_b] = O\left(\tau\sqrt{\log m}\right) + m \cdot \exp\left(-\Omega(1/\tau^2)\right).$$

By our choice of τ as set in Section 4.1, we get the desired overall bound of $O(\tau\sqrt{\log m})$ and the proof is complete. □

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 ($\mathscr{G}$ Fools Bentkus's Mollifier). *Let $\mathscr{G}$ 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 $b \in \mathbb{R}^m$,*

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\{-1,1\}^n} \left[\widetilde{O}_{b,\lambda}(A\boldsymbol{u})\right] - \mathop{\mathbf{E}}_{\boldsymbol{z}\sim\mathscr{G}_{\mathrm{MZ}}} \left[\widetilde{O}_{b,\lambda}(A\boldsymbol{z})\right] \right| = O(\delta).$$

At a very high level, in line with the usual Lindeberg approach, Theorem 8.1 is proved by hybridizing between $\boldsymbol{u}$ and $\boldsymbol{z}$ via a sequence of intermediate distributions. In our setting there are $L+1$ such distributions, the first of which is $\boldsymbol{u}$ and the last of which is $\boldsymbol{z}$, and the ℓth of which may be viewed as "filling in buckets $\ell,\dots,L$ according to $\boldsymbol{u}$ and filling in buckets $1,\dots,\ell-1$ according to $\boldsymbol{z}$," 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 $[n] = \textsc{Head}_i \sqcup \textsc{Tail}_i$ such that $|\textsc{Head}_i| \le k$ and $(A_i)_{\textsc{Tail}_i}$ is τ-regular with 2-norm $\|(A_i)_{\textsc{Tail}_i}\|_2$ equal to 1. Therefore, we may write A as $H+T$ where

$$H_{ij} = A_{ij}\cdot \mathbb{1}[\,j \in \textsc{Head}_i\,] \quad \text{and} \quad T_{ij} = A_{ij}\cdot \mathbb{1}[\,j \in \textsc{Tail}_i\,]$$

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 $\boldsymbol{u},\boldsymbol{y}$ be random variables over $\{-1,1\}^B$, where $\boldsymbol{u}$ is uniform and $\boldsymbol{y}$ δ_{CNF}-fools the class of width-w CNFs. For all $b \in \mathbb{R}^m$, $\lambda > 0$, and all integers $d \ge 2$*

$$\begin{aligned} &\left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(H^B\boldsymbol{u}+T^B\boldsymbol{u})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(H^B\boldsymbol{y}+T^B\boldsymbol{y})\right]\right| \\ &= \delta_{\mathrm{CNF}}\cdot m^{d-1}\cdot O_d\!\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} + O_d\!\left(\frac{\sqrt{\log m}}{\lambda}\right)^{d}\left(\mathbf{E}\left[\|T^B\boldsymbol{u}\|_\infty^d\right] + \mathbf{E}\left[\|T^B\boldsymbol{y}\|_\infty^d\right]\right). \end{aligned} \tag{21}$$

As we will see later, Equation (21) 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 (21) 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 $|\alpha| = c$,*

$$\left| \mathbf{E}\left[\partial_\alpha \widetilde{O}_{b,\lambda}(H^B \boldsymbol{u}) \cdot (T^B \boldsymbol{u})^\alpha\right] - \mathbf{E}\left[\partial_\alpha \widetilde{O}_{b,\lambda}(H^B \boldsymbol{y}) \cdot (T^B \boldsymbol{y})^\alpha\right] \right| = \delta_{\mathrm{CNF}} \cdot O_c\left(\frac{\sqrt{n}}{\lambda}\right)^c. \tag{22}$$

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 "$|\mathbf{E}[\Upsilon(\boldsymbol{v} + \Delta)] - \mathbf{E}[\Upsilon(\boldsymbol{v}' + \Delta')]|$" 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 $\widetilde{O}_b$ by using pseudorandom generators for small-width CNF formulas.

Before proving Claim 8.4, we observe that Lemma 8.3 follows as a consequence:

Proof of Lemma 8.3 Assuming Claim 8.4. By the multidimensional Taylor expansion (Fact 3.2) applied twice to $\widetilde{O}_{b,\lambda}$, we have

$$\begin{aligned}
(21) &\leq \left| \sum_{0 \leq |\alpha| \leq d-1} \frac{1}{\alpha!} \mathbf{E}\left[\partial_\alpha \widetilde{O}_{b,\lambda}(H^B \boldsymbol{u}) \cdot (T^B \boldsymbol{u})^\alpha\right] - \frac{1}{\alpha!} \mathbf{E}\left[\partial_\alpha \widetilde{O}_{b,\lambda}(H^B \boldsymbol{y}) \cdot (T^B \boldsymbol{y})^\alpha\right] \right| \\
&\quad + \mathbf{E}\left[\left|\mathrm{err}(H^B \boldsymbol{u}, T^B \boldsymbol{u})\right|\right] + \mathbf{E}\left[\left|\mathrm{err}(H^B \boldsymbol{y}, T^B \boldsymbol{y})\right|\right] \\
&\leq \sum_{0 \leq |\alpha| \leq d-1} \left| \mathbf{E}\left[\partial_\alpha \widetilde{O}_{b,\lambda}(H^B \boldsymbol{u}) \cdot (T^B \boldsymbol{u})^\alpha\right] - \mathbf{E}\left[\partial_\alpha \widetilde{O}_{b,\lambda}(H^B \boldsymbol{y}) \cdot (T^B \boldsymbol{y})^\alpha\right] \right| \qquad (23) \\
&\quad + \sup_{v \in \mathbb{R}^m} \left\{ \sum_{|\alpha| = d} |\partial_\alpha \widetilde{O}_{b,\lambda}(v)| \right\} \cdot \left(\mathbf{E}\left[\|T^B \boldsymbol{u}\|_\infty^d\right] + \mathbf{E}\left[\|T^B \boldsymbol{y}\|_\infty^d\right] \right).
\end{aligned}$$

By Claim 8.4, each of the $O(m^{d-1})$ summands of Equation (23) is at most $\delta_{\mathrm{CNF}} \cdot O(\sqrt{n}/\lambda)^{d-1}$. This along with the bound on $\widetilde{O}_{b,\lambda}$'s derivatives given by Theorem 6.5,

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

yields Lemma 8.3. □

8.1.1 *Proof of Claim 8.4.*

Definition 8.6. We say that a function $\xi : \{-1,1\}^B \to \mathbb{R}$ is *Boolean* if its range is contained in $\{0,1\}$. For $\xi_1, \ldots, \xi_m : \{-1,1\}^B \to \mathbb{R}$, 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 $\xi : \{-1,1\}^B \to [0,1]$ is a weight-1 combination of Boolean functions.*
(2) A function $\xi : \{-1,1\}^B \to [-W, W]$ 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_1W_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_1W_2) combination of Boolean product functions.

