THEORY OF COMPUTING, Volume 10 (1), 2014, pp. 1–26
www.theoryofcomputing.org

SPECIAL ISSUE: ANALYSIS OF BOOLEAN FUNCTIONS

Bounding the Sensitivity of Polynomial Threshold Functions

Prahladh Harsha* Adam Klivans Raghu Meka†

Received July 16, 2012; Revised November 11, 2013; Published March 25, 2014

Abstract: We give the first nontrivial upper bounds on the average sensitivity and noise sensitivity of polynomial threshold functions. More specifically, for a Boolean function f on n variables equal to the sign of a real, multivariate polynomial of total degree d, we prove

- The average sensitivity of f is at most $O(n^{1-1/(4d+6)})$. (We also give a combinatorial proof of the bound $O(n^{1-1/2^d})$.)
- The noise sensitivity of f with noise rate δ is at most $O(\delta^{1/(4d+6)})$.

Previously, only bounds for the degree $d = 1$ case were known ($O(\sqrt{n})$ and $O(\sqrt{\delta})$, for average and noise sensitivity, respectively).

We highlight some applications of our results in learning theory where our bounds immediately yield new agnostic learning algorithms and resolve an open problem of Klivans, O'Donnell and Servedio (FOCS'08).

An extended abstract of this result, which was proved independently by two groups (the authors of this paper and Diakonikolas et al. [11]), appeared in the Proc. 42nd ACM Symposium on Theory of Computing, 2010 [10].

*Work done while the author was visiting the University of Texas at Austin.

†Work done while the author was at the University of Texas at Austin.

ACM Classification: F.2.2, G.3, I.2.6

AMS Classification: 68Q25, 68Q32

Key words and phrases: polynomial threshold functions, average sensitivity

 DOI: 10.4086/toc.2014.v010a001

The proof of our results use (i) the invariance principle of Mossel, O'Donnell and Oleszkiewicz (2010), (ii) the anti-concentration properties of polynomials in Gaussian space due to Carbery and Wright (2001) and (iii) new structural theorems about random restrictions of polynomial threshold functions obtained via hypercontractivity.

These structural results may be of independent interest, as they provide a generic template for transforming problems related to polynomial threshold functions defined on the Boolean hypercube to polynomial threshold functions defined in Gaussian space.

1 Introduction

1.1 Background

Let P be a real, multivariate polynomial of degree d, and let $f = \mathsf{sign}(P)$. We say that the Boolean function f is a polynomial threshold function (PTF) of degree d. PTFs play an important role in computational complexity with applications in circuit complexity [2, 4], learning theory [26, 24], communication complexity [35, 36], and quantum computing [3]. While many interesting properties (e. g., Fourier spectra, influence, sensitivity) have been characterized for the case $d = 1$ of linear threshold functions (LTFs), very little is known for degrees 2 and higher. Gotsman and Linial [13] conjectured, for example, that the average sensitivity of a degree-d polynomial is $O(d\sqrt{n})$. In this work, we take a step towards resolving this conjecture and give the first nontrivial bounds on the average sensitivity and noise sensitivity of degree-d PTFs (Theorem 1.6).

Average sensitivity [5] and noise sensitivity [17, 6] are two fundamental quantities that arise in the analysis of Boolean functions. Roughly speaking, the average sensitivity of a Boolean function f measures the expected number of bit positions that change the sign of f for a randomly chosen input, and the noise sensitivity of f measures the probability over a randomly chosen input x that f changes sign if each bit of x is flipped independently with probability δ (we give formal definitions below).

Bounds on the average and noise sensitivity of Boolean functions have direct applications in hardness of approximation [14, 23], hardness amplification [31], circuit complexity [27], the theory of social choice [19], and quantum complexity [37]. In this paper, we focus on applications in learning theory, where it is known that bounds on the noise sensitivity of a class of Boolean functions yield learning algorithms for the class that succeed in harsh noise models (i. e., work in the agnostic model of learning) [18]. We obtain the first efficient algorithms for agnostically learning PTFs with respect to the uniform distribution on the hypercube. We also give efficient algorithms for agnostically learning ellipsoids in $\mathbb{R}^n$ with respect to the Gaussian distribution, resolving an open problem of Klivans et al. [25]. We discuss these learning theory applications in Section 2.

