Polynomial Threshold Functions: Structure, Approximation and Pseudorandomness

Ido Ben-Eliezer[*]   Shachar Lovett[†]   Ariel Yadin[‡]

November 29, 2009

Abstract

We study the computational power of polynomial threshold functions, that is, threshold functions of real polynomials over the boolean cube. We provide two new results bounding the computational power of this model.

Our first result shows that low-degree polynomial threshold functions cannot approximate any function with many influential variables. We provide a couple of examples where this technique yields tight approximation bounds.

Our second result relates to constructing pseudorandom generators fooling low-degree polynomial threshold functions. This problem has received attention recently, where Diakonikolas *et al* [13] proved that k-wise independence suffices to fool linear threshold functions. We prove that any low-degree polynomial threshold function, which can be represented as a function of a small number of linear threshold functions, can also be fooled by k-wise independence. We view this as an important step towards fooling general polynomial threshold functions, and we discuss a plausible approach achieving this goal based on our techniques.

Our results combine tools from real approximation theory, hyper-contractive inequalities and probabilistic methods. In particular, we develop several new tools in approximation theory which may be of independent interest.

---

[*]School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Email: idobene@tau.ac.il

[†]Weizmann Institute of Science, Rehovot, Israel. Email: shachar.lovett@weizmann.ac.il. Research supported by the Israel Science Foundation (grant 1300/05)

[‡]Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK. Email: a.yadin@statslab.cam.ac.uk

1 Introduction

A boolean function $h : \{-1, 1\}^n \to \{-1, 1\}$ is a threshold (or sign) function of a real function $f : \{-1, 1\}^n \to \mathbb{R}$ if

$$h(x_1, \ldots, x_n) = \text{sgn}(f(x_1, \ldots, x_n)).$$

In this work we study thresholds of low-degree polynomials, or Polynomial Threshold Functions (PTFs). There is a long line of research that study the case of linear functions, i.e. degree 1 polynomials, which are commonly called Linear Threshold Functions (LTFs), or *halfspaces* (see, e.g., [18, 8, 13] and their references within). A key example for an LTF is the *majority* function which can be defined as

$$\text{M}aj(x_1, \ldots, x_n) = \text{sgn}(x_1 + \ldots + x_n - \lceil n/2 \rceil).$$

The main challenge that we tackle in our work is bounding the computational power of low-degree PTFs. We consider two main problems. Constructing explicit pseudorandom distributions that fool low-degree PTFs, and providing lower bounds for the computation and approximation capabilities of PTFs.

Pseudorandom generators for PTFs An important question is whether k-wise independence fools PTFs for small values of k. In particular it is interesting whether k can be independent of the number of variables n.

A boolean function $h : \{-1, 1\}^n \to \{-1, 1\}$ is ε-fooled by k-wise independence if for any k-wise independent distribution K taking values in $\{-1, 1\}^n$ we have

$$|\mathbb{P}_{x \in K}[h(x) = 1] - \mathbb{P}_{x \in U}[h(x) = 1]| \le \varepsilon,$$

where U denotes the uniform distribution over $\{-1, 1\}^n$. We say that a k-wise independence fools degree-d polynomials if it fools any threshold function $h(x) = \text{sgn}(f(x) - t)$ for $t \in \mathbb{R}$), for any degree-d real polynomial. This notion can be extended to fooling real functions.

The problem of whether k-wise independence fools LTFs was first addressed by Benjamini et al. [8], who proved that k-wise independence fools the majority function, and subsequently by Diakonikolas et al. [13] who proved that k-wise independence fools LTFs. In both cases $k = \text{polylog}(\varepsilon) \cdot \varepsilon^{-2}$ was required to achieve error ε.

Our first result extends the result of Diakonikolas et al. [13] to thresholds of low-degree polynomials which depend on a small number of linear functions. We see it as an important step towards building pseudorandom generators fooling general PTFs. For a real polynomial $p(x) = \sum p_I \prod_{i \in I} x_i$ define its *weight* as the sum of the absolute values of the coefficients, excluding the constant coefficient, that is

$$wt(p) = \sum_{I \ne \emptyset} |p_I|$$

2

Theorem 1. *Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree d polynomial, which can be decomposed as a function of m linear functions. That is, there exist linear functions $g_1, \ldots, g_m : \{-1, 1\}^n \to \mathbb{R}$ and a degree-d polynomial $p : \mathbb{R}^m \to \mathbb{R}$ such that*

$$f(x) = p(g_1(x), \ldots, g_m(x))$$

for all $x \in \{-1, 1\}^n$. Assume that $g_1, \ldots, g_m$ are normalized such that $\mathbb{E}[g_1^2] = \ldots = \mathbb{E}[g_t^2] = 1$. Then k-wise independence ε-fools $f(x)$ for

$$k = exp(O(d/\varepsilon)^d) + poly((\log m \cdot d/\varepsilon)^d, m, wt(p)).$$

Lower bounds for approximation by PTFs   A boolean function $g : \{-1, 1\}^n \to \{-1, 1\}$ is said to be ε-approximated by degree d PTFs, if there exists a degree d PTF $h(x)$ s.t. $\mathbb{P}_{x \in U}[h(x) = g(x)] \geq 1 - \varepsilon$.

We prove that functions whose variables have high influence cannot be approximated by low-degree PTFs, where the influence of a variable x_i in g is defined as the probability that flipping x_i changes the value of g, i.e.

$$\mathrm{Inf}_i(g) = \mathbb{P}_x[g(x) \neq g(x \oplus e_i)],$$

where e_i is the i-th unit vector. We prove

Theorem 2. *Let $g : \{-1, 1\}^n \to \{-1, 1\}$ be a boolean function, such that $\mathrm{Inf}_i(g) \geq \tau$ for at least n^α variables. Then for any degree-d polynomial threshold function h we have*

$$\mathbb{P}_x[h(x) = g(x)] \leq 1 - \frac{\tau}{2} + \eta$$

where $\eta = O(d/(\alpha \log n)^{1/8d})$.

We illustrate the power of Theorem 2 by showing two examples. The first one shows that MOD_m function cannot be approximated by low degree PTFs, while the second result shows that any low-degree polynomials over $\mathbb{F}_2$ cannot be approximated by low-degree PTFs much better than the best trivial approximation. Let define the MOD_m function as

$$\mathrm{MOD}_m(x_1, \ldots, x_n) = \begin{cases} 1 & \sum_{i=1}^n \frac{x_i+1}{2} \equiv 0 \pmod{m} \\ -1 & \sum_{i=1}^n \frac{x_i+1}{2} \not\equiv 0 \pmod{m} \end{cases}$$

Note that as $\frac{x_i+1}{2} \in \{0, 1\}$, this definition is essentially equivalent to the common one. We have the following.

Corollary 3. *Let $h : \{-1, 1\}^n \to \{-1, 1\}$ be a degree-d polynomial threshold function for $d \leq O(\log \log n / \log \log \log n)$. Then*

$$\mathbb{P}[h(x) = \mathrm{MOD}_m(x)] \leq 1 - \frac{1}{m} + o(1).$$

3

This result is tight in the sense that trivially the MOD_m function admits an $1 - \frac{1}{m}$ approximation by the constant -1 function (which is also a degree-0 PTF).

Corollary 4. *Let $q : \{-1,1\}^n \to \{-1,1\}$ be a degree-r polynomial over $\mathbb{F}_2$ depending on all variables. Let $h : \{-1,1\}^n \to \{-1,1\}$ be a degree-d polynomial threshold function for $d \leq O(\log \log n / \log \log \log n)$. Then*

$$\mathbb{P}[h(x) = q(x)] \leq 1 - 2^{-r} + o(1).$$

This result is essentially tight, as if q is a product of r linear forms, then the constant 1 function gives an $1 - 2^{-r}$ approximation of q.

1.1 Tools

Approximation tools and k-wise independence. Several recent works used the method of approximating by real polynomials to show that certain families of functions are fooled by k-wise independent distributions. This method can be described as follows. In order to show that k-wise independence ε-fools a certain family of functions, one has to show that for every function f in that family, there is a degree k polynomial p_l and degree k polynomial p_u, such that for every $x \in \{-1,1\}^n$ we have $p_l(x) \leq f(x) \leq p_u(x)$, and such that $\mathbb{E}_x[p_u(x) - p_l(x)] \leq \varepsilon$. Using this technique, Bazzi [7] proved in a breakthrough paper that logarithmic-wise independence fools DNF and CNF formulas. Later, Braverman [10] proved that polylogarithmic-wise independence fools small constant depth circuits, settling a conjecture of Linial and Nisan [20].

In this work we use the method of approximating polynomials for the problem of fooling low degree PTFs. We introduce a general method of obtaining polynomials which are both bounding and approximating for any function which depends on a small number of subfunctions whose tail distribution 'behaves nicely'. In our case we apply it for functions of a few linear functions, but we believe that these methods should have independent interest.

Our starting point is the multidimensional Jackson's theorem, which states that every Lipschitz function f on m variables admits an ε-approximation by a degree-d polynomial, where d depends only on ε, m and the Lipschitz constant of f. We then use several additional techniques to show that f admits a polynomial approximation p which is a good approximation in a multidimensional box near the origin, and above f everywhere. Finally, we apply these techniques as well as some concentration and anti-concentration results to show that p is a good approximation for f.

Finally, we apply these techniques to show that any threshold of a function of a few linear functions (or a function of a few linear PTF's) can be fooled by k-wise independence, for k that is independent of the number of variables.

Decision trees and approximation of PTF. Our first tool is a new structural result about PTFs. Given a polynomial threshold function p, we show that it has a small set of variables, on which *most* of their possible assignments we obtain a function with no influential variable. More precisely, the partial assignments are given by a small depth decision tree.

Let D be a decision tree on the variables $x_1, \ldots, x_n$. Each internal node of D is labeled by some variable and has two outgoing edges, corresponding to the possible assignments to this variable. The set of leaves of the decision tree correspond to partial assignments to the variables. The set of the leaves of D is denoted by $L(D)$, and for any $\ell \in L(D)$ and a function $f(x_1, \ldots, x_n)$ we denote by $f|_\ell$ the function restricted to the partial assignment given by ℓ. For more precise definitions see Section 2. We prove the following result.

Lemma 5. *Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree-d polynomial, and let $h(x) = \mathrm{sgn}(f(x))$. For any $\epsilon, \delta > 0$, there exists a decision tree D of depth at most $2^{ed/\delta} \cdot \log(1/\epsilon)$, such that*

$$\mathbb{P}_{\ell \in L(D)}[\mathrm{Inf}_\infty(f|_\ell) > \delta] < \epsilon$$

and

$$\mathbb{P}_{\ell \in L(D)}[\mathrm{Inf}_\infty(h|_\ell) > \delta'] < \epsilon$$

for $\delta' = O(d \cdot \delta^{1/8d})$.

We sketch the proof of Theorem 2. If a function g approximates a PTF h, then after most partial assignments of variables, g still approximates h. We show that under most of these assignments, our obtained PTF does not have any influential variable, and therefore cannot approximate functions with many influential variables.

Independently of our work, Diakonikolas et al. [16] and Harsha et al. [19] proved similar results. We state their results in our terminology.