We are now ready to prove Claim 8.4.

Proof of Claim 8.4. We define the function $G_\alpha : \{-1,1\}^B \to \mathbb{R}$,

$$\begin{aligned} G_\alpha(x) &:= \partial_\alpha \widetilde{O}_{b,\lambda}(H^B x) \cdot (T^B x)^\alpha \\ &= \left(\prod_{i \notin S} \widetilde{\mathbb{1}}_{b_i,\lambda}(H_i^B x) \prod_{i \in S} \widetilde{\mathbb{1}}^{(\alpha_i)}_{b_i,\lambda}(H_i^B x) \right) \cdot \prod_{i \in S} (T_i^B x)^{\alpha_i}, \end{aligned} \tag{24}$$

where S denotes $\mathrm{supp}(\alpha) = \{i \in [m] \colon \alpha_i > 0\}$. (Equation (24) crucially relies on the product structure of $\widetilde{O}_{b,\lambda} : \mathbb{R}^m \to (0,1)$; recall Equation (5).)

Note that Claim 8.4 is equivalent to the claim that $\boldsymbol{y}$ δ-fools G_α for $\delta = \delta_{\mathrm{CNF}} \cdot O_c(\sqrt{n}/\lambda)^c$. We analyze the three types of functions in Equation (24) in turn:

- Recalling the assumptions of Lemma 8.3, by Item 1 of Fact 8.8, the function $x \mapsto \widetilde{\mathbb{1}}_{b_i,\lambda}(H_i^B x)$ is a weight-1 combination of Boolean functions. Furthermore, since $|\mathrm{supp}(H_i^B)| \le w$, it is in fact a weight-1 combination of Boolean w-juntas.
- Similarly, by Item 2 of Fact 8.8, the function $x \mapsto \widetilde{\mathbb{1}}^{(\alpha_i)}_{b_i,\lambda}(H_i^B x)$ is a weight-$(2\|\widetilde{\mathbb{1}}^{(\alpha_i)}_{b_i,\lambda}\|_\infty)$ combination of Boolean w-juntas.
- Since $\|T_i^B\|_1 \le \sqrt{B} \cdot \|T_i^B\|_2 \le \sqrt{B} \le \sqrt{n}$ 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 $G_\alpha : \{-1,1\}^B \to \mathbb{R}$ is a weight-W combination of Boolean product functions $\Xi : \{-1,1\}^B \to \{0,1\}$, where

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

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 $\boldsymbol{y}$ δ_{CNF}-fools the class of width-w CNFs, we conclude that $\boldsymbol{y}$ δ-fools G_α where $\delta = \delta_{\mathrm{CNF}} \cdot W$. This completes the proof of Claim 8.4. □

8.2 Bounding the Error Terms

We will use the following technical result:

Claim 8.9 (Rosenthal's Inequality). *Let $\beta \in [0,1]$ and let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be independent $\{0, \pm 1\}$-valued random variables, each being 0 with probability $1-\beta$ and ± 1 with probability $\beta/2$ each. Let $w \in \mathbb{R}^n$ be a τ-regular vector of 2-norm 1. Then for any $q \ge 2$,*

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

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

Proof. This is an almost immediate consequence of a refinement of an inequality due to Rosenthal [53]. The exact version we use is due to Nagaev and Pinelis [41] (see also Reference [[51], (4)]); in our context, it states that

$$\begin{aligned}\mathbf{E}[|w \cdot \boldsymbol{x}|^q] &\le 2^{O(q)} \cdot \left(q^q \sum_{j=1}^n \mathbf{E}[|w_j \boldsymbol{x}_j|^q] + q^{q/2} \left(\sum_{j=1}^n \mathbf{E}[(w_j \boldsymbol{x}_j)^2] \right)^{q/2} \right) \\ &\le 2^{O(q)} \cdot \left(q^q \beta \sum_{j=1}^n |w_j|^q + (q\beta)^{q/2} \right).\end{aligned}$$

Since $\beta \sum_j |w_j|^q \le \beta \left(\sum_j w_j^2 \right) \cdot \tau^{q-2} = \beta \tau^{q-2}$, using $x^q + y^q \le (x+y)^q$ for positive x, y we get the claimed bound. □

The following lemma will be used to bound the expectations on the right-hand side of Equation (21):

Lemma 8.10. *Let $L, r_{\text{hash}}, r_{\text{bucket}}$, and τ be as set in Section 4.1. Let $\boldsymbol{h} : [n] \to [L]$ be an r_{hash}-wise uniform hash function, and fix a bucket $\ell \in [L]$. Let $\boldsymbol{y} \sim \{-1, 1\}^n$ 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 $d \ge 2$,*

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

Proof. Let q be the largest even integer smaller than both r_{hash} and r_{bucket}; note that $q = \Theta(\log(m/\delta))$. For notational brevity, we let $\boldsymbol{X}$ denote the $\mathbb{R}^m$-valued random variable $\boldsymbol{X} := T^{\boldsymbol{h}^{-1}(\ell)} \boldsymbol{y}_{\boldsymbol{h}^{-1}(\ell)}$. Since $r_{\text{bucket}}, r_{\text{hash}} \ge q$, we can express $\boldsymbol{X}$ as $\sum_{j=1}^n \boldsymbol{x}_j T^j$ where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \sim \{-1, 0, 1\}$ are q-wise independent random variables distributed as in Claim 8.9, with $\beta = 1/L$.

Since $q > d$ for sufficiently large m, we have that

$$\mathbf{E}\left[\|\boldsymbol{X}\|_\infty^d \right] \le \mathbf{E}\left[\|\boldsymbol{X}\|_q^d \right] \le \mathbf{E}\left[\|\boldsymbol{X}\|_q^q \right]^{d/q} = \left(\sum_{i=1}^m \mathbf{E}[\boldsymbol{X}_i^q] \right)^{d/q}.$$

Applying Claim 8.9 to bound each $\mathbf{E}[\boldsymbol{X}_i^q]$, we conclude that

$$\begin{aligned}\mathbf{E}\left[\|\boldsymbol{X}\|_\infty^d \right] &= \left(m \cdot O\left(q\tau \cdot (1/L\tau^2)^{1/q} + \sqrt{q/L} \right)^q \right)^{d/q} \\ &= m^{d/q} \cdot O\left(q\tau + \sqrt{q/L} \right)^d \\ &= O_d\left(\tau \log(m/\delta) + \sqrt{\frac{\log(m/\delta)}{L}} \right)^d,\end{aligned}$$

where the second inequality uses the fact that $(\frac{1}{L\tau^2})^{1/q} = (\frac{\delta}{\log m})^{O(1/q)} = O(1)$. This completes the proof of Lemma 8.10. □

