THEORY OF COMPUTING, Volume 10 (2), 2014, pp. 27–53
www.theoryofcomputing.org

A Regularity Lemma and Low-Weight Approximators for Low-Degree Polynomial Threshold Functions

Ilias Diakonikolas * Rocco A. Servedio† Li-Yang Tan‡
Andrew Wan§

Received September 8, 2012; Revised September 27, 2013; Published March 25, 2014

Abstract: We give a "regularity lemma" for degree-d polynomial threshold functions (PTFs) over the Boolean cube $\{-1,1\}^n$. Roughly speaking, this result shows that every degree-d PTF can be decomposed into a constant number of subfunctions such that almost all of the subfunctions are close to being regular PTFs. Here a "regular" PTF is a PTF $\mathrm{sign}(p(x))$ where the influence of each variable on the polynomial $p(x)$ is a small fraction of the total influence of p.

A conference version of this paper appeared in the *Proc. 25th Ann. IEEE Conf. on Computational Complexity, CCC 2010* [11].

*`ilias.d@ed.ac.uk`. Supported in part by a Simons Postdoctoral Fellowship. Most of this work was done at Columbia University supported by NSF grant CCF-0728736, and by an Alexander S. Onassis Foundation Fellowship. Part of this research was done while visiting IBM Almaden.

†`rocco@cs.columbia.edu`. Supported by NSF grants CNS-0716245, CCF-0915929, and CCF-1115703.

‡`liyang@cs.columbia.edu`. Supported by DARPA award no. HR0011-08-1-0069 and NSF CyberTrust grant no. CNS-0716245.

§`atw12@eecs.berkeley.edu`. This work was done at Columbia University supported by NSF CyberTrust award CNS-0716245.

ACM Classification: F.1.3, G.2.0

AMS Classification: 68Q15, 68R01

Key words and phrases: complexity theory, Boolean functions, polynomials, threshold functions

 DOI: 10.4086/toc.2014.v010a002

As an application of this regularity lemma, we prove that for any constants $d \geq 1, \varepsilon > 0$, every degree-$d$ PTF over n variables can be approximated to accuracy ε by a constant-degree PTF that has integer weights of total magnitude $O_{\varepsilon,d}(n^d)$. This weight bound is shown to be optimal up to logarithmic factors.

1 Introduction

A polynomial threshold function (henceforth PTF) is a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$,

$$f(x) = \mathrm{sign}(p(x)),$$

where $p : \{-1,1\}^n \to \mathbb{R}$ is a polynomial with real coefficients. If p has degree d, we say that f is a *degree-d* PTF. Low-degree PTFs are a natural generalization of linear threshold functions (the case $d = 1$) and hence are of significant interest in complexity theory, see e. g., [1, 4, 29, 28, 9, 13, 16, 25, 32, 34] and many other works.

The *influence* of coordinate i on a function $g : \{-1,1\}^n \to \mathbb{R}$ measures the extent to which x_i affects the output of g. More precisely, we have $\mathrm{Inf}_i(g) = \sum_{S \ni i} \widehat{g}(S)^2$, where $\sum_{S \subseteq [n]} \widehat{g}(S)\chi_S(x)$ is the Fourier expansion of g. The *total influence* of g is the sum of all n coordinate influences, $\mathrm{Inf}(g) = \sum_{i=1}^{n} \mathrm{Inf}_i(g)$. See [27, 18] for background on influences.

We say that a polynomial $p : \{-1,1\}^n \to \mathbb{R}$ is "τ-regular" if the influence of each coordinate on p is at most a τ fraction of p's total influence (see Section 3 for a more detailed definition). A PTF f is said to be τ-regular if $f = \mathrm{sign}(p)$, where p is τ-regular. Roughly speaking, regular polynomials and PTFs are useful because they inherit some nice properties of PTFs and polynomials over Gaussian (rather than Boolean) inputs; this intuition can be made precise using the "invariance principle" of Mossel et al. [26]. This point of view has been useful in the $d = 1$ case for constructing pseudorandom generators [6], low-weight approximators [33, 10], and other results for LTFs [30, 24].

1.1 Our results

A regularity lemma for degree-d PTFs A number of useful results in different areas, loosely referred to as "regularity lemmas," show that for various types of combinatorial objects an arbitrary object can be approximately decomposed into a constant number of "pseudorandom" sub-objects. The best-known example of such a result is Szemerédi's classical regularity lemma for graphs [35], which (roughly) says that any graph can be decomposed into a constant number of subsets such that almost every pair of subsets induces a "pseudorandom" bipartite graph. Another example is Green's recent regularity lemma for Boolean functions [14]. Results of this sort are useful because different properties of interest are sometimes easier to establish for pseudorandom objects, and via regularity lemmas it may be possible to prove the corresponding theorems for general objects. We note also that results of this sort play an important part in the "structure versus randomness" paradigm that has been prominent in recent work in combinatorics and number theory, see e. g., [36].

We prove a structural result about degree-d PTFs which follows the above pattern; we thus refer to it as a "regularity lemma for degree-d PTFs." Our result says that any low-degree PTF can be decomposed as a small depth decision tree, most of whose leaves are close to regular PTFs:

Theorem 1.1. *Let $f(x) = \mathrm{sign}(p(x))$ be any degree-d PTF. Fix any $\tau > 0$. Then f is equivalent to a decision tree $\mathcal{T}$, of depth*

$$\mathrm{depth}(d,\tau) := \frac{1}{\tau} \cdot \left(d \log \frac{1}{\tau}\right)^{O(d)}$$

with variables at the internal nodes and a degree-d PTF $f_\rho = \mathrm{sign}(p_\rho)$ at each leaf ρ, with the following property: with probability at least $1-\tau$, a random path[1] *from the root reaches a leaf ρ such that either (i) p_ρ is τ-regular, or (ii) f_ρ is τ-close to a constant function.*

Regularity is a natural way to capture the notion of pseudorandomness for PTFs, and results of interest can be easier to establish for regular PTFs than for arbitrary PTFs (this is the case for our main application, constructing low-weight approximators, as we describe below). Our regularity lemma provides a general tool to reduce questions about arbitrary PTFs to regular PTFs; it has been used in this way as an essential ingredient in the recent proof that bounded independence fools all degree-2 PTFs [8] and (subsequently to the conference publication of this work [11]) degree-d PTFs [19] (see Section 9). We note that the recent construction of pseudorandom generators for degree-d PTFs of [25] also crucially uses a decomposition result which is very similar to our regularity lemma; we discuss the relation between our work and [25] in more detail below and in Section 1.2. Finally, Kane's recent work establishing the correct exponent in the Gotsman-Linial conjecture also uses a decomposition result which is very similar to our regularity lemma [21].

Application: Every low-degree PTF has a low-weight approximator [33] showed that every linear threshold function (LTF) over $\{-1,1\}^n$ can be ε-approximated by an LTF with integer weights $w_1,\dots,w_n$ such that $\sum_i w_i^2 = n \cdot 2^{\tilde{O}(1/\varepsilon^2)}$. (Here and throughout the paper we say that g is an ε-approximator for f if f and g differ on at most $\varepsilon 2^n$ inputs from $\{-1,1\}^n$.) This result and the tools used in its proof found several subsequent applications in complexity theory and learning theory, see e. g., [6, 30].

We apply our regularity lemma for degree-d PTFs to prove an analogue of the [33] result for low-degree polynomial threshold functions. Our result implies that for any constants d,ε, any degree-d PTF has an ε-approximating PTF of constant degree and (integer) weight $O(n^d)$.

When we refer to the *weight* of a PTF $f = \mathrm{sign}(p(x))$, we assume that all the coefficients of p are integers; by "weight" we mean the sum of the squares of p's coefficients. We prove

Theorem 1.2. *Let $f(x) = \mathrm{sign}(p(x))$ be any degree-d PTF. Fix any $\varepsilon > 0$. Then there is a polynomial $q(x)$ of degree $D = (d/\varepsilon)^{O(d)}$ and weight $2^{(d/\varepsilon)^{O(d)}} \cdot n^d$ such that $\mathrm{sign}(q(x))$ is ε-close to f.*

A result on the existence of low-weight ε-approximators for PTFs is implicit in the recent work [9]. They show that any degree-d PTF f has Fourier concentration $\sum_{|S|>1/\varepsilon^{O(d)}} \widehat{f}(S)^2 \le \varepsilon$, and this easily implies that f can be ε-approximated by a PTF with integer weights. (Indeed, recall that the learning algorithm of [22] works by constructing such a PTF as its hypothesis.) The above Fourier concentration bound implies that there is a PTF of degree $1/\varepsilon^{O(d)}$ and weight $n^{1/\varepsilon^{O(d)}}$ which ε-approximates f. In contrast, our Theorem 1.2 can give a weaker degree bound (if $d = 1/\varepsilon^{\omega(1)}$), but always gives a much stronger weight bound in terms of the dependence on n. We mention here that Podolskii [31] has shown that for every constant $d \ge 2$, there is a degree-d PTF for which any *exact* representation requires weight $n^{\Omega(n^d)}$.

[1] A random path corresponds to the standard uniform random walk on the tree.

We also prove lower bounds showing that weight $\widetilde{\Omega}(n^d)$ is required to ε-approximate degree-d PTFs for sufficiently small constant ε; see Section 4.3.

Techniques An important ingredient in our proof of Theorem 1.1 is a case analysis based on the "critical index" of a degree-d polynomial (see Section 3 for a formal definition). The critical index measures the first point (going through the variables from most to least influential) at which the influences "become small;" it is a natural generalization of the definition of the "critical index" of a linear form [33] that has been useful in several subsequent works [30, 6, 10]. Roughly speaking we show that:

- If the critical index of p is large, then a random restriction fixing few variables (the variables with largest influence in p) causes $\mathrm{sign}(p)$ to become a close-to-constant function with non-negligible probability; see Section 3.1.

- If the critical index of p is positive but small, then a random restriction as described above causes p to become regular with non-negligible probability; see Section 3.2.

- If the critical index of p is zero, then p is already a regular polynomial as desired.

Structure of the paper Section 2 contains notation and background. In Section 3 we prove our main result (Theorem 1.1). In Section 4 we prove our upper bound regarding integer-weight approximations (Theorem 1.2) and also give a nearly matching lower bound for any constant accuracy $\varepsilon > 0$.

1.2 Related work

Theorem 1.1 and the results in Sections 3.1 and 3.2 strengthen earlier results with a similar flavor that appeared in the work of Diakonikolas et al. [7]. Furthermore, similar structural results were proven simultaneously and independently by Ben-Eliezer et al. [23] and by Harsha et al. and Meka et al. [16, 25]. We describe each of these works below and their structural results for PTFs as they compare to our Theorem 1.1.

The results in [7] were obtained simultaneously and independently by Diakonikolas et al. [9] and Harsha et al. [16], and [7] is the resulting merge (which followed the approach of [9]). A regularity lemma as in Theorem 1.1 is not explicitly derived in [7]; using their Lemmas 5.10 and 5.9 and the ideas present here, one may obtain the conclusion of Theorem 1.1, but with $\mathrm{depth}(d,\tau) := (1/\tau)(d\log n\log(1/\tau))^{O(d)}$. Note the dependence on n; eliminating this dependence is essential for our low-weight approximator application and for the applications in [8, 19]. We eliminate the dependence on the dimension partly by developing dimension-independent versions of Lemmas 5.10 and 5.9, which we do in Lemmas 3.5 and 3.9, respectively.

Harsha et al. [16] give a result which is very similar to Lemma 3.14, the main component in our proof of Theorem 1.1. By applying the result from [16], Meka and Zuckerman [25] give a dimension-independent regularity lemma which is quite similar to our Theorem 1.1. They obtain a small-depth decision tree such that most leaves are ε-close to being ε-regular under a *stronger* definition of regularity, which we will call "ε-regularity in ℓ_2" to distinguish it from our notion.

