Bounded Independence Fools Degree–2 Threshold Functions

Ilias Diakonikolas*
Department of Computer Science
Columbia University
New York, NY 10027
ilias@cs.columbia.edu

Daniel M. Kane†
Department of Mathematics
Harvard University
Cambridge, MA 02138
dankane@math.harvard.edu

Jelani Nelson‡
CSAIL
Massachusetts Institute of Technology
Cambridge, MA 02139
minilek@mit.edu

Abstract— **For an n-variate degree–2 real polynomial p, we prove that $\mathbf{E}_{x \sim \mathcal{D}}[\mathrm{sign}(p(x))]$ is determined up to an additive ε as long as $\mathcal{D}$ is a k-wise independent distribution over $\{-1, 1\}^n$ for $k = \mathrm{poly}(1/\varepsilon)$. This gives a broad class of explicit pseudorandom generators against degree-2 boolean threshold functions, and answers an open question of Diakonikolas et al. (FOCS 2009).**

Keywords-derandomization; k-wise independence; polynomial threshold functions

I. INTRODUCTION

This paper is concerned with the power of limited independence to fool low-degree polynomial threshold functions. A *degree-d polynomial threshold function* (henceforth d-PTF), is a boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ expressible as $f(x) = \mathrm{sign}(p(x))$, where p is an n-variate degree-d real polynomial, and $\mathrm{sign} : \mathbb{R} \to \{-1, 1\}$ is -1 for negative arguments and 1 otherwise. PTFs have played an important role in computer science since the early perceptron work of Minsky and Papert [26], and have since been extensively investigated in circuit complexity, communication complexity, learning theory, and voting theory.

A distribution $\mathcal{D}$ on $\{-1, 1\}^n$ is said to ε-fool a function $f : \{-1, 1\}^n \to \{-1, 1\}$ if

$$|\mathbf{E}_{x \sim \mathcal{D}}[f(x)] - \mathbf{E}_{x \sim \mathcal{U}_n}[f(x)]| \leq \varepsilon$$

where $\mathcal{U}_n$ is the uniform distribution on $\{-1, 1\}^n$. We say that $\mathcal{D}$ ε-fools a class $\mathcal{F}$ of functions if $\mathcal{D}$ ε-fools every $f \in \mathcal{F}$. A distribution $\mathcal{D}$ on $\{-1, 1\}^n$ is k-wise independent if every restriction of $\mathcal{D}$ to k coordinates is uniform on $\{-1, 1\}^k$. Despite their simplicity, k-wise independent distributions have been a surprisingly

powerful and versatile derandomization tool, fooling classes such as AC^0 [2], [7], [30] and halfspaces [11].

Our Contributions. The problem we study is the following: How large must $k = k(n, d, \varepsilon)$ be in order for *every* k-wise independent distribution on $\{-1, 1\}^n$ to ε-fool the class of d-PTFs? The $d = 1$ case of this problem was recently considered in [11], where it was shown that $k(n, 1, \varepsilon) = \tilde{\Theta}(1/\varepsilon^2)$ independent of n, where the tilde notation hides $\mathrm{polylog}(1/\varepsilon)$ factors. An open problem in [11] was to identify $k = k(n, d, \varepsilon)$ for $d \geq 2$. In this work, we make progress on this question by proving the following:

Theorem I.1 (Main Theorem). *Any $\tilde{\Omega}(\varepsilon^{-9})$-wise independent distribution on $\{-1, 1\}^n$ ε-fools all 2-PTFs.*

Prior to this work, for $d > 1$ it was not even known whether $o(n)$-wise independence suffices for constant ε. Standard explicit constructions of k-wise independent distributions over $\{-1, 1\}^n$ have seed length $O(k \cdot \log n)$ [1], [10], which is optimal up to constant factors. As a consequence, Theorem I.1 gives a general class of explicit pseudorandom generators (PRGs) for 2-PTFs with seed length $\log n \cdot \tilde{O}(\varepsilon^{-9})$.

Another consequence of Theorem I.1 is that bounded independence suffices for the invariance principle of [27] in the case of degree-2 polynomials. Roughly, this says that for a "low influence" degree-2 polynomial p the distribution of $p(x)$ is essentially invariant if x is drawn from a k-wise distribution over n uniform random signs versus a k-wise distribution over n standard Gaussians.

The techniques we employ to obtain our main result are quite robust. Our approach yields for example that Theorem I.1 holds not only over the hypercube, but also over the n-variate Gaussian distribution. The proof also readily extends to show that the intersection of m halfspaces, or even m degree-2 threshold functions, is ε-fooled by $\mathrm{poly}(1/\varepsilon)$-wise independence for any constant m (over both the hypercube and the multivariate Gaussian). As a special case of the latter result, we ob-

tain that the Goemans-Williamson hyperplane rounding scheme [16] can be derandomized using $\Omega(1/\varepsilon^2)$-wise independence.[1]

A key component in our proof is a generic method known as *FT-mollification* [22] for smoothing functions while keeping control on high-order derivatives. In this work, we both refine the FT-mollification construction of [22] as well as generalize it to the multivariate setting. Our new construction turns out to have independent applications: for example, it yields a simple proof of Jackson's theorem in approximation theory [9], as well as one of its multivariate generalizations to the unit ℓ_2 ball by Nathan and Shapiro [28].

Motivation and Related Work. The literature is rich with explicit generators for various natural classes of functions. In recent years, there has been much interest in not only constructing PRGs for natural complexity classes, but also in doing so with as broad and natural a family of PRGs as possible. One example is the recent work of Bazzi [2] on fooling depth-2 circuits (simplified by Razborov [30]), and of Braverman [7] on fooling AC^0, with bounded independence.

During the past year there has been a flurry of results on constructing PRGs against threshold functions [29], [11], [25], [17], [19], [23], [4]. Most directly related to the results of this paper is the work of [25]. Independently and concurrently to this work, they constructed PRGs against d-PTFs with seed length $\log n \cdot 2^{O(d)} \cdot (1/\varepsilon)^{8d+3}$ [25]. For $d = 2$ their seed-length is qualitatively similar to ours. Their PRG is not based on k-wise independence alone.

II. NOTATION

Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial and $p(x) = \sum_{S \subseteq [n]} \widehat{p}_S \chi_S(x)$ be its Fourier-Walsh expansion, where $\chi_S(x) = \prod_{i \in S} x_i$. The *variance* of p is $\mathbf{Var}[p] = \sum_{|S|>0} \widehat{p}_S^2$. The *influence* of variable i on p is $\mathrm{Inf}_i(p) = \sum_{S \ni i} \widehat{p}_S^2$, and the *total influence* of p is $\mathrm{Inf}(p) = \sum_{i=1}^n \mathrm{Inf}_i(p)$. If $\mathrm{Inf}_i(p) \leq \tau \cdot \mathrm{Inf}(p)$ for all i, we say that the polynomial p is τ-*regular*. If $f(x) = \mathrm{sign}(p(x))$, where p is τ-regular, we say that f is a τ-*regular* PTF.

For $R \subseteq \mathbb{R}^d$ denote by $I_R : \mathbb{R}^d \to \{0,1\}$ its characteristic function, i.e. $I_R(x) = 1$ iff $x \in R$. It will be convenient in some of the proofs to phrase our results in terms of ε-fooling $I_{[0,\infty)}(p(x))$ as opposed to $\mathrm{sign}(p(x))$; by linearity of expectation, these two tasks are equivalent up to changing ε by a factor of 2.

We frequently use $A \approx_\varepsilon B$ to denote that $|A - B| = O(\varepsilon)$, and we let the function $d_2(x, R)$ denote the ℓ_2 distance from some $x \in \mathbb{R}^d$ to a region $R \subseteq \mathbb{R}^d$.

Finally, we familiarize the reader with some multi-index notation. A d-dimensional multi-index is a vector $\beta \in \mathbb{N}^d$ (here $\mathbb{N}$ is the nonnegative integers). For $\alpha, \beta \in \mathbb{N}^d$, we say $\alpha \leq \beta$ if the inequality holds coordinate-wise, and for such α, β we define $|\beta| = \sum_i \beta_i$, $\binom{\beta}{\alpha} = \prod_{i=1}^d \binom{\beta_i}{\alpha_i}$, and $\beta! = \prod_{i=1}^d \beta_i!$. For $x \in \mathbb{R}^d$ we use x^β to denote $\prod_{i=1}^d x_i^{\beta_i}$, and for $f : \mathbb{R}^d \to \mathbb{R}$ we use $\partial^\beta f$ to denote $\frac{\partial^{|\beta|}}{\partial x_1^{\beta_1} \cdots \partial x_d^{\beta_d}} f$.