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 $h : [n] \to [L]$ is *H-good* if

$$|h^{-1}(\ell) \cap \mathrm{supp}(H_i)| \leq w := \frac{2k}{L} \tag{25}$$

for all buckets $\ell \in [L]$ and rows $i \in [m]$. Equivalently, for all $\ell \in [L]$, every row of the the submatrix $H^{h^{-1}(\ell)}$ 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 $\boldsymbol{h} : [n] \to [L]$ is r_{hash}-wise uniform where $r_{\mathrm{hash}} \geq C_1 \log(Lm/\delta)$, then*

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

Proof. Fix any $\ell \in [L]$ and $i \in [m]$. The quantity $|\boldsymbol{h}^{-1}(\ell) \cap \mathrm{supp}(H_i)|$ is a sum of $|\mathrm{supp}(H_i)| \leq k$ many r_{hash}-wise independent $\{0,1\}$-valued random variables, each of which takes the value 1 with probability $1/L$. To bound the probability that $|\boldsymbol{h}^{-1}(\ell) \cap \mathrm{supp}(H_i)|$ is larger than w, we apply the well-known tail bounds for sums of limited-independence random variables due to Schmidt, Siegel, and Srinivasan [54], specifically their Theorem 5(I)(a). Taking the "δ" of their paper to be 1 and observing that their "μ" is our k/L and their "k" is our $r_{\mathrm{hash}} = \Theta(\log(Lm/\delta))$, we get that $\mathbf{Pr}[|\boldsymbol{h}^{-1}(\ell) \cap \mathrm{supp}(H_i)| > w] \leq \delta/(Lm)$. The proposition follows by a union bound over all $\ell \in [L]$ and $i \in [m]$. □

We are now ready to prove Theorem 8.1, which we restate here for convenience:

Theorem 8.12. *Let $\mathscr{G}$ 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 $b \in \mathbb{R}^m$,*

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{u} \sim \{-1,1\}^n} \left[\widetilde{O}_{b,\lambda}(A\boldsymbol{u}) \right] - \mathop{\mathbf{E}}_{z \sim \mathscr{G}_{\mathrm{MZ}}} \left[\widetilde{O}_{b,\lambda}(Az) \right] \right| = O(\delta).$$

Proof. Let $\boldsymbol{h}, \boldsymbol{y}^1, \ldots, \boldsymbol{y}^L, \tilde{\boldsymbol{y}}^1, \ldots, \tilde{\boldsymbol{y}}^L, \check{\boldsymbol{y}}$, and $\boldsymbol{y}^\star$ be the random hash function and random variables associated with our generator $\mathscr{G}$, as defined in Definition 4.2. Recall that a draw from $\boldsymbol{z} \sim \mathscr{G}$ is $\boldsymbol{z} := \check{\boldsymbol{y}} \oplus \boldsymbol{y}^\star$. We will show that in fact $\check{\boldsymbol{y}}$ alone satisfies:

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{u} \sim \{-1,1\}^n} \left[\widetilde{O}_{b,\lambda}(A\boldsymbol{u}) \right] - \mathbf{E} \left[\widetilde{O}_{b,\lambda}(A\check{\boldsymbol{y}}) \right] \right| = O(\delta). \tag{26}$$

Since $\boldsymbol{y}^\star$ and $\check{\boldsymbol{y}}$ are independent, Theorem 8.1 follows as a consequence of Equation (26).

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

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

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

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

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

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

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

$$\begin{aligned}
\text{LHS of Equation (26)} &= \left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{u})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\check{\boldsymbol{y}})\right]\right| \\
&= \left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{\boldsymbol{h},0})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{\boldsymbol{h},L})\right]\right| \\
&\le \mathop{\mathbf{E}}_{\boldsymbol{h}}\left[\left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{\boldsymbol{h},0})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{\boldsymbol{h},L})\right]\right|\right] \\
&\le \mathop{\mathbf{E}}_{\boldsymbol{h}}\left[\left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{\boldsymbol{h},0})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{\boldsymbol{h},L})\right]\right| \cdot \mathbb{1}\left[\boldsymbol{h} \text{ is } H\text{-good}\right]\right] \\
&\quad + \mathbf{Pr}\left[\boldsymbol{h} \text{ is not } H\text{-good}\right] \\
&\le \underbrace{\mathop{\mathbf{E}}_{\boldsymbol{h}}\left[\left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{\boldsymbol{h},0})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{\boldsymbol{h},L})\right]\right| \cdot \mathbb{1}\left[\boldsymbol{h} \text{ is } H\text{-good}\right]\right]}_{\heartsuit} + \delta.
\end{aligned}$$

The penultimate inequality uses the fact that $\widetilde{O}_{b,\lambda}$ 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 $r_{\text{hash}} \ge C_1 \log(Lm/\delta)$).

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

$$\left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{h,0})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{h,L})\right]\right| \le \sum_{\ell=1}^{L}\left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{h,\ell-1})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{h,\ell})\right]\right|. \tag{27}$$

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

$$\left|\mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{h,\ell-1})\right] - \mathbf{E}\left[\widetilde{O}_{b,\lambda}(A\boldsymbol{x}^{h,\ell})\right]\right|. \tag{28}$$