Theorem 6 (Theorem 1 in [16]). *Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree-d polynomial, and let $h(x) = \mathrm{sgn}(f(x))$. For any $\tau > 0$, there exists a decision tree D of depth $\frac{1}{\tau} \cdot (d \log \frac{1}{\tau})^{O(d)}$ such that with probability $1 - \tau$ over a random leaf $\ell \in L(D)$, the function $h|_\ell$ is either τ-close to being constant, or has $\mathrm{Inf}_\infty(h) < \tau$.*

Theorem 7 (Lemmas 5.1 and 5.2 in [19]). *Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree-d polynomial, and let $h(x) = \mathrm{sgn}(f(x))$. For any $\tau > 0$, there exists a decision tree D of depth $\frac{\mathrm{polylog}(\tau)}{\tau^2} \cdot \exp(d)$ such that with probability $1 - \tau$ over a random leaf $\ell \in L(D)$, the function $h|_\ell$ is either τ-close to being constant, or has $\mathrm{Inf}_\infty(h) < \tau$.*

We note that using Theorem 7 instead of Lemma 22 one can get an improvement in the dependence on the degree in Theorem 2. In particular, Corollaries 3 and 4 hold for degrees $d \leq O(\log n / \log \log n)$.

1.2 Towards fooling low degree PTFs

We propose a general method for proving that k-wise independence fools low degree PTFs. This is a high level approach and currently we are able to prove only a special case.

Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a real function. We say that f is δ-normal if the distribution of $f(x)$ over uniform input is δ-close to the standard normal distribution. That is,

$$|\mathbb{P}_{x \in U}[f(x) \geq t] - \mathbb{P}[N \geq t]| < \delta$$

for any $t \in \mathbb{R}$, where $N \sim N(0, 1)$ is a standard normal variable. In what follows we let $f(x)$ be a degree d polynomial, $h(x) = \mathrm{sgn}(f(x))$ a PTF and $\varepsilon > 0$ the required error.

5

(*i*). **Reduction to low-influence PTF**: It is enough to prove that k-wise independence fools PTFs with small influences. We prove this in Lemma 22 and Claim 12. The important properties of PTFs with low influences is that their distribution is not concentrated around any specific value (see Lemma 19), which can later be used to build approximating polynomials for such functions.

(*ii*). **δ-normal polynomials**: Assume that $f(x)$ is a degree-d polynomial with low influences which is $\delta(\varepsilon)$-regular. Then $h(x) = \text{sgn}(f(x))$ is fooled by $k(\varepsilon)$-wise independence. This can be proved using the same proof technique of Diakonikolas et al. [13], using the approximating polynomials for the sgn functions they construct, when replacing the tail bounds for linear polynomials by the normal distribution.

(*iii*). **Functions of a few δ-normal polynomials**: Assume that $f(x)$ is a degree-d polynomial with low influences, which can be decomposed as a function of m polynomials $g_1, \ldots, g_m$, each is $\delta(m, \varepsilon)$-normal. Then $h(x) = \text{sgn}(f(x))$ is fooled by $k(m, \varepsilon)$-wise independence. Our proofs can be slightly altered to prove this, again replacing tail bounds for linear polynomials by the normal distribution. This can be also extended when allowing a small error term.

(*iv*). **Regularization of degree-d polynomials**: We conjecture that for every $\delta, \tau > 0$, any degree d polynomial $f : \{-1, 1\}^n \to \mathbb{R}$ can be regularized in the following way. There exist a small number $t = t(d, \delta, \tau)$ of variables $x_{i_1}, \ldots, x_{i_t}$, and a small number $m = m(d, \delta, \tau)$ of δ-normal polynomials $g_1, \ldots, g_m : \{-1, 1\}^n \to \mathbb{R}$, a low-degree polynomial $p : \mathbb{R}^m \to \mathbb{R}$ and an error polynomial $e : \{-1, 1\}^n \to \mathbb{R}$ with $\|e\|_2 < \tau$, such that

$$f(x) = p(x_{i_1}, \ldots, x_{i_t}, g_1(x), \ldots, g_m(x)) + e(x).$$

For linear polynomials, this can be proved using the tools of Diakonikolas et al. [13]. We were able to prove this conjecture also for quadratic polynomials, and conjecture that the same holds for all constant degrees d.

(*v*). **Putting everything together**: Let $f(x)$ be a degree d PTF. We start by reducing it to a PTF with low influences using a partial assignment for a small number of variables. We use the conjecture to decompose it as a function of a small number of δ-normal PTFs, and use this decomposition to prove that k-wise independence to fool f.

So where does this fail? The critical point of failure is in the dependence of the number of functions m used in the decomposition of f, and the required distance δ between their distribution and the normal distribution. We can prove that if f can be decomposed into a function of m δ-normal functions for small enough δ then the proof follows through. The problem is that δ has to be very small; in particular $\delta < exp(-m^5)$. On the other hand in the regularization conjecture, the number of components m depend on δ. We can prove the regularization conjecture for quadratic polynomials for $m \geq 1/\delta^2$. These two requirements have no common solution.

We note the independently of our work, Meka and Zuckerman [24] constructed an explicit pseudorandom generator fooling all degree-d PTFs. Their construction involves partitions the set of inputs into a small number of buckets (using a pairwise independent hash function), and then applying k-wise independent distribution to each bucket independently.

1.3 More related Work

The study of distributions that fool low-degree polynomials and related functions has received considerable attention. For example, fooling linear polynomials over finite fields [25, 4], which has a numerous number of applications and extensions, pseudorandom generators for low degree polynomials [9, 21, 27, 3] and fooling modular sums [22].

Bruck [11] studied polynomial threshold functions, and proved that such functions can be computed by depth-2 polynomial sized circuits with unbounded fan-in linear threshold gates. Aspnes et al. [6] studied the approximation of boolean functions by some threshold functions. Namely, they study the best possible approximation for the parity function and other symmetric functions by low-degree PTF, and proved that for every degree-k PTF p, we have

$$\mathbb{P}_x[p(x) \neq PARITY(x)] \geq \frac{\sum_{i=0}^{\lfloor (n-k-1)/2 \rfloor} \binom{n}{i}}{2^n},$$

and this bound is tight. However, their bounds for other functions are not fully explicit and are not tight.

A few recent results consider the problem of constructing pseudorandom generators for threshold functions. This problem has a natural geometrical interpretation. Rabani and Shpilka [26] provided a construction of ε-net for halfspaces. Namely, a set of points S for which for every halfspace h that satisfies $\varepsilon \leq \mathbb{P}_{x \in \{-1,1\}^n}[h(x) = 1] \leq 1 - \varepsilon$ there are two points $s_1, s_2 \in S$ such that $h(s_1) = -1$ and $h(s_2) = 1$. The size of their construction is polynomial in n and $\frac{1}{\varepsilon}$. [13] proved that any k-wise distribution fools halfspaces, for k that is polynomial in $\frac{1}{\varepsilon}$. Their dependence on k is nearly optimal, as shown by Benjamini et al. [8].

A subsequent work of Diakonikolas et al. [14] show that k-wise independence fools quadratic threshold functions, and intersections of such functions.

The rest of our paper is organized as follows. We introduce some preliminary definitions and tools in Section 2. This section includes definitions and results that are related to k-wise independence, decision trees, concentration of multivariate polynomials and some other analytical tools. In Section 3 we present our new structural results on low-degree PTF, and present our application that shows that certain functions cannot be approximated by low degree PTF. Finally, in Section 4 we present our new tools from approximation theory, and show that k-wise independence fools thresholds of functions of a few linear polynomials.

Throughout this work we do not try to optimize constants. Also, we omit floor and ceiling signs whenever these are not crucial.

2   Preliminaries

In this section we provide some necessary definitions that will be widely used throughout the work, including definitions and tools related to k-wise independent distributions, decision trees, analytical tools, and concentration bounds for multivariate polynomials.

2.1 k-wise independent distributions and polynomials

A distribution D on the boolean cube $\{-1, 1\}^n$ is k-wise independent if the marginal distribution of any k coordinates is the uniform distribution. There are explicit constructions of such distributions of size $O(n^{\lceil k/2 \rceil})$, and these constructions are essentially optimal [2].

Given a class of functions $\mathbb{S}$ from the boolean cube to $\{-1, 1\}$, a distribution D ε-fools $\mathbb{S}$ if for every $\varphi \in \mathbb{S}$, we have

$$|\mathbb{P}_{x \in U}[\varphi(x) = 1] - \mathbb{P}_{x \in D}[\varphi(x) = 1]| \leq \varepsilon.$$

Combining these two definitions, for simplicity we define the following.

Definition 8 (k-wise independence fooling boolean functions). A boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ is said to be fooled by k-wise independence with error ε, if for any k-wise independent distribution K,

$$|\mathbb{P}_{x \in U}[f(x) = 1] - \mathbb{P}_{x \in K}[f(x) = 1]| \leq \varepsilon.$$

The following claim is sufficient for k-wise distributions to ε-fool a boolean function.

Claim 9. *Let* $f : \{-1, 1\}^n \to \{-1, 1\}$. *Assume there are two degree-k polynomials* $p_u, p_l : \{-1, 1\}^n \to \mathbb{R}$ *such that*

- $p_l(x) \leq f(x) \leq p_u(x)$ *for all* $x \in \{-1, 1\}^n$.

- $\mathbb{E}_{x \in U}[p_u(x) - p_l(x)] \leq \varepsilon$.

Then k-wise independence fools f with error ε.

The proof of this claim is simple, and can be found for example in [7]. It is worth noting that Bazzi [7] also proved that the condition is necessary using linear programming duality.

Our next definition extends the notion of fooling boolean functions, and defines it for real functions as well.

Definition 10 (k-wise independence fooling real functions). Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a function. We say that k-wise distributions fool f with error ε, if for any k-wise distribution K over $\{-1, 1\}^n$, and any $t \in \mathbb{R}$,

$$|\mathbb{P}_{x \in U}[f(x) \leq t] - \mathbb{P}_{x \in K}[f(x) \leq t]| \leq \epsilon$$

A real function $f(x_1, \ldots, x_n)$ is a degree-d polynomial if it can be represented as

$$f(x) = \sum_{k=0}^{d} \sum_{i_1 \leq \ldots \leq i_k \in [n]} \alpha_{i_1, \ldots, i_k} x_{i_1} \ldots x_{i_k}.$$

A polynomial is *multilinear* if each variable appears in every monomial at most once. Equivalently, it can be represented as

$$f(x) = \sum_{k=0}^{d} \sum_{i_1 < \ldots < i_k \in [n]} \alpha_{i_1, \ldots, i_k} x_{i_1} \ldots x_{i_k}.$$

Each function $f : \{-1, 1\}^n \to \mathbb{R}$ can be uniquely represented by a multilinear polynomial. We will interchangeably regard f both as a boolean function and as a multilinear polynomial.

2.2 Decision trees

A *Decision Tree* over binary variables $x_1, \ldots, x_n$ is a binary tree, where each internal node v is labeled by one of the variables x_v, such that the labels along any path from the root to a leaf are distinct. Also, the two (directed) edges that leave each node are labeled by -1 and 1. Therefore, given a path P from the root to a leaf, for every variable x that appears along the path we can uniquely define a value $x_P \in \{-1, 1\}$ to be the label of the edge in P that leaves the node labeled by x.

A path P from the root to a leaf ℓ defines a partial assignment A_ℓ by assigning every variable that appears on x by x_P. All the variables that do not appear on P remain unassigned.

We denote the set of variables labeling the vertices in the path to ℓ by $\text{var}(\ell)$. We denote the set of leaves of a decision tree D by $L(D)$.