For $A \in \mathbb{R}^{n \times n}$, $\|A_F\| = (\sum_{i,j=1}^{n,n} A_{i,j}^2)^{1/2}$ denotes the *Frobenius norm* of A, $\mathrm{tr}(A)$ denotes the trace of A, and $\|A\|_2$ denotes the *operator norm* of A, i.e. $\sup_{\|x\|_2=1} \|Ax\|_2$, which also equals the largest magnitude of an eigenvalue of A if A is real and symmetric. We use $\lambda_{\min}(A)$ to denote the smallest magnitude of an eigenvalue of A.

III. OVERVIEW OF OUR PROOF OF THEOREM I.1

Our proof outline follows the strategy set forth in [11]: we first prove that bounded independence fools the class of *regular* 2-PTFs (Step 1), then reduce the general case to the regular case (Step 2) to show that bounded independence fools all 2-PTFs. The bulk of our proof is to establish Step 1. Step 2 is achieved by adapting the recent results of [12].

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a boolean function. To show that f is fooled by k-wise independence, it suffices – and is in fact necessary – to prove the existence of two degree-k "sandwiching" polynomials $q_u, q_l : \{-1,1\}^n \to \{-1,1\}$ that approximate f in ℓ_1-norm (see e.g. [2], [5]). Even though this is an n-dimensional approximation problem, it may be possible to exploit the additional structure of the function under consideration to reduce it to a low-dimensional problem. This is exactly what is done in [11] (and subsequently in [22]) for the case of regular halfspaces.

We now briefly explain the approach of [11]. Let $f(x) = \mathrm{sign}(\langle w, x \rangle)$ be an ε^2-regular halfspace, i.e. $\|w\|_2 = 1$ and $\max_i |w_i| \leq \varepsilon$. The works of [11], [22] use the Berry-Esséen theorem, which states that the random variable $\langle w, x \rangle$ behaves approximately like a standard Gaussian and hence can be treated as if it was one-dimensional. Thus, both [11] and [22] construct (implicitly in the latter) a (different in each case) univariate polynomial $P : \mathbb{R} \to \mathbb{R}$ that is a good "upper sandwich" ℓ_1-approximation to the sign function under the normal distribution in $\mathbb{R}$. The desired n-variate sandwiching polynomials are then obtained (roughly) by setting $q_u(x) = P(\langle w, x \rangle)$ and $q_l(x) = -P(-\langle w, x \rangle)$.

---

[1]Concurrent independent work of [17] also implies $\Omega(\mathrm{polylog}(1/\varepsilon)/\varepsilon^2)$-independence suffices. Other derandomizations of GW-rounding are known with better ε-dependence, though not solely via k-wise independence [23], [24], [31].

That is, the n-dimensional approximation problem is reduced to a 1-dimensional one. It turns out that this approach suffices for the case of halfspaces. In [11] the polynomial P is constructed using classical approximation theory tools. In [22] it is obtained by taking a truncated Taylor expansion of a certain smooth approximation to the sign function, constructed via a method dubbed "Fourier Transform mollification" (henceforth FT-mollification).

Let $f(x) = \text{sign}(p(x))$ be a regular 2-PTF. A first natural attempt to handle this case would be to again use some *univariate* polynomial approximation Q to the sign function – potentially allowing its degree to increase – and then take $q_u(x) = Q(p(x))$, as before. Such an approach turns out to fail for both constructions outlined above, for essentially the same reason (namely, that a degree–2 polynomial with variance 1 is only guaranteed to satisfy a tail bound of $\exp(-O(t))$ as opposed to the $\exp(-\Omega(t^2))$ of the degree-1 case). In fact, it is conjectured [14] that *no* univariate ℓ_1 ε-approximating polynomial for the sign function (i.e., without even requiring the sandwiching condition) can have degree $2^{o(1/\varepsilon^2)}$ (see Section 10.2 of [15] for related lower bounds).

We now describe FT-mollification and our departure from the univariate approach.

A. FT-mollification

FT-mollification is a general procedure to obtain a smooth function with bounded derivatives that approximates some bounded function f. The univariate version of the method in the context of derandomization was introduced in [22]. In this paper we refine the technique and generalize it to the multivariate setting, and later use it to prove our main theorem. We remark here that the FT-mollification construction given in the current work is not only a generalization of that in [22], but is redone from scratch and is simpler, while also yielding improved bounds even in univariate applications.

For the univariate case, where $f : \mathbb{R} \to \mathbb{R}$, [22] defined $\tilde{f}^c(x) = (c \cdot \hat{b}(c \cdot t) * f(t))(x)$ for a parameter c, where $\hat{b}$ has unit integral and is the Fourier transform of a smooth function b of compact support (a so-called *bump function*). Here "$*$" denotes convolution. The idea of smoothing functions via convolution with a smooth approximation of the Dirac delta function is old, dating back to "Friedrichs mollifiers" [13] in 1944. Indeed, the only difference between Friedrichs mollification and FT-mollification is that in the former, one convolves f with the scaled bump function, and not its Fourier transform. The switch to the Fourier transform is made to have better control on the high-order derivatives of

the resulting smooth function, which turns out to be crucial in making our proofs work.

The method can be illustrated as follows. Let $X = \sum_i a_i X_i$ for independent X_i. Suppose we would like to argue that $\mathbf{E}[f(X)] \approx_\varepsilon \mathbf{E}[f(Y)]$, where $Y = \sum_i a_i Y_i$ for k-wise independent Y_i's that are individually distributed as the X_i. Let $\tilde{f}^c$ be the FT-mollified version of f. If the parameter $c = c(\varepsilon)$ is appropriately selected, we can guarantee that $|f(x) - \tilde{f}^c(x)| < \varepsilon$ "almost everywhere", and furthermore have "good" upper bounds on the high-order derivatives of $\tilde{f}^c$. We could then hope to show the following chain of inequalities: $\mathbf{E}[f(X)] \approx_\varepsilon \mathbf{E}[\tilde{f}^c(X)] \approx_\varepsilon \mathbf{E}[\tilde{f}^c(Y)] \approx_\varepsilon \mathbf{E}[f(Y)]$. To justify the first inequality, f and $\tilde{f}^c$ are close almost everywhere, and so it suffices to argue that X is sufficiently anti-concentrated in the small region where they are not close. The second inequality would use Taylor's theorem, bounding the error via upper bounds on moment expectations of X and the high-order derivatives of $\tilde{f}^c$. Showing the final inequality would be similar to the first, except that one needs to justify that even under k-wise independence the distribution of Y is sufficiently anti-concentrated. The argument outlined above was used in [22] to provide an alternative proof that bounded independence fools regular halfspaces, and to optimally derandomize Indyk's moment estimation algorithm in data streams [20].

We now describe our switch to multivariate FT-mollification. Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be arbitrary, and let $S = f^{-1}(1) \subseteq \mathbb{R}^n$. Then, fooling $f(x)$ and fooling $I_S(x)$ are equivalent. A natural attempt to this end would be to generalize FT-mollification to n dimensions, then FT-mollify I_S and argue as above using the multivariate Taylor's theorem. Such an approach is perfectly valid, but as one might expect, there is a penalty for working over high dimensions. Both our quantitative bounds on the error introduced by FT-mollifying, and the error coming from the multivariate Taylor's theorem, increase with the dimension. Our approach is then to find a *low-dimensional representation* of such a region S which allows us to obtain the desired bounds. We elaborate below on how this can be accomplished in our setting.

B. Our Approach

Let $f = \text{sign}(p)$ be a regular multilinear degree-2 PTF with variance 1 (wlog). Let us assume for simplicity that p is a quadratic form; handling the additive linear form and constant is easier. Our approach is now as follows. We decompose p as $p_1 - p_2 + p_3$, where p_1, p_2 are positive semidefinite quadratic forms with no small nonzero eigenvalues and p_3 is indefinite with all

eigenvalues small in magnitude; such a decomposition follows from elementary linear algebra.