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

$$\begin{aligned}
(28) &= \left|\mathop{\mathbf{E}}_{\boldsymbol{s}}\left[\mathop{\mathbf{E}}_{\boldsymbol{u}}\left[\widetilde{O}_{b,\lambda}(A^{\overline{B}}\boldsymbol{s} + A^{B}\boldsymbol{u}_B)\right] - \mathop{\mathbf{E}}_{\boldsymbol{y}^{\ell},\tilde{\boldsymbol{y}}^{\ell}}\left[\widetilde{O}_{b,\lambda}(A^{\overline{B}}\boldsymbol{s} + A^{B}(\boldsymbol{y}^{\ell} \oplus \tilde{\boldsymbol{y}}^{\ell})_B)\right]\right]\right| \\
&\le \mathop{\mathbf{E}}_{\boldsymbol{s}}\left[\left|\mathop{\mathbf{E}}_{\boldsymbol{u}}\left[\widetilde{O}_{b,\lambda}(A^{\overline{B}}\boldsymbol{s} + A^{B}\boldsymbol{u}_B)\right] - \mathop{\mathbf{E}}_{\boldsymbol{y}^{\ell},\tilde{\boldsymbol{y}}^{\ell}}\left[\widetilde{O}_{b,\lambda}(A^{\overline{B}}\boldsymbol{s} + A^{B}(\boldsymbol{y}^{\ell} \oplus \tilde{\boldsymbol{y}}^{\ell})_B)\right]\right|\right] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{s}}\left[\left|\mathop{\mathbf{E}}_{\boldsymbol{u}}\left[\widetilde{O}_{b-A^{\overline{B}}\boldsymbol{s},\lambda}(A^{B}\boldsymbol{u}_B)\right] - \mathop{\mathbf{E}}_{\boldsymbol{y}^{\ell},\tilde{\boldsymbol{y}}^{\ell}}\left[\widetilde{O}_{b-A^{\overline{B}}\boldsymbol{s},\lambda}(A^{B}(\boldsymbol{y}^{\ell} \oplus \tilde{\boldsymbol{y}}^{\ell})_B)\right]\right|\right] \qquad \text{(Fact 6.3)} \\
&= \mathop{\mathbf{E}}_{\boldsymbol{s}}\left[\left|\mathop{\mathbf{E}}_{\boldsymbol{u}}\left[\widetilde{O}_{b-A^{\overline{B}}\boldsymbol{s},\lambda}(H^{B}\boldsymbol{u}_B + T^{B}\boldsymbol{u}_B)\right] - \mathop{\mathbf{E}}_{\boldsymbol{y}^{\ell},\tilde{\boldsymbol{y}}^{\ell}}\left[\widetilde{O}_{b-A^{\overline{B}}\boldsymbol{s},\lambda}(H^{B}(\boldsymbol{y}^{\ell} \oplus \tilde{\boldsymbol{y}}^{\ell})_B + T^{B}(\boldsymbol{y}^{\ell} \oplus \tilde{\boldsymbol{y}}^{\ell})_B)\right]\right|\right].
\end{aligned}$$

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 s of $\boldsymbol{s}$, and we get that this quantity is at most

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

and therefore

$$\text{RHS of Equation (27)} \le L \cdot \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} + O\left(\frac{\sqrt{\log m}}{\lambda}\right)^{d} \cdot \sum_{\ell=1}^{L} \left(\mathop{\mathbf{E}}_{\boldsymbol{u}} \left[\|T^{h^{-1}(\ell)} \boldsymbol{u}_{h^{-1}(\ell)}\|_\infty^d \right] + \mathop{\mathbf{E}}_{\boldsymbol{y}^\ell, \tilde{\boldsymbol{y}}^\ell} \left[\|T^{h^{-1}(\ell)} (\boldsymbol{y}^\ell \oplus \tilde{\boldsymbol{y}}^\ell)_{h^{-1}(\ell)}\|_\infty^d \right] \right). \tag{29}$$

Since Equation (29) holds for every H-good hash h, we have shown that

$$\begin{aligned}
\heartsuit &\le \mathop{\mathbf{E}}_{\boldsymbol{h}} \left[(\text{RHS of Equation (27)}) \cdot \mathbb{1}[\, \boldsymbol{h} \text{ is } H\text{-good}\,] \right] \\
&\le \mathop{\mathbf{E}}_{\boldsymbol{h}} \left[(\text{RHS of Equation (29)}) \cdot \mathbb{1}[\, \boldsymbol{h} \text{ is } H\text{-good}\,] \right] \\
&\le \mathop{\mathbf{E}}_{\boldsymbol{h}} \left[(\text{RHS of Equation (29)}) \right] \\
&= L \cdot \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} \\
&\quad + O\left(\frac{\sqrt{\log m}}{\lambda}\right)^{d} \cdot \underbrace{\sum_{\ell=1}^{L} \left(\mathop{\mathbf{E}}_{\boldsymbol{h}, \boldsymbol{u}} \left[\|T^{\boldsymbol{h}^{-1}(\ell)} \boldsymbol{u}_{\boldsymbol{h}^{-1}(\ell)}\|_\infty^d \right] + \mathop{\mathbf{E}}_{\boldsymbol{h}, \boldsymbol{y}^\ell, \tilde{\boldsymbol{y}}^\ell} \left[\|T^{\boldsymbol{h}^{-1}(\ell)} (\boldsymbol{y}^\ell \oplus \tilde{\boldsymbol{y}}^\ell)_{\boldsymbol{h}^{-1}(\ell)}\|_\infty^d \right] \right)}_{\diamond}.
\end{aligned}$$

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

$$\begin{aligned}
\heartsuit &\le L \cdot \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} + O\left(\frac{\sqrt{\log m}}{\lambda}\right)^{d} \cdot 2L \cdot O\left(\tau \log(m/\delta) + \sqrt{\frac{\log(m/\delta)}{L}}\right)^{d} \\
&= L \cdot \delta_{\text{CNF}} \cdot m^{d-1} \cdot O\left(\frac{\sqrt{n}}{\lambda}\right)^{d-1} + L \cdot O\left(\frac{\tau\sqrt{\log(m)}\log(m/\delta)}{\lambda} + \frac{\sqrt{\log(m)\log(m/\delta)}}{\lambda\sqrt{L}}\right)^{d}.
\end{aligned} \tag{30}$$

By our choice of parameters as set in Section 4.1,

$$(30) = O(\delta) + \frac{(\log m)^5}{\delta^{2+\varepsilon}} \cdot O\left(\delta^{\varepsilon} \cdot \frac{\log(m)(\log(m/\delta))^{1.5+\varepsilon}}{(\log m)^{2.5+\varepsilon}} + \delta^{\varepsilon/2} \cdot \frac{\log(m)\log(m/\delta)}{(\log m)^{2.5}}\right)^{d}.$$

Taking d to be sufficiently large relative to ε, the above expression can be bounded by $O(\delta)$. This establishes Equation (26), and the proof of Theorem 8.1 is complete. □

9 PROOF OF THEOREM 5.3

Having completed both steps of the two-step program described at the end of Section 6, we are finally ready to prove Theorem 5.3, which we restate here for convenience:

Theorem 9.1. *Let $\mathscr{G}$ be our generator with parameters as set in Section 4.1. For all m-facet (k,τ)-standardized polytopes $Ax \le b$,*

$$\left| \Pr_{\boldsymbol{u} \sim \{-1,1\}^n} \left[A\boldsymbol{u} \in O_b \right] - \Pr_{\boldsymbol{z} \sim \mathscr{G}} \left[A\boldsymbol{z} \in O_b \right] \right| = O(\delta).$$