The *depth* of a leaf is the length of the path from the root to it, and the depth of a decision tree is the maximal depth of a leaf.

With a slight abuse of notation, we define a random leaf in a decision tree to be the result of the following procedure. We start at the root, and at each step we move to one of his children, uniformly and independently of the other choices. When we arrive a leaf ℓ we output it. Equivalently, we choose each leaf ℓ with probability $2^{-\text{depth}(\ell)}$.

We now can define the restriction of a function with respect to a certain leaf ℓ and with respect to a decision tree D.

Definition 11. Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a function, D be a decision tree on $x_1, \ldots, x_n$ and ℓ be a leaf in D. We define the restriction of f to ℓ, denoted by $f|_\ell$, to be the function obtained by f after assigning the variables $x_1, \ldots, x_n$ according to A_ℓ. Namely, the domain of $f|_\ell$ is $\{-1, 1\}^{[n] \setminus \text{var}(\ell)}$, and the range of $f|_\ell$ is $\mathbb{R}$.

Similarly, given a distribution $\mathcal{D}$, define its restriction to ℓ, $\mathcal{D}|_\ell$ to be the the distribution obtained from D conditioning on the partial assignment A_ℓ.

We define a random function $f|_D$ by choosing a random leaf ℓ of D and restricting f to ℓ.

We will need the following easy claim.

Claim 12. *Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a function, and D a decision tree, such that*

$$\mathbb{P}_{\ell \in L(D)} \left[k\text{-wise independent distributions fool } f|_\ell \text{ with error } \varepsilon\right] \geq 1 - \delta.$$

Then $(k + \operatorname{depth}(D))$-wise independent distributions fool f with error $\varepsilon + \delta$.

Proof. Let K be some k'-wise independent distribution for $k' = k + \operatorname{depth}(D)$. For any leaf $\ell \in L(D)$, the restriction $K|_\ell$ of K given by ℓ is k-wise independent.

Let $\ell \in L(D)$ be a random leaf of D. Say ℓ is *good* if k-wise independent distributions fool $f|_\ell$ with error ε. By our assumption ℓ is good with probability at least $1 - \delta$.

Let $t \in \mathbb{R}$. For any good leaf we have

$$\left| \mathbb{P}_{x \in U|_\ell}[f(x) \leq t] - \mathbb{P}_{x \in K|_\ell}[f(x) \leq t] \right| < \varepsilon.$$

For any other leaf we can bound

$$\left| \mathbb{P}_{x \in U|_\ell}[f(x) \leq t] - \mathbb{P}_{x \in K|_\ell}[f(x) \leq t] \right| \leq 1.$$

Hence we get

$$\left| \mathbb{P}_{x \in U}[f(x) \leq t] - \mathbb{P}_{x \in K}[f(x) \leq t] \right| \leq \mathbb{E}_{\ell \in L(D)} \left| \mathbb{P}_{x \in U|_\ell}[f(x) \leq t] - \mathbb{P}_{x \in K|_\ell}[f(x) \leq t] \right| \leq \varepsilon + \delta.$$

$\square$

We will also require a bound on the L_2 norm of linear functions, under a partial restriction given by a decision tree.

Lemma 13. *Let $g : \{-1, 1\}^n \to \mathbb{R}$ be a linear function with $\mathbb{E}[g^2] = 1$. Let D be a decision tree. Then*

$$\mathbb{P}_{\ell \in L(D)}[\mathbb{E}[(g|_\ell)^2] \geq t] \leq 3e^{-t/8}.$$

Proof. We will need the following variant of the Azuma-Hoeffding inequality. Let $X_1, \ldots, X_n$ be random variables, such that $X_i = c_i(X_1, \ldots, X_{i-1})$ or $X_i = -c_i(X_1, \ldots, X_{i-1})$, each with probability $1/2$, where $c_i : \{-1, 1\}^{i-1} \to \mathbb{R}$ is some deterministic function, such that a.s. $X_1^2 + \ldots + X_n^2 \leq 1$. We will prove that

$$\mathbb{P}[X_1 + \ldots + X_n \geq t] \leq e^{-t^2/2}.$$

First we show how we apply this inequality. Let $g(x) = a + \sum a_i x_i$ where $\sum a_i^2 + a = 1$. Let ℓ be a leaf of D. Notice that $g|_\ell(x) = (a + \sum_{i \in \operatorname{var}\ell} a_i x_i|_\ell) + \sum_{i \notin \operatorname{var}\ell} a_i x_i$. Hence, to bound the probability that $\mathbb{E}[(g|_\ell)^2]$ is large, we need to bound the probability that $\sum_{i \in \operatorname{var}\ell} a_i x_i|_\ell$ is large. We will assume w.l.o.g that $t \geq 8$ since otherwise the required inequality holds immediately.

Define a sequence of random variables $X_1, X_2, \ldots$. Let i_1 be the index of the first variable queried by D. Define $X_1 = \pm a_{i_1}$. Given the value of x_{i_1}, let i_2 be the index of the second

variable queried by D. Define $X_2 = \pm a_{i_2}$. Notice that in fact $X_2 = \pm c_2(X_1)$. Let i_3 be the index of the third variable queried by D, and define $X_3 = \pm a_{i_3}$. Again, $X_3 = \pm c_3(X_1, X_2)$, and we continue until we reach a leaf. If X_d is a leaf of D, we define the remaining variables $X_{d+1}, \ldots, X_n$ to be 0. Let $X = \sum X_i$. Notice that

$$g|_\ell(x) = (a + X) + \sum_{i \notin \text{var}(\ell)} a_i x_i.$$

Since the conditions of the inequality hold for $X_1, \ldots, X_n$, we get that $\mathbb{P}[X \geq t] \leq e^{-t^2/2}$. We wish to bound the probability over $\ell \in L(D)$ that $\mathbb{E}[(g|_\ell)^2] \geq t]$. If this event occurs, then we must have $X \geq \sqrt{t} - 1$. Since we assume $t \geq 8$ this gives $X \geq \sqrt{t}/2$, which gives

$$\mathbb{P}_{\ell \in L(D)}[\mathbb{E}[g|_\ell^2] \geq t] \leq \mathbb{P}[X \geq \sqrt{t}/2] \leq e^{-t/8}.$$

We now turn to prove the modification of the Azuma-Hoeffding inequality. Set $\lambda > 0$ to be determined later, and consider $E = \mathbb{E}[e^{\lambda(X_1 + \ldots + X_n)}]$. We can decompose $E = \prod_{i=1}^{k} \mathbb{E}[e^{\lambda X_i} | X_1, \ldots, X_{i-1}]$. We have

$$\mathbb{E}[e^{\lambda X_i} | X_1, \ldots, X_{i-1}] = \frac{1}{2} e^{\lambda c_i(X_1, \ldots, X_{i-1})} + \frac{1}{2} e^{-\lambda c_i(X_1, \ldots, X_{i-1})}.$$

Using the inequality $\frac{1}{2}(e^x + e^{-x}) \leq e^{x^2/2}$ we get

$$\mathbb{E}[e^{\lambda X_i} | X_1, \ldots, X_{i-1}] \leq e^{\lambda^2 c_i(X_1, \ldots, X_{i-1})^2/2}.$$

Hence

$$E \leq \mathbb{E}_{X_1, \ldots, X_n}[e^{\lambda^2/2 \cdot (c_1^2 + c_2(X_1)^2 + \ldots + c_n(X_1, \ldots, X_{n-1})^2)}] = \mathbb{E}_{X_1, \ldots, X_n}[e^{\lambda^2/2 \cdot (X_1^2 + \ldots + X_n^2)}] \leq e^{\lambda^2/2}.$$

Thus we get

$$\mathbb{P}[X_1 + \ldots + X_n \geq t] \leq e^{\lambda^2/2 - \lambda t}.$$

Setting $\lambda = t$ gives the required inequality. $\qquad\square$

2.3 Analytical tools

The Lipschitz constant of a function bounds the change in the function value when the inputs are perturbed. It will be convenient for us to measure distance in the L_∞ norm. Recall that for $z = (z_1, \ldots, z_m) \in \mathbb{R}^m$, its L_∞ norm is defined as the maximal absolute value of its coordinates, i.e.

$$\|z\|_\infty = \max\{|z_i| : i \in [m]\}.$$

Definition 14 (Lipschitz constant). Let $F : \mathbb{R}^m \to \mathbb{R}$ be a function. The Lipschitz constant of F, denoted by $L(F)$, is defined as

$$L(F) = \sup_{z', z'' \in \mathbb{R}^m} \frac{|F(z') - F(z'')|}{\|z' - z''\|_\infty}.$$

11

The function F is said to be *Lipschitz* if $L(F) < \infty$.

Let C be a convex subset of $\mathbb{R}^m$. The Lipschitz constant of F restricted to C, denoted $L_C(F)$, is defined as

$$L_C(F) = \sup_{z', z'' \in C} \frac{|F(z') - F(z'')|}{\|z' - z''\|_\infty}.$$

.

We will use restricted Lipschitz constant only for cubes.

Definition 15. The cubic ε-neighborhood of a point $z \in \mathbb{R}^m$ is defined as

$$\mathcal{C}(z, \varepsilon) = \{z' \in \mathbb{R}^m : \|z - z'\|_\infty \leq \varepsilon\}.$$

For a set $S \subset \mathbb{R}^m$, the cube ε-neighborhood of S is defined as

$$\mathcal{C}(S, \varepsilon) = \bigcup_{z \in S} \mathcal{C}(z, \varepsilon).$$

2.4 Tail estimates for polynomials

In this subsection we prove two results about the concentration of degree-d multilinear polynomials. The first result gives a tail estimate on the probability that a degree-d polynomial is very large, and the second result provides a lower bound on the probability it is concentrated near a certain value. In both results we apply techniques based on hyper-contractivity [23].

2.4.1 Tail bounds

We prove in this subsection a general tail estimate on multilinear polynomials, which holds both under the uniform distribution over $\{-1, 1\}^n$ and under the standard multi-normal distribution. Namely, we show that for any degree-d multilinear polynomial $f(x_1, \ldots, x_n)$, the probability that $|f(x)| \geq t$ is bounded by $exp(-t^{2/d})$. We observe that this is tight by considering the polynomial obtained by multilinearizing $f(x) = (x_1 + \ldots + x_n)^d$. Our main result follows.

Lemma 16. *Let $f(x_1, \ldots, x_n)$ be a multilinear degree-d polynomial with $\mathbb{E}[f^2] = 1$. Then for every $t \geq 1$,*

$$\mathbb{P}_{x \in U}[|f(x)| \geq t] \leq 2^{-\frac{d}{4} \cdot t^{2/d}}$$

and

$$\mathbb{P}_{x \in N}[|f(x)| \geq t] \leq 2^{-\frac{d}{4} \cdot t^{2/d}}.$$

Let X be a real random variable. Denote $\|X\|_q = (\mathbb{E}[|X|^q])^{1/q}$. Following the notation from [23], we say that X is $(2, q, \eta)$ hyper-contractive if for every $a \in \mathbb{R}$,

$$\|a + \eta X\|_q \leq \|a + X\|_2.$$

We use the following two theorems from [23].