Then, as suggested by the aforementioned, we would like to identify a region $R \subseteq \mathbb{R}^d$ for small d such that $I_{\{z:p(z) \geq 0\}}(x)$ can be written as $I_R(F(x))$ for some $F : \{-1, 1\}^n \to \mathbb{R}^d$ that depends on the p_i, then FT-mollify I_R. The region R is selected as follows: note we can write $p_3(x) = x^T A_{p_3} x$, where A_{p_3} is a real symmetric matrix with trace Υ. We consider the region $R = \{z : z_1^2 - z_2^2 + z_3 + \Upsilon \geq 0\} \subseteq \mathbb{R}^3$ and define $F(x) = (\sqrt{p_1(x)}, \sqrt{p_2(x)}, p_3(x) - \Upsilon)$, then observe that $I_R(F(x)) = 1$ iff $p(x) \geq 0$. (Recall that p_1, p_2 are positive-semidefinite, hence the first two coordinates are always real.) We then prove via FT-mollification that $\mathbf{E}[I_R(F(x))]$ is preserved to within ε by bounded independence. Due to our choice of F, when applying Taylor's theorem our error grows only like $2^{O(k)} \cdot c^k \cdot (\mathbf{E}[\sqrt{p_1(x)}^k] + \mathbf{E}[\sqrt{p_2(x)}^k] + \mathbf{E}[(p_3(x) - \Upsilon)^k])/k^k$ for some (non-constant) c in our proof, and we want this error to be ε. Essentially, these square roots save us since kth moments of quadratic forms can grow like k^k, which would nullify the k^k in the denominator of Taylor's theorem; by having square roots, we only have to deal with $(k/2)$th moments. To handle p_3, we use the Hanson-Wright inequality for quadratic forms with small eigenvalues [18]. The fact that we need p_1, p_2 to not only be positive semidefinite, but to also have no small eigenvalues, is needed because quadratic forms with no small nonzero eigenvalues satisfy good tail bounds. This is relevant because $\tilde{I}_R^c(F(x))$ and $I_R(F(x))$ are *not* close for $F(x)$ near the boundary of R, and we can show that the probability of this event is small when p_1, p_2 satisfy good tail bounds.

IV. MULTIVARIATE FT-MOLLIFICATION

We now state and sketch the proof of our FT-mollification theorem, which yields generic smoothing guarantees for arbitrary bounded functions mapping $\mathbb{R}^d$ to $\mathbb{R}$. In our proof of Theorem I.1, we are concerned with $d = 4$. In some of the other applications of our technique mentioned in Section VII, d can be a growing parameter, e.g. the number of halfspaces when fooling intersections of halfspaces. In what follows, we refer to $\tilde{F}^c$ as the *FT-mollification* of F.

Theorem IV.1. Let $F : \mathbb{R}^d \to \mathbb{R}$ be bounded, $c > 0$ arbitrary. There exists $\tilde{F}^c : \mathbb{R}^d \to \mathbb{R}$ satisfying

  i. $\|\partial^\beta \tilde{F}^c\|_\infty \leq \|F\|_\infty \cdot (2c)^{|\beta|}$ for all $\beta \in \mathbb{N}^d$.
  ii. Fix some $x \in \mathbb{R}^d$. Then if $|F(x) - F(y)| \leq \varepsilon$ whenever $\|x - y\|_2 \leq \delta$ for some $\varepsilon, \delta \geq 0$, then $|\tilde{F}^c(x) - F(x)| \leq \varepsilon + \|F\|_\infty \cdot O(d^2/(c^2\delta^2))$.
  iii. $\tilde{F}^c$ is nonnegative if F is nonnegative.

Proof (Sketch). In our full version we show the existence of a probability density B on $\mathbb{R}^d$ satisfying $\mathbf{E}_{x \sim B}[\|x\|_2^2] = O(d^2)$, and $\|\partial^\beta B\|_1 \leq 2^{|\beta|}$ for all $\beta \in \mathbb{N}^d$. This density B is obtained by taking a "smooth enough" function $b : \mathbb{R}^d \to \mathbb{R}$ of compact support with $\int_{\mathbb{R}^d} b^2(y)dy = 1$, then letting B be the square of its Fourier transform. We then define $B_c(x) = c^d \cdot B(cx)$, and $\tilde{F}^c(x) = (B_c * F)(x) = \int_{\mathbb{R}^d} B_c(y)F(x - y)dy$.

For (i), using basic properties of convolution we show $\|\tilde{F}^c\|_\infty \leq \|F\|_\infty \cdot c^{|\beta|} \cdot \|\partial^\beta B\|_1$, at which point we use our bounds on $\|\partial^\beta B\|_1$. For (ii), since B is a probability density we have $\int_{\mathbb{R}^d} B_c(y)dy = 1$ for all c. Thus, $\int_{\mathbb{R}^d} B_c(x - y)F(y)dy = F(x) + \int_{\mathbb{R}^d}(F(y) - F(x))B_c(x - y)dy$. We then split the domain of integration into the regions $\|x - y\|_2 < \delta$ and $\|x - y\|_2 \geq \delta$. The integral over the first region is bounded by ε, and over the second region by the product of $\|F\|_\infty$ and a tail bound for B, which we can obtain by the second moment method since B has bounded variance. Item (iii) follows since for F nonnegative, $\tilde{F}^c$ is the convolution of two nonnegative functions. ∎

The following theorem is a corollary of Theorem IV.1 in the case F is the indicator function of a subset $R \subseteq \mathbb{R}^d$. In Theorem IV.2, and in later invocations of the theorem, we use the following notation: for $R \subset \mathbb{R}^d$, we let ∂R denote the boundary of R (specifically, ∂R denotes the set of points $x \in \mathbb{R}^d$ such that for every $\varepsilon > 0$, the ball about x of radius ε intersects both R and $\mathbb{R}^d \backslash R$).

Theorem IV.2. For any $R \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$, $|I_R(x) - \tilde{I}_R^c(x)| \leq \min\{1, O((\frac{d}{c \cdot d_2(x, \partial R)})^2)\}$.

V. WARMUP: FOOLING REGULAR HALFSPACES

As a warmup to our main result, we show how to use Theorem IV.2 to provide a simple proof that $\Omega(1/\varepsilon^2)$-wise independence fools the class of ε^2-regular halfspaces, i.e. halfspaces $\{x : \langle w, x \rangle \geq \theta\} \subseteq \{-1, 1\}^n$ where $|w_i| \leq \varepsilon$ for all i and $\|w\|_2 = 1$. This improves upon the bounds of [11], [22] by $\text{polylog}(1/\varepsilon)$ factors, and is optimal up to constant factors [11].

Theorem V.1. Let $H_{w,\theta} = \{x : \langle w, x \rangle \geq \theta\} \subseteq \{-1, 1\}^n$ such that $|w_i| \leq \varepsilon$ for all $i \in [n]$ with $\|w\|_2 = 1$, i.e. $H_{w,\theta}$ is ε^2-regular. Suppose $x_1, \ldots, x_n$ are independent Bernoulli, and $y_1, \ldots, y_n$ are k-wise independent Bernoulli for $k \geq C/\varepsilon^2$ for a sufficiently large even constant C. For $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, $|\mathbf{Pr}[x \in H_{w,\theta}] - \mathbf{Pr}[y \in H_{w,\theta}]| = O(\varepsilon)$.

Proof: Let $X = \langle w, x \rangle, Y = \langle w, y \rangle$. It is equivalent to show $|\mathbf{E}[I_{[\theta, \infty)}(X)] - \mathbf{E}[I_{[\theta, \infty)}(Y)]| = O(\varepsilon)$. We

show the following chain of inequalities for $c = 1/\varepsilon$:

$$\mathbf{E}[I_{[\theta,\infty)}(X)] \approx_{\varepsilon} \mathbf{E}[\tilde{I}^c_{[\theta,\infty)}(X)]$$
$$\approx_{\varepsilon} \mathbf{E}[\tilde{I}^c_{[\theta,\infty)}(Y)] \approx_{\varepsilon} \mathbf{E}[I_{\theta,\infty)}(Y)]$$

Here $\tilde{I}^c_{[\theta,\infty)}$ is as in Theorem IV.2, where $R = [\theta,\infty)$ and $d = 1$. Note then $d_2(z, \partial R)$ is just $|z - \theta|$.