Let $p : \{-1,1\}^n \to \mathbb{R}$ be a polynomial and $\varepsilon > 0$. We say that the polynomial p is "ε-regular in ℓ_2" if

$$\sqrt{\sum_{i=1}^{n} \mathrm{Inf}_i(p)^2} \leq \varepsilon \cdot \sum_{i=1}^{n} \mathrm{Inf}_i(p)\,.$$

Recall that in our definition of regularity, instead of upper bounding the ℓ_2-norm of the influence vector $I = (\mathrm{Inf}_1(p),\ldots,\mathrm{Inf}_n(p))$ by ε times the total influence of p (i. e., the ℓ_1 norm of I), we upper bound the ℓ_∞ norm (i. e., the maximum influence). We may thus call our notion "ε-regularity in ℓ_∞."

Note that if a polynomial is ε-regular in ℓ_2, then it is also ε-regular in ℓ_∞. (And this implication is easily seen to be essentially tight, e. g., if we have many variables with tiny influence and one variable with an ε-fraction of the total influence.) For the other direction, if a polynomial is ε-regular in ℓ_∞, then it is $\sqrt{\varepsilon}$-regular in ℓ_2. (This is also tight if we have $1/\varepsilon$ many variables with influence ε.)

Meka and Zuckerman prove the following.

Theorem 1.3 ([25]). *Every degree-d PTF $f = \mathrm{sign}(p)$ can be expressed as a decision tree of depth $2^{O(d)} \cdot (1/\varepsilon^2)\log^2(1/\varepsilon)$ with variables at the internal nodes and a degree-d PTF $f_\rho = \mathrm{sign}(p_\rho)$ at each leaf ρ, such that with probability $1-\varepsilon$, a random root-to-leaf path reaches a leaf ρ such that f_ρ is ε-close to being ε-regular in ℓ_2. (In particular, for a "good" leaf ρ, either p_ρ will be ε-regular in ℓ_2 or f_ρ will be ε-close to a constant.)*

Theorem 1.1 shows exactly the same statement as the one above if we replace "ℓ_2" by "ℓ_∞" and the depth of the tree by $(1/\varepsilon)\cdot(d\log(1/\varepsilon))^{O(d)}$.

Since ε-regularity in ℓ_2 implies ε-regularity in ℓ_∞, the result of [25] implies a version of Theorem 1.1 which has depth of $2^{O(d)} \cdot (1/\varepsilon^2)\log^2(1/\varepsilon)$. Hence the latter result and our result are quantitatively incomparable to each other. Roughly, if d is a constant (independent of ε), then our Theorem 1.1 is asymptotically better when ε becomes small. This range of parameters is quite natural in the context of pseudo-random generators. In particular, in the proof that $\mathrm{poly}(1/\varepsilon)$-wise independence ε-fools degree-2 PTFs [8], using [25] instead of Theorem 1.1, would give a worse bound on the degree of independence (namely, $\widetilde{O}(\varepsilon^{-18})$ as opposed to $\widetilde{O}(\varepsilon^{-9})$). On the other hand, if $d = \widetilde{\Omega}(\log(1/\varepsilon))$, then the result of [25] is better.

Finally, Ben-Eliezer et al. [23] establish the existence of a decision tree such that most leaves are τ-regular (as opposed to τ-close to being τ-regular):

Theorem 1.4 ([23]). *Every degree-d PTF $f = \mathrm{sign}(p)$ can be expressed as a decision tree of depth $2^{(d/\tau)^{O(d)}} \cdot \log(1/\tau)$ with variables at the internal nodes and a degree-d PTF $f_\rho = \mathrm{sign}(p_\rho)$ at each leaf ρ, such that with probability $1-\tau$, a random root-to-leaf path reaches a leaf ρ such that f_ρ is τ-regular.*

Note that the depth of their tree is exponential in $1/\tau$.

2 Preliminaries

We start by establishing some basic notation. We write $[n]$ to denote $\{1,2,\ldots,n\}$ and $[k,\ell]$ to denote $\{k,k+1,\ldots,\ell\}$. We write $\mathbf{E}[X]$ and $\mathbf{Var}[X]$ to denote expectation and variance of a random variable X, where the underlying distribution will be clear from the context. For $x \in \{-1,1\}^n$ and $A \subseteq [n]$ we write x_A to denote $(x_i)_{i\in A}$.

For a function $f : \{-1,1\}^n \to \mathbb{R}$ and $q \geq 1$, we denote by $\|f\|_q$ its ℓ_q norm, i. e.,

$$\|f\|_q \stackrel{\text{def}}{=} \mathbf{E}_x\left[|f(x)|^q\right]^{1/q},$$

where the intended distribution over x will always be uniform over $\{-1,1\}^n$. For Boolean-valued functions $f,g : \{-1,1\}^n \to \{-1,1\}$ the distance between f and g, denoted $\operatorname{dist}(f,g)$, is $\mathbf{Pr}_x[f(x) \neq g(x)]$ where the probability is over uniform $x \in \{-1,1\}^n$.

We assume familiarity with the basic elements of Fourier analysis over $\{-1,1\}^n$; see Section 2.1 for a concise review. Our proofs will use various analytic and probabilistic bounds, which we collect for easy reference in Section 2.1 below. We call the reader's attention in particular to Theorem 2.3; throughout the paper, C (which will be seen to play an important role in our proofs) denotes C_0^2, where C_0 is the universal constant from that theorem.

2.1 Useful background results

Fourier Analysis over $\{-1,1\}^n$ and influences The survey by de Wolf [37] is an excellent source of background material on Fourier analysis over $\{-1,1\}^n$, including material on the Hypercontractive Inequality; here we recall only the very basics that we will require. We consider functions $f : \{-1,1\}^n \to \mathbb{R}$ (though we often focus on Boolean-valued functions which map to $\{-1,1\}$), and we think of the inputs x to f as being distributed according to the uniform probability distribution. The set of such functions forms a 2^n-dimensional inner product space with inner product given by $\langle f,g\rangle = \mathbf{E}[f(x)g(x)]$. The set of functions $(\chi_S)_{S\subseteq[n]}$ defined by $\chi_S(x) = \prod_{i\in S} x_i$ forms a complete orthonormal basis for this space. Given a function $f : \{-1,1\}^n \to \mathbb{R}$ we define its *Fourier coefficients* by

$$\widehat{f}(S) \stackrel{\text{def}}{=} \mathbf{E}[f(x)\chi_S(x)],$$

and we have that $f(x) = \sum_S \widehat{f}(S)\chi_S(x)$. We refer to the maximum $|S|$ over all nonzero $\widehat{f}(S)$ as the *Fourier degree* of f.

As an easy consequence of orthonormality we have *Plancherel's identity* $\langle f,g\rangle = \sum_S \widehat{f}(S)\widehat{g}(S)$, which has as a special case *Parseval's identity*, $\mathbf{E}[f(x)^2] = \sum_S \widehat{f}(S)^2$. From this it follows that for every $f : \{-1,1\}^n \to \{-1,1\}$ we have $\sum_S \widehat{f}(S)^2 = 1$. We recall the well-known fact (see e. g., [17]) that the total influence $\operatorname{Inf}(f)$ of any Boolean function equals $\sum_S \widehat{f}(S)^2|S|$. Note that, in this setting, the expectation and the variance can be expressed in terms of the Fourier coefficients of f by $\mathbf{E}[f] = \widehat{f}(\emptyset)$ and $\mathbf{Var}[f] = \sum_{\emptyset\neq S\subseteq[n]} \widehat{f}(S)^2$.

Let $f : \{-1,1\}^n \to \mathbb{R}$ and $f(x) = \sum_S \widehat{f}(S)\chi_S(x)$ be its Fourier expansion. The *influence* of variable i on f is $\operatorname{Inf}_i(f) \stackrel{\text{def}}{=} \sum_{S\ni i} \widehat{f}(S)^2$, and the *total influence* of f is $\operatorname{Inf}(f) = \sum_{i=1}^n \operatorname{Inf}_i(f)$.

Useful probability bounds We first recall the following moment bound for low-degree polynomials, which is equivalent to the well-known hypercontractive inequality of [3, 15]:

Theorem 2.1. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial and $q > 2$. Then*

$$\|p\|_q \leq (q-1)^{d/2}\|p\|_2 .$$

The following concentration bound for low-degree polynomials, a simple corollary of hypercontractivity, is well known (see e. g., [12, 2]):

Theorem 2.2. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial. For any $t > e^d$, we have*

$$\mathbf{Pr}_x[|p(x)| \geq t\|p\|_2] \leq \exp(-\Omega(t^{2/d}))\,.$$

We will also need the following weak anti-concentration bound for low-degree polynomials over the cube:

Theorem 2.3 ([12, 2])**.** *There is a universal constant $C_0 > 1$ such that for any non-zero degree-d polynomial $p : \{-1,1\}^n \to \mathbb{R}$ with $\mathbf{E}[p] = 0$, we have*

$$\mathbf{Pr}_x[p(x) > C_0^{-d} \cdot \|p\|_2] > C_0^{-d}\,.$$

Throughout this paper, we let $C = C_0^2$, where C_0 is the universal constant from Theorem 2.3. Note that since $C > C_0$, Theorem 2.3 holds for C as well.

We denote by $\mathcal{N}^n$ the standard n-dimensional Gaussian distribution $\mathcal{N}(0,1)^n$. The following two facts will be useful in the proof of Theorem 1.2, in particular in the analysis of the regular case. The first fact is a powerful anti-concentration bound for low-degree polynomials over independent Gaussian random variables: $\mathbf{E}_{\mathcal{G}\sim\mathcal{N}^n}[p(\mathcal{G})^2]^{1/2}$):

Theorem 2.4 ([5])**.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a nonzero degree-d multilinear polynomial. For all $\varepsilon > 0$ and $t \in \mathbb{R}$ we have*

$$\mathbf{Pr}_{\mathcal{G}\sim\mathcal{N}^n}\left[|p(\mathcal{G}) - t| \leq \varepsilon \cdot \sqrt{\mathbf{Var}[p(\mathcal{G})]}\right] \leq O(d\varepsilon^{1/d})\,.$$

We note that the above bound is essentially tight.

The second fact is a version of the invariance principle of Mossel, O'Donnell and Oleszkiewicz, specifically Theorem 3.19 under hypothesis **H4** in [26]:

Theorem 2.5 ([26])**.** *Let*

$$p(x) = \sum_{S\subseteq[n],|S|\leq d} \widehat{p}(S)\chi_S(x)$$

be a degree-d multilinear polynomial with $\mathbf{Var}[p] = 1$. Suppose each coordinate $i \in [n]$ has $\mathrm{Inf}_i(p) \leq \tau$. Then,

$$\sup_{t\in\mathbb{R}} |\mathbf{Pr}_x[p(x) \leq t] - \mathbf{Pr}_{\mathcal{G}\sim\mathcal{N}^n}[p(\mathcal{G}) \leq t]| \leq O(d\tau^{1/(8d)})\,.$$

3 Main result: a regularity lemma for low-degree PTFs

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a degree-d PTF. Fix a representation $f(x) = \mathrm{sign}(p(x))$, where $p : \{-1,1\}^n \to \mathbb{R}$ is a degree-d polynomial which (w. l. o. g.) we may take to have $\mathbf{Var}[p] = 1$. We assume w. l. o. g. that the variables are ordered in such a way that $\mathrm{Inf}_i(p) \geq \mathrm{Inf}_{i+1}(p)$ for all $i \in [n-1]$.

We now define the notion of the τ-critical index of a polynomial [9] and state its basic properties.

Definition 3.1. Let $p:\{-1,1\}^n \to \mathbb{R}$ and $\tau > 0$. Assume the variables are ordered such that $\mathrm{Inf}_j(p) \geq \mathrm{Inf}_{j+1}(p)$ for all $j \in [n-1]$. The *τ-critical index* of p is the least i such that:

$$\mathrm{Inf}_{i+1}(p) \leq \tau \cdot \sum_{j=i+1}^{n} \mathrm{Inf}_j(p)\,. \tag{3.1}$$

If (3.1) does not hold for any i we say that the τ-critical index of p is $+\infty$. If p has τ-critical index 0, we say that p is *τ-regular*.