Lemma 17 (Theorem 3.13 in [23]). *If X is uniform on $\{-1, 1\}$, or a standard normal random variable $N(0, 1)$, then for every $q \geq 2$, X is $(2, q, \eta)$ hyper-contractive with $\eta = (q-1)^{-1/2}$.*

Lemma 18 (Proposition 3.12 in [23]). *Let X be $(2, q, \eta)$ hyper-contractive. Let $f(x_1, \ldots, x_n)$ be a multilinear degree-d polynomial. Let $Q = f(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are i.i.d and distributed according to X. Then*

$$\|Q\|_q \leq \eta^{-d} \|Q\|_2$$

Proof of Lemma 16. Let X be either a uniform random variable over $\{-1, 1\}$ or standard normal random variable $N(0, 1)$. Let $Q = f(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are i.i.d and distributed according to X. In either case we have $\|Q\|_2 = \mathbb{E}[f^2]^{1/2} = 1$. Fix $q \geq 2$ to be determined later. By Lemma 17, X is $(2, q, \eta)$ for $\eta = (q-1)^{-1/2}$. Thus, by Lemma 18 we have

$$\mathbb{E}_{x \in X^n}[|f(x)|^q] \leq (q-1)^{dq/2}.$$

Thus by Markov's inequality

$$\mathbb{P}_{x \in X^n}[|f(x)| \geq t^{d/2}] \leq \left(\frac{q-1}{t}\right)^{qd/2}.$$

Since $t \geq 1$ we can set $q = t/2 + 1$ and get

$$\mathbb{P}_{x \in X^n}[|f(x)| \geq t^{d/2}] \leq 2^{-td/4}.$$

Hence we conclude

$$\mathbb{P}_{x \in X^n}[|f(x)| \geq t] \leq 2^{-\frac{d}{4} \cdot t^{2/d}}.$$

$\square$

2.4.2 Concentration lower bounds

The main result of this subsection is the following lemma.

Lemma 19. *There exist constants $c_1, c_2 > 0$ such that the following holds. Let $f(x_1, \ldots, x_n)$ be a polynomial of degree d such that $\mathrm{Var}[f] = 1$. For $\varepsilon > 0$ let $\alpha = (c_1 \cdot \epsilon/d)^d$ and $\tau = (c_2 \cdot \epsilon/d)^{8d}$. If $\mathrm{Inf}_\infty(f) \leq \tau$, then for every $t \in \mathbb{R}$,*

$$\mathbb{P}_{x \in U}[|f(x) - t| \leq \alpha] \leq \epsilon.$$

We use the following two theorems.

Lemma 20 (Theorem 2.1 in [23]). *Let $f(x_1, \ldots, x_n)$ be a multilinear degree d polynomial, such that $\mathrm{Inf}_\infty(f) \leq \tau$. Then for every $t \in \mathbb{R}$*

$$|\mathbb{P}_{x \in U}[f(x) \leq t] - \mathbb{P}_{x \in N}[f(x) \leq t]| \leq O(d\tau^{1/8d}).$$

13

The following is an immediate corollary of Theorem 8 in Carbery and Wright [12], which is also stated as Corollary 3.23 in [23].

Lemma 21. *Let $f(x_1, \ldots, x_n)$ be a multilinear degree d polynomial such that $\mathrm{Var}[f] = 1$. Then for every $t \in \mathbb{R}$,*
$$\mathbb{P}_{x \in N}[|f(x) - t| \leq \alpha] \leq O(d\alpha^{1/d}).$$

Proof of Lemma 19. Let f be a degree-d polynomial such that $\mathrm{Inf}_\infty(f) \leq \tau$. By Lemma 20 we have:
$$\mathbb{P}_{x \in U}[|f(x) - t| \leq \alpha] \leq \mathbb{P}_{x \in N}[|f(x) - t| \leq \alpha] + O(d\tau^{1/8d}).$$

By Lemma 21 we have
$$\mathbb{P}_{x \in N}[|f(x) - t| \leq \alpha] \leq O(d\alpha^{1/d})$$

Combing the two results we get:

$$\mathbb{P}_{x \in U}[|f(x) - t| \leq \alpha] \leq O(d \cdot (\tau^{1/8d} + \alpha^{1/d})).$$

Setting $\alpha = (c_1 \cdot \epsilon/d)^d$ and $\tau = (c_2 \cdot \epsilon/d)^{8d}$ for some absolute constants $c_1, c_2 > 0$ we get

$$\mathbb{P}_{x \in U}[|f(x) - t| \leq \alpha] \leq \epsilon.$$

$\square$

3   The effect of partial assignments

We prove in this section that functions with many influential variables cannot be non-trivially approximated by low-degree PTFs. The proof depends on a new general structural result for polynomials and polynomial threshold functions. We show that for every such function there exists a small depth decision tree D, such that $f|_D$ has low influence with high probability.

Lemma 22. *Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree-d polynomial, and let $h(x) = \mathrm{sgn}(f(x))$. For every $\epsilon, \delta > 0$, there exists a decision tree D of depth at most $2^{ed/\delta} \cdot \log(1/\epsilon)$, such that*

$$\mathbb{P}_{\ell \in L(D)}[\mathrm{Inf}_\infty(f|_\ell) > \delta] < \epsilon$$

and

$$\mathbb{P}_{\ell \in L(D)}[\mathrm{Inf}_\infty(h|_\ell) > \delta'] < \epsilon$$

for $\delta' = O(d \cdot \delta^{1/8d})$.

The proof of Lemma 22 appears in Subsection 3.1.

We apply Lemma 22 in order to prove our main result of this section, that functions with many influential variables cannot be approximated by low-degree PTFs. We restate Theorem 2 for the convenience of the reader.

Theorem 23 (Theorem 2, restated). *Let $g : \{-1,1\}^n \to \{-1,1\}$ be a boolean function, such that $\mathrm{Inf}_i(g) \geq \tau$ for at least n^α variables. Then for any degree-d polynomial threshold function h we have*

$$\mathbb{P}_x[h(x) = g(x)] \leq 1 - \frac{\tau}{2} + \eta$$

where $\eta = O(d/(\alpha \log n)^{1/8d})$.

Before proving Theorem 23, we give a couple of examples for its application. We show that low-degree PTFs do not admit a non-trivial approximation for the MOD_m function, or low degree polynomials over $\mathbb{F}_2$.

Corollary 24 (Corollary 3, restated). *Let $h : \{-1,1\}^n \to \{-1,1\}$ be a degree-d polynomial threshold function for $d = O(\log \log n / \log \log \log n)$. Then*

$$\mathbb{P}[h(x) = \mathrm{MOD}_m(x)] \leq 1 - \frac{1}{m} + o(1).$$

Proof. It is straightforward to verify that $\mathrm{Inf}_i(\mathrm{MOD}_m) = \frac{2}{m}$ for all $i \in [n]$, the proof now follows by Theorem 23. $\qquad\square$

Corollary 25 (Corollary 4, restated). *Let $q : \{-1,1\}^n \to \{-1,1\}$ be a degree-r polynomial over $\mathbb{F}_2$ depending on all variables. Let $h : \{-1,1\}^n \to \{-1,1\}$ be a degree-d polynomial threshold function for $d \leq O(\log \log n / \log \log \log n)$. Then*

$$\mathbb{P}[h(x) = q(x)] \leq 1 - 2^{-r} + o(1).$$

Proof. We will prove $\mathrm{Inf}_i(q) \geq 2^{1-r}$ for all $i \in [n]$. Let $q(x) = (-1)^{q'(x')}$, where $q' : \mathbb{F}_2^n \to \mathbb{F}_2$ and $x' \in \mathbb{F}_2^n$ set by $x_i = (-1)^{x'_i}$. We will in fact show that $\mathbb{P}[q'(x') \neq q'(x' \oplus e_i)] \geq 2^{1-r}$. write $q'(x') = x'_i q'_1(x') + q'_2(x')$. As q'_1 is a non-zero polynomial of degree at most $r - 1$, we have $\mathbb{P}[q'_1(x') = 1] \geq 2^{1-r}$. $\qquad\square$

We now return to prove Theorem 23.

Proof of Theorem 23. Let $g : \{-1,1\}^n \to \{-1,1\}$ be a boolean function for which $\mathrm{Inf}_i(g) \geq \tau$ for at least $n' = n^\alpha$ variables. We will provide a lower bound on $q = \mathbb{P}[g(x) \neq h(x)]$,

Set $\delta > 0$ and $\varepsilon > 0$ to be determined later. Set $m = 2^{ed/\delta} \log 1/\varepsilon$ and $\delta' = O(d \cdot \delta^{1/8d})$. Using Lemma 22 we get that there exists a decision tree D of depth at most m, such that

$$\mathbb{P}_{\ell \in L(D)}[\mathrm{Inf}_\infty(h|_\ell) > \delta'] < \varepsilon.$$

In each path in D there are at most m variables. Thus, there exists a variable x_i for which $\mathrm{Inf}_i(g) \geq \tau$ which appears in at most m/n' of the paths. Equivalently, a random leaf $\ell \in L(D)$ assigns a value to x_i with probability at most m/n'. We get

$$\begin{aligned}
\mathbb{P}[g(x) \neq g(x \oplus e_i)] &\leq \mathbb{E}_{\ell \in L(D)}\left[\mathbb{P}_x[g|_\ell(x) \neq g|_\ell(x \oplus e_i)]\right] + m/n' \\
&\leq \mathbb{E}_{\ell \in L(D)}[\mathbb{P}_x[g|_\ell(x) \neq h|_\ell(x)] + \mathbb{P}_x[h|_\ell(x) \neq h|_\ell(x \oplus e_i)] \\
&\quad + \mathbb{P}_x[h|_\ell(x \oplus e_i) \neq g|_\ell(x \oplus e_i)]] + m/n' \\
&= 2\,\mathbb{P}[g(x) \neq h(x)] + \mathbb{E}_{\ell \in L(D)}[\mathrm{Inf}_i(h|_\ell)] + m/n' \\
&= 2q + \delta' + \varepsilon + m/n'
\end{aligned}$$

15

On the other hand, by assumption we have $\mathbb{P}[g(x) \neq g(x \oplus e_i)] \geq \tau$. Combining the two bounds we get that

$$\mathbb{P}[g(x) \neq h(x)] = q \geq \frac{1}{2}(\tau - \varepsilon - \delta' - m/n')$$

$$\geq \frac{\tau}{2} - O(\varepsilon + d\delta^{1/8d} + 2^{ed/\delta}\log(1/\varepsilon)/n')$$

Setting $\delta = O(d/\log n')$ and ε small enough (for example $\varepsilon = 1/n'$) gives

$$q = \mathbb{P}[g(x) \neq h(x)] \geq \frac{\tau}{2} - \eta$$

for $\eta = O(\frac{d}{(\alpha \log n)^{1/8d}})$. □

3.1 Proof of Lemma 22

The proof of Lemma 22 will be conducted in three steps. First we show that for every low-degree polynomial there exists a partial assignment of a small set of variables under which we get a polynomial with low influences. We then argue that if a polynomial has low influences, then so does its threshold. We then conclude by showing that if there is a single good assignment, then by taking larger set of variables we get that most of the assignments are good. The first step is accomplished by the following lemma.

Lemma 26. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial. For every $\delta > 0$ there exist a set of variables $x_{i_1}, \ldots, x_{i_k}$ and assignments for these variables $b_{i_1}, \ldots, b_{i_k} \in \{-1,1\}$, such that*

$$\mathrm{Inf}_\infty(f|_{x_{i_1}=b_{i_1},\ldots,x_{i_k}=b_{i_k}}) \leq \delta$$

and $k \leq ed/\delta$.