(**A**) $\mathbf{E}[\mathbf{I}_{[\theta,\infty)}(\mathbf{X})] \approx_{\varepsilon} \mathbf{E}[\tilde{\mathbf{I}}^{\mathbf{c}}_{[\theta,\infty)}(\mathbf{X})]$:

$$|\mathbf{E}[I_{[\theta,\infty)}(X)] - \mathbf{E}[\tilde{I}^c_{[\theta,\infty)}(X)]|$$
$$\leq \mathbf{E}[|I_{[\theta,\infty)}(X) - \tilde{I}^c_{[\theta,\infty)}(X)|]$$
$$\leq \mathbf{Pr}[|X - \theta| < \varepsilon]$$
$$+ \sum_{s=0}^{\infty} \mathbf{Pr}[2^s\varepsilon \leq |X - \theta| < 2^{s+1}\varepsilon]$$
$$\times O(c^{-2}2^{-2s}\varepsilon^{-2})$$
$$\leq O(\varepsilon) + \sum_{s=0}^{\infty} 2^{-2s} \cdot \mathbf{Pr}[|X - \theta| < 2^{s+1}\varepsilon]$$
$$= O(\varepsilon)$$

since $\mathbf{Pr}[|X - \theta| \leq t] = O(t + \varepsilon)$ for any $t > 0$, by ε^2-regularity and the Berry-Esséen Theorem.

(**B**) $\mathbf{E}[\tilde{\mathbf{I}}^{\mathbf{c}}_{[\theta,\infty)}(\mathbf{X})] \approx_{\varepsilon} \mathbf{E}[\tilde{\mathbf{I}}^{\mathbf{c}}_{[\theta,\infty)}(\mathbf{Y})]$: By Taylor's theorem, $\tilde{I}^c_{[\theta,\infty)}(z) = P_{k-1}(z) \pm \|(\tilde{I}^c_{[\theta,\infty)})^{(k)}\|_{\infty} \cdot |z|^k/k!$ for $z \in \mathbb{R}$ and $f^{(k)}$ being the kth derivative of f, where P_{k-1} is a degree-$(k - 1)$ polynomial. By k-wise independence, $\mathbf{E}[P_{k-1}(X)] = \mathbf{E}[P_{k-1}(Y)]$ and $\mathbf{E}[X^k] = \mathbf{E}[Y^k]$. For k even, $|z|^k = z^k$. Hence,

$$|\mathbf{E}[\tilde{I}^c_{[\theta,\infty)}(X)] - \mathbf{E}[\tilde{I}^c_{[\theta,\infty)}(Y)]|$$
$$\leq 2 \cdot \frac{\|(\tilde{I}^c_{[\theta,\infty)})^{(k)}\|_{\infty} \cdot \mathbf{E}[X^k]}{k!} \leq 2^{O(k)} \cdot \frac{c^k \cdot k^{k/2}}{k^k},$$

which is $O(\varepsilon)$ since $k = \Omega(c^2)$. The last inequality used Theorem IV.2 to bound $\|(\tilde{I}^c_{[\theta,\infty)})^{(k)}\|_{\infty}$, and Khintchine's inequality gives $\mathbf{E}[X^k] \leq k^{k/2}$.

(**C**) $\mathbf{E}[\mathbf{I}_{[\theta,\infty)}(\mathbf{Y})] \approx_{\varepsilon} \mathbf{E}[\tilde{\mathbf{I}}^{\mathbf{c}}_{[\theta,\infty)}(\mathbf{Y})]$: This is argued identically as in the first inequality, but we now must show that even under $\Omega(1/\varepsilon^2)$-wise independence we still have $\mathbf{Pr}[|Y - \theta| \leq \varepsilon] = O(\varepsilon)$. Suppose we had a function $f : \mathbb{R} \to \mathbb{R}$ such that (1) $f \geq I_{[\theta-\varepsilon,\theta+\varepsilon]}$ on $\mathbb{R}$ (implying for example $\mathbf{E}[f(Y)] \geq \mathbf{E}[I_{[\theta-\varepsilon,\theta+\varepsilon]}(Y)]$), (2) $\mathbf{E}[f(X)] = O(\varepsilon)$, and (3) $\|f^{(\ell)}\|_{\infty} \leq O(1/\varepsilon)^{\ell}$ for all $\ell \geq 0$. Given (2) and (3), we can apply Taylor's theorem just as above to show $|\mathbf{E}[f(X)] - \mathbf{E}[f(Y)]| = O(\varepsilon)$, i.e. $\mathbf{E}[f(Y)] = O(\varepsilon)$. Using (1) then gives our desired upper bound on $\mathbf{E}[I_{[\theta-\varepsilon,\theta+\varepsilon]}(Y)] = \mathbf{Pr}[|Y - \theta| \leq \varepsilon]$.

It only remains to exhibit such an f: we take $f = 2 \cdot \tilde{I}^{c'}_{[\theta-2\varepsilon,\theta+2\varepsilon]}$ for c' a sufficiently large constant times $1/\varepsilon$. For (1), if $x \notin [\theta - \varepsilon, \theta + \varepsilon]$ then

$I_{[\theta-\varepsilon,\theta+\varepsilon]} = 0$, whereas $f \geq 0$. If $x \in [\theta - \varepsilon, \theta + \varepsilon]$, then $\min\{x - (\theta - 2\varepsilon), x - (\theta + 2\varepsilon)\} \geq \varepsilon$, implying $f(x) \geq 1$ by Theorem IV.2, choice of c', and the fact that $I_{[\theta-2\varepsilon,\theta+2\varepsilon]}(x) = 1$. Item (2) follows by applying (A) above; (3) follows from (i) of Theorem IV.1 and Taylor's theorem (as in (B) above). ∎

The proof structure of Theorem V.1 is similar to that in [22]. In particular, both use the same chain of inequalities. However, due to differences in the FT-mollification guarantees of [22], the proof there gave a worse bound on k by a $\mathrm{polylog}(1/\varepsilon)$ factor. The main reason for this is that the FT-mollification construction of [22] gave an $\tilde{I}^c_{[a,b]}$ approximating $I_{[a,b]}$ such that the guarantee was only that the two functions were within ε for x "far" from $\{a, b\}$, and differed by at most a constant for x "close" to the boundary. Meanwhile, in our current FT-mollification construction, the quality of $\tilde{I}^c_{[a,b]}$ gracefully degrades as x approaches the boundary. Furthermore, the proof of (C) given here is arguably more intuitive than the argument in [22], which relied on some complex analysis.

One consequence of Theorem V.1 is that the Berry-Esséen theorem is derandomized by $\Omega(1/\varepsilon^2)$-independence, which is asymptotically optimal [11]. Specifically, Theorem V.1 implies, after also carrying out the same argument under the Gaussian measure, that $\sup_{t\in\mathbb{R}} |\mathbf{Pr}[\langle w, x \rangle \leq t] - \mathbf{Pr}[\langle w, g \rangle \leq t]| \leq \varepsilon$ as long as the x_i and g_i are each $\Omega(1/\varepsilon^2)$-wise independent and $\|w\|_{\infty} \leq \varepsilon$, where the x_i are Bernoulli and the g_i are Gaussian. The original Berry-Esséen theorem required independent x_i and g_i, and [11], [22] required $\mathrm{polylog}(1/\varepsilon)/\varepsilon^2$-wise independence.

VI. Proof of Theorem I.1

We now give our proof of Theorem I.1. In Section VI-A we analyze the regular case of our main theorem, and Section VI-B reduces the general case to the regular case.

A. Fooling regular degree-2 threshold functions

In this section we show the following.