1.2 Main definitions and results

We begin by defining the (Boolean) noise sensitivity of a Boolean function:

Definition 1.1 (Boolean noise sensitivity). Let f be a Boolean function $f : \{1,-1\}^n \to \{1,-1\}$. For any $\delta \in (0,1)$, let X be a random element of the hypercube $\{1,-1\}^n$ and Z a δ-perturbation of X defined as follows: for each i independently, Z_i is set to X_i with probability $1-\delta$ and $-X_i$ with probability δ. The noise sensitivity of f, denoted $\mathbb{NS}_\delta(f)$, for noise δ is then defined as follows:

$$\mathbb{NS}_\delta(f) = \Pr[f(X) \neq f(Z)] .$$

Intuitively, the Boolean noise sensitivity of f measures the probability that f changes value when a random input to f is perturbed slightly. In order to analyze Boolean noise sensitivity, we will also need to analyze the Gaussian noise sensitivity, which is defined similarly, but the random variables X and Z are drawn from a multivariate Gaussian distribution. Let $\mathcal{N} = \mathcal{N}(0,1)$ denote the univariate Gaussian distribution on $\mathbb{R}$ with mean 0 and variance 1.

Definition 1.2 (Gaussian noise sensitivity)**.** Let $f : \mathbb{R}^n \to \{-1,1\}$ be any Boolean function on $\mathbb{R}^n$. Let X,Y be two independent random variables drawn from the multivariate Gaussian distribution $\mathcal{N}^n$ and Z a δ-perturbation of X defined by $Z = (1-\delta)X + \sqrt{2\delta-\delta^2}Y$. The Gaussian noise sensitivity of f, denoted $\mathbb{GNS}_\delta(f)$, for noise δ is defined as follows:

$$\mathbb{GNS}_\delta(f) = \Pr[f(X) \neq f(Z)] .$$

It is well known that the Boolean and Gaussian noise sensitivity of LTFs are at most $O(\sqrt{\delta})$. Our results give the first nontrivial bounds for degrees 2 and higher in both the Gaussian and Boolean cases, with the Gaussian case being considerably easier to handle than the Boolean case.

Theorem 1.3 (Boolean noise sensitivity)**.** *For any degree-d PTF $f : \{1,-1\}^n \to \{1,-1\}$ and $0 < \delta < 1$,*

$$\mathbb{NS}_\delta(f) = 2^{O(d)} \cdot \left(\delta^{1/(4d+6)}\right) .$$

For the Gaussian case, we get a slightly better dependence on the degree d.

Theorem 1.4 (Gaussian noise sensitivity)**.** *For any degree-d polynomial P such that P is either multilinear or corresponds to an ellipsoid, the following holds for the corresponding PTF $f = \mathsf{sign}(P)$. For all $0 < \delta < 1$,*

$$\mathbb{GNS}_\delta(f) = 2^{O(d)} \cdot \left(\delta^{1/(2d+1)}\right) .$$

Diakonikolas et al. [11] prove that a similar bound holds for all degree-d PTFs. Our next set of results bound the average sensitivity or total influence of degree-d PTFs.

Definition 1.5 (average sensitivity)**.** Let f be a Boolean function, and let X be a random element of the hypercube $\{1,-1\}^n$. Let $X^{(i)} \in \{1,-1\}^n$ be such that $X_i^{(i)} = -X_i$ and $X_j^{(i)} = X_j$ for $j \neq i$. Then, the influence of the i^{th} variable is defined by

$$\mathbb{I}_i(f) = \Pr\left[f(X) \neq f\left(X^{(i)}\right)\right] .$$

The sum of all the influences is referred to as the average sensitivity of the function f,

$$\mathbb{AS}(f) = \sum_i \mathbb{I}_i(f) .$$

Clearly, for any function f, $\mathbb{AS}(f)$ is at most n. It is well known that the average sensitivity of "unate" functions (functions monotone in each coordinate), and thus of LTFs in particular is $O(\sqrt{n})$. This bound is tight as the Majority function has average sensitivity $\Theta(\sqrt{n})$. As mentioned before, Gotsman and Linial [13] conjectured in 1994 that the average sensitivity of any degree-d PTF f is $O(d\sqrt{n})$. We are not aware of any progress on this conjecture until now, with no $o(n)$ bounds known.