Note that if p is a τ-regular polynomial then $\max_i \mathrm{Inf}_i(p) \leq d\tau$ since the total influence of p is at most d. If $f(x) = \mathrm{sign}(p(x))$, we say f is τ-regular when p is τ-regular, and we take the τ-critical index of f to be that of p.[2] The following lemma says that the total influence $\sum_{i=j+1}^{n} \mathrm{Inf}_i(p)$ goes down geometrically as a function of j prior to the critical index:

Lemma 3.2. *Let $p:\{-1,1\}^n \to \mathbb{R}$ and $\tau > 0$. Let k be the τ-critical index of p. For $j \in [0,k]$ we have*

$$\sum_{i=j+1}^{n} \mathrm{Inf}_i(p) \leq (1-\tau)^j \cdot \mathrm{Inf}(p)\,.$$

Proof. The lemma trivially holds for $j=0$. In general, since j is at most k, we have that $\mathrm{Inf}_j(p) \geq \tau \cdot \sum_{i=j}^{n} \mathrm{Inf}_i(p)$, or equivalently $\sum_{i=j+1}^{n} \mathrm{Inf}_i(p) \leq (1-\tau)\cdot\sum_{i=j}^{n} \mathrm{Inf}_i(p)$ which yields the claimed bound. □

We will use the fact that in expectation, the influence of an unrestricted variable in a polynomial does not change under random restrictions:

Lemma 3.3. *Let $p:\{-1,1\}^n \to \mathbb{R}$. Consider a random assignment ρ to the variables $x_1,\dots,x_k$ and fix $\ell \in [k+1,n]$. Then $\mathbf{E}_\rho[\mathrm{Inf}_\ell(p_\rho)] = \mathrm{Inf}_\ell(p)$.*

Proof. To prove Lemma 3.3, we first recall an observation about the expected value of Fourier coefficients under random restrictions (see e. g., [22]):

Fact 3.4. *Let $p:\{-1,1\}^n \to \mathbb{R}$. Consider a random assignment ρ to the variables $x_1,\dots,x_k$. Fix any $S \subseteq [k+1,n]$. Then we have $\widehat{p_\rho}(S) = \sum_{T\subseteq[k]} \widehat{p}(S\cup T)\rho_T$ and therefore $\mathbf{E}_\rho[\widehat{p_\rho}(S)^2] = \sum_{T\subseteq[k]} \widehat{p}(S\cup T)^2$.*

In words, the above fact says that all the Fourier weight on sets of the form $S \cup \{$any subset of restricted variables$\}$ "collapses" down onto S in expectation. Consequently, the influence of an unrestricted variable does not change in expectation under random restrictions:

We thus have

$$\begin{aligned}\mathbf{E}_\rho\left[\mathrm{Inf}_\ell(p_\rho)\right] &= \mathbf{E}_\rho\left[\sum_{\ell\in S\subseteq[k+1,n]} \widehat{p_\rho}(S)^2\right] = \sum_{T\subseteq[k]}\sum_{\ell\in S\subseteq[k+1,n]} \widehat{p}(S\cup T)^2\\ &= \sum_{\ell\in U\subseteq[n]} \widehat{p}(U)^2 = \mathrm{Inf}_\ell(p)\,.\end{aligned}$$
□

[2] Strictly speaking, τ-regularity is a property of a particular representation and not of a PTF f, which could have many different representations. The particular representation we are concerned with will always be clear from context.

Notation: For $\mathcal{S} \subseteq [n]$, we write "$\rho$ fixes $\mathcal{S}$" to indicate that $\rho \in \{-1,1\}^{|\mathcal{S}|}$ is a restriction mapping $x_{\mathcal{S}}$, i. e., each coordinate in $\mathcal{S}$, to either -1 or 1 and leaving coordinates not in $\mathcal{S}$ unrestricted.

3.1 The large critical index case

The main result of this section is Lemma 3.5, which says that if the critical index of f is large, then a noticeable fraction of restrictions ρ of the high-influence variables cause f_ρ to become close to a constant function.

Lemma 3.5. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a degree-d PTF $f = \operatorname{sign}(p)$. Fix $\beta > 0$ and suppose that f has τ-critical index at least $K \stackrel{\text{def}}{=} \alpha/\tau$, where $\alpha = \Omega(d \log\log(1/\beta) + d \log d)$. Then, for at least a $1/(2C^d)$ fraction of restrictions ρ fixing $[K]$, the function f_ρ is β-close to a constant function.*

proof Partition the coordinates into a "head" part $H \stackrel{\text{def}}{=} [K]$ (the high-influence coordinates) and a "tail" part $T = [n] \setminus H$. We can write $p(x) = p(x_H, x_T) = p'(x_H) + q(x_H, x_T)$, where $p'(x_H)$ is the truncation of p comprising only the monomials all of whose variables are in H, i. e., $p'(x_H) = \sum_{S \subseteq H} \widehat{p}(S)\chi_S(x_H)$.

Now consider a restriction ρ of H and the corresponding polynomial $p_\rho(x_T) = p(\rho, x_T)$. It is clear that the constant term of this polynomial is exactly $p'(\rho)$. To prove the lemma, we will show that for at least a $1/(2C^d)$ fraction of all $\rho \in \{-1,1\}^K$, the (restricted) degree-d PTF $f_\rho(x_T) = \operatorname{sign}(p_\rho(x_T))$ satisfies $\mathbf{Pr}_{x_T}[f_\rho(x_T) \neq \operatorname{sign}(p'(\rho))] \le \beta$. Let us define the notion of a *good restriction*:

Definition 3.6. A restriction $\rho \in \{-1,1\}^K$ that fixes H is called *good* iff the following two conditions are simultaneously satisfied:

$$\text{(i)}\ \ |p'(\rho)| \ge t^* \stackrel{\text{def}}{=} 1/(2C^d), \quad \text{and} \quad \text{(ii)}\ \ \|q(\rho, x_T)\|_2 \le t^* \cdot \big(\Theta(\log(1/\beta))\big)^{-d/2}.$$

Intuitively condition (i) says that the constant term $p'(\rho)$ of p_ρ has "large" magnitude, while condition (ii) says that the polynomial $q(\rho, x_T)$ has "small" ℓ_2-norm. We claim that if ρ is a good restriction then the degree-d PTF f_ρ satisfies $\mathbf{Pr}_{x_T}[f_\rho(x_T) \neq \operatorname{sign}(p'(\rho))] \le \beta$. To see this claim, note that for any fixed ρ we have $f_\rho(x_T) \neq \operatorname{sign}(p'(\rho))$ only if $|q(\rho, x_T)| \ge |p'(\rho)|$, so to show this claim it suffices to show that if ρ is a good restriction then $\mathbf{Pr}_{x_T}[|q(\rho, x_T)| \ge |p'(\rho)|] \le \beta$. But for ρ a good restriction, by conditions (i) and (ii) we have

$$\mathbf{Pr}_{x_T}\left[|q(\rho, x_T)| \ge |p'(\rho)|\right] \le \mathbf{Pr}_{x_T}\left[|q(\rho, x_T)| \ge \|q(\rho, x_T)\|_2 \cdot \big(\Theta(\log(1/\beta))\big)^{d/2}\right],$$

which is at most β by the concentration bound (Theorem 2.2), as desired. So the claim holds: if ρ is a good restriction then $f_\rho(x_T)$ is β-close to $\operatorname{sign}(p'(\rho))$. Thus to prove Lemma 3.5 it remains to show that at least a $1/(2C^d)$ fraction of all restrictions ρ to H are good.

We prove this in two steps. First we show (Lemma 3.7) that the polynomial p' is not too concentrated: with probability at least $1/C^d$ over ρ, condition (i) of Definition 3.6 is satisfied. We then show (Lemma 3.8) that the polynomial q is highly concentrated: the probability (over ρ) that condition (ii) is *not* satisfied is at most $1/(2C^d)$. Lemma 3.5 then follows by a union bound.

Lemma 3.7. *We have that* $\mathbf{Pr}_\rho\left[|p'(\rho)| \ge t^*\right] \ge 1/C^d$.

Proof. Using the fact that the critical index of p is large, we will show that the polynomial p' has large variance (close to 1), and hence we can apply the anti-concentration bound Theorem 2.3.

We start by establishing that $\mathbf{Var}[p']$ lies in the range $[1/2, 1]$. To see this, first recall that for $g : \{-1,1\}^n \to \mathbb{R}$ we have $\mathbf{Var}[g] = \sum_{\emptyset \neq S \subseteq [n]} \widehat{g}^2(S)$. It is thus clear that $\mathbf{Var}[p'] \leq \mathbf{Var}[p] = 1$. To establish the lower bound we use the property that the "tail" T has "very small" influence in p, which is a consequence of the critical index of p being large. More precisely, Lemma 3.2 yields

$$\sum_{i \in T} \mathrm{Inf}_i(p) \leq (1-\tau)^K \cdot \mathrm{Inf}(p) = (1-\tau)^{\alpha/\tau} \cdot \mathrm{Inf}(p) \leq d \cdot e^{-\alpha} \tag{3.2}$$

where the last inequality uses the fact that $\mathrm{Inf}(p) \leq d$. Therefore, we have:

$$\mathbf{Var}[p'] = \mathbf{Var}[p] - \sum_{T \cap S \neq \emptyset, S \subseteq [n]} \widehat{p}(S)^2 \geq 1 - \sum_{i \in T} \mathrm{Inf}_i(p) \geq 1 - de^{-\alpha} \geq 1/2$$

where the first inequality uses the fact that $\mathrm{Inf}_i(p) = \sum_{i \in S \subseteq [n]} \widehat{p}(S)^2$, the second follows from (3.2) and the third from our choice of α. We have thus established that indeed $\mathbf{Var}[p'] \in [1/2, 1]$.

At this point, we would like to apply Theorem 2.3 for p'. Note however that $\mathbf{E}[p'] = \mathbf{E}[p] = \widehat{p}(\emptyset)$ which is not necessarily zero. To address this minor technical point we apply Theorem 2.3 twice: once for the polynomial $p'' = p' - \widehat{p}(\emptyset)$ and once for $-p''$. (Clearly, $\mathbf{E}[p''] = 0$ and $\mathbf{Var}[p''] = \mathbf{Var}[p'] \in [1/2, 1]$.) We thus get that, independent of the value of $\widehat{p}(\emptyset)$, we have $\mathbf{Pr}_\rho[|p'(\rho)| > 2^{-1/2} \cdot C^{-d}] \geq C^{-d}$, as desired. $\square$

Lemma 3.8. *We have that* $\mathbf{Pr}_\rho\left[\|q(\rho, x_T)\|_2 > t^* \cdot \left(\Theta(\log(1/\beta)\right)^{-d/2}\right] \leq 1/(2C^d)$.

Proof. To obtain the desired concentration bound we must show that the degree-$2d$ polynomial $Q(\rho) = \|q(\rho, x_T)\|_2^2$ has "small" variance. The desired bound then follows by an application of Theorem 2.2.

We thus begin by showing that $\|Q\|_2 \leq 3^d de^{-\alpha}$. To see this, we first note that $Q(\rho) = \sum_{\emptyset \neq S \subseteq T} \widehat{p_\rho}(S)^2$. Hence an application of the triangle inequality for norms and hypercontractivity (Theorem 2.1) yields:

$$\|Q\|_2 \leq \sum_{\emptyset \neq S \subseteq T} \|\widehat{p_\rho}(S)\|_4^2 \leq 3^d \sum_{\emptyset \neq S \subseteq T} \|\widehat{p_\rho}(S)\|_2^2.$$

We now proceed to bound from above the RHS term by term:

$$\begin{aligned}\sum_{\emptyset \neq S \subseteq T} \|\widehat{p_\rho}(S)\|_2^2 &= \sum_{\emptyset \neq S \subseteq T} \mathbf{E}_\rho[\widehat{p_\rho}(S)^2] = \mathbf{E}_\rho\Big[\sum_{\emptyset \neq S \subseteq T} \widehat{p_\rho}(S)^2\Big] \leq \mathbf{E}_\rho\left[\mathrm{Inf}(p_\rho)\right] = \mathbf{E}_\rho\Big[\sum_{i \in T} \mathrm{Inf}_i(p_\rho)\Big] \\ &= \sum_{i \in T} \mathbf{E}_\rho\left[\mathrm{Inf}_i(p_\rho)\right] = \sum_{i \in T} \mathrm{Inf}_i(p) \leq de^{-\alpha} \end{aligned} \tag{3.3}$$

where the first inequality uses the fact $\mathrm{Inf}(p_\rho) \geq \sum_{\emptyset \neq S \subseteq T} \widehat{p_\rho}(S)^2$, the equality in (3.3) follows from Lemma 3.3, and the last inequality is equation (3.2). We have thus shown that $\|Q\|_2 \leq 3^d de^{-\alpha}$.