Proof. We construct a sequence of assignments for the variables of f, assigning a value to a single variable at each step, that will lead eventually to a polynomial $f|_{x_{i_1}=b_{i_1},\ldots,x_{i_k}=b_{i_k}}$ whose influence is bounded by δ.

Every degree-d polynomial f can be uniquely represented as

$$f(x) = \sum_{I \subset [n], |I| \leq d} f_I \prod_{i \in I} x_i.$$

For $\alpha \geq 0$ define operator $V_\alpha(f)$ to be

$$V_\alpha(f) = \sum_{I \subset [n], |I| \leq d} |f_I|^2 (1+\alpha)^{|I|}.$$

Note that $V_0(f) = \mathbb{E}[f^2]$.

Fix a variable x_i, and let $f(x) = x_i f_1(x') + f_2(x')$ where $x' = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. We have $f|_{x_i=1} = f_1 + f_2$ and $f|_{x_i=-1} = -f_1 + f_2$. Notice that $V_0(f_1) = Inf_i(f) \cdot V_0(f)$.

16

We first claim that

$$\tfrac{1}{2}(V_\alpha(f|_{x_i=1}) + V_\alpha(f|_{x_i=-1})) = V_\alpha(f) - \alpha V_\alpha(f_1) \tag{3.1}$$

To prove it, write $f_1(x') = \sum f_{1,I} \prod_{i\in I} x'_i$ and $f_2(x') = \sum f_{2,I} \prod_{i\in I} x'_i$. We have

$$V_\alpha(f|_{x_i=1}) + V_\alpha(f|_{x_i=-1}) = V_\alpha(f_1 + f_2) + V_\alpha(-f_1 + f_2) =$$

$$\sum_I (f_{1,I} + f_{2,I})^2 (1+\alpha)^{|I|} + \sum_I (-f_{1,I} + f_{2,I})^2 (1+\alpha)^{|I|} =$$

$$2 \cdot \sum_I (f_{1,I}^2 + f_{2,I}^2)(1+\alpha)^{|I|} =$$

$$2 \cdot \sum_I (f_{1,I}^2 (1+\alpha)^{|I|+1} + f_{2,I}^2 (1+\alpha)^{|I|}) - 2\alpha \cdot \sum_I f_{1,I}^2 (1+\alpha)^{|I|} =$$

$$2 \cdot (V_\alpha(f) - \alpha V_\alpha(f_1))$$

This proves (3.1). In particular for $\alpha = 0$ we get

$$\tfrac{1}{2}(V_0(f|_{x_i=1}) + V_0(f|_{x_i=-1})) = V_0(f). \tag{3.2}$$

and for $\alpha > 0$ we have

$$\tfrac{1}{2}(V_\alpha(f|_{x_i=1}) + V_\alpha(f|_{x_i=-1})) \le V_\alpha(f) - \alpha \cdot Inf_i(f) \cdot V_0(f), \tag{3.3}$$

since $V_\alpha(f_1) \ge V_0(f_1) = Inf_i(f) \cdot V_0(f)$.
Define $S_\alpha(f) = \frac{V_\alpha(f)}{V_0(f)}$. We next prove that

$$\min\left(S_\alpha(f|_{x_i=1}), S_\alpha(f|_{x_i=-1})\right) \le S_\alpha(f) - \alpha \cdot Inf_i(f) \tag{3.4}$$

By combining (3.1) and (3.2) we get

$$S_\alpha(f) = \frac{V_\alpha(f)}{V_0(f)} = \frac{V_\alpha(f|_{x_i=1}) + V_\alpha(f|_{x_i=-1})}{V_0(f|_{x_i=1}) + V_0(f|_{x_i=-1})} + \frac{\alpha V_\alpha(f_1)}{V_0(f)} \ge \tag{3.5}$$

$$\min\left(\frac{V_\alpha(f|_{x_i=1})}{V_0(f|_{x_i=1})}, \frac{V_\alpha(f|_{x_i=-1})}{V_0(f|_{x_i=-1})}\right) + \frac{\alpha V_0(f_1)}{V_0(f)} = \tag{3.6}$$

$$\min\left(S_\alpha(f|_{x_i=1}), S_\alpha(f|_{x_i=-1})\right) + \alpha \cdot Inf_i(f) \tag{3.7}$$

Consider the polynomial f. We first bound $S_\alpha(f)$,

$$S_\alpha(f) = \frac{V_\alpha(f)}{V_0(f)} = \frac{\sum_I |f_I|^2 (1+\alpha)^{|I|}}{\sum_I |f_I|^2} \le (1+\alpha)^d.$$

Note that either $Inf_\infty(f) \le \delta$, or there exists a variable x_{i_1}, such that

$$\min\left(S_\alpha(f|_{x_{i_1}=1}), S_\alpha(f|_{x_{i_1}=-1})\right) \le S_\alpha(f) - \alpha \cdot \delta$$

Consider the restriction $f_{x_{i_1}=b_{i_1}}$ for $b_{i_1} \in \{-1,1\}$ minimizing $S_\alpha(f_{x_{i_1}=b_{i_1}})$. Either $\mathrm{Inf}_\infty(f_{x_{i_1}=b_{i_1}}) \leq \delta$, or otherwise we could find another variable x_{i_2} such that

$$\min\left(S_\alpha(f|_{x_{i_1}=b_{i_1},x_{i_2}=1}), S_\alpha(f|_{x_{i_1}=b_{i_1},x_{i_2}=-1})\right) \leq S_\alpha(f|_{x_{i_1}=b_{i_1}}) - \alpha \cdot \delta$$

Continuing in this fashion, since $S_\alpha \geq 0$, we must reach after at most $k \leq \frac{(1+\alpha)^d}{\alpha\delta}$ steps a polynomial $f|_{x_{i_1}=b_{i_1},\ldots,x_{i_k}=b_{i_k}}$ such that $\mathrm{Inf}_\infty(f|_{x_{i_1}=b_{i_1},\ldots,x_{i_k}=b_{i_k}}) \leq \delta$. Choosing optimally $\alpha = \frac{1}{d-1}$ we get $k \leq e \cdot d/\delta$. $\qquad\square$

We now show that if a polynomial has low influences, then so does its threshold.

Lemma 27. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial such that $\mathrm{Inf}_\infty(f) = \delta$. Let $h(x) = \mathrm{sgn}(f(x))$. Then*

$$\mathrm{Inf}_\infty(h) \leq O(d \cdot \delta^{1/8d}).$$

Proof. Assume w.l.o.g $\mathrm{Var}[f] = 1$, and we will bound $\mathrm{Inf}_i(h)$ for all $i = 1, \ldots, n$.

We first argue that if $\mathbb{E}[f^2]$ is large, then h has low influences. Let $f(x) = c + f_0(x)$, where c is the free coefficient of f. We have $\mathrm{Var}[f] = \mathbb{E}[f_0^2] = 1$ and $\mathbb{E}[f^2] = 1 + c^2$. The probability that $h(x) = h(0)$ is bounded by

$$\mathbb{P}[h(x) = h(0)] \leq \mathbb{P}[|f_0(x)| \geq c] \leq \frac{\mathbb{E}[f_0^2]}{c^2} = \frac{1}{c^2}.$$

Thus for large c we get a bound on the influence of h, since

$$\mathrm{Inf}_i(h) = \mathbb{P}[h(x) \neq h(x \oplus e_i)] \leq \mathbb{P}[h(x) \neq h(0)] + \mathbb{P}[h(x \oplus e_i) \neq h(0)] \leq 2/c^2.$$

In particular if $c > \delta^{-1/4}$ we get that $\mathrm{Inf}_i(h) \leq O(\delta^{1/2})$ and we are done. We thus assume from now on that $c \leq \delta^{-1/4}$.

Let $f(x) = x_i f_1(x) + f_2(x)$, where f_1, f_2 do not depend on x_i. By our assumption on the influences,

$$\mathbb{E}_x[f_1^2] = \mathrm{Inf}_i(f) \cdot \mathbb{E}[f^2] \leq \delta(1 + c^2) \leq 2\delta^{1/2}.$$

Set $a = \delta^{1/8}$ and consider the following two cases.

(*i*). $|f(x)| \leq a$

(*ii*). $|f_1(x)| \geq a$

If neither of these cases occur, then flipping x_i does not change the sign of f. Thus we can bound

$$\mathrm{Inf}_i(h) \leq \mathbb{P}[|f(x)| \leq a] + \mathbb{P}[|f_1(x)| \geq a].$$

We first estimate the first summand. By Lemma 19. Set $\tilde{\delta} \geq \max(\frac{d}{c_1}a^{1/d}, \frac{d}{c_2}\delta^{1/8d})$ where c_1, c_2 are the constants in Lemma 19. We get

$$\mathbb{P}[|f(x)| \leq a] \leq \tilde{\delta} = O(d \cdot \delta^{1/8d}).$$

18

We proceed by estimating the second summand. By Markov inequality and get

$$\mathbb{P}[|f_1(x)| \geq a] \leq \frac{E[f_1^2]}{a^2} \leq 2\delta^{1/4}.$$

Combining the two estimations we get that

$$\mathrm{Inf}_i(h) \leq O(d \cdot \delta^{1/8d}),$$

as desired.

$\square$

We next prove Lemma 22. Using Lemma 26 we prove the existence of a small depth decision tree, such that for most of its leaves, the polynomial restricted to the leaf has low influences. We use Lemma 27 to argue that when this happens also the threshold function has low influences.

Proof of Lemma 22. We first prove the theorem for a polynomial f, and then for a PTF h. We build a decision tree D in steps. At every step, some of the leaves of D will be *open*, and some will be *closed*. If a leaf ℓ is closed then $\mathrm{Inf}_\infty(f|_\ell) \leq \delta$. A leaf is open if it is not closed. Initially, our tree consists a single vertex, the root, which is open.

Let ℓ be an open leaf, and consider the polynomial $f|_\ell$. By Lemma 26, there exist a set of variables $x_{i_1}, \ldots, x_{i_k}$, $k \leq \frac{ed}{\delta}$ and an assignment to these variables $b_{i_1}, \ldots, b_{i_k} \in \{-1, 1\}$, such that

$$\mathrm{Inf}_\infty(f|_{\ell, x_{i_1} = b_{i_1}, \ldots, x_{i_k} = b_{i_k}}) \leq \delta.$$

We add under a ℓ a subtree whose leaves correspond to all the 2^k possible assignments of $x_{i_1}, \ldots, x_{i_k}$. Note that at least one of the leaves in the new tree is closed, and the other leaves may be either closed or open. Therefore, a random walk of length k that starts at ℓ will end at a closed leaf with probability at least 2^{-k}.

This process defines a tree D' of depth at most n, as every variable appears in every path at most once. Let $D(t)$ be the tree obtained by truncating D' at depth $t \cdot 2^k$. Namely, the depth of $D(t)$ is $t \cdot 2^k$. The probability that a random walk that start from the root will end at open leaf is at most $(1 - 2^{-k})^t \leq e^{-2^{-k}t}$. Thus, setting, $t = \log(1/\epsilon) \cdot 2^{ed/\delta}$ will guarantee that a random leaf in D is closed with probability at least $1 - \epsilon$, as required.