Theorem VI.1. Let $0 < \varepsilon < 1$ be given. Let $X_1, \ldots, X_n$ be independent Bernoulli and $Y_1, \ldots, Y_n$ be $2k$-wise independent Bernoulli for k a sufficiently large multiple of $1/\varepsilon^8$. If p is multilinear and of degree 2 with $\sum_{|S|>0} \hat{p}_S^2 = 1$, and $\mathrm{Inf}_i(p) \leq \tau$ for all i, then

$$\mathbf{E}[\mathrm{sign}(p(X))] - \mathbf{E}[\mathrm{sign}(p(Y))] = O(\varepsilon + \tau^{1/9}).$$

Throughout this section, p always refers to the polynomial of Theorem VI.1, and τ refers to the maximum influence of any variable in p. Observe p (over the

5

hypercube) can be written as $q + p_4 + C$, where q is a multilinear quadratic form, p_4 is a linear form, and C is a constant. For a quadratic form q, we can write a real symmetric matrix A_q such that $q(x) = x^T A_q x$, where x^T denotes the transpose of x. Since we can assume the sum of squared coefficients in p (ignoring C) is 1, this implies $\|A_q\|_F \leq 1/2$ and $\sum_S \widehat{p_4}_S^2 \leq 1$. Using the spectral theorem for real symmetric matrices, we write $p = p_1 - p_2 + p_3 + p_4 + C$ where p_1, p_2, p_3 are quadratic forms satisfying $\lambda_{\min}(A_{p_1}), \lambda_{\min}(A_{p_2}) \geq \delta$, $\|A_{p_3}\|_2 < \delta$, and $\|A_{p_i}\|_F \leq 1/2$ for $1 \leq i \leq 3$, and also with p_1, p_2 positive semidefinite. Such a decomposition follows by writing $A_q = Q^T \Lambda_q Q$ for some diagonal matrix Λ_q and orthogonal Q, then writing $\Lambda_q = \Lambda_{p_1} - \Lambda_{p_2} + \Lambda_{p_3}$, where Λ_{p_1} contains the eigenvalues of Λ_q above δ, Λ_{p_2} contains the negation of those below $-\delta$, and Λ_{p_3} contains the remaining eigenvalues. Then, set $A_{p_i} = Q\Lambda_{p_i}Q^T$. Throughout this section we let $p_1, \ldots, p_4, C, \delta$ be as discussed here. We use Υ to denote $\operatorname{tr}(A_{p_3})$. The value δ will be set later in the proof of Theorem VI.1.

It will be convenient to define the map $M_p : \mathbb{R}^n \to \mathbb{R}^4$ for $M_p(x) = (\sqrt{p_1(x)}, \sqrt{p_2(x)}, p_3(x) - \Upsilon, p_4(x))$. Note the the first two coordinates of $M_p(x)$ are indeed real since p_1, p_2 are positive semidefinite. To show Theorem VI.1, we follow the template of Section V, by showing that $\mathbf{E}[I_R(M_p(X))]$ is determined by k-wise independence for $R = \{z : z_1^2 - z_2^2 + z_3 + z_4 + C + \Upsilon \geq 0\} \subset \mathbb{R}^4$ (note $I_R(M_p(x))$ iff $p(x) \geq 1$).

Before giving the proof of Theorem VI.1, we first state Lemma VI.3, which says that for $F : \mathbb{R}^4 \to \mathbb{R}$, $F(M_p(x))$ is fooled by bounded independence as long as F is even in x_1, x_2 and certain technical conditions are satisfied.

The proof of Lemma VI.3 crucially uses the following moment bound for quadratic forms:

Theorem VI.2 (Hanson-Wright inequality [18]). Let $A \in \mathbb{R}^{n \times n}$ be symmetric and $x \in \{-1, 1\}^n$ be random. Then for all $k \geq 2$, $\mathbf{E}[|(x^T A x) - \operatorname{tr}(A)|^k] \leq C^k \cdot \max\{\sqrt{k}\|A\|_F, k\|A\|_2\}^k$ for C an absolute constant.

We note that while [18] give a tail bound, the above moment bound can be easily derived via integration. In the full version of our paper, we also provide a new proof of Theorem VI.2.

Lemma VI.3. Let $\varepsilon > 0$ be arbitrary. Let $F : \mathbb{R}^4 \to \mathbb{R}$ be even in each of its first two arguments such that $\|\partial^\beta \tilde{F}^c\|_\infty = O(\alpha^{|\beta|})$ for all multi-indices $\beta \in \mathbb{N}^4$ and some $\alpha > 1$. Suppose $1/\delta \geq B\alpha$ for a sufficiently large constant B. Let $X_1, \ldots, X_n$ be independent Bernoulli, and $Y_1, \ldots, Y_n$ be k'-independent Bernoulli for $k' = 2k$ with $k \geq \max\{\log(1/\varepsilon), B\alpha/\sqrt{\delta}, B\alpha^2\}$ an even inte-

ger. Write $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$. Then $|\mathbf{E}[F(M_p(X))] - \mathbf{E}[F(M_p(Y))]| < \varepsilon$.

Proof: We Taylor-expand F to obtain a polynomial P_{k-1} containing all monomials up to degree $k-1$. Since $F(x)$ is even in x_1, x_2, we can assume P_{k-1} is a polynomial in x_1^2, x_2^2, x_3, x_4. Let $x \in \mathbb{R}^4$ be arbitrary. We apply Taylor's theorem to bound $R(x) = |F(x) - P_{k-1}(x)|$. Define $x_* = \max_i\{|x_i|\}$. Then

$$
\begin{aligned}
R(x) &\leq \alpha^k \cdot \sum_{|\beta|=k} \frac{|x_1|^{\beta_1} \cdot |x_2|^{\beta_2} \cdot |x_3|^{\beta_3} \cdot |x_4|^{\beta_4}}{\beta_1! \cdot \beta_2! \cdot \beta_3! \cdot \beta_4!} \\
&\leq \alpha^k x_*^k \cdot \sum_{|\beta|=k} \frac{1}{\beta_1! \cdot \beta_2! \cdot \beta_3! \cdot \beta_4!} \\
&= \alpha^k x_*^k \cdot \frac{1}{k!} \cdot \sum_{|\beta|=k} \binom{k}{\beta_1, \ldots, \beta_4} \\
&\leq \alpha^k 4^k \cdot \frac{x_1^k + x_2^k + x_3^k + x_4^k}{k!}, \qquad \text{(VI.1)}
\end{aligned}
$$

with the absolute values unnecessary in the last inequality since k is even. We now observe

$$
\begin{aligned}
&|\mathbf{E}[F(M_p(X))] - \mathbf{E}[F(M_p(Y))]| \\
&\leq \alpha^k 2^{O(k)} k^{-k} \cdot (\mathbf{E}[(p_1(X))^{k/2}] + \mathbf{E}[(p_2(X))^{k/2}] \\
&\quad + \mathbf{E}[(p_3(X) - \Upsilon)^k] + \mathbf{E}[(p_4(X))^k])
\end{aligned}
$$

since (a) every term in $P_{k-1}(M_p(X))$ is a monomial of degree at most $2k-2$ in the X_i, by evenness of P_{k-1} in x_1, x_2, and is thus determined by $2k$-independence, (b) $\sqrt{p_1(X)}, \sqrt{p_2(X)}$ are real by positive semidefiniteness of p_1, p_2 (note that we are only given that the high order partial derivatives are bounded by $O(\alpha^k)$ on the reals; we have no guarantees for complex arguments), and (c) the moment expectations above are equal for X and Y since they are determined by $2k$-independence.

We now bound the error term above. We have

$$
\mathbf{E}[(p_1(X))^{k/2}] = 2^{O(k)}(k^{k/2} + \delta^{-k/2})
$$

by Lemma A.1 and Lemma A.3, with the same bound holding for $\mathbf{E}[(p_2(X))^{k/2}]$. We also have

$$
\mathbf{E}[(p_3(X) - \Upsilon)^k] \leq 2^{O(k)} \cdot \max\left\{\sqrt{k}, (\delta k)\right\}^k
$$

by Theorem VI.2. We finally have $\mathbf{E}[(p_4(X))^k] \leq k^{k/2}$ by Khintchine's inequality. Thus in total,

$$
\begin{aligned}
&|\mathbf{E}[F(M_p(X))] - \mathbf{E}[F(M_p(Y))]| \\
&\leq 2^{O(k)} \cdot ((\alpha/\sqrt{k})^k + (\alpha/(k\sqrt{\delta}))^k + (\alpha\delta)^k),
\end{aligned}
$$

which is at most ε for sufficiently large B by our lower bounds on k and $1/\delta$. ∎

In proving Theorem VI.1, we will need a lemma which states that p is anticoncentrated even when evaluated on Bernoulli random variables which are k-wise independent. We show this in Lemma VI.5, whose proof invokes Lemma VI.4. We defer the proof of Lemma VI.4 to the full version, which relies on the invariance principle [27], Gaussian anticoncentration [8], and Theorem A.2.

Lemma VI.4. Let $\eta, \eta' \geq 0, t \in \mathbb{R}$ be given, and let $X_1, \ldots, X_n$ be independent Bernoulli. Then

$$\mathbf{Pr}[|p(X) - t| \leq \eta \cdot (\sqrt{p_1(X)} + \sqrt{p_2(X)} + 1) + \eta']$$
$$= O(\sqrt{\eta'} + (\eta^2/\delta)^{1/4} + \tau^{1/9} + \exp(-\Omega(1/\delta))).$$

Lemma VI.5. For $\varepsilon' > 0$, let $k \geq D/(\varepsilon')^4$ for a universal constant $D > 0$. Let $Y_1, \ldots, Y_n$ be k-wise independent Bernoulli, and let $t \in \mathbb{R}$ be arbitrary. Then $\mathbf{Pr}[|p(Y) - t| < \varepsilon'] \leq O(\sqrt{\varepsilon'} + \tau^{1/9})$.