We now upper bound

$$\mathbf{Pr}_\rho\left[Q(\rho) > (t^*)^2 \cdot \Theta(\log(1/\beta))^{-d}\right].$$

Since $\|Q\|_2 \leq 3^d de^{-\alpha}$, Theorem 2.2 implies that for all $t > e^d$ we have $\mathbf{Pr}_\rho[Q(\rho) > t \cdot 3^d de^{-\alpha}] \leq \exp(-\Omega(t^{1/d}))$. Taking t to be $\Theta(d^d \ln^d C)$ this upper bound is at most $1/(2C^d)$. Our choice of the parameter α gives $t \cdot d3^d \cdot e^{-\alpha} \leq (t^*)^2 \cdot \Theta(\log(1/\beta))^{-d}$. This completes the proof of Lemma 3.8, and thus also the proof of Lemma 3.5. $\square$

3.2 The small critical index case

In this section we show that if the critical index of p is "small," then a random restriction of "few" variables causes p to become regular with non-negligible probability. We do this by showing that no matter what the critical index is, a random restriction of all variables up to the τ-critical index causes p to become τ'-regular, for some τ' not too much larger than τ, with probability at least $1/(2C^d)$. More formally, we prove:

Lemma 3.9. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial with τ-critical index $k \in [n]$. Let ρ be a random restriction that fixes $[k]$, and let $\tau' = (C' \cdot d \ln d \cdot \ln \frac{1}{\tau})^d \cdot \tau$ for some suitably large absolute constant C'. With probability at least $1/(2C^d)$ over the choice of ρ, the restricted polynomial p_ρ is τ'-regular.*

Proof. We must show that with probability at least $1/(2C^d)$ over ρ the restricted polynomial p_ρ satisfies

$$\mathrm{Inf}_\ell(p_\rho) / \sum_{j=k+1}^{n} \mathrm{Inf}_j(p_\rho) \le \tau' \tag{3.4}$$

for all $\ell \in [k+1,n]$. Note that before the restriction, we have $\mathrm{Inf}_\ell(p) \le \tau \cdot \sum_{j=k+1}^n \mathrm{Inf}_j(p)$ for all $\ell \in [k+1,n]$ because the τ-critical index of p is k.

Let us give an intuitive explanation of the proof. We first show (Lemma 3.10) that with probability at least C^{-d} the denominator in (3.4) does not decrease under a random restriction. This is an anti-concentration statement that follows easily from Theorem 2.3. We then show (Lemma 3.11) that with probability at least $1 - C^{-d}/2$ the numerator in (3.4) does not increase by much under a random restriction, i. e., *no* variable influence $\mathrm{Inf}_\ell(p_\rho)$, $\ell \in [k+1,n]$, becomes too large. Thus both events occur (and p_ρ is τ'-regular) with probability at least $C^{-d}/2$.

We note that while each individual influence $\mathrm{Inf}_\ell(p_\rho)$ is indeed concentrated around its expectation (see Claim 3.12), we need a concentration statement for $n-k$ such influences. This might seem difficult to achieve since we require bounds that are independent of n. We get around this difficulty by a "bucketing" argument that exploits the fact (at many different scales) that all but a few influences $\mathrm{Inf}_\ell(p)$ must be "very small."

It remains to state and prove Lemmas 3.10 and 3.11. Consider the event

$$\mathcal{E} \stackrel{\text{def}}{=} \left\{ \rho \in \{-1,1\}^k \mid \sum_{\ell=k+1}^{n} \mathrm{Inf}_\ell(p_\rho) \ge \sum_{\ell=k+1}^{n} \mathrm{Inf}_\ell(p) \right\}.$$

We first show:

Lemma 3.10. $\mathbf{Pr}_\rho[\mathcal{E}] \ge C^{-d}$.

Proof. It follows from Fact 3.4 and the Fourier expression of $\mathrm{Inf}_\ell(p_\rho)$ that $A(\rho) \stackrel{\text{def}}{=} \sum_{\ell=k+1}^n \mathrm{Inf}_\ell(p_\rho)$ is a degree-$2d$ polynomial. By Lemma 3.3 we get that $\mathbf{E}_\rho[A] = \sum_{\ell=k+1}^n \mathrm{Inf}_\ell(p) > 0$. Also observe that $A(\rho) \ge 0$ for all $\rho \in \{-1,1\}^k$. We may now apply Theorem 2.3 to the polynomial $A' = A - \mathbf{E}_\rho[A]$, to obtain:

$$\mathbf{Pr}_\rho[\mathcal{E}] = \mathbf{Pr}[A' \ge 0] \ge \mathbf{Pr}[A' \ge C_0^{-2d} \cdot \sigma(A')] > C_0^{-2d} = C^{-d}. \qquad \square$$

We now turn to Lemma 3.11. Consider the event

$$\mathcal{J} \stackrel{\text{def}}{=} \left\{ \rho \in \{-1,1\}^k \;\middle|\; \max_{\ell \in [k+1,n]} \mathrm{Inf}_\ell(p_\rho) > \tau' \sum_{j=k+1}^{n} \mathrm{Inf}_j(p) \right\}.$$

We show:

Lemma 3.11. $\mathbf{Pr}_\rho[\mathcal{J}] \le (1/2) \cdot C^{-d}$.

The rest of this subsection consists of the proof of Lemma 3.11. A useful intermediate claim is that the influences of individual variables do not increase by a lot under a random restriction (note that this claim does not depend on the value of the critical index):

Claim 3.12. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial. Let ρ be a random restriction fixing $[j]$. Fix any $t > e^{2d}$ and any $\ell \in [j+1,n]$. With probability at least $1 - \exp(-\Omega(t^{1/d}))$ over ρ, we have $\mathrm{Inf}_\ell(p_\rho) \le 3^d t \,\mathrm{Inf}_\ell(p)$.*

Proof. The identity $\mathrm{Inf}_\ell(p_\rho) = \sum_{\ell \in S \subseteq [j+1,n]} \widehat{p_\rho}(S)^2$ and Fact 3.4 imply that $\mathrm{Inf}_\ell(p_\rho)$ is a degree-$2d$ polynomial in ρ. Hence the claim follows from the concentration bound, Theorem 2.2, assuming we can appropriately upper bound the ℓ_2-norm of the polynomial $\mathrm{Inf}_\ell(p_\rho)$. So, to prove Claim 3.12 it suffices to show that

$$\|\mathrm{Inf}_\ell(p_\rho)\|_2 \le 3^d \mathrm{Inf}_\ell(p)\,. \tag{3.5}$$

The proof of equation (3.5) is similar to the argument establishing that $\|Q\|_2 \le 3^d d e^{-\alpha}$ in Section 3.1. The triangle inequality tells us that we may bound the ℓ_2-norm of each squared-coefficient separately:

$$\|\mathrm{Inf}_\ell(p_\rho)\|_2 \le \sum_{\ell \in S \subseteq [j+1,n]} \|\widehat{p}_\rho(S)^2\|_2\,.$$

Since $\widehat{p}_\rho(S)$ is a degree-d polynomial, Theorem 2.1 yields that

$$\|\widehat{p}_\rho(S)^2\|_2 = \|\widehat{p}_\rho(S)\|_4^2 \le 3^d \|\widehat{p}_\rho(S)\|_2^2\,,$$

hence

$$\|\mathrm{Inf}_\ell(p_\rho)\|_2 \le 3^d \sum_{\ell \in S \subseteq [j+1,n]} \|\widehat{p}_\rho(S)\|_2^2 = 3^d \mathrm{Inf}_\ell(p)\,,$$

where the last equality is a consequence of Fact 3.4. Thus equation (3.5) holds, and Claim 3.12 is proved. □

Claim 3.12 says that for any given coordinate, the probability that its influence after a random restriction increases by a t factor decreases exponentially in t. Note that Claim 3.12 and a naive union bound over all coordinates in $[k+1,n]$ does not suffice to prove Lemma 3.11. Instead, we proceed as follows: We partition the coordinates in $[k+1,n]$ into "buckets" according to their influence in the tail of p. In particular, the i-th bucket ($i \ge 0$) contains all variables $\ell \in [k+1,n]$ such that

$$\frac{\mathrm{Inf}_\ell(p)}{\sum_{j=k+1}^{n} \mathrm{Inf}_j(p)} \in [\tau/2^{i+1}, \tau/2^i]\,.$$

We analyze the effect of a random restriction ρ on the variables of each bucket i separately and then conclude by a union bound over all the buckets.

So fix a bucket i. Note that, by definition, the number of variables in the i-th bucket is at most $2^{i+1}/\tau$. We bound from above the probability of the event $\mathcal{B}(i)$ that there exists a variable ℓ in bucket i that violates the regularity constraint, i. e., such that $\mathrm{Inf}_\ell(p_\rho) > \tau' \sum_{\ell=k+1}^n \mathrm{Inf}_\ell(p)$. We will do this by a combination of Claim 3.12 and a union bound over the variables in the bucket. We will show:

Claim 3.13. *We have that* $\mathbf{Pr}_\rho[\mathcal{B}(i)] \le 2^{-(i+2)} \cdot C^{-d}$.

The above claim completes the proof of Lemma 3.11 by a union bound across buckets. Indeed, assuming the claim, the probability that *any* variable $\ell \in [k+1,n]$ violates the condition $\mathrm{Inf}_\ell(p_\rho) \le \tau' \sum_{\ell=k+1}^n \mathrm{Inf}_\ell(p)$ is at most

$$\sum_{i=0}^{\infty} \mathbf{Pr}_\rho[\mathcal{B}(i)] \le C^{-d} 2^{-2} \sum_{i=0}^{\infty} (1/2)^i = (1/2) \cdot C^{-d}.$$

It thus remains to prove Claim 3.13. Fix a variable ℓ in the i-th bucket. We apply Claim 3.12 selecting a value of

$$t = \widetilde{t} \stackrel{\text{def}}{=} \left(\ln \frac{C^d 4^{i+2}}{\tau} \right)^d.$$

It is clear that $\widetilde{t} \le c'^d (d + i + \ln(1/\tau))^d$ for some absolute constant c'. As a consequence, there is an absolute constant C' such that for every i,

$$\widetilde{t} \le 3^{-d} C'^d 2^i \left(d \ln d \ln \frac{1}{\tau} \right)^d. \tag{3.6}$$

(To see this, note that for $i \le 10 d \ln d$ we have $d + i + \ln(1/\tau) < 11 d \ln d \ln(1/\tau)$, from which the claimed bound is easily seen to hold. For $i > 10 d \ln d$, we use $d + i + \ln(1/\tau) < d i \ln(1/\tau)$ and the fact that $i^d < 2^i$ for $i > 10 d \ln d$.)