We proceed by proving the second item. Let h be a PTF as stated, and observe that by Lemma 27, for any leaf ℓ for which $\mathrm{Inf}_\infty(f|_\ell) \leq \delta$ we have that $\mathrm{Inf}_\infty(\mathrm{sgn}(f|_\ell)) \leq O(d\delta^{1/8d}) = \delta'$. Since $\mathrm{sgn}(f|_\ell) = \mathrm{sgn}(f)|_\ell = h|_\ell$, we get

$$\mathbb{P}_{\ell \in L(D)}[\mathrm{Inf}_\infty(h|_\ell) > \delta'] < \varepsilon.$$

$\square$

4   Fooling threshold of polynomials depending on a few linear functions

Recall that the weight of a polynomial $G : \mathbb{R}^m \to \mathbb{R}$ is the sum of the absolute values of the coefficients of its monomials, excluding the free coefficient. Our main result in this section is Theorem 28, which is stated below.

Theorem 28 (Theorem 1, restated). *Fix $\varepsilon > 0$. Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree-d polynomial, which can be decomposed as $f(x) = G(g_1(x), \ldots, g_m(x))$ where*

(i). The functions $g_1, \ldots, g_m$ are linear with $\mathbb{E}[g_1^2] = \ldots = \mathbb{E}[g_m^2] = 1$.

(ii). G is a degree-d polynomial.

Then k-wise distributions ε-fool $\mathrm{sgn}(f)$ for $k = exp(O(d/\varepsilon)^d) + poly((\log m \cdot d/\varepsilon)^d, m, wt(G))$.

The main lemma shows that any multivariate Lipschitz function F admits a polynomial p with the following two properties. The polynomial p bounds F from above everywhere, and p approximates F in a cube around the origin.

Lemma 29. *Let $F : \mathbb{R}^m \to [-1, 1]$ be a Lipschitz function. Let $A > 0$ and $0 < \varepsilon < 1$ be arbitrary. There exists a degree-k polynomial $p(z_1, \ldots, z_m)$ such that*

(i). For every $z \in \mathbb{R}^m$, $p(z) \geq F(z)$.

(ii). For every $z \in [-A, A]^m$, $p(z) \leq F(z) + \varepsilon$.

and $k \leq O\left(\frac{A \cdot m^{3/2} \cdot L(F)}{\epsilon}\right)$.

The proof of Lemma 29 appears in Subsection **??**

We next apply Lemma 29 to show that k-wise distributions fool any boolean function $f : \{-1, 1\}^n \to \mathbb{R}$ with the following properties. The function f be decomposed as $f(x) = G(g_1(x), \ldots, g_m(x))$, where $g_1, \ldots, g_m$ are linear functions, the polynomial G is Lipschitz, and the distribution of f is not too concentrated around any specific value.

Lemma 30. *Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a function which can be decomposed as $f(x) = G(g_1(x), \ldots, g_m(x))$ where*

(i). The functions $g_1, \ldots, g_m : \{-1, 1\}^n \to \mathbb{R}$ are linear with $\mathbb{E}[g_1^2], \ldots, \mathbb{E}[g_m^2] \leq 1$.

(ii). The function $G : \mathbb{R}^m \to \mathbb{R}$ is continuous and Lipschitz on the cube $[-C, C]^m$, for $C = 100\sqrt{\log(m/\varepsilon)}$.

(iii). The function f is anti-concentrated, $\mathbb{P}_x[|f(x)| < \alpha] < \varepsilon/100$ for some α depending on ε.

Then there exists a degree-k polynomial $p : \{-1, 1\}^n \to \mathbb{R}$ such that

- $p(x) \geq \operatorname{sgn}(f(x))$ *for all* $x \in \{-1, 1\}^n$.

- $\mathbb{E}_{x \in U}[p(x) - \operatorname{sgn}(f(x))] \leq \varepsilon$.

where $k = O(\frac{dm^5 L^2}{\alpha^2 \varepsilon^2} \ln(mL/\alpha\varepsilon))$ *and* $L = \max(L_{[-C,C]^m}(g), 1)$.

The following claim bounds the Lipschitz constant of degree-d polynomials.

Claim 31. *Let* $G : \mathbb{R}^m \to \mathbb{R}$ *be a degree-d polynomial. The Lipschitz constant of G on* $[-C, C]^m$ *is bounded by* $dC^{d-1} \cdot wt(G)$.

Proof. We start by bounding the Lipschitz constant of monomials on $[-C, C]^m$. We then will get the result for G by the additivity of the Lipschitz constant.

Let M be a monomial $M(z_1, \ldots, z_d) = \prod z_i$. Let $z, z' \in [-C, C]^m$ such that $\|z - z'\|_\infty \leq \varepsilon$. Let $z'_i = z_i + e_i$ where $|e_i| \leq \varepsilon$. We have

$$|M(z') - M(z)| = | \sum_{k=1}^{d} \left(\prod_{i=1}^{k}(z_i + e_i) \prod_{i=k+1}^{d} z_i - \prod_{i=1}^{k-1}(z_i + e_i) \prod_{i=k}^{d} z_i \right) | \leq$$

$$\sum_{k=1}^{d} \prod_{i=1}^{k-1} |z_i| \prod_{i=k+1}^{d} |z_i + e_i| e_i \leq$$

$$dC^{d-1}\varepsilon.$$

Hence $L_{[-C,C]^m}(M) \leq dC^{d-1}$.

Write $G(z) = \sum_{I \subset [m], |I| \leq d} \alpha_I M_I(z)$ where M_I are monomials. The Lipschitz constant of G on $[-C, C]^m$ is thus bounded by $\sum |\alpha_I| L_{[-C,C]^m}(M_I) \leq dC^{d-1} \cdot wt(G)$. $\square$

We proceed to the proof of Theorem 28.

Proof of Theorem 28 . Let $f(x)$ be a degree-d polynomial, which can be decomposed as $f(x) = G(g_1(x), \ldots, g_m(x))$ where $g_1, \ldots, g_m$ are linear and $\mathbb{E}[g_1^2] = \ldots = \mathbb{E}[g_m^2] = 1$. Set $\delta = O(\varepsilon/d)^{8d}$. By Lemma 22 there exists a decision tree D of depth at most $exp(O(d^{8d+1}/\varepsilon^{8d})$ such that

$$\mathbb{P}_{\ell \in L(D)}[\operatorname{Inf}_\infty(f|_\ell) > \delta] < \varepsilon/100.$$

By Lemma 13 we have for each linear function g_i

$$\mathbb{P}_{\ell \in L(D)}[\mathbb{E}[((g_i)|_\ell)^2] \geq t] \leq \varepsilon/100m$$

for $t = O(\log \varepsilon/m)$. Thus with probability $1 - \varepsilon/100$, we have both that $\operatorname{Inf}_\infty(f|_\ell) \leq \delta$ and $\mathbb{E}[((g_i)|_\ell)^2] \leq t$ for all $i \in [m]$. Fix such ℓ. Since $f|_\ell$ has low influences, Lemma 19 gives

$$\mathbb{P}_{x \in U}[|f|_\ell(x)| \leq \alpha] < \varepsilon/1000.$$

for $\alpha = O(\varepsilon/d)^d$.

Let g'_i be a normalization of $(g_i)_\ell$ such that $\mathbb{E}[(g'_i)^2] = 1$. We can write $f_\ell(x) = G'(g'_1(x), \ldots, g'_m(x))$ where $wt(G') \leq wt(G) \cdot t$. By Claim 31 we have $L_{[-C,C]^m}(G) \leq$

$dC^{d-1} \cdot wt(G)$ for $C = 100\sqrt{\log(m/100\varepsilon)}$. Applying Lemma 30 we get there exists a degree-k polynomial $p_u(x)$ such that both $p_u(x) \geq \mathrm{sgn}(f|_\ell(x))$ for all $x \in \{-1,1\}^n$, and $\mathbb{E}_{x \in U}[p_u(x) - \mathrm{sgn}(f|_\ell(x))] \leq \varepsilon/10$. Applying the same reasoning on the polynomial $-f(x)$ we get there exists a degree-k polynomial $p_l(x)$ such that both $p_l(x) \leq \mathrm{sgn}(f|_\ell(x))$ for all $x \in \{-1,1\}^n$ and $\mathbb{E}_{x \in U}[\mathrm{sgn}(f|_\ell(x)) - p_l(x)] \leq \varepsilon/10$. Combining the two bounds we conclude that k-wise distributions $\varepsilon/10$-fool $f|_\ell$. Since this holds for $1 - \varepsilon/100$ fraction of the leaves ℓ, we get by Claim 12 that $k' = k + \mathrm{depth}(D)$ independence ε-fool f.

We conclude by bounding k and k'. We have $k = O(\frac{dm^5 L^2}{\alpha^2 \varepsilon^2} \log(mL/\alpha\varepsilon)) = O(d/\varepsilon)^d \cdot m^5 wt(G)^2 \log(m/\varepsilon)^d \cdot O(\log(d \cdot m \cdot wt(G)/\varepsilon))$, and $\mathrm{depth}(D) = exp((d/\varepsilon)^{O(d)}$, hence we have $k' = exp((d/\varepsilon)^{O(d)} + poly(O(\log m \cdot d/\varepsilon)^d, m, wt(G))$, as claimed. $\qquad\square$

4.1   Proof of Lemma 29

Our starting point is a fundamental result in the theory of approximation theory. Roughly speaking, it says that any Lipschitz function can be well approximated by a low-degree polynomial on a bounded region. Explicitly we use the following result of Ganzburg [17].

Lemma 32 (Multidimensional Jackson-type theorem, Theorem 1 in [17]). *Let $F : \mathbb{R}^m \to \mathbb{R}$ be a Lipschitz function. For every k there is a degree-k polynomial $p_k(z_1, \ldots, z_m)$, such that*

$$\sup_{z \in [-1,1]^m} |F(z) - p_k(z)| \leq C \cdot \frac{m^{3/2} L(F)}{k}$$

where C is an absolute constant.

We get the following corollary.

Corollary 33. *Let $F : \mathbb{R}^m \to \mathbb{R}$ be a Lipschitz function. For every $\varepsilon > 0$ there exists $k = O(m^{3/2} L(F)/\varepsilon)$ and a degree k polynomial p_k such that*

- $p_k(z) \geq F(z)$ for all $z \in [-1,1]^m$

- $p_k(z) - F(z) \leq \varepsilon$ for all $z \in [-1,1]^m$.

Proof. Let p_k be the polynomial obtained by Lemma 32 such that $\sup_{z \in [-1,1]^m} |F(z) - p_k(z)| < \varepsilon/2$, and take $p'_k(z) = p_k(z) + \varepsilon/2$. $\qquad\square$

We also need the following bound on the growth of real polynomials.

Lemma 34. *let $g(w)$ be a univariate degree-k polynomial. Then for every $w \in \mathbb{R}$,*

$$|g(w)| \leq (\max_{w \in [-1,1]} |g(w)|) \cdot |w + \sqrt{w^2 - 1}|^k.$$

We will need the following corollary of Lemma 34.