Proof: The proof in spirit works similarly to step (C) in the proof of Theorem V.1. We define the region $T_{t,\varepsilon'} = \{z : |z_1^2 - z_2^2 + z_3 + z_4 + C + \Upsilon - t| < \varepsilon'\} \subset \mathbb{R}^4$ and note $\mathbf{Pr}[|p(Y) - t| < \varepsilon'] = \mathbf{E}[I_{T_{t,\varepsilon'}}(M_p(Y))]$. Then, just as when proving Theorem V.1, we would like a smooth function f which upper bounds $I_{T_{t,\varepsilon'}}$ and has small expectation under full independence, so that we may apply Taylor's theorem (specifically, Lemma VI.3) to show that its expectation is also small under bounded independence. To accomplish this, we define the region $S_{\rho,t,\varepsilon'} = \{z : d_2(z, T_{t,\varepsilon'}) \leq \rho\}$ then take f to be $2 \cdot \tilde{I}^c_{S_{\rho,t,\varepsilon'}}$ for some $\rho > 0$ and $c = \Omega(1/\rho)$.

Noting $\mathbf{Pr}[|p(Z) - t| < \varepsilon'] = \mathbf{E}[I_{T_{t,\varepsilon'}}(M_p(Z))]$ for any random variable $Z = (Z_1, \ldots, Z_n)$,

$$\mathbf{Pr}[|p(Z) - t| \leq \varepsilon'] \leq 2 \cdot \mathbf{E}[\tilde{I}^c_{S_{\rho,t,\varepsilon'}}(M_p(Z))]. \quad \text{(VI.2)}$$

We now proceed in two steps. We first show $\mathbf{E}[\tilde{I}^c_{S_{\rho,t,\varepsilon'}}(M_p(X))] = O(\sqrt{\varepsilon'} + \tau^{1/9})$, then show $\mathbf{E}[\tilde{I}^c_{S_{\rho,t,\varepsilon'}}(M_p(Y))] \approx_{\varepsilon'} \mathbf{E}[\tilde{I}^c_{S_{\rho,t,\varepsilon'}}(M_p(X))]$ by applying Lemma VI.3, then conclude via Eq. (VI.2).

$\mathbf{E}[\tilde{\mathbf{I}}^{\mathbf{c}}_{\mathbf{S}_{\rho,t,\varepsilon'}}(\mathbf{M_p}(\mathbf{X}))] = \mathbf{O}(\sqrt{\varepsilon'} + \tau^{1/9})$: For $x \notin T_{t,\varepsilon'}$,

$$d_2(x, T_{t,\varepsilon'})$$
$$\geq \frac{1}{2} \cdot \min\left\{ \frac{|x_1^2 - x_2^2 + x_3 + x_4 + C + \Upsilon - t| - \varepsilon'}{2(|x_1| + |x_2| + 1)}, \right.$$
$$\left. \sqrt{|x_1^2 - x_2^2 + x_3 + x_4 + C + \Upsilon - t| - \varepsilon'} \right\}.$$

This is because by adding a vector v to x, we can change each individual coordinate of x by at most $\|v\|_2$, and can thus change the value of $|x_1^2 - x_2^2 + x_3 + x_4 + C + \Upsilon - t| - \varepsilon'$ by at most $2\|v\|_2 \cdot (|x_1| + |x_2| + 1) + \|v\|_2^2$.

Now let $X \in \{-1, 1\}^n$ be uniformly random. We thus have that, for any particular $w > 0$,

$$\mathbf{Pr}[0 < d_2(M_p(X), T_{t,\varepsilon'}) \leq w]$$
$$\leq \mathbf{Pr}\left[\min\left\{ \frac{|p(X) - t| - \varepsilon'}{2(\sqrt{p_1(X)} + \sqrt{p_2(X)} + 1)}, \right. \right.$$
$$\left. \left. \sqrt{|p(X) - t| - \varepsilon'} \right\} \leq 2w \right]$$
$$\leq \mathbf{Pr}[|p(X) - t| \leq 4w \cdot (\sqrt{p_1}(X) + \sqrt{p_2(X)} + 1) + \varepsilon']$$
$$+ \mathbf{Pr}[|p(X) - t| \leq 4w^2 + \varepsilon']$$
$$= O(\sqrt{\varepsilon'} + w + \sqrt{w} + (w^2/\delta)^{1/4} + \tau^{1/9} + \exp(-\Omega(1/\delta)))$$
$$\text{(VI.3)}$$

with the last inequality holding by Lemma VI.4.

By Theorem IV.1 and our setting of c, $\tilde{I}^c_{S_{\rho,t,\varepsilon'}}(x) = \max\{1, O((c \cdot d_2(x, T_{t,\varepsilon'}))^{-2})\}$ when $d_2(x, T_{t,\varepsilon'}) \geq 2\rho$. Then,

$$\mathbf{E}[\tilde{I}^c_{S_{\rho,t,\varepsilon'}}(M_p(X))]$$
$$\leq \mathbf{Pr}[d_2(M_p(X), T_{t,\varepsilon'}) \leq 2\rho]$$
$$+ O\left(\sum_{s=1}^{\infty} 2^{-2s} \cdot \mathbf{Pr}[d_2(M_p(X), T_{t,\varepsilon'}) \leq 2^{s+1}\rho] \right)$$
$$\leq O(\sqrt{\varepsilon'} + \sqrt{\rho} + (\rho^2/\delta)^{1/4} + \tau^{1/9} + \exp(-\Omega(1/\delta))$$
$$+ O\left(\sum_{s=1}^{\infty} 2^{-2s} \cdot (\sqrt{\varepsilon'} + 2^{s+1}\rho + \sqrt{2^{s+1}\rho} \right.$$
$$\left. + (2^{2s+2}\rho^2/\delta)^{1/4} + \tau^{1/9} + \exp(-\Omega(1/\delta))) \right)$$
$$= O(\sqrt{\varepsilon'} + \sqrt{\rho} + (\rho^2/\delta)^{1/4} + \tau^{1/9} + \exp(-\Omega(1/\delta))$$
$$\text{(VI.4)}$$

We now make the settings $\rho = (\varepsilon')^2, 1/\delta = 2Bc$, where $B > 1$ is the sufficiently large constant in Lemma VI.3. Thus Eq. (VI.4) is now $O(\sqrt{\varepsilon'} + \tau^{1/9})$. (We remark that a different δ is used when proving Theorem VI.1.)

$\mathbf{E}[\tilde{\mathbf{I}}^{\mathbf{c}}_{\mathbf{S}_{\rho,t,\varepsilon'}}(\mathbf{M_p}(\mathbf{Y}))] \approx_{\varepsilon'} \mathbf{E}[\tilde{\mathbf{I}}^{\mathbf{c}}_{\mathbf{S}_{\rho,t,\varepsilon'}}(\mathbf{M_p}(\mathbf{X}))]$: We remark that $\tilde{I}^c_{S_{\rho,t,\varepsilon'}}$ can be assumed to be even in both x_1, x_2. If not, then consider the symmetrization

$$(1/4) \cdot (\tilde{I}^c_{S_{\rho,t,\varepsilon'}}(x_1, x_2, x_3, x_4) + \tilde{I}^c_{S_{\rho,t,\varepsilon'}}(-x_1, x_2, x_3, x_4)$$
$$+ \tilde{I}^c_{S_{\rho,t,\varepsilon'}}(x_1, -x_2, x_3, x_4) + \tilde{I}^c_{S_{\rho,t,\varepsilon'}}(-x_1, -x_2, x_3, x_4)).$$
$$\text{(VI.5)}$$

The inequality follows by Lemma VI.3, given our choice of k, δ. This completes our proof by applying Eq. (VI.2) with $Z = Y$. ∎

The following lemma follows from Lemma VI.4 and Lemma VI.5. The proof is in the full version.