Inequality (3.6) can be rewritten as $3^d \cdot \widetilde{t} \cdot \tau / 2^i \le \tau'$. Hence, our assumption on the range of $\mathrm{Inf}_\ell(p)$ gives

$$3^d \cdot \widetilde{t} \cdot \mathrm{Inf}_\ell(p) \le \tau' \cdot \sum_{j=k+1}^{n} \mathrm{Inf}_j(p).$$

Therefore, by Claim 3.12, the probability that coordinate ℓ violates the condition

$$\mathrm{Inf}_\ell(p_\rho) \le \tau' \sum_{j=k+1}^{n} \mathrm{Inf}_j(p)$$

is at most $\tau/(C^d 4^{i+2})$ by our choice of $\widetilde{t}$. Since bucket i contains at most $2^{i+1}/\tau$ coordinates, Claim 3.13 follows by a union bound. Hence Lemma 3.11, and thus Lemma 3.9, is proved. □

3.3 Putting everything together: proof of Theorem 1.1

The following lemma combines the results of the previous two subsections:

Lemma 3.14. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial and $0 < \widetilde{\tau}, \beta < 1/2$. Fix*

$$\alpha = \Theta(d \log\log(1/\beta) + d \log d) \quad \text{and} \quad \widetilde{\tau}' = \widetilde{\tau} \cdot (C' d \ln d \ln(1/\widetilde{\tau}))^d ,$$

where C' is a universal constant. (We assume w. l. o. g. that the variables are ordered s.t. $\mathrm{Inf}_i(p) \geq \mathrm{Inf}_{i+1}(p)$, $i \in [n-1]$.) One of the following statements holds true:

1. *The polynomial p is $\widetilde{\tau}$-regular.*
2. *With probability at least $1/(2C^d)$ over a random restriction ρ fixing the first $\alpha/\widetilde{\tau}$ (most influential) variables of p, the function $\mathrm{sign}(p_\rho)$ is β-close to a constant function.*
3. *There exists a value $k \leq \alpha/\widetilde{\tau}$, such that with probability at least $1/(2C^d)$ over a random restriction ρ fixing the first k (most influential) variables of p, the polynomial p_ρ is $\widetilde{\tau}'$-regular.*

Proof. The proof is by case analysis based on the value ℓ of the $\widetilde{\tau}$-critical index of the polynomial p. If $\ell = 0$, then by definition p is $\widetilde{\tau}$-regular, hence the first statement of the lemma holds. If $\ell > \alpha/\widetilde{\tau}$, then we randomly restrict the first $\alpha/\widetilde{\tau}$ many variables. Lemma 3.5 says that for a random restriction ρ fixing these variables, with probability at least $1/(2C^d)$ the (restricted) degree-d PTF $\mathrm{sign}(p_\rho)$ is β-close to a constant. Hence, in this case, the second statement holds. To handle the case $\ell \in [1, \alpha/\widetilde{\tau}]$, we apply Lemma 3.9. This lemma says that with probability at least $1/(2C^d)$ over a random restriction ρ fixing variables $[\ell]$, the polynomial p_ρ is $\widetilde{\tau}'$-regular, so the third statement of Lemma 3.14 holds. □

Proof of Theorem 1.1. We begin by observing that any function f on $\{-1,1\}^n$ is equivalent to a decision tree where each internal node of the tree 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 subfunction f_ρ. Given an arbitrary degree-d PTF $f = \mathrm{sign}(p)$, we will construct a decision tree $\mathcal{T}$ of the form described in Theorem 1.1. It is clear that in any such tree every leaf function f_ρ will be a degree-d PTF; we must show that $\mathcal{T}$ has depth $\mathrm{depth}(d,\tau)$ and that with probability $1-\tau$ over the choice of a random root-to-leaf path ρ, the restriction ρ is such that either $f_\rho = \mathrm{sign}(p_\rho)$ is τ-close to a constant function, or p_ρ is τ-regular.

For a tree T computing $f = \mathrm{sign}(p)$, we denote by $N(T)$ its set of internal nodes and by $L(T)$ its set of leaves. We call a leaf $\rho \in L(T)$ "good" if either f_ρ is τ-close to a constant function or p_ρ is τ-regular. We call a leaf "bad" otherwise. Let $GL(T)$ and $BL(T)$ be the sets of "good" and "bad" leaves in T respectively.

The basic approach for the proof is to invoke Lemma 3.14 repeatedly in a sequence of at most $2C^d \ln(1/\tau)$ stages. In the first stage we apply Lemma 3.14 to f itself; this gives us an initial decision tree. In the second stage we apply Lemma 3.14 to the restricted subfunctions f_ρ corresponding to leaves of the initial decision tree that are bad; this "grows" our initial decision tree. Subsequent stages continue similarly; we will argue that after at most $2C^d \ln(1/\tau)$ stages, the resulting tree satisfies the required properties for $\mathcal{T}$. In every application of Lemma 3.14 the parameters β and $\widetilde{\tau}'$ are both taken to be τ; note that taking $\widetilde{\tau}'$ to be τ sets the value of $\widetilde{\tau}$ in Lemma 3.14 to a value that is less than τ.

We now provide the details. In the first stage, the initial application of Lemma 3.14 results in a tree T_1. This tree T_1 may consist of a single leaf node that is $\widetilde{\tau}$-regular (if f is $\widetilde{\tau}$-regular to begin with—in this case, since $\widetilde{\tau} < \tau$, we are done), or a complete decision tree of depth $\alpha/\widetilde{\tau}$ (if f had large critical index), or a complete decision tree of depth $k < \alpha/\widetilde{\tau}$ (if f had small critical index). Note that in each case the depth of T_1 is at most $\alpha/\widetilde{\tau}$. Lemma 3.14 guarantees that:

$$\mathbf{Pr}_{\rho \in T_1}[\rho \in BL(T_1)] \leq 1 - 1/(2C^d),$$

where the probability is over a random root-to-leaf path ρ in T_1.

In the second stage, the "good" leaves $\rho \in GL(T_1)$ are left untouched; they will be leaves in the final tree $\mathcal{T}$. For each "bad" leaf $\rho \in BL(T_1)$, we order the unrestricted variables in decreasing order of their influence in the polynomial p_ρ, and we apply Lemma 3.14 to f_ρ. This "grows" T_1 at each bad leaf by replacing each such leaf with a new decision tree; we call the resulting overall decision tree T_2.

A key observation is that the probability that a random path from the root reaches a "bad" leaf is significantly smaller in T_2 than in T_1; in particular

$$\mathbf{Pr}_{\rho \in T_2}[\rho \in BL(T_2)] \leq (1 - 1/(2C^d))^2.$$

We argue this as follows: Let ρ be any fixed "bad" leaf in T_1, i. e., $\rho \in BL(T_1)$. The function f_ρ is not $\widetilde{\tau}'$-regular and consequently not $\widetilde{\tau}$-regular. Thus, either statement (2) or (3) of Lemma 3.14 must hold when the Lemma is applied to f_ρ. The tree that replaces ρ in T_0 has depth at most $\alpha/\widetilde{\tau}$, and a random root-to-leaf path ρ_1 *in this tree* reaches a "bad" leaf with probability at most $1 - 1/(2C^d)$. So the overall probability that a random root-to-leaf path in T_2 reaches a "bad" leaf is at most $(1 - 1/(2C^d))^2$.

Continuing in this fashion, in the i-th stage we replace all the bad leaves of T_{i-1} by decision trees according to Lemma 3.14 and we obtain the tree T_i. An inductive argument gives that

$$\mathbf{Pr}_{\rho \in T_i}[\rho \in BL(T_i)] \leq (1 - 1/(2C^d))^i,$$

which is at most τ for $i^* \stackrel{\text{def}}{=} 2C^d \ln(1/\tau)$.

The depth of the overall tree will be the maximum number of stages $(2C^d \ln(1/\tau))$ times the maximum depth added in each stage (at most $\alpha/\widetilde{\tau}$, since we always restrict at most this many variables), which is at most $(\alpha/\widetilde{\tau}) \cdot i^*$. Since $\beta = \tau$, we get $\alpha = \Theta(d \log\log(1/\tau) + d \log d)$. Recalling that $\widetilde{\tau}'$ in Lemma 3.14 is set to τ, we see that

$$\widetilde{\tau} = \frac{\tau}{(C' d \ln d \ln(1/\tau))^{O(d)}}.$$

By substitution we get that the depth of the tree is upper bounded by $d^{O(d)} \cdot (1/\tau) \cdot \log(1/\tau)^{O(d)}$ which concludes the proof of Theorem 1.1. $\square$

4 Every degree-d PTF has a low-weight approximator

In this section we apply Theorem 1.1 to prove Theorem 1.2, which we restate below:

Theorem 1.2. *Let $f(x) = \operatorname{sign}(p(x))$ be any degree-d PTF. Fix any $\varepsilon > 0$ and let $\tau = (\Theta(1) \cdot \varepsilon/d)^{8d}$. Then there is a polynomial $q(x)$ of degree $D = d + \operatorname{depth}(d,\tau)$ and weight $n^d \cdot 2^{4\operatorname{depth}(d,\tau)} \cdot (d/\varepsilon)^{O(d)}$, which is such that the PTF $\operatorname{sign}(q(x))$ is ε-close to f.*

To prove Theorem 1.2, we first show that any sufficiently regular degree-d PTF over n variables has a low-weight approximator, of weight roughly n^d. Theorem 1.1 asserts that almost every leaf ρ of $\mathcal{T}$ is a regular PTF or close to a constant function; at each regular leaf ρ we use the low-weight approximator of the previous sentence to approximate f_ρ. Finally, we combine all of these low-weight polynomials to get an overall PTF of low weight which is a good approximator for f. We give details below.

4.1 Low-weight approximators for regular PTFs

In this subsection we prove that every sufficiently regular PTF has a low-weight approximator of degree d:

Lemma 4.1. *Given $\varepsilon > 0$, let $\tau = (\Theta(1)\cdot\varepsilon/d)^{8d}$. Let $p : \{-1,1\}^n \to \mathbb{R}$ be a τ-regular degree-d polynomial with $\mathbf{Var}[p] = 1$. There exists a degree-d polynomial $q : \{-1,1\}^n \to \mathbb{R}$ of weight $n^d \cdot (d/\varepsilon)^{O(d)}$ such that $\operatorname{sign}(q(x))$ is an ε-approximator for $\operatorname{sign}(p(x))$.*

Proof. The polynomial q is obtained by rounding the weights of p to an appropriate granularity, similar to the regular case in [33] for the $d = 1$ case. To show that this works, we use the fact that regular PTFs have very good anti-concentration. In particular we will use the following claim, which follows easily by combining the invariance principle [26] and Gaussian anti-concentration [5]:

Claim 4.2. *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a τ-regular degree-d polynomial with $\mathbf{Var}[p] = 1$. Then*

$$\mathbf{Pr}_x[|p(x)| \le \tau] \le O(d\tau^{1/8d}).$$

Proof. We recall that, since $\mathbf{Var}[p] = 1$ and p is of degree d, it holds $\mathrm{Inf}(p) \le d$. Thus, since p is τ-regular, we have that $\max_{i\in[n]} \mathrm{Inf}_i(p) \le d\tau$. An application of the invariance principle (Theorem 2.5) in tandem with anti-concentration in gaussian space (Theorem 2.4) yields

$$\begin{aligned}\mathbf{Pr}_x[|p(x)| \le \tau] &\le O(d\cdot(d\tau)^{1/8d}) + \mathbf{Pr}_{\mathcal{G}\sim\mathcal{N}^n}[|p(\mathcal{G})| \le \tau] \\ &\le O(d\tau^{1/8d}) + O(d\tau^{1/d}) = O(d\tau^{1/8d}),\end{aligned}$$

and the claim follows. □

We turn to the detailed proof of Lemma 4.1. We first note that if the constant coefficient $\widehat{p}(\emptyset)$ of P has magnitude greater than $(O(\log(1/\varepsilon)))^{d/2}$, then Theorem 2.2 (applied to $p(x) - \widehat{p}(\emptyset)$) implies that $\operatorname{sign}(p(x))$ agrees with $\operatorname{sign}(\widehat{p}(\emptyset))$ for at least a $1-\varepsilon$ fraction of inputs x. So in this case $\operatorname{sign}(p(x))$ is ε-close to a constant function, and the conclusion of the Lemma certainly holds. Thus we henceforth assume that $|\widehat{p}(\emptyset)|$ is at most $(O(\log(1/\varepsilon)))^{d/2}$.