We give two upper bounds on the average sensitivity of degree-d PTFs. We first use a simple translation lemma for bounding average sensitivity in terms of noise sensitivity of a Boolean function and Theorem 1.3 to obtain the following bound.

Theorem 1.6 (average sensitivity)**.** *For a degree-d PTF $f : \{1,-1\}^n \to \{1,-1\}$,*

$$\mathbb{AS}(f) = 2^{O(d)} \cdot \left(n^{1-1/(4d+6)}\right).$$

We also give an elementary combinatorial argument, to show that the average sensitivity of any degree-d PTF is at most $3n^{1-1/2^d}$. The combinatorial proof is based on the following lemma for general Boolean functions that may prove useful elsewhere. For $x \in \{1,-1\}^n$, and $i \in [n]$, let $x_{-i} = (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$.

Lemma 1.7. *For Boolean functions $f_i : \{1,-1\}^n \to \{1,-1\}$ with f_i not depending on the i'th coordinate x_i, and $X \in_u \{1,-1\}^n$,*

$$\mathop{\mathbb{E}}_X \left[\left| \sum_i X_i f_i(X_{-i}) \right| \right]^2 \le 2 \sum_i \mathbb{AS}(f_i) + n\,.$$

We believe that when the functions f_i in the above lemma are LTFs, the above bound can be improved to $O(n)$, which in turn would imply the Gotsman-Linial conjecture for quadratic threshold functions.

1.3 Random restrictions of PTFs—a structural result

An important ingredient of our sensitivity bounds for PTFs are new structural theorems about random restrictions of PTFs obtained via *hypercontractivity*. The structural results we obtain can be seen as part of the high level "randomness vs. structure" paradigm that has played a fundamental role in many recent breakthroughs in additive number theory and combinatorics. Specifically, we obtain the following structural result (Lemmas 5.1 and 5.2): for any PTF, there exists a small set of variables such that with at least a constant probability, any random restriction of these variables satisfies one of the following: (1) the restricted polynomial is "regular" in the sense that no single variable has large influence or (2) the sign of the restricted polynomial is a very biased function.

We remark that our structural results, though motivated by similar results of Servedio [34] and Diakonikolas et al. [9] for the simpler case of LTFs, do not follow from a generalization of their arguments for LTFs to PTFs. The structural results for random restrictions of low-degree PTFs provide a reasonably generic template for reducing problems involving arbitrary PTFs to ones on regular PTFs. In fact, these structural properties are used precisely for the above reason both in this work and in a parallel work by one of the authors, Meka and Zuckerman [29] to construct pseudorandom generators for PTFs.

1.4 Related and subsequent work

Independent of this work, Diakonikolas, Raghavendra, Servedio, and Tan [11] have obtained nearly identical results to ours for both the average and noise sensitivity of PTFs. The broad outline of their proof is also similar to ours. In our proof, we first obtain bounds on noise sensitivity and then move to average sensitivity using a translation lemma. On the other hand, Diakonikolas et al. [11] first obtain bounds on the average sensitivity of PTFs and then use a generalization of Peres' argument [33] for LTFs to move from average sensitivity to noise sensitivity.

Regarding our structural result described in Section 1.3, Diakonikolas, Servedio, Tan and Wan [12] have independently obtained similar results to ours. As an application, they prove the existence of low-weight approximators for polynomial threshold functions. In the context of approximating functions with many influential variables by low-degree PTFs, Ben-Eliezer, Lovett and Yadin [28] obtained similar structural results (with weaker parameters) independent of both our work and that of [12].