Lemma VI.6. Let $\eta, \eta' \geq 0$ be given, and let $Y_1, \ldots, Y_n$ be k-independent Bernoulli for k as in Lemma VI.5 with $\varepsilon' = \min\{\eta/\sqrt{\delta}, \eta'\}$. Also assume $k \geq \lceil 2/\delta \rceil$. Then

$$\mathbf{Pr}[|p(X) - t| \leq \eta \cdot (\sqrt{p_1(X)} + \sqrt{p_2(X)} + 1) + \eta']$$
$$= O(\sqrt{\eta'} + (\eta^2/\delta)^{1/4} + \tau^{1/9} + \exp(-\Omega(1/\delta))).$$

We are now ready to prove the main theorem of this section.

Proof (of Theorem VI.1). Consider the region $R \subset \mathbb{R}^4$ defined by $R = \{z : z_1^2 - z_2^2 + z_3 + z_4 + C + \Upsilon \geq 0\}$. Then note that $I_{[0,\infty)}(p(x)) = 1$ if and only if $I_R(M_p(x)) = 1$. It thus suffices to show that I_R is fooled in expectation by bounded independence.

We set $\rho = \varepsilon^4$, $c = 1/\rho$, and $1/\delta = 2Bc$ for B the constant in the statement of Lemma VI.3. We now show a chain of inequalities to give our theorem:

$$\mathbf{E}[I_R(M_p(X))] \approx_{\varepsilon + \tau^{1/9}} \mathbf{E}[\tilde{I}_R^c(M_p(X))]$$
$$\approx_\varepsilon \mathbf{E}[\tilde{I}_R^c(M_p(Y))] \approx_{\varepsilon + \tau^{1/9}} \mathbf{E}[I_R(M_p(Y))].$$

$\mathbf{E}[\mathbf{I_R}(\mathbf{M_p}(\mathbf{X}))] \approx_{\varepsilon + \tau^{1/9}} \mathbf{E}[\mathbf{\tilde{I}_R^c}(\mathbf{M_p}(\mathbf{X}))]$: We have

$$\mathbf{Pr}[d_2(M_p(X), \partial R) \leq w]$$
$$\leq \mathbf{Pr}[|p(X)| \leq 4w \cdot (\sqrt{p_1(X)} + \sqrt{p_2(X)} + 1)]$$
$$+ \mathbf{Pr}[|p(X)| \leq 4w^2]$$
$$= O(w + \sqrt{w} + (w^2/\delta)^{1/4} + \tau^{1/9} + \exp(-\Omega(1/\delta))),$$

as in Eq. (VI.3). Then,

$$|\mathbf{E}[I_R(M_p(X))] - \mathbf{E}[\tilde{I}_R^c(M_p(X))]|$$
$$\leq \mathbf{E}[|I_R(M_p(X))] - \tilde{I}_R^c(M_p(X))|]$$
$$\leq \mathbf{Pr}[d_2(M_p(X), \partial R) \leq 2\rho]$$
$$+ O\left(\sum_{s=1}^{\infty} 2^{-2s} \cdot \mathbf{Pr}[d_2(M_p(X), \partial R) \leq 2^{s+1}\rho]\right)$$
$$\leq O(\sqrt{\rho} + (\rho^2/\delta)^{1/4} + \tau^{1/9} + \exp(-\Omega(1/\delta))$$
$$+ O\left(\sum_{s=1}^{\infty} 2^{-2s} \cdot (\sqrt{2^{s+1}\rho} + (2^{2s+2}\rho^2/\delta)^{1/4} + \tau^{1/9}\right.$$
$$\left. + \exp(-\Omega(1/\delta)))\right) = O(\varepsilon + \tau^{1/9})$$

by choice of ρ, δ, and Theorem IV.2 and Lemma VI.6.

$\mathbf{E}[\mathbf{\tilde{I}_R^c}(\mathbf{M_p}(\mathbf{X}))] \approx_\varepsilon \mathbf{E}[\mathbf{\tilde{I}_R^c}(\mathbf{M_p}(\mathbf{Y}))]$: We can assume $\tilde{I}_R^c$ is even in x_1, x_2 as in Eq. (VI.5). We apply

Lemma VI.3 with $\alpha = 2c$, noting that $1/\delta = B\alpha$ and that our setting of k is sufficiently large.

$\mathbf{E}[\mathbf{\tilde{I}_R^c}(\mathbf{M_p}(\mathbf{Y}))] \approx_{\varepsilon + \tau^{1/9}} \mathbf{E}[\mathbf{I_R}(\mathbf{M_p}(\mathbf{Y}))]$: The argument is identical as with the first inequality. We remark that we do have sufficient independence to apply Lemma VI.6 since, mimicking our analysis of the first inequality, we have

$$\mathbf{Pr}[|p(Y)| \leq 4\rho \cdot (\sqrt{p_1(Y)} + \sqrt{p_2(Y)} + 1)]$$
$$+ \mathbf{Pr}[|p(Y)| \leq 4\rho^2]$$
$$\leq \mathbf{Pr}[|p(Y)| \leq 4\rho \cdot (\sqrt{p_1(Y)} + \sqrt{p_2(Y)} + 1)]$$
$$+ \mathbf{Pr}[|p(Y)| \leq \varepsilon^2] \qquad \text{(VI.6)}$$

since $\rho^2 = o(\varepsilon^2)$. We can apply Lemma VI.6 to Eq. (VI.6) since $k \geq \lceil 2/\delta \rceil$ and $k = \Omega(1/(\varepsilon'')^4)$ for $\varepsilon'' = \min\{\rho/\sqrt{\delta}, \varepsilon^2\} = \varepsilon^2$. Thus, Eq. (VI.6) is $O(\varepsilon + \tau^{1/9})$.

∎

Our main theorem of this Section (Theorem VI.1) also holds under the case that the X_i, Y_i are standard normal, and without any error term depending on τ. See our full version for a proof.

B. Reduction to the regular case

In this section, we complete the proof of Theorem I.1. We accomplish this by providing a reduction from the general case to the regular case. In fact, such a reduction can be shown to hold for any degree $d \geq 1$ and establishes the following:

Theorem VI.7. Suppose K-wise independence ε-fools the class of τ-regular d-PTFs, for some parameter $0 < \tau \leq \varepsilon$. Then $(K + L)$-wise independence ε-fools all d-PTFs, where $L = L(d, \tau) = (1/\tau) \cdot (d \log(1/\tau))^{O(d)}$.

Noting that τ-regularity implies that the maximum influence of any particular variable is at most $d \cdot \tau$, Theorem VI.1 yields that 2-PTFs that are τ-regular, for $\tau = O(\varepsilon^9)$, are ε-fooled by $\Omega(\varepsilon^{-8})$-wise independence. By plugging in $\tau = O(\varepsilon^9)$ in the above theorem we obtain Theorem I.1. The proof of Theorem VI.7 is obtained by a simple adaptation of the regularity lemma in [12][2].

Proof (Sketch). (of Theorem VI.7). Any boolean function f on $\{-1, 1\}^n$ can be expressed as a binary decision tree where each internal node is labeled by a variable, every root-to-leaf path corresponds to a restriction ρ that fixes the variables as they are set on the path, and every leaf is labeled with the restricted

---

[2]We note that [25] prove a very similar regularity lemma to obtain their PRGs for d- PTFs. One could alternatively use this instead of [12]. For $d = 2$ this would give a worse bound of $\tilde{\Omega}(\varepsilon^{-18})$.

subfunction f_ρ. The main claim is that, if f is a d-PTF, then it has such a decision-tree representation with certain strong properties. In particular, by [12], an arbitrary d-PTF $f = \mathrm{sign}(p)$ can be represented as a decision tree $\mathcal{T}$ of depth $L(d, \tau)$, so that with probability $1-\tau$ over the choice of a uniformly random root-to-leaf path ρ, the restricted subfunction (leaf) $f_\rho = \mathrm{sign}(p_\rho)$ is either a τ-regular d-PTF or is τ-close to a constant function.