Let

$$\alpha = \frac{\tau}{(Kn\cdot\ln(4/\varepsilon))^{d/2}}$$

where $K > 0$ is an absolute constant (specified later). For each $S \ne \emptyset$ let $\widehat{q}(S)$ be the value obtained by rounding $\widehat{p}(S)$ to the nearest integer multiple of α, and let $\widehat{q}(\emptyset)$ equal $\widehat{p}(\emptyset)$. This defines a degree-d polynomial $q(x) = \sum_S \widehat{q}(S)\chi_S(x)$. It is easy to see that rescaling by α, all of the non-constant coefficients of $q(x)/\alpha$ are integers. Since each coefficient $\widehat{q}(S)$ has magnitude at most twice that of $\widehat{p}(S)$, we may bound the sum of squares of coefficients of $q(x)/\alpha$ by

$$\frac{\widehat{p}(\emptyset)^2}{\alpha^2} + \frac{\sum_{S\neq\emptyset} 4\widehat{p}(S)^2}{\alpha^2} \leq \frac{(O(\log(1/\varepsilon)))^d}{\alpha^2} \leq n^d \cdot (d/\varepsilon)^{O(d)}.$$

We now observe that the constant coefficient $\widehat{p}(\emptyset)$ of $q(x)$ can be rounded to an integer multiple of α without changing the value of $\mathrm{sign}(q(x))$ for any input x. Doing this, we obtain a polynomial $q'(x)/\alpha$ with all integer coefficients, weight $n^d \cdot (d/\varepsilon)^{O(d)}$, and which has $\mathrm{sign}(q'(x)) = \mathrm{sign}(q(x))$ for all x.

In the rest of our analysis we shall consider the polynomial $q(x)$ (recall that the constant coefficient of $q(x)$ is precisely $\widehat{p}(\emptyset)$). It remains to show that $\mathrm{sign}(q)$ is an ε-approximator for $\mathrm{sign}(p)$. For each $S \neq \emptyset$ let $\widehat{e}(S)$ equal $\widehat{p}(S) - \widehat{q}(S)$. This defines a polynomial (with constant term 0) $e(x) = \sum_S \widehat{e}(S) x_S$, and we have $q(x) + e(x) = p(x)$. (The coefficients $\widehat{e}(S)$ are the "errors" induced by approximating $\widehat{p}(S)$ by $\widehat{q}(S)$.)

Recall that $\tau = (\Theta(1) \cdot \varepsilon/d)^{8d}$. For any input x, we have that $\mathrm{sign}(q(x)) \neq \mathrm{sign}(p(x))$ only if either

(i) $|e(x)| \geq \tau$, or

(ii) $|p(x)| \leq \tau$.

Since each coefficient of $e(x)$ satisfies

$$|\widehat{e}(S)| \leq \frac{\alpha}{2} \leq \frac{\tau}{2(Kn \cdot \ln(4/\varepsilon))^{d/2}},$$

the sum of squares of all (at most n^d) coefficients of e is at most

$$\sum_S \widehat{e}(S)^2 \leq \frac{\tau^2}{4(K \ln(4/\varepsilon))^d}, \quad \text{and thus} \quad \|e\|_2 \leq \frac{\tau}{2(K \ln(4/\varepsilon))^{d/2}}.$$

Applying Theorem 2.2, we get that $\mathbf{Pr}_x[|e(x)| \geq \tau] \leq \varepsilon/2$ (for a suitable absolute constant choice of K), so we have upper bounded the probability of (i).

For (ii), we use the anti-concentration bound for regular polynomials, Claim 4.2. This directly gives us that $\mathbf{Pr}_x[|p(x)| \leq \tau] \leq O(d\tau^{1/8d}) \leq \varepsilon/2$.

Thus the probability, over a random x, that either (i) or (ii) holds is at most ε. Consequently $\mathrm{sign}(q)$ is an ε-approximator for $\mathrm{sign}(p)$, and Lemma 4.1 is proved. □

4.2 Proof of Theorem 1.2

Let $f = \mathrm{sign}(p)$ be an arbitrary degree-d PTF over n Boolean variables, and let $\varepsilon > 0$ be the desired approximation parameter. We invoke Theorem 1.1 with its "τ" parameter set to $\tau = (\Theta(1) \cdot (\varepsilon/2)/d)^{8d}$ (i. e., our choice of τ is obtained by plugging in "$\varepsilon/2$" for ε in the first sentence of Lemma 4.1). Each leaf ρ of the tree T as described in Theorem 1.1 (we call these "good" leaves) is either a τ-regular degree-d PTF or τ-approximated by a constant function. By Lemma 4.1, for each regular leaf ρ there is a degree-d polynomial of weight $n^d \cdot (d/\varepsilon)^{O(d)}$, which we denote $q^{(\rho)}$, such that f_ρ is $\varepsilon/2$-close to $\mathrm{sign}(q^{(\rho)})$. For each of the other leaves in T (which are reached by at most a τ fraction of all inputs to T—we call these "bad" leaves), for which f_ρ is not τ-close to any τ-regular degree-d PTF, let $q^{(\rho)}$ be the constant-1 function.

For each leaf ρ of depth r in T, let $P_\rho(x)$ be the unique multilinear polynomial of degree r which outputs 2^r iff x reaches ρ and outputs 0 otherwise. (As an example, if ρ is a leaf which is reached by the path "$x_3 = -1$, $x_6 = 1$, $x_2 = 1$" from the root in T, then $P_\rho(x)$ would be $(1-x_3)(1+x_6)(1+x_2)$.) Our final PTF is

$$g(x) = \operatorname{sign}(Q(x)), \quad \text{where} \quad Q(x) = \sum_\rho P_\rho(x) q^{(\rho)}(x).$$

It is easy to see that on any input x, the value $Q(x)$ equals $2^{|\rho_x|} \cdot q^{(\rho_x)}(x)$, where we write ρ_x to denote the leaf of T that x reaches and $|\rho_x|$ to denote the depth of that leaf. Thus $\operatorname{sign}(Q(x))$ equals $\operatorname{sign}(q^{(\rho_x)}(x))$ for each x, and from this it follows that $\mathbf{Pr}_x[g(x) \neq f(x)]$ is at most $\tau + \tau + \varepsilon/2 < \varepsilon$. Here the first τ is because a random input x may reach a bad leaf with probability up to τ, and the $\tau + \varepsilon/2$ is because for each good leaf ρ, the function is either τ-approximated by a constant function or $\varepsilon/2$-approximated by $\operatorname{sign}(q^{(\rho)})$.

Since T has depth $\operatorname{depth}(d,\tau)$, it is easy to see that Q has degree at most $\operatorname{depth}(d,\tau) + d$. It is clear that the coefficients of Q are all integers, so it remains only to bound the sum of squares of these coefficients. Each polynomial addend $P_\rho(x) q^{(\rho)}(x)$ in the sum is easily seen to have sum of squared coefficients

$$\sum_S \widehat{P_\rho q^{(\rho)}}(S)^2 = \mathbf{E}[(P_\rho \cdot q^{(\rho)})^2] \leq \left(\max_x P_\rho(x)^2\right) \cdot \mathbf{E}[q^{(\rho)}(x)^2] \leq 2^{2\operatorname{depth}(d,\tau)} \cdot n^d \cdot (d/\varepsilon)^{O(d)}. \tag{4.1}$$

Since T has depth $\operatorname{depth}(d,\tau)$, the number of leaves ρ is at most $2^{\operatorname{depth}(d,\tau)}$, and hence for each S by Cauchy-Schwarz we have

$$\widehat{Q}(S)^2 = \left(\sum_\rho \widehat{P_\rho q^{(\rho)}}(S)\right)^2 \leq 2^{\operatorname{depth}(d,\tau)} \cdot \sum_\rho \widehat{P_\rho q^{(\rho)}}(S)^2. \tag{4.2}$$

This implies that the total weight of Q is

$$\begin{aligned}
\sum_S \widehat{Q}(S)^2 &\leq 2^{\operatorname{depth}(d,\tau)} \cdot \sum_{\rho,S} \widehat{P_\rho q^{(\rho)}}(S)^2 && \text{(using (4.2))} \\
&\leq 2^{2\operatorname{depth}(d,\tau)} \max_\rho \left(\sum_S \widehat{P_\rho q^{(\rho)}}(S)^2\right) \\
&\leq 2^{4\operatorname{depth}(d,\tau)} \cdot n^d \cdot (d/\varepsilon)^{O(d)}, && \text{(using (4.1))}
\end{aligned}$$

and Theorem 1.2 is proved. □

4.3 Degree-d PTFs require $\widetilde{\Omega}(n^d)$-weight approximators

In this section we give two lower bounds on the weight required to ε-approximate certain degree-d PTFs. (We use the notation $\Omega_d()$ below to indicate that the hidden constant of the big-Omega depends on d.)

Theorem 4.3. *For all sufficiently large n, there is a degree-d n-variable PTF $f(x)$ with the following property: Let $K(d)$ be any positive-valued function depending only on d. Suppose that $g(x) = \mathrm{sign}(q(x))$ is a degree-$K(d)$ PTF with integer coefficients $\widehat{q}(S)$ such that $\mathrm{dist}(f,g) \leq \varepsilon^\star$ where $\varepsilon^\star \stackrel{\mathrm{def}}{=} C^{-d}/2$. Then the weight of q is $\Omega_d(n^d/\log n)$.*

Theorem 4.4. *For all sufficiently large n, there is a degree-d n-variable PTF $f(x)$ with the following property: Suppose that $g(x) = \mathrm{sign}(q(x))$ is any PTF (of any degree) with integer coefficients $\widehat{q}(S)$ such that $\mathrm{dist}(f,g) \leq \varepsilon^\star$ where $\varepsilon^\star \stackrel{\mathrm{def}}{=} C^{-d}/2$. Then the weight of q is $\Omega_d(n^{d-1})$.*

Viewing d and ε as constants, Theorem 4.3 implies that the $O(n^d)$ weight bound of our ε-approximator from Theorem 1.2 (of constant degree) is essentially optimal for any constant-degree ε-approximator. Theorem 4.4 says that there is only small room for improving our weight bound even if arbitrary-degree PTFs are allowed as approximators.

Theorems 4.3 and 4.4 are both consequences of the following theorem:

Theorem 4.5. *There exists a set $\mathcal{C} = \{f_1,\ldots,f_M\}$ of $M = 2^{\Omega_d(n^d)}$ degree-d PTFs f_i such that for any $1 \leq i < j \leq M$, we have $\mathrm{dist}(f_i, f_j) \geq C^{-d}$.*

Proof of Theorems 4.3 and 4.4 assuming Theorem 4.5. First we prove Theorem 4.3. We begin by claiming that there are at most

$$\left(3\binom{n}{\leq K(d)}\right)^A$$

many integer-weight PTFs of degree $K(d)$ and weight at most A. This is because any such PTF can be obtained by making a sequence of A steps, where at each step either -1, 0, or 1 is added to one of the $\binom{n}{\leq K(d)}$ many monomials of degree at most $K(d)$. Each step can be carried out in $3\binom{n}{\leq K(d)}$ ways, giving the claimed bound.

By Theorem 4.5, there are M distinct degree-d PTFs $f_1,\ldots,f_M$ any two of which are C^{-d}-far from each other. Consequently any Boolean function (in particular, any weight-A degree-$K(d)$ PTF g) can have $\mathrm{dist}(g, f_i) \leq C^{-d}/2$ for at most one f_i. Since there are only

$$\left(3\binom{n}{\leq K(d)}\right)^A$$

many weight-A degree-$K(d)$ PTFs, which is less than M for some $A = \Omega_d(n^d/\log n)$, it follows that some f_i must have distance at least $C^{-d}/2$ from every weight-A, degree-$K(d)$ PTF. This gives Theorem 4.3.

The proof of Theorem 4.4 is nearly identical. We now use the fact that there are at most $(3\cdot 2^n)^A$ many integer-weight PTFs of weight at most A (and any degree), and use the fact that $(3\cdot 2^n)^A$ is less than M for some $A = \Omega_d(n^{d-1})$. □

It remains to prove Theorem 4.5.