22

Lemma 35. *Let $p(z_1, \ldots, z_m)$ be a degree-k polynomial, such that $p(z) \leq c$ for all $z \in [-1, 1]^m$. If $|z_i| \geq |z_j|$ for every $1 \leq i, j \leq m$, then*

$$|p(z)| \leq c \cdot \max(|2z_i|^k, 1).$$

Proof. Assume w.l.o.g that $|z_1| \geq |z_i|$ for every $i \in \{1, \ldots, m\}$. If $|z_1| \leq 1$ that $(z_1, \ldots, z_m) \in [-1, 1]^m$ and by assumption $p(z) \leq c$. Otherwise consider the following univariate polynomial $g(w)$ that is obtained by restricting p to the line passing through zero and z, defined as

$$g(w) = p(w, wz_2/z_1, \ldots, wz_m/z_1).$$

When $w \in [-1, 1]$, we have $(w, wz_2/z_1, \ldots, wz_m/z_1) \in [-1, 1]^m$. Hence $\max_{w \in [-1,1]} g(w) \leq c$. Applying Lemma 34 we get that

$$|p(z)| = |g(z_1)| \leq c \cdot |z_1 + \sqrt{z_1^2 - 1}|^k \leq c \cdot |2z_1|^k.$$

□

We are now ready to state and prove the main lemma that will be used to prove Lemma 29.

Lemma 36. *Let $F : \mathbb{R}^m \to [-1, 1]$ be a Lipschitz function. For every $0 < \varepsilon < 1$ there exists a degree-k polynomial p'_k such that*

- $p'_k(z) \geq F(z)$ *for all $z \in \mathbb{R}^m$.*

- $p'_k(z) - F(z) \leq \varepsilon$ *for all $z \in [-1/4, 1/4]^m$.*

where $k = O(m^{3/2} L(F)/\varepsilon)$.

Proof. Let p_k be the polynomial guaranteed by Corollary 33 for error $\varepsilon/2$. Set $k' \geq \max(k, 4m/\varepsilon)$ be an even integer, and define

$$p'_k(z_1, \ldots, z_m) = p_k(z_1, \ldots, z_m) + 4\left((2x_1)^{k'} + \ldots + (2x_m)^{k'}\right).$$

We will prove that $p'_k(z) \geq F(z)$ for all $z \in \mathbb{R}^m$, and $p'_k(z) \leq F(z) + \varepsilon$ for $z \in [-1/4, 1/4]^m$.

Let $z \in \mathbb{R}^m$ be arbitrary. If $z \in [-1, 1]^m$ we already have that $p'_k(z) \geq p_k(z) \geq F(z)$. Otherwise, assume w.l.o.g that $|z_1| \geq \max(|z_2|, \ldots, |z_m|)$, and hence $|z_1| > 1$.

Since p_k approximates F with error $\varepsilon < 1$ on $[-1, 1]^m$, we have that $|p_k(z)| \leq 2$ for all $z \in [-1, 1]^m$. Applying Lemma 35 we get that

$$p_k(z) \leq 2|2z_1|^k.$$

Thus in particular, $p_k(z) \geq -2|2z_1|^k$. By our definition of $p'_k(z)$ we get that

$$
\begin{aligned}
p'_k(z) &= p_k(z) + 4\left((2x_1)^{k'} + \ldots + (2x_m)^{k'}\right) \\
&\geq -2|2z_1|^k + 4((2z_1)^{k'} + \ldots + (2z_m)^{k'}) \\
&\geq -2|2z_1|^k + 4(2z_1)^{k'} \\
&= -2|2z_1|^k + 4|2z_1|^{k'} \\
&\geq -2|2z_1|^k + 4|2z_1|^k \\
&= 2|2z_1|^k \geq 1.
\end{aligned}
$$

23

and in particular we get that
$$p'_k(z) \geq F(z).$$

We next estimate the obtained approximation of p'_k in $[-1/4, 1/4]^m$. Observe that for $z \in [-1/4, 1/4]^m$,
$$|p'_k(z) - p_k(z)| \leq 4m2^{-k'}$$

and by our choice of k', we have that $|p'_k(z) - P_k(z)| \leq \varepsilon/2$. Since p_k approximates F on $[-1, 1]^m$ with error $\varepsilon/2$, it does so in particular in $[-1/4, 1/4]^m$. Hence we get

$$\max_{z \in [-1/4, 1/4]^m} p'_k(z) - F(z) \leq \varepsilon.$$

$\square$

The proof of Lemma 29 now follows as an immediate corollary of Lemma 36.

Proof of Lemma 29. Let $F : \mathbb{R}^m \to [-1, 1]$ be a Lipschitz function. Define $F'(z) = F(z/4A)$, and apply Lemma 36 on F' to obtain a polynomial p'_k such that $p'_k(z) \geq F'(z)$ for all $z \in \mathbb{R}^m$ and $p'_k(z) \leq F'(z) + \varepsilon$ for $z \in [-1/4, 1/4]^m$. The polynomial $p(z) = p'_k(4A \cdot z)$ is the desired approximation polynomial for F. The bound on the degree follows from Lemma 36 since $L(F') = 4A \cdot L(F)$. $\square$

4.2   Proof of Lemma 30

We start with the following definition.

Definition 37 (zero-set). For $G : \mathbb{R}^m \to \mathbb{R}$ we define its zero-set, denoted $\mathcal{Z}(G)$ to be

$$\mathcal{Z}(G) = \{z \in \mathbb{R}^m : G(z) = 0\}.$$

Lemma 38. *Let $G : \mathbb{R}^m \to \mathbb{R}$ be a continuous real function. For every $\tau > 0$ there exists a function $\tilde{G} : \mathbb{R}^m \to [-1, 1]$ such that*

- $\tilde{G}(z) \geq \operatorname{sgn}(G(z))$ *for all $z \in \mathbb{R}^m$.*

- *For every $z \notin \mathcal{C}(\mathcal{Z}(G), \tau)$, $\tilde{G}(z) = \operatorname{sgn}(G(z))$.*

- $L(\tilde{G}) \leq O(m/\tau)$.

Proof. Set $\tau' = \tau/2$ and define

$$G'(z) = \max_{z' \in \mathcal{C}(z, \tau')} \operatorname{sgn}(G(z))$$

and

$$\tilde{G}(z) = \frac{1}{|\mathcal{C}(z, \tau')|} \int_{z' \in \mathcal{C}(z, \tau')} G'(z') dz'.$$

24

First we argue that $\tilde{G}(z) \geq \text{sgn}(G(z))$ for all $z \in \mathbb{R}^m$. Since for every $z' \in \mathcal{C}(z, \tau')$, $G'(z') \geq \text{sgn}(G(z))$. By definition, $\tilde{G}(z)$ is defined as the average of $G'(z')$ over $z' \in \mathcal{C}(z, \tau')$, we get that $\tilde{G}(z) \geq \text{sgn}(G(z))$.

We continue by showing that $\tilde{G}(z) = \text{sgn}(G(z))$ for $z \notin \mathcal{C}(\mathcal{Z}(G), \tau)$. For $z' \in \mathcal{C}(z, \tau)$ we have $\text{sgn}(G(z')) = \text{sgn}(G(z))$, since G is continuous and has no zeros in $\mathcal{C}(z, \tau)$. As $\mathcal{C}(z', \tau') \subset \mathcal{C}(z, \tau)$, we have $G'(z') = \text{sgn}(G(z))$, and hence we conclude that $\tilde{G}(z) = \text{sgn}(G(z))$.

We next bound $L(G)$. Let $z', z'' \in \mathbb{R}^m$. We consider the following two cases. If $\|z' - z''\|_\infty \geq \tau'$ then since $\tilde{G}$ is bounded, i.e. $|\tilde{G}|_\infty \leq 1$, we have

$$\frac{|\tilde{G}(z') - \tilde{G}(z'')|}{\|z' - z''\|_\infty} \leq 2/\tau'.$$

Otherwise, if $\|z' - z''\|_\infty < \tau'$, we have

$$|\tilde{G}(z') - \tilde{G}(z'')| = |\frac{1}{(2\tau')^m} \left(\int_{t \in \mathcal{C}(z', \tau')} G'(t)dt - \int_{t \in \mathcal{C}(z'', \tau')} G'(t)dt \right)| \leq$$

$$\frac{1}{\tau^m} \int_{t \in \mathcal{C}(z', \tau') \triangle \mathcal{C}(z'', \tau')} |G'(t)|dt \leq$$

$$\frac{|\mathcal{C}(z', \tau') \triangle \mathcal{C}(z'', \tau')|}{\tau^m}$$

where $\triangle$ denotes the symmetric difference between two sets.

A straight forward calculation shows that

$$|\mathcal{C}(z', \tau') \triangle \mathcal{C}(z'', \tau')| \leq O(m(2\tau')^{m-1}\|z' - z''\|_\infty).$$

Hence we get

$$\frac{|\tilde{G}(z') - \tilde{G}(z'')|}{\|z' - z''\|_\infty} \leq O(m/\tau').$$

$\square$

Lemma 39. *Let $f(x) = G(g_1(x), \ldots, g_m(x))$ as in the definition of Lemma 30 and assume that the assumptions of Lemma 30 hold. Then*

$$\mathbb{P}_{x \in \{-1,1\}^n}[(g_1(x), \ldots, g_m(x)) \in \mathcal{C}(\mathcal{Z}(G), \tau)] \leq \varepsilon/10$$

for $\tau = \alpha/L$.

Proof. We consider two cases, the first when $(g_1(x), \ldots, g_m(x)) \in [-(C - \tau), C - \tau]^m$, and the second when $(g_1(x), \ldots, g_m(x)) \notin [-(C - \tau), C - \tau]^m$.

In the first case, let $x \in \{-1,1\}^n$ be such that $(g_1(x), \ldots, g_m(x)) \in [-(C - \tau), C - \tau]^m \cap \mathcal{C}(\mathcal{Z}(G), \tau)$. We will prove that $|f(x)| < \alpha$, and by our assumption the probability over all $\{-1,1\}^n$ that $|f(x)| < \alpha$ is bounded by $\varepsilon/10$. To show that $|f(x)| < \alpha$, let $z = (g_1(x), \ldots, g_m(x)) \in \mathbb{R}^m$. z is in L_∞ distance of at most τ from a zero z_0 of G, and since

25

$z \in [-(C-\tau), C-\tau]^m$, we get that $z_0 \in [-C, C]^m$. Since G is Lipschitz on $[-C, C]^m$, we conclude that

$$G(z) \le G(z_0) + L_{[-C,C]^m} \cdot \|z - z_0\|_\infty \le L_{[-C,C]^m} \cdot \tau \le \alpha.$$

We now consider the second case, that $(g_1(x), \ldots, g_m(x)) \notin [-(C-\tau), C-\tau]^m$. We will bound the probability that this event occurs. By our construction $\tau \le 1$, hence it is enough to bound the probability that $(g_1(x), \ldots, g_m(x)) \notin [-(C-1), C-1]^m$, i.e. $|g_i(x)| \ge C - 1$ for some $i \in [m]$. Since we assumed each g_i is δ-normal, we get that

$$\mathbb{P}[|g_i(x)| \ge C - 1] \le 2(\delta + \mathbb{P}[N \ge C - 1])$$

where $N \sim N(0, 1)$ is a standard normal variable. Using standard normal estimations and setting $C = O(\sqrt{\log(m/\varepsilon)})$ gives

$$\mathbb{P}[|g_i(x)| \ge C - 1] \le 2(\delta + \varepsilon/100m)$$

since $\delta < \varepsilon/100m$ we get that $\mathbb{P}[|g_i(x)| \ge C - 1] \le \varepsilon/10m$, and using the union bound over all $g_1, \ldots, g_m$ we get that the total error is bounded by $\varepsilon/10$. $\qquad\square$

The following lemma bounds the tail moments of linear functions, and is somewhat similar to Lemma 4.2 in [13].