Proof. 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 $\widetilde{O}_{b^{\text{in}},\lambda}, \widetilde{O}_{b^{\text{out}},\lambda}$ 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 $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$ being $A\boldsymbol{u}$ and $A\boldsymbol{z}$, respectively,

using Theorem 8.1 to show that Equation (7) is satisfied for both $\widetilde{O}_{b^{\text{in}},\lambda}$ and $\widetilde{O}_{b^{\text{out}},\lambda}$ with $\gamma = O(\delta)$. We conclude that:

$$\begin{aligned}
&\left| \mathop{\mathbf{Pr}}_{\boldsymbol{u}\sim\{-1,1\}^n}[A\boldsymbol{u} \in O_b] - \mathop{\mathbf{Pr}}_{\boldsymbol{z}\sim\mathscr{G}}[A\boldsymbol{z} \in O_b] \right| \\
&= O(\delta) + \mathbf{Pr}[A\boldsymbol{u} \in \eth_{\pm\Lambda} O_b] && \text{(Lemma 6.9 and Theorem 8.1)} \\
&= O(\delta) + O\Big(\Lambda\sqrt{\log m}\Big) && \text{(Theorem 7.1; note that } \Lambda \geq \tau \text{ is indeed satisfied)} \\
&= O(\delta) + O\left(\lambda\sqrt{\log(m/\delta)\log m}\right) \\
&= O(\delta). && (31)
\end{aligned}$$

This completes the proof of Theorem 5.3. □

APPENDICES

A

In this section, we prove Theorem 5.1. The proof uses the "critical index" theory for Boolean halfspaces, introduced in Reference [55] and used in several subsequent works on halfspaces.

Definition A.1 (Critical Index). Let $w \in \mathbb{R}^n$ and assume for notational simplicity that $|w_1| \geq |w_2| \geq \cdots \geq |w_n|$. The *$\tau$-critical index* of w is the least j such that the "tail" $(w_j, w_{j+1}, \ldots, w_n)$ is τ-regular, or ∞ if no such j exists.

Given A as in Theorem 5.1, the rows that are already (k,τ)-regular pose no difficulty as a simple rescaling of any such row (and the corresponding entry of b) makes it (k,τ)-standardized. The remaining rows A_i have τ-critical index exceeding k. The critical index theory [49, 55] says that such halfspaces $\mathbb{1}[A_i x \leq b_i]$ are very close to k-juntas, and in fact Reference [9] shows that this is true even under $(k+2)$-wise uniform distributions (for a slightly larger choice of k as alluded to in Remark 5.2). We tweak the quantitative aspects of these arguments below to work for the choice of k given in Equation (3). It will be convenient to follow the treatment in Reference [18].

The first lemma below says that if the "head" variables are set uniformly, then the resulting random variable has good anticoncentration at the scale of the two-norm of the tail:

Lemma A.2. *Let $\tau \in (0,1)$, $\varepsilon \in (0,1/2)$, $s > 1$. Then for a certain $\ell = O(\log(s)\log(1/\varepsilon)/\tau^2)$ the following holds: If $w \in \mathbb{R}^n$ as in Definition A.1 has τ-critical index at least ℓ, then for all $\theta \in \mathbb{R}$,*

$$\mathop{\mathbf{Pr}}_{\substack{\boldsymbol{u}\sim\{-1,1\}^\ell \\ \text{uniform}}}[|w_1\boldsymbol{u}_1 + \cdots + w_\ell \boldsymbol{u}_\ell - \theta| \leq s\cdot\sigma] \leq \varepsilon + O(\log(1/\varepsilon)\exp(-s^2/2)),$$

where $\sigma := \sqrt{w_{\ell+1}^2 + \cdots + w_n^2}$.

Proof. We refer directly to the proof of the almost identical Reference [18, Theorem 5.3] in the full version of that paper. In that proof we may take "δ" to be τ^2, and "η" to be $1/\sqrt{3}$, since we work with uniform ± 1 bits (see Fact 3.3.5 therein). The only change needed in the proof occurs before "inequality (10)." That inequality uses the fact that a certain random variable $\boldsymbol{z}$ satisfies the tail bound $\mathbf{Pr}[|\boldsymbol{z}| \geq s\rho] \leq O(1/s^4)$ when ρ is at most the standard deviation of $\boldsymbol{z}$. But in our current setting, the random variable $\boldsymbol{z}$ equals $w_1\boldsymbol{u}_1 + \cdots + w_\ell\boldsymbol{u}_\ell$, i.e., it is a weighted sum of independent uniform ± 1 bits, and so we have the improved tail bound $2\exp(-s^2/2)$ using Hoeffding. Carrying through the remainder of the proof with this change yields the conclusion of Lemma A.2. □

Lemma A.3. *Let $\tau \in (0,1)$ and let $\varepsilon \in (0,1/2)$. Then for a certain $k = O(\log(1/\varepsilon)\log\log(1/\varepsilon)/\tau^2)$ and $r = O(\log(1/\varepsilon))$, the following holds for every $w \in \mathbb{R}^n$ that is* not *(k,τ)-regular:*

Let $H \subseteq [n]$ be the set of k coordinates i for which $|w_i|$ is largest and let $T = [n] \setminus H$. Assume $w' \in \mathbb{R}^n$ has $w'_H = w_H$ and $\|w'_T\|_2 \leq \|w_T\|_2$. Then for any $\theta \in \mathbb{R}$,

$$\Pr_{\boldsymbol{y}}[\mathbb{1}[w \cdot \boldsymbol{y} \leq \theta] \neq \mathbb{1}[w' \cdot \boldsymbol{y} \leq \theta]] = O(\varepsilon)$$

provided $\boldsymbol{y} \sim \{-1,1\}^n$ is $(k+r)$-wise uniform.

Proof. Suppose w is not (k,τ)-regular. By reordering coordinates we may assume that $H = [k]$; then the non-(k,τ)-regularity of w means the τ-critical index of w exceeds k. We may therefore apply Lemma A.2 with $s = O(\sqrt{\log(1/\varepsilon)})$. Using the fact that $\boldsymbol{y}_H$ is fully uniform, we get

$$\mathbf{Pr}[|w_H \cdot \boldsymbol{y}_H - \theta| \leq s \cdot \|w_T\|_2] = O(\varepsilon) \qquad \text{(and note that } w'_H \cdot \boldsymbol{y}_H = w_H \cdot \boldsymbol{y}_H\text{)}. \tag{32}$$

Conditioned on any outcome of $\boldsymbol{y}_H$, the distribution of $\boldsymbol{y}_T$ remains r-wise uniform. We claim that it remains to show the following:

$$\mathbf{Pr}[|w'_T \cdot \boldsymbol{y}_T| \geq s \cdot \|w_T\|_2] = O(\varepsilon). \tag{33}$$

To see that this suffices, observe that by Equation (32) we have that $|w_H \cdot \boldsymbol{y}_H - \theta| = |w'_H \cdot \boldsymbol{y}_H - \theta| > s \cdot \|w_T\|_2$ except with probability $O(\varepsilon)$. Also, by applying Equation (33) with w' and with $w' = w$, we get both $|w_T \cdot \boldsymbol{y}_T|, |w'_T \cdot \boldsymbol{y}_T| \leq s \cdot \|w_T\|_2$ except with another probability at most $O(\varepsilon)$. When all of these events occur, $\mathbb{1}[w \cdot \boldsymbol{y} \leq \theta]$ and $\mathbb{1}[w' \cdot \boldsymbol{y} \leq \theta]$ agree.