4.3.1 Proof of Theorem 4.5

The proof is by the probabilistic method. We define the following distribution $\mathcal{D}$ over n-variable degree-d polynomials. A draw of $p(x) = \sum_{S\subset[n],|S|=d} \widehat{p}(S)\chi_S(x)$ from $\mathcal{D}$ is obtained in the following way: each of the $\binom{n}{d}$ coefficients $\widehat{p}(S)$ is independently and uniformly selected from $\{-1,1\}$.

We will prove Theorem 4.5 using Lemma 4.6, which says that it is extremely likely the polynomial c—the product of two independent draws a and b from $\mathcal{D}$—will have both small bias and large variance.

Lemma 4.6. *Let $a(x)$ and $b(x)$ be two degree-d polynomials drawn independently from $\mathcal{D}$, and let $c(x) = a(x)b(x)$. Then with probability at least $1-2^{-\Omega_d(n^d)}$ we have:*

1. $|\widehat{c}(\emptyset)| \le \frac{1}{4}C^{-d}\binom{n/2}{d}$, *and*
2. $\mathbf{Var}[c] = \sum_{|S|>0}\widehat{c}(S)^2 \ge \frac{1}{12}\binom{n/2}{d}^2$.

Suppose that Lemma 4.6 holds. Let $a(x)$ and $b(x)$ be independent draws from $\mathcal{D}$ and let $c(x) = a(x)b(x)$ which satisfies the conclusions of the lemma. Then the constant term $\widehat{c}(\emptyset)$ is small compared with the variance of $c(x)$. Let us rescale so the variance is 1; i. e., define the polynomial

$$e(x) \stackrel{\text{def}}{=} \frac{c(x)}{\mathbf{Var}[c]^{1/2}}$$

so $\mathbf{Var}[e] = 1$ and $|\widehat{e}(\emptyset)| < C^{-d}$. We now apply Theorem 2.3 to the degree-$2d$ polynomial $q(x) = -e(x)+\widehat{e}(\emptyset)$, and we see that with probability at least C^{-d} (over a random uniform draw of x) we have $-e(x)+\widehat{e}(\emptyset) > C^{-d}$, and hence $\mathbf{Pr}_x[\mathrm{sign}(e(x)) < 0] > C^{-d}$.

We now observe that $\mathrm{sign}(e(x)) < 0$ if and only if $\mathrm{sign}(a(x)) \ne \mathrm{sign}(b(x))$, and consequently $\mathbf{Pr}_x[\mathrm{sign}(e(x)) < 0]$ is precisely $\mathrm{dist}(\mathrm{sign}(a),\mathrm{sign}(b))$. We thus have that for $a(x), b(x)$ drawn from $\mathcal{D}$ as described above, the probability that $\mathrm{dist}(\mathrm{sign}(a),\mathrm{sign}(b))$ is less than C^{-d} is at most $2^{-\alpha_d n^d}$ for some absolute constant $\alpha_d > 0$ (depending only on d).

Now let us consider

$$M = 2^{(\alpha_d/2)n^d}$$

many independent draws of polynomials $a_1, a_2, \ldots, a_M$ from $\mathcal{D}$. A union bound over all the

$$\binom{M}{2} < 2^{\alpha_d n^d}$$

pairs (i, j) with $1 \le i < j \le M$ gives that with nonzero probability, every a_i, a_j pair satisfies

$$\mathrm{dist}(\mathrm{sign}(a_i),\mathrm{sign}(a_j)) \ge C^{-d}\,.$$

Thus there must be some outcome for the polynomials $a_1, a_2, \ldots, a_M$ such that

$$\mathrm{dist}(\mathrm{sign}(a_i),\mathrm{sign}(a_j)) \ge C^{-d}$$

for all $1 \le i < j \le M$. Setting $f_i = \mathrm{sign}(a_i)$ for this outcome, Theorem 4.5 is proved. □

It remains only for us to prove Lemma 4.6.

4.3.2 Proof of Lemma 4.6

Let us consider $a(x) = \sum_{|S|=d} \widehat{a}(S)\chi_S(x)$ and $b(x) = \sum_{|S|=d} \widehat{b}(S)\chi_S(x)$ drawn independently from $\mathcal{D}$. We will show that the bias of the polynomial $c(x) = a(x)b(x)$ fails to satisfy the bound in item 1 with probability $2^{-\Omega_d(n^d)}$. Then we show the variance of c fails to satisfy item 2 with probability $2^{-\Omega_d(n^d)}$, and the lemma follows from a union bound.

To bound the bias of c, we begin by noting that:

$$\widehat{c}(\emptyset) = \sum_{S\subseteq[n]} \widehat{a}(S)\widehat{b}(S)\,.$$

Each term $\widehat{a}(S)\widehat{b}(S)$ in the summand is uniform, i.i.d in $\{-1,1\}$. Define the random variable $X_S = 1/2-(1/2)\widehat{a}(S)\widehat{b}(S)$. Then $\sum_{S\subseteq[n]} X_S$ is binomially distributed and setting

$$t = \frac{1}{4}C^{-d}\left(\frac{n}{2d}\right)^d,$$

we may apply the Chernoff bound to obtain:

$$\mathbf{Pr}\left[\widehat{c}(\emptyset) < -\frac{1}{4}C^{-d}\left(\frac{n}{2d}\right)^d\right] = \mathbf{Pr}[X > \mathbf{E}[X]+t] \le \exp\left(-2\frac{t^2}{\binom{n}{d}}\right) = 2^{-\Omega_d(n^d)}\,.$$

The same analysis gives a bound on the magnitude in the negative direction. Since

$$\binom{n/2}{d} \ge \left(\frac{n}{2d}\right)^d,$$

this concludes the analysis for the first item of the lemma.

Now we show that item 2 of the lemma also fails with very small probability. The following terminology will be useful. Let $T\subset[n]$ be a subset of size exactly $|T| = 2d$ (we think of T as the set of variables defining some monomial of degree $2d$). For such a T we let

$$\mathrm{first}(T) \stackrel{\mathrm{def}}{=} T\cap[n/2] \qquad\text{and}\qquad \mathrm{second}(T) \stackrel{\mathrm{def}}{=} T\cap[n/2+1,n]\,.$$

We say that such a T is *balanced* if $|\mathrm{first}(T)| = |\mathrm{second}(T)| = d$. Note that there are exactly $\binom{n/2}{d}^2$ many balanced subsets T.

We say that a subset $U\subset[n]$, $|U| = d$ is *pure* if U is contained entirely in $[n/2+1,n]$.

Let us consider

$$a(x) = \sum_{|S|=d}\widehat{a}(S)\chi_S(x) \qquad\text{and}\qquad b(x) = \sum_{|S|=d}\widehat{b}(S)\chi_S(x)$$

drawn independently from $\mathcal{D}$. Fix any outcome for a (i. e., for the values of all $\binom{n}{d}$ coefficients $\widehat{a}(S)$), and fix any outcome for $\widehat{b}(U)$ for every U which is *not* pure. Thus the only "remaining randomness" is the value (drawn uniformly from $\{-1,1\}\}$) for each of the $\binom{n/2}{d}$ coefficients $\widehat{b}(U)$ for pure U. We will show that with probability at least $1-2^{-\Omega_d(n^d)}$ over the remaining randomness, at least $1/6$ of the

$$\binom{n/2}{d}^2$$

many balanced subsets T have $\widehat{c}(T) \neq 0$. Since each value $\widehat{c}(T)$ which is nonzero is at least 1 in magnitude, this suffices to prove the lemma.

Consider any fixed pure subset $U \subset [n], |U| = d$ (for example $U = \{n-d+1,\ldots,n\}$). Let T be a balanced subset of n (so $|T| = 2d$) such that second(T) equals U. (There are precisely $\binom{n/2}{d}$ balanced subsets T with this property; let $\mathcal{T}_U$ denote the collection of all $\binom{n/2}{d}$ of them.) Consider the value $\widehat{c}(T)$: this is

$$\widehat{c}(T) = \sum_{S \subseteq T, |S|=d} \widehat{a}(S)\widehat{b}(T-S)\,.$$

The only "not-yet-fixed" part of the above expression is the single coefficient $\widehat{b}(U)$; everything else has been fixed. Since the coefficient $\widehat{a}(T-U)$ of $\widehat{b}(U)$ is a nonzero integer, there are two possible outcomes for the value of $\widehat{c}(T)$, depending on whether $\widehat{b}(U)$ is set to +1 or -1. These two possible values differ by 2; consequently, there is at most one possible outcome of $\widehat{b}(U)$ that will cause $\widehat{c}(T)$ to be zero. (Note that it may well be the case that no outcome for $\widehat{b}(U)$ would cause $\widehat{c}(T)$ to become zero.)

Let us say that an outcome of $\widehat{b}(U)$ is *pernicious* if it has the following property: at most $1/3$ of the $\binom{n/2}{d}$ elements $T \in \mathcal{T}_U$ have $\widehat{c}(T)$ take a nonzero value under that outcome of $\widehat{b}(U)$. (Equivalently, at least $2/3$ of the $\binom{n/2}{d}$ elements $T \in \mathcal{T}_U$ have $\widehat{c}(T)$ become zero under that outcome of $\widehat{b}(U)$.) It may be the case that neither outcome in $\{-1,1\}$ for $\widehat{b}(U)$ is pernicious (e. g., if each outcome makes at least 95% of the $\widehat{c}(T)$ values come out nonzero). It cannot be the case that both outcomes $\{-1,1\}$ for $\widehat{b}(U)$ are pernicious (for if there were two pernicious outcomes, this would mean that at least $1/3$ of the $\widehat{c}(T)$ values evaluate to 0 under both outcomes for $\widehat{b}(U)$, but it is impossible for any $\widehat{c}(T)$ to evaluate to 0 under two outcomes for $\widehat{b}(U)$). Consequently we have

$$\mathbf{Pr}\big[\text{the outcome of } \widehat{b}(U) \text{ is pernicious}\big] \leq 1/2\,.$$

This is true independently for each of the $\binom{n/2}{d}$ many pure subsets U. As a result, a simple analysis gives

$$\mathbf{Pr}\left[\text{at least } 3/4 \text{ of the } \binom{n/2}{d} \text{ pure subsets } U \text{ have a pernicious outcome}\right] \leq 2^{-\Omega_d(n^d)}\,.$$

Thus we may assume that fewer than $3/4$ of the $\binom{n/2}{d}$ pure subsets U have a pernicious outcome. So at least $1/4$ of the pure subsets U are non-pernicious. For each such non-pernicious U, more than $1/3$ of the $\binom{n/2}{d}$ elements in $\mathcal{T}_U$ have $\widehat{c}(T)$ take a nonzero value. Consequently, at least

$$\frac{1}{12}\binom{n/2}{d}^2$$

many balanced subsets T overall have $\widehat{c}(T) \neq 0$. This proves the lemma. □

5 Conclusions and open problems

In this paper, we gave a regularity lemma for low-degree PTFs. Our lemma roughly says that any low-degree PTF can be decomposed as a shallow binary decision tree $\mathcal{T}$, such that "most" leaves of $\mathcal{T}$ correspond to "regular" low-degree PTFs (or constant functions). As an application, we showed that any constant-degree PTF is well-approximated by a constant-degree PTF with low integer weights. Our result has also been used as an essential ingredient in the recent line of work [8, 19] establishing that bounded independence fools low-degree PTFs.

We conclude this paper by mentioning a few relevant open problems. At the quantitative level, an obvious question is whether the depth of the tree $\mathcal{T}$ must depend exponentially on the degree parameter d. It is plausible to conjecture that such a dependence on d is inherent. It would also be very interesting to devise qualitatively improved "regularity lemmas" for PTFs (e. g., using a more refined notion of regularity) that may result in improved bounds for applications; see [20] for a first important step in this direction.