Lemma 40. *Let* $g : \{-1, 1\}^n \to \mathbb{R}$ *be a linear function with* $\mathbb{E}[g^2] = 1$. *Let* $c > 0$ *and* $A \ge 2c$. *Then*

$$\mathbb{E}_{x \in \{-1,1\}^n}[|g(x)|^{cA} 1_{|g(x)| \ge A}] \le 3e^{2cA \ln(A) + 2c^2 - \frac{1}{2}(A - 2c)^2}.$$

Proof. Define $E_t = \mathbb{E}_{x \in \{-1,1\}^n}[|g(x)|^{cA} 1_{i \le |g(x)| < i+1}]$. We have to bound $E = \sum_{i \ge A} E_i$. By Hoeffding bound (see, e.g., [5]),

$$\mathbb{P}_{x \in U}[|g(x)| \ge i] \le 2e^{-i^2/2}.$$

Hence we get $E_i \le 2e^{-i^2/2}(i+1)^{cA}$. Therefore

$$(i+1)^{cA} \le i^{2cA} = A^{2cA} \cdot (i/A)^{2cA} \le A^{2cA} \cdot e^{2ci}$$

where we used the fact that $x \le e^x$ for $x = i/A$. Summing over $i \ge A$ we get

$$E \le A^{2cA} \sum_{i \ge A} e^{-i^2/2 + 2ci} =$$

$$A^{2cA} \sum_{i \ge A} e^{-\frac{1}{2}(i - 2c)^2 + 2c^2} \le$$

$$3A^{2cA} e^{2c^2} e^{-\frac{1}{2}(A - 2c)^2}).$$

where we used the fact that $\sum_{i \ge C} e^{-\frac{1}{2}i^2} \le \sum_{i \ge C^2} e^{-\frac{1}{2}i} \le 3e^{-C^2/2}$. $\qquad\square$

We are now ready to prove Lemma 30.

Proof of Lemma 30. Set $A > 1$ to be determined later. Let $\tilde{G} : \mathbb{R}^m \to [-1, 1]$ be the Lipschitz function approximating and bounding $\text{sgn}(G)$ guaranteed by Lemma 38. Let $p : \mathbb{R}^m \to \mathbb{R}$ be the polynomial guaranteed by Lemma 29 approximating $\tilde{G}$ on $[-A, A]^m$ with error $\varepsilon/10$. The degree of p is $k_1 = O(\frac{Am^{3/2}L(\tilde{G})}{\varepsilon}) = A \cdot \phi(\varepsilon)$, where $\phi(\varepsilon) = O(\frac{m^{5/2}L}{\alpha\varepsilon})$ is independent of our choice of A. Set $p^* : \{-1, 1\}^n \to \mathbb{R}$ to be defined as

$$p^*(x) = p(g_1(x), \ldots, g_m(x)).$$

We have that

- The polynomial p^* is of degree at most $A \cdot \phi(\varepsilon)$.

- For all $x \in \{-1, 1\}^n$, $p^*(x) \geq \text{sgn}(f(x))$.

- For all $x \in [-A, A]^m$ such that $(g_1(x), \ldots, g_m(x)) \notin \mathcal{C}(\mathcal{Z}(G), \tau)$ we have $p^*(x) \leq \text{sgn}(f(x)) + \varepsilon/10$.

- For all $x \in [-A, A]^m$ such that $(g_1(x), \ldots, g_m(x)) \in \mathcal{C}(\mathcal{Z}(G), \tau)$ we have $p^*(x) \leq 2$.

To conclude the proof we have to show that $\mathbb{E}_x[p^*(x) - \text{sgn}(f(x))] \leq \varepsilon$. We consider three ranges of values for x.

(i). $x \in \{-1, 1\}^n$ such that $(g_1(x), \ldots, g_m(x)) \in [-A, A]^m \setminus \mathcal{C}(\mathcal{Z}(G), \tau)$.

(ii). $x \in \{-1, 1\}^n$ such that $(g_1(x), \ldots, g_m(x)) \in [-A, A]^m \bigcap \mathcal{C}(\mathcal{Z}(G), \tau)$.

(iii). $x \in \{-1, 1\}^n$ such that $(g_1(x), \ldots, g_m(x)) \notin [-A, A]^m$.

To bound (i), we use the fact that for all x such that $(g_1(x), \ldots, g_m(x)) \in [-A, A]^m \setminus \mathcal{C}(\mathcal{Z}(G), \tau)$ we know that $p^*(x) - \text{sgn}(f(x)) \leq \varepsilon/10$, hence the total contributed error is bounded by $\varepsilon/10$.

To bound (ii), we use Lemma 39 to conclude that the probability over $x \in \{-1, 1\}^n$ that $(g_1(x), \ldots, g_m(x)) \in \mathcal{C}(\mathcal{Z}(G), \tau)$ is bounded by $\varepsilon/10$. Since we know that for such x we have $p^*(x) \leq 2$ and $\text{sgn}(f(x)) \geq -1$, we can bound the total error by $3/10\varepsilon$.

Finally, let ε_3 be the error in (iii). Namely,

$$\varepsilon_3 = \mathbb{E}_x\left[(p(g_1(x), \ldots, g_m(x)) - \text{sgn}(f(x))) \cdot 1_{(g_1(x), \ldots, g_m(x)) \notin [-A, A]^m}\right].$$

We bound ε_3 by the union bound over which of $g_1(x), \ldots, g_m(x)$ is maximal.

$$\varepsilon_3 \leq \sum_{i=1}^m \mathbb{E}_x\left[(p(g_1(x), \ldots, g_m(x)) - \text{sgn}(f(x))) \cdot 1_{(g_1(x), \ldots, g_m(x)) \notin [-A, A]^m} \cdot 1_{g_i(x) = \max(g_1(x), \ldots, g_m(x))}\right].$$

Since $|p(z)| \leq 2$ for $z \in [-1, 1]^m$ and $|\mathrm{sgn}(f(x))| = 1$, by Lemma 35 we get

$$\varepsilon_3 \leq \sum_{i=1}^{m} \mathbb{E}_x \left[(2|2g_i(x)|^{\deg(p)} + 1) \cdot 1_{|g_i(x)| \geq A} \right]$$

$$\leq 2^{\deg(p)+2} \sum_{i=1}^{m} \mathbb{E}_x \left[|g_i(x)|^{\deg(p)} \cdot 1_{|g_i(x)| \geq A} \right]$$

Recall that $\deg(p) = k_1 = A \cdot \phi(\varepsilon)$. Using Lemma 40 we get the bound

$$\varepsilon_3 \leq 3m2^{cA} e^{2cA \ln A + 2c^2 - \frac{1}{2}(A-2c)^2}$$

where $c = \phi(\varepsilon)$. Recall that $\phi(\varepsilon) > m/\varepsilon$, hence we get that picking $A = \Omega(c \ln c) = \Omega(\phi(\varepsilon) \ln(\phi(\varepsilon)))$ will yield $\varepsilon_3 \leq \varepsilon/10$. $\qquad\square$

References

[1] M. Ajtai and A. Wigderson, *Determinstic simulation of probabilistic constant depth circuits*, Proceedings of the 26th Annual Symposium on Foundations of Computer Science (FOCS), 1985, 11–19.

[2] N. Alon, L. Babai and A. Itai, *A fast and simple randomized algorithm for the maximal indpendent set problem*, J. of algorithms, 7:657-583, 1986.

[3] N. Alon, I. Ben-Eliezer and M. Krivelevich, *Small sample spaces cannot fool low degree polynomials*, proceedings of the 12th International Workshop on Randomization and Computation (RANDOM 2008), 266–275.

[4] N. Alon, O. Goldreich, J. Håstad and R. Peralta,*Simple constructions of almost k-wise independent random variables*, Random Structures and Algorithms 3 (1992), 289–304.

[5] N. Alon and J. Spencer, **The probabilistic method**, Wiley, 2008.

[6] . J. Aspnes, R. Beigel, M. Furst and S. Rudich, *The expressive power of voting polynomials*, The 23th ACM Symposium on Theory of Computing (STOC), pages 402–409.

[7] L. M. J. Bazzi, *polylogarithmic indepencdence can fool DNF formulas*, Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS), 2007, pp. 63–73.

[8] I. Benjamini, O. Gurel-Gurevich and R. Peled, *K-wise Independent Distributions, boolean Functions and Percolation*, manuscript.

[9] A. Bogdanov and E. Viola, *Pseudorandom bits for polynomials*, Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS), 2007, 41–51.

[10] M. Braverman, *poly-logarithmic independene fools AC^0 circuits*, Proceedings of the 40th Annual Symposium on Foundations of Computer Science (FOCS), 2009, to appear.

[11] , J. Bruck, *Harmonic Analysis of Polynomial Threshold Functions*, SIAM J. Discrete Math. Volume 3, Issue 2, pp. 168-177,1990.

[12] , A. Carbery and J. Wright, *Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$*, Math. Res. Lett., 8(3), 233-248, 2001.

[13] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio and E. Viola, *Bounded independence fools halfspaces*, to appear in Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2009, to appear.

[14] I. Diakonikolas, D.M. Kane and J. Nelson, *Bounded Independence Fools Degree-2 Threshold Functions*, preprint. Arxiv:0911.3389.

[15] I. Diakonikolas, P. Raghavendra R. Servedio and L. Tan, *Average sensitivity and noise sensitivity of polynomial threshold functions*, manuscript.

[16] I. Diakonikolas, R. Servedio, L. Tan and A. Wan, *A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions*, manuscript.

[17] M.I. Ganzburg, *The theorems of Jackson and Berstein in $\mathcal{R}^m$*, Russian Mathematical Surveys, 34 221-222, 1979.

[18] A. Hajnal, W. Maass, P. Pudlak, M. Szegedy, and G. Turan, *Threshold circuits of bounded depth.* Journal of Computer and System Sciences, 46:129-154, 1993.

[19] P. Harsha, A. Klivans and R. Makhu, *Bounding the sensitivity of polynomial threshold functions*, manuscript.

[20] N. Linial and N. Nisan, *Approximate inclusion-exclusion*, Combinatorica, 10(1990) 349-365.

[21] S. Lovett, *Unconditional pseudorandom generators for low degree polynomials*, Proceedings of the 40th Annual ACM Symposium, STOC 2008, 557-562.

[22] S. Lovett, O. Reingold, L. Trevisan and S. Vadhan, *Pseudorandom Bit Generators that Fool Modular Sums*, proceedings of the 13th International Workshop on Randomization and Computation (RANDOM), 2009, pp. 615-630.

[23] E. Mossel, R. O'Donnell and K. Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality*, Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2005, pp. 21-30.

[24] R. Meka and D. Zuckerman, *Pseudorandom generators for polynomial threshold functions*, manuscript.

[25] J. Naor and M. Naor,*Small bias probability spaces: efficient constructions and applications*, Proceedings of the 22th Annual ACM Symposium (STOC), 1990, pp. 213-223.

[26] Yuval Rabani and Amir Shpilka,*Explicit construction of a small epsilon-net for linear threshold functions*, Proceedings of the 41th Annual ACM Symposium (STOC), 2009, pp. 649-658.

[27] E. Viola,*The sum of d small-bias generators fools polynomials of degree d*, Proceedings of the 23th IEEE Conference on Computational Complexity (CCC), 2008, pp. 124-127.