Finally, we can establish Equation (33) by appealing to, e.g., Reference [47, Theorem 9.23]. That theorem (with $k = 1$) shows that for $t \geq \sqrt{2e}$, any linear form $f(\boldsymbol{x})$ in uniform ± 1 random variables $\boldsymbol{x}$ has $\mathbf{Pr}[|f(\boldsymbol{x})| \geq t\|f\|_2] \leq \exp(-O(t^2))$. If we could directly apply this to the linear form $w'_T \cdot \boldsymbol{y}_T$, then we would be done by taking $t = s$ and using $\|w'_T\|_2 \leq \|w_T\|_2$. We cannot directly apply this theorem, because the bits $\boldsymbol{y}_T$ are not uniformly random. However, inspecting the proof of Reference [47, Theorem 9.23] shows that it suffices for those bits to be $O(t^2)$-wise uniform, which they are provided that $r = O(\log(1/\varepsilon)) = O(s^2) = O(t^2)$. The reason that this suffices is because the proof only uses $(2, q, 1/\sqrt{q-1})$-hypercontractivity of $f(\boldsymbol{x})$ for $q = O(t^2)$, and (for even integer q) this condition only involves the first q moments of $f(\boldsymbol{x})$, which do not change if $\boldsymbol{x}$ is assumed to be merely q-wise uniform rather than truly uniform. □

We can now prove Theorem 5.1:

Proof. We will use Lemma A.3 with $\varepsilon = c\delta/m$ for small constant $c > 0$. This leads to the choice of k in the statement of Theorem 5.1; also, $r \ll k$ and so $2k \geq r + k$.

Given $A \in \mathbb{R}^{m \times n}$, as noted earlier the rows that are (k,τ)-regular are not a problem, so we consider all rows A_i that are *not* (k,τ)-regular. For these rows, we apply Lemma A.3, taking A'_i to agree with A_i on the appropriate "head" coordinates H_i, and taking A'_i to simply be 0 on the remaining "tail" coordinates. Note that A'_i is now trivially (k,τ)-regular. By Lemma A.3, we have that

$$\Pr_{\boldsymbol{y}}[\mathbb{1}[A_i \cdot \boldsymbol{y} \leq b_i] \neq \mathbb{1}[A'_i \cdot \boldsymbol{y} \leq b_i]] \leq \delta/m.$$

Taking $b'_i = b_i$ for these i's, and union-bounding over the at most m of them, we are almost at the point of establishing Equation (4) from the theorem statement. We now have that *all* A'_i are (k,τ)-regular; the only deficiency is that the "tail" of each row need not have 2-norm 1, as required.

Whenever the "tail" of A'_i has nonzero 2-norm, we can simply scale A'_i and b' by the same positive factor to make the tail of A'_i have 2-norm 1; this scaling does not change the Boolean function $\mathbb{1}[A'_i \cdot x \leq b'_i]$ at all. The only (very minor) difficulty now remaining is that some of the rows A'_i may have tail with 2-norm zero. It is well known, however, that one can always slightly

perturb the coefficients and threshold in a halfspace without changing it as a Boolean function.[3] We can perturb in such a way that the tail coefficients all become equal to some sufficiently small $\eta > 0$. After this perturbation, the row A'_i *is* (k, τ)-regular (this holds, recalling that $k \le n/2$, since $n - k \ge k \ge 1/\tau^2$) and its tail has positive 2-norm. Now we can scale up (A'_i, b'_i) as before to make the tail have 2-norm 1. □

B

We recall Claim 7.11:

Claim B.1 (Claim 7.11 Restated). *For $2 \le m \le 2^n$, there is a matrix $A \in \{-1,1\}^{m\times n}$ and a vector $b \in \mathbb{Z}^m$ such that*

$$\mathbf{Pr}[A\boldsymbol{u} \in \partial O_b] = \Omega\left(\frac{\sqrt{\ln m}}{\sqrt{n}}\right).$$

Proof. The proof is a simple probabilistic existence argument that follows the approach used to prove Theorem 2 in Reference [26]. For a polytope $\mathcal{K} = \mathbb{1}[Ax \le b]$ we define

$$\text{Inside}(\mathcal{K}) = \{x \in \{-1,1\}^n : Ax \in O_b \setminus \partial O_b, \quad \text{i.e., } A_i x < b_i \text{ for all } i \in [m]\},$$
$$\text{Surface}(\mathcal{K}) = \{x \in \{-1,1\}^n : Ax \in \partial O_b, \quad \text{i.e., } Ax \le b \text{ and } A_i x = b_i \text{ for some } i \in [m]\}.$$

Given $2 \le m \le 2^n$, if $m < 10$, then the one-facet polytope $\mathbb{1}[x_1 + \cdots + x_n \le 0]$ does the job (more formally, we take A to be the $m \times n$ all-1's matrix and b to be the zero vector in $\mathbb{R}^m$). It is also clear that proving our result for $m \le 2^{n/10}$ also proves it for $2^{n/10} \le m \le 2^n$. So, we henceforth assume that $10 \le m \le 2^{n/10}$. Let $k \ge n/2$ be an integer to be chosen later, and let $F : \{-1,1\}^n \to \{0,1\}$ denote the halfspace $F(x) = \mathbb{1}[x_1 + \cdots + x_n \le k]$. Now define the following quantities:

$$p_{\mathrm{I}} := |\text{Inside}(F)|/2^n = \binom{n}{<k}/2^n, \qquad p_{\mathrm{S}} := |\text{Surface}(F)|/2^n = \binom{n}{k}/2^n.$$

Let $\overline{\boldsymbol{\sigma}} = (\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^m)$ where each $\boldsymbol{\sigma}^i$ is an independent uniform string in $\{-1,1\}^n$. Let $\boldsymbol{A} \in \{-1,1\}^{m\times n}$ be the matrix whose ith row is $\boldsymbol{\sigma}^i$, and let b be the vector $(k, \ldots, k) \in \mathbb{Z}^m$. It is easy to see that to prove our result it suffices to show that there is a fixed outcome A of $\boldsymbol{A}$ such that

$$\mathbf{Pr}[A\boldsymbol{u} \in \partial O_b] = \Omega\left(\frac{\sqrt{\log m}}{\sqrt{n}}\right), \tag{34}$$

and this is what we show below. Towards this end, for each $i \in [m]$, let us define the matrix $\boldsymbol{A}^{\setminus i} \in \{-1,1\}^{(m-1)\times n}$ obtained by removing the ith row of $\boldsymbol{A}$, and further define $b' = (k, \ldots, k) \in \mathbb{Z}^{m-1}$.