We believe that our regularity lemma may find other applications. For example, it may be useful in the problem of algorithmically reconstructing a degree-d PTF from its Fourier coefficients of degree at most d (see [30] for a solution to the $d = 1$ version of this problem). Finally, regarding integer-weight approximations, say that an integer weight approximator to a degree-d PTF is *proper* if its degree is (at most) d. Clearly, the approximators that we obtain are not proper; constructing proper such approximations is left as an interesting open problem.

References

[1] JAMES ASPNES, RICHARD BEIGEL, MERRICK L. FURST, AND STEVEN RUDICH: The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994. Preliminary version in STOC'91. [doi:10.1007/BF01215346] 28

[2] PER AUSTRIN AND JOHAN HÅSTAD: Randomly supported independence and resistance. *SIAM J. Comput.*, 40(1):1–27, 2011. Preliminary version in STOC'09. [doi:10.1137/100783534] 33

[3] ALINE BONAMI: Étude des coefficients Fourier des fonctions de $L^p(G)$. *Annales de l'Institute Fourier*, 20(2):335–402, 1970. EuDML. 32

[4] JEHOSHUA BRUCK: Harmonic analysis of polynomial threshold functions. *SIAM J. Discrete Math.*, 3(2):168–177, 1990. [doi:10.1137/0403015] 28

[5] ANTHONY CARBERY AND JAMES WRIGHT: Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Mathematical Research Letters*, 8(3):233–248, 2001. [doi:10.4310/MRL.2001.v8.n3.a1] 33, 42

[6] ILIAS DIAKONIKOLAS, PARIKSHIT GOPALAN, RAGESH JAISWAL, ROCCO A. SERVEDIO, AND EMANUELE VIOLA: Bounded independence fools halfspaces. *SIAM J. Comput.*, 39(8):3441–3462, 2010. Preliminary version in FOCS'09. See also at ECCC. [doi:10.1137/100783030] 28, 29, 30

[7] ILIAS DIAKONIKOLAS, PRAHLADH HARSHA, ADAM KLIVANS, RAGHU MEKA, PRASAD RAGHAVENDRA, ROCCO A. SERVEDIO, AND LI-YANG TAN: Bounding the average sensitivity and noise sensitivity of polynomial threshold functions. In *Proc. 42nd STOC*, pp. 533–542. ACM Press, 2010. [doi:10.1145/1806689.1806763] 30

[8] ILIAS DIAKONIKOLAS, DANIEL M. KANE, AND JELANI NELSON: Bounded independence fools degree-2 threshold functions. In *Proc. 51st FOCS*, pp. 11–20. IEEE Comp. Soc. Press, 2010. See also at ECCC and arXiv. [doi:10.1109/FOCS.2010.8] 29, 30, 31, 49

[9] ILIAS DIAKONIKOLAS, PRASAD RAGHAVENDRA, ROCCO SERVEDIO, AND LI-YANG TAN: Average sensitivity and noise sensitivity of polynomial threshold functions. Technical report, 2009. To appear in SIAM J. Comput. [arXiv:0909.5011] 28, 29, 30, 33

[10] ILIAS DIAKONIKOLAS AND ROCCO A. SERVEDIO: Improved approximation of linear threshold functions. *Comput. Complexity*, 22(3):623–677, 2013. Preliminary version in CCC'09. [doi:10.1007/s00037-012-0045-5] 28, 30

[11] ILIAS DIAKONIKOLAS, ROCCO A. SERVEDIO, LI-YANG TAN, AND ANDREW WAN: A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions. In *Proc. 25th IEEE Conf. on Computational Complexity (CCC'10)*, pp. 211–222. IEEE Comp. Soc. Press, 2010. [doi:10.1109/CCC.2010.28] 27, 29

[12] IRIT DINUR, EHUD FRIEDGUT, GUY KINDLER, AND RYAN O'DONNELL: On the Fourier tails of bounded functions over the discrete cube. *Israel J. Math.*, 160(1):389–412, 2007. Preliminary version in STOC'06. [doi:10.1007/s11856-007-0068-9] 33

[13] CRAIG GOTSMAN AND NATHAN LINIAL: Spectral properties of threshold functions. *Combinatorica*, 14(1):35–50, 1994. [doi:10.1007/BF01305949] 28

[14] BEN GREEN: A Szemerédi-type regularity lemma in abelian groups, with applications. *Geometric and Functional Analysis (GAFA)*, 15(2):340–376, 2005. See also at arXiv. [doi:10.1007/s00039-005-0509-8] 28

[15] LEONARD GROSS: Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975. JSTOR. 32

[16] PRAHLADH HARSHA, ADAM KLIVANS, AND RAGHU MEKA: Bounding the sensitivity of polynomial threshold functions. *Theory of Computing*, 10(1):1–26, 2014. See also at arXiv. [doi:10.4086/toc.2014.v010a001] 28, 30

[17] JEFF KAHN, GIL KALAI, AND NATHAN LINIAL: The influence of variables on Boolean functions. In *Proc. 29th FOCS*, pp. 68–80. IEEE Comp. Soc. Press, 1988. [doi:10.1109/SFCS.1988.21923] 32

[18] GIL KALAI AND SHMUEL SAFRA: Threshold phenomena and influence. In *Computational Complexity and Statistical Physics*, pp. 25–60. Oxford Univ. Press, 2006. Available at author's home page. 28

[19] DANIEL M. KANE: k-independent Gaussians fool polynomial threshold functions. In *Proc. 26th IEEE Conf. on Computational Complexity (CCC'11)*, pp. 252–261. IEEE Comp. Soc. Press, 2011. See also at arXiv. [doi:10.1109/CCC.2011.13] 29, 30, 49

[20] DANIEL M. KANE: A structure theorem for poorly anticoncentrated Gaussian chaoses and applications to the study of polynomial threshold functions. In *Proc. 53rd FOCS*, pp. 91–100. IEEE Comp. Soc. Press, 2012. [doi:10.1109/FOCS.2012.52] 49

[21] DANIEL M. KANE: The correct exponent for the Gotsman-Linial conjecture. In *Proc. 28th IEEE Conf. on Computational Complexity (CCC'13)*, pp. 56–64, 2013. See also at arXiv. [doi:10.1109/CCC.2013.15] 29

[22] NATHAN LINIAL, YISHAY MANSOUR, AND NOAM NISAN: Constant depth circuits, Fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993. Preliminary version in FOCS'89. [doi:10.1145/174130.174138] 29, 34

[23] SHACHAR LOVETT, IDO BEN-ELIEZER, AND ARIEL YADIN: Polynomial threshold functions: Structure, approximation and pseudorandomness. *Electron. Colloq. on Comput. Complexity (ECCC)*, 16:118, 2009. ECCC. See also at arXiv. 30, 31

[24] KEVIN MATULEF, RYAN O'DONNELL, RONITT RUBINFELD, AND ROCCO A. SERVEDIO: Testing halfspaces. *SIAM J. Comput.*, 39(5):2004–2047, 2010. Preliminary version in SODA'09. See also at ECCC, and related paper in Property Testing. [doi:10.1137/070707890] 28

[25] RAGHU MEKA AND DAVID ZUCKERMAN: Pseudorandom generators for polynomial threshold functions. *SIAM J. Comput.*, 42(3):1275–1301, 2013. Preliminary version in STOC'10. See also at arXiv. [doi:10.1137/100811623] 28, 29, 30, 31

[26] ELCHANAN MOSSEL, RYAN O'DONNELL, AND KRZYSZTOF OLESZKIEWICZ: Noise stability of functions with low influences: invariance and optimality. *Ann. of Math.*, 171(1):295–341, 2010. Preliminary version in FOCS'05. [doi:10.4007/annals.2010.171.295] 28, 33, 42

[27] RYAN O'DONNELL: *Analysis of Boolean Functions*. Cambridge Univ. Press, 2014. See online version here. 28

[28] RYAN O'DONNELL AND ROCCO A. SERVEDIO: Extremal properties of polynomial threshold functions. *J. Comput. System Sci.*, 74(3):298–312, 2008. Preliminary version in CCC'03. [doi:10.1016/j.jcss.2007.06.021] 28

[29] RYAN O'DONNELL AND ROCCO A. SERVEDIO: New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010. Preliminary version in STOC'03. [doi:10.1007/s00493-010-2173-3] 28

[30] RYAN O'DONNELL AND ROCCO A. SERVEDIO: The Chow parameters problem. *SIAM J. Comput.*, 40(1):165–199, 2011. Preliminary version in STOC'08. [doi:10.1137/090756466] 28, 29, 30, 49

[31] VLADIMIR V. PODOLSKII: Perceptrons of large weight. *Problems of Information Transmission*, 45(1):46–53, 2009. Preliminary version in CSR'07. [doi:10.1134/S0032946009010062] 29

[32] MICHAEL SAKS: Slicing the hypercube. In KEITH WALKER, editor, *Surveys in Combinatorics 1993*, pp. 211–256. Cambridge Univ. Press, 1993. [doi:10.1017/CBO9780511662089.009] 28

[33] ROCCO A. SERVEDIO: Every linear threshold function has a low-weight approximator. *Comput. Complexity*, 16(2):180–209, 2007. Preliminary version in CCC'06. [doi:10.1007/s00037-007-0228-7] 28, 29, 30, 42

[34] ALEXANDER A. SHERSTOV: The intersection of two halfspaces has high threshold degree. In *Proc. 50th FOCS*, pp. 343–362. IEEE Comp. Soc. Press, 2009. See also at ECCC. [doi:10.1109/FOCS.2009.18] 28

[35] ENDRE SZEMERÉDI: Regular partitions of graphs. *Colloq. Internat. CNRS: Problèmes combinatoires et théorie des graphes*, 260:399–401, 1978. 28

[36] TERENCE TAO: Structure and randomness in combinatorics. In *Proc. 48th FOCS*, pp. 3–15. IEEE Comp. Soc. Press, 2007. [doi:10.1109/FOCS.2007.17] 28

[37] RONALD DE WOLF: *A Brief Introduction to Fourier Analysis on the Boolean Cube*. Number 1 in Graduate Surveys. Theory of Computing Library, 2008. [doi:10.4086/toc.gs.2008.001] 32

AUTHORS

Ilias Diakonikolas
School of Informatics
University of Edinburgh
ilias.d@ed.ac.uk
http://www.iliasdiakonikolas.org

Rocco A. Servedio
Department of Computer Science
Columbia University
rocco@cs.columbia.edu
http://www.cs.columbia.edu/~rocco

Li-Yang Tan
Department of Computer Science
Columbia University
liyang@cs.columbia.edu
http://www.cs.columbia.edu/~liyang

Andrew Wan
Simons Institute for the Theory of Computing
University of California, Berkeley
atw12@eecs.berkeley.edu
http://people.seas.harvard.edu/~atw12

ABOUT THE AUTHORS

ILIAS DIAKONIKOLAS is an assistant professor in the School of Informatics at the University of Edinburgh. He obtained his Ph. D. from Columbia University advised by Mihalis Yannakakis. He is interested in algorithms, learning, statistics, and game theory.

ROCCO SERVEDIO is an associate professor in the Department of Computer Science at Columbia University. He graduated from Harvard University, where his Ph. D. was supervised by Les Valiant. He is interested in computational learning theory, computational complexity, and property testing, with the study of Boolean functions as an underlying theme tying these topics together. He enjoys spending time with his family and hopes to share a bottle of Cutty Sark with H.M. in the afterlife.

LI-YANG TAN is a Ph. D. student at Columbia University where he is fortunate to be advised by Rocco Servedio. His research interests lie in complexity theory, with a focus on discrete Fourier analysis, the complexity of Boolean functions, and computational learning theory.

ANDREW WAN is currently a research fellow at the Simons Institute for the Theory of Computing at UC Berkeley. Before this appointment, he was an assistant professor at IIIS, Tsinghua University. He received his doctorate from Columbia University in 2010 under the supervision of Tal Malkin and Rocco Servedio. His interests include complexity theory, cryptography, and computational learning theory. Before graduate school, he was a student of philosophy at Columbia University and enjoyed playing the piano, the trumpet, and the accordion. Although he still enjoys playing music, the PAC model rarely affords him the time.