Our proof of Theorem VI.7 is based on the above structural lemma. Under the uniform distribution, there is some particular distribution on the leaves (the tree is not of uniform height); then conditioned on the restricted variables the variables still undetermined at the leaf are still uniform. With $(K + L)$-wise independence, a random walk down the tree arrives at each leaf with the same probability as in the uniform case (since the depth of the tree is at most L). Hence, the probability mass of the "bad" leaves is at most $\tau \leq \varepsilon$ even under bounded independence. Furthermore, the induced distribution on each leaf (over the unrestricted variables) is K-wise independent. Consider a good leaf. Either the leaf is τ-regular, in which case we can apply Theorem VI.1, or it is τ-close to a constant function. At this point though we arrive at a technical issue. The statement and proof in [12] concerning "close-to-constant" leaves holds only under the uniform distribution, though we observe that a simple modification of their proof (in particular, Lemmas 3 and 5 in [12]) shows that the statement holds even under $O(d \cdot \log(1/\tau))$-wise independence; see the full version for details. ∎

VII. Other Applications

We briefly sketch several other applications of our techniques here; details are in the full version. Our approach implies that $\mathrm{poly}(m)/\varepsilon^2$-wise independence ε-fools intersections of m halfspaces under the Gaussian measure. If the halfspaces are $H_i = \{x : \langle a_i, x \rangle \geq \theta_i\}$, then one simply needs to fool $I_R(F(x))$ for $R = \{z : z_i \geq \theta_i\} \subset \mathbb{R}^m$, and $F(x) = (\langle a_i, x \rangle, \ldots, \langle a_m, x \rangle)$. This is carried out as in the proof of Theorem V.1, but using a union bound to bound the probability of $F(x)$ being near ∂R, and using the multivariate Taylor's theorem. Note this implies that the randomized hyperplane rounding scheme of Goemans and Williamson [16] only requires that the coefficient vector defining the hyperplane need only have $\Omega(1/\varepsilon^2)$-wise independent entries. Also, one can generalize our proof in Section VI-A to show that $\mathrm{poly}(m/\varepsilon)$-wise independence fools the intersection of m degree-2 threshold functions.

Our new FT-mollification construction, which refines that of [22], also improves a bound given in [22].

Namely, plugging our construction into their argument shows that $\Omega(1/\varepsilon^p)$-wise independence suffices to fool Indyk's median estimator for moment estimation in data streams, improving their bound by $\mathrm{polylog}(1/\varepsilon)$ factors.

Our FT-mollification also recovers a generalization due to [28] of Jackson's theorem in approximation theory to the higher-dimensional unit ℓ_2 ball.

Theorem VII.1 ([28]). For $F : \mathbb{R}^m \to \mathbb{R}$ define
$$\omega(F, \delta) = \sup_{\substack{\|x\|_2, \|y\|_2 \leq 1 \\ \|x-y\|_2 \leq \delta}} |F(x) - F(y)|.$$
For any $k \geq 1$ there exists a polynomial p_k of degree k with $\sup_{\|x\|_2 \leq 1} |F(x) - p_k(x)| = O(\omega(F, m/k))$.

Our proof is simple: to obtain p_k, FT-mollify F then Taylor-expand to degree k; details are in the full version.

Finally, FT-mollification followed by Taylor expansion shows that there exists a degree-k polynomial p_k for $k = O(1/\varepsilon^2)$ that ε-approximates in ℓ_1 the sign function under the Gaussian distribution on the real line, i.e. such that $\mathbf{E}_{x \sim N(0,1)}[|\mathrm{sign}(x) - p_k(x)|] \leq \varepsilon$.

Using the framework of [21] the aforementioned implies that halfspaces can be agnostically learned with error ε under the Gaussian distribution in time $\mathrm{poly}(n^k/\varepsilon)$, improving the previously best known achievable k [11] by a $\log^2(1/\varepsilon)$ factor.

VIII. Conclusions

By a probabilistic argument, there exist generators with seed-length $O(d \log n + \log(1/\varepsilon))$ for degree-d PTFs (see [25]). Hence, there is still a substantial gap between probabilistic and explicit constructions, and resolving this issue even for $d = 1$ remains an open problem.

References

[1] N. Alon, L. Babai, and A. Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms*, 7(4):567–583, 1986.

[2] L. Bazzi. Polylogarithmic independence can fool DNF formulas. In *FOCS*, pages 63–73, 2007.

[3] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102(1):159–182, 1975.

[4] I. Ben-Eliezer, S. Lovett, and A. Yadin. Polynomial threshold functions: Structure, approximation and pseudorandomness. *CoRR*, abs/0911.3473, 2009.

[5] I. Benjamini, O. Gurel-Gurevich, and R. Peled. On k-wise independent distributions and boolean functions. Available at http://www.wisdom.weizmann.ac.il/~origurel/, 2007.

[6] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier*, 20:335–402, 1970.

[7] M. Braverman. Poly-logarithmic independence fools AC^0 circuits. In *CCC*, pages 3–8, 2009.

[8] A. Carbery and J. Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Mathematical Research Letters*, 8(3):233–248, 2001.

[9] E. W. Cheney. *Introduction to Approximation Theory*. McGraw-Hill, 1966.

[10] B. Chor and O. Goldreich. On the power of two-point based sampling. *Journal of Complexity*, 5(1):96–106, Mar. 1989.

[11] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. A. Servedio, and E. Viola. Bounded independence fools halfspaces. In *FOCS*, pages 171–180, 2009.

[12] I. Diakonikolas, R. A. Servedio, L.-Y. Tan, and A. Wan. A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions. In *CCC*, pages 211–222, 2010.

[13] K. O. Friedrichs. The identity of weak and strong extensions of differential operators. *Transactions of the American Mathematical Society*, 55(1):132–151, 1944.

[14] M. I. Ganzburg. personal communication.

[15] M. I. Ganzburg. *Limit theorems of polynomial approximation with exponential weights*. Memoirs of the American Mathematical Society, 2008.

[16] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42:1115–1145, 1995.

[17] P. Gopalan, R. O'Donnell, Y. Wu, and D. Zuckerman. Fooling functions of halfspaces under product distributions. In *CCC*, pages 223–234, 2010.

[18] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, 42(3):1079–1083, 1971.

[19] P. Harsha, A. Klivans, and R. Meka. An invariance principle for polytopes. In *STOC*, pages 543–552, 2010.

[20] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006.

[21] A. T. Kalai, A. R. Klivans, Y. Mansour, and R. A. Servedio. Agnostically learning halfspaces. *SIAM J. Comput.*, 37(6):1777–1805, 2008.

[22] D. M. Kane, J. Nelson, and D. P. Woodruff. On the exact space complexity of sketching and streaming small norms. In *SODA*, pages 1161–1178, 2010.

[23] Z. Karnin, Y. Rabani, and A. Shpilka. Explicit dimension reduction and its applications. *ECCC*, TR09-121, 2009.

[24] S. Mahajan and R. Hariharan. Derandomizing semidefinite programming based approximation algorithms. In *FOCS*, pages 162–169, 1995.

[25] R. Meka and D. Zuckerman. Pseudorandom generators for polynomial threshold functions. In *STOC (see also CoRR abs/0910.4122)*, pages 427–436, 2010.

[26] M. A. Minsky and S. L. Papert. *Perceptrons*. MIT Press, Cambridge, MA, 1969 (expanded edition 1988).

[27] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010.

[28] D. Newman and H. Shapiro. Jackson's theorem in higher dimensions. *On approximation theory (Proc. Conf. Oberwolfach 1963)*, MR 32(310):208–219, 1964.

[29] Y. Rabani and A. Shpilka. Explicit construction of a small epsilon-net for linear threshold functions. In *STOC*, pages 649–658, 2009.

[30] A. A. Razborov. A simple proof of Bazzi's theorem. *ACM Transactions on Computation Theory*, 1(1), 2009.

[31] D. Sivakumar. Algorithmic derandomization via complexity theory. In *STOC*, pages 619–626, 2002.

APPENDIX

Lemma A.1. Let $A \in \mathbb{R}^{n \times n}$ be symmetric with $\lambda_{\min}(A) > 0$. Then $|\text{tr}(A)| \leq \|A\|_F^2 / \lambda_{\min}(A)$.

For $f : \{-1, 1\}^n \to \mathbb{R}$ and $r \geq 1$, we denote $\|f\|_r = (\mathbf{E}_{x \sim \mathcal{U}_n}[|f(x)|^r])^{1/r}$.

Theorem A.2 (Hypercontractivity [3], [6]). If $f : \{-1, 1\}^n \to \mathbb{R}$ is a degree-d polynomial and $1 \leq r < q \leq \infty$, then $\|f\|_q \leq ((q-1)/(r-1))^{d/2} \cdot \|f\|_r$.

Lemma A.3. Let $f(x)$ be a quadratic form. Then, for $X = (X_1, \ldots, X_n)$ a vector of independent Bernoullis,

$$\mathbf{E}[|f(X)|^k] \leq 2^k(\|A_f\|_F k^k + |\text{tr}(A_f)|^k).$$