For each fixed $z \in \{-1,1\}^n$ and each $i \in [m]$ we have

$$\mathop{\mathbf{Pr}}_{\overline{\sigma}}\left[z \in \text{Inside}(\mathbb{1}[\boldsymbol{A}^{\setminus i} x \le b'])\right] = p_{\mathrm{I}}^{m-1}$$

and

$$\mathop{\mathbf{Pr}}_{\overline{\sigma}}\left[z \in \text{Surface}(\mathbb{1}[\boldsymbol{\sigma}^i \cdot x \le k])\right] = p_{\mathrm{S}}.$$

Since $\boldsymbol{\sigma}^i$ and $(\boldsymbol{\sigma}^{i'})_{i' \in [m]\setminus\{i\}}$ are independent for each $i \in [m]$, it follows that

$$\mathop{\mathbf{Pr}}_{\overline{\sigma}}\left[z \in \text{Inside}(\mathbb{1}[\boldsymbol{A}^{\setminus i} x \le b']) \ \&\ z \in \text{Surface}(\mathbb{1}[\boldsymbol{\sigma}^i \cdot x \le k])\right] = p_{\mathrm{S}} \cdot p_{\mathrm{I}}^{m-1},$$

[3] Given a halfspace $\mathbb{1}[w \cdot x \le \theta]$, there is a smallest value $\theta' > \theta$ achievable as $w \cdot x$ for $x \in \{-1,1\}^n$; first perturb θ upward to $(\theta + \theta')/2$. Now no input x achieves $w \cdot x = \theta$ exactly, so we can perturb the coefficients of w by sufficiently small amounts.

and since the events

$$z \in \text{Inside}(\mathbb{1}[\boldsymbol{A}^{\setminus i} x \le b']) \ \&\ z \in \text{Surface}(\mathbb{1}[\boldsymbol{\sigma}^{i} \cdot x \le k])$$

and

$$z \in \text{Inside}(\mathbb{1}[\boldsymbol{A}^{\setminus i'} x \le b']) \ \&\ z \in \text{Surface}(\mathbb{1}[\boldsymbol{\sigma}^{i'} \cdot x \le k])$$

are mutually exclusive for $i \neq i' \in [m]$, by a union bound we have that

$$\underset{\overline{\sigma}}{\mathbf{Pr}}\left[z \in \text{Surface}(\mathbb{1}[\boldsymbol{A}x \le b'])\right] = \underset{\overline{\sigma}}{\mathbf{Pr}}[\boldsymbol{A}z \in \partial O_{b'}] = m \cdot p_{\text{S}} \cdot p_{\text{I}}^{m-1}.$$

It follows that there is an outcome of $\overline{\boldsymbol{\sigma}}$ such that the resulting matrix $A \in \mathbb{R}^{m \times n}$ has at least an $m \cdot p_{\text{S}} \cdot p_{\text{I}}^{m-1}$ fraction of all points in $\{-1, 1\}^n$ satisfying $Az \in \partial O_b$; i.e.,

$$\mathbf{Pr}[A\boldsymbol{u} \in \partial O_b] \ge m \cdot p_{\text{S}} \cdot p_{\text{I}}^{m-1}. \tag{35}$$

It remains only to argue that for any $10 \le m \le 2^{n/10}$, there is a value k such that, for $p_{\text{I}} = \binom{n}{<k}$ and $p_{\text{S}} = \binom{n}{k}$, we have

$$m \cdot p_{\text{S}} \cdot p_{\text{I}}^{m-1} = \Omega\left(\sqrt{\log m}/\sqrt{n}\right).$$

Towards this end we choose k to be the largest integer such that $\binom{n}{<k}/2^n \le 1 - \frac{1}{m}$. Recalling that $10 \le m \le 2^{n/10}$, we have that $n/2 \le k \le 0.99n$, and hence $\binom{n}{k}$ and $\binom{n}{k-1}$ are within an absolute constant multiplicative factor of each other. It follows that

$$p_{\text{I}} = \binom{n}{<k}/2^n = 1 - \Theta(1/m),$$

which implies that

$$p_{\text{I}}^{m-1} = \Omega(1).$$

Writing $k = n/2 + (\sqrt{n}/2)t$, we have the elementary binomial tail lower bound $1 - p_{\text{I}} \ge \exp(-O(t^2))$ (see, e.g., Reference [36, inequality (4.2)])); hence $t \ge \Omega(\sqrt{\ln m})$. The desired bound

$$p_{\text{S}} = \binom{n}{k}/2^n = \Omega(t/(m\sqrt{n}))$$

now follows from asymptotically tight estimates (up to universal constants for all $0 \le t \le \sqrt{n}$) for the Mills ratio of the binomial distribution; see Reference [38]. □

ACKNOWLEDGMENTS

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. *Adv. Comput. Res.* 5 (1989), 199–222.
[2] Srinivasan Arunachalam and Penghui Yao. 2021. Positive spectrahedrons: Geometric properties, invariance principles and pseudorandom generators. *CoRR* abs/2101.08141 (2021).
[3] Louay Bazzi. 2009. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.* 38, 6 (2009), 2220–2272.
[4] Richard Beigel, Nick Reingold, and Daniel Spielman. 1995. PP is closed under intersection. *J. Comput. Syst. Sci.* 50, 2 (1995), 191–202.
[5] 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.
[6] Avrim Blum and Ravi Kannan. 1997. Learning an intersection of a constant number of halfspaces under a uniform distribution. *J. Comput. Syst. Sci.* 54, 2 (1997), 371–380.

[7] Mei-Chu Chang. 2002. A polynomial bound in Freiman's theorem. *Duke Math. J.* 113, 3 (2002), 399–419. DOI: https://doi.org/10.1215/S0012-7094-02-11331-3

[8] Eshan Chattopadhyay, Anindya De, and Rocco Servedio. 2018. Simple and efficient pseudorandom generators from Gaussian processes. Retrieved from https://eccc.weizmann.ac.il/report/2018/100/.

[9] 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.

[10] 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.

[11] Devdatt Dubhashi and Alessandro Panconesi. 2009. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, Cambridge.

[12] Paul Erdős. 1945. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* 51 (1945), 898–902.

[13] William Feller. 1968. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons.

[14] Péter Frankl and Zoltán Füredi. 1988. Solution of the Littlewood–Offord problem in high dimensions. *Ann. Math.* 128, 2 (1988), 259–270.

[15] 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.

[16] 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)*.

[17] Parikshit Gopalan, Raghu Meka, and Omer Reingold. 2013. DNF sparsification and a faster deterministic counting algorithm. *Comput. Complex.* 22, 2 (2013), 275–310. DOI: https://doi.org/10.1007/s00037-013-0068-6

[18] 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. Retrieved from https://arxiv.org/abs/1001.1593.

[19] Prahladh Harsha, Adam R. Klivans, and Raghu Meka. 2012. An invariance principle for polytopes. *J. ACM* 59, 6 (2012), 29:1–29:25. DOI: https://doi.org/10.1145/2395116.2395118

[20] Pooya Hatami, William Hoza, Avishay Tal, and Roei Tell. 2021. Fooling constant-depth threshold circuits. *Electron. Colloquium Comput. Complex.* 28 (2021), 2.

[21] Russell Impagliazzo, Cristopher Moore, and Alexander Russell. 2014. An entropic proof of Chang's inequality. *SIAM J. Discrete Math.* 28, 1 (2014), 173–176. DOI: https://doi.org/10.1137/120877982

[22] 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.

[23] Valentine Kabanets, Sajin Koroth, Zhenjian Lu, Dimitrios Myrisiotis, and Igor Oliveira. 2020. Algorithms and lower bounds for de Morgan formulas of low-communication leaf gates. In *Proceedings of the 35th Computational Complexity Conference (CCC'20) (LIPIcs)*, Vol. 169. 15:1–15:41.

[24] Daniel Kane. 2011. k-Independent Gaussians fool polynomial threshold functions. In *Proceedings of the 26th IEEE Conference on Computational Complexity (CCC)*. 252–261.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] Zohar Shay Karnin, Yuval Rabani, and Amir Shpilka. 2012. Explicit dimension reduction and its applications. *SIAM J. Comput.* 41, 1 (2012), 219–249.

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

[31] Daniel Kleitman. 1970. On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors. *Adv. Math.* 5, 1 (1970), 155–157.

[32] Adam Klivans, Ryan O'Donnell, and Rocco A. Servedio. 2004. Learning intersections and thresholds of halfspaces. *J. Comput. Syst. Sci.* 68, 4 (2004), 808–840.

[33] 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.

[34] 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.

[35] Pravesh 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.
[36] Michel Ledoux and Michel Talagrand. 1991. *Probability in Banach Spaces*. Springer.
[37] 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.
[38] Brendan D. McKay. 1989. On Littlewood's estimate for the binomial distribution. *Adv. Appl. Probab.* 21, 2 (1989), 475–478. DOI: https://doi.org/10.2307/1427172
[39] Raghu Meka and David Zuckerman. 2013. Pseudorandom generators for polynomial threshold functions. *SIAM J. Comput.* 42, 3 (2013), 1275–1301.
[40] Marvin Minsky and Seymour Papert. 1968. *Perceptrons: An Introduction to Computational Geometry*. The MIT Press, Cambridge, MA.
[41] Sergey Nagaev and Iosif Pinelis. 1978. Some inequalities for the distribution of sums of independent random variables. *Theor. Probab. Applic.* 22, 2 (1978), 248–256.
[42] 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.
[43] Noam Nisan. 1991. Pseudorandom bits for constant depth circuits. *Combinatorica* 11, 1 (1991), 63–70.
[44] Noam Nisan. 1992. Pseudorandom generators for space-bounded computations. *Combinatorica* 12, 4 (1992), 449–461.
[45] Noam Nisan and Avi Wigderson. 1994. Hardness vs. randomness. *J. Comput. Syst. Sci.* 49, 2 (1994), 149–167. DOI: https://doi.org/10.1016/S0022-0000(05)80043-1
[46] Noam Nisan and David Zuckerman. 1996. Randomness is linear in space. *J. Comput. Syst. Sci.* 52, 1 (1996), 43–52. DOI: https://doi.org/10.1006/jcss.1996.0004
[47] Ryan O'Donnell. 2014. *Analysis of Boolean Functions*. Cambridge University Press. Retrieved from http://analysisofbooleanfunctions.net/.
[48] Ryan O'Donnell and Rocco Servedio. 2010. New degree bounds for polynomial threshold functions. *Combinatorica* 30, 3 (2010), 327–358.
[49] Ryan O'Donnell and Rocco A. Servedio. 2011. The chow parameters problem. *SIAM J. Comput.* 40, 1 (2011), 165–199.
[50] Ryan O'Donnell, Rocco A. Servedio, and Li-Yang Tan. 2019. Fooling polytopes. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC'19)*. ACM, 614–625.
[51] Iosif Pinelis and Sergei Utev. 1985. Estimates of the moments of sums of independent random variables. *Theor. Probab. Applic.* 29, 3 (1985), 574–577.
[52] Alexander Razborov. 2009. A simple proof of Bazzi's theorem. *ACM Trans. Comput. Theor.* 1, 1 (2009), 3.
[53] 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.
[54] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. 1995. Chernoff-Hoeffding bounds for applications with limited independence. *SIAM J. Discrete Math.* 8, 2 (1995), 223–250. DOI: https://doi.org/10.1137/S089548019223872X
[55] Rocco A. Servedio. 2007. Every linear threshold function has a low-weight approximator. *Comput. Complex.* 16, 2 (2007), 180–209.
[56] 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.
[57] 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.
[58] Alexander A. Sherstov. 2013. The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.* 42, 6 (2013), 2329–2374.
[59] Alexander A. Sherstov. 2013. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica* 33, 1 (2013), 73–96.
[60] Michel Talagrand. 1996. How much are increasing sets positively correlated? *Combinatorica* 16, 2 (1996), 243–258. DOI: https://doi.org/10.1007/BF01844850
[61] Terence Tao. 2010. 254A Notes: Topics in random matrix theory. Retrieved from https://terrytao.wordpress.com/tag/lindeberg-replacement-trick/.
[62] Terence Tao and Van Vu. 2012. The Littlewood–offord problem in high dimensions and a conjecture of Frankl and Füredi. *Combinatorica* 32, 3 (2012), 363–372.
[63] Salil P. Vadhan. 2012. Pseudorandomness. *Found. Trends Theoret. Comput. Sci.* 7, 1-3 (2012), 1–336.
[64] Santosh Vempala. 2010. A random-sampling-based algorithm for learning intersections of halfspaces. *J. ACM* 57, 6:32 (2010).

Received July 2020; revised March 2021; accepted April 2021