There has been considerable progress towards the resolution of the Gotsman-Linial conjecture since our work. In particular, Kane proved optimal upper bounds for the average sensitivity of PTFs in the Gaussian setting [20] and an upper bound of $O_{d,\varepsilon}(n^{5/6+\varepsilon})$ in the Boolean setting [21]. Building on this final work, Kane recently showed an upper bound of $\sqrt{n}(\log n)^{O(d\log d)}2^{O(d^2\log d)}$. These later works have a similar broad outline as ours (use regularity lemma to reduce to Gaussian case) and use a much stronger structure theorem for PTFs called *diffused decompositions* by Kane [21].

1.5 Proof outline

The proofs of our theorems are inspired by the use of the *invariance principle* in the proof of the "Majority is Stablest" theorem [30]. As in the proof of the "Majority is Stablest" theorem, our main technical tools are the invariance principle and the anti-concentration bounds (also called small ball probabilities) of Carbery and Wright [8].

Bounding the probability that a threshold function changes value either when it is perturbed slightly (in the case of noise sensitivity) or when a variable is flipped (average sensitivity) involves bounding probabilities of the form $\Pr[|Q(X)| \leq |R(X)|]$ where $Q(X), R(X)$ are low-degree polynomials and R has small ℓ_2-norm relative to that of Q. The event $|Q(X)| \leq |R(X)|$ implies that either $|Q(X)|$ is small or $|R(X)|$ is large. In other words, for every γ

$$\Pr[|Q(X)| \leq |R(X)|] \leq \Pr[|Q(X)| \leq \gamma] + \Pr[|R(X)| > \gamma].$$

Since R has small norm, the second quantity in the above expression can be easily bounded using a tail bound (even Markov's inequality suffices). Bounding the first quantity is trickier. Our first observation is that if the random variable X were distributed according to the Gaussian distribution as opposed to the uniform distribution on the hypercube, bounds on probabilities of the form $\Pr[|Q(X)| \leq \gamma]$ immediately follow from the anti-concentration bounds of Carbery and Wright [8]. We then transfer these bounds to the Boolean setting using the invariance principle.

Unfortunately, the invariance principle holds only for regular polynomials (i.e., polynomials in which no single variable has large influence). We thus obtain the required bounds on noise sensitivity and average sensitivity for the special case of regular PTFs. We then extend these results to an arbitrary PTF f using our structural results on random restrictions of the PTF f. The structural results state that either the restricted PTF is a regular polynomial or is a very biased function. In the former case, we resort to the above argument for regular PTFs and bound the noise sensitivity of the given PTF. In the latter case, we merely note that the noise sensitivity of a biased function can be easily bounded. This in turn lets us extend the results for regular PTFs to all PTFs.

2 Learning theory applications

In this section we briefly elaborate on the learning theory applications of our results. Our bounds on Boolean and Gaussian noise sensitivity imply learning results in the challenging *agnostic* model of learning of Haussler [15] and Kearns, Schapire and Sellie [22] which we define below.

Definition 2.1. Let $\mathcal{D}$ be an arbitrary distribution on $\mathcal{X}$ and $\mathcal{C}$ a class of Boolean functions $f : \mathcal{X} \to \{-1,1\}$. For $\delta, \varepsilon \in (0,1)$, we say that algorithm A is a (δ,ε)-agnostic learning algorithm for $\mathcal{C}$ with respect to $\mathcal{D}$ if the following holds. For any distribution $\mathcal{D}'$ on $\mathcal{X} \times \{-1,1\}$ whose marginal over $\mathcal{X}$ is $\mathcal{D}$, if A is given access to a set of labeled examples (x,y) drawn from $\mathcal{D}'$, then with probability at least $1-\delta$ algorithm A outputs a hypothesis $h : \mathcal{X} \to \{-1,1\}$ such that

$$\Pr_{(x,y)\sim\mathcal{D}'}[h(x) \neq y] \leq \mathsf{opt} + \varepsilon$$

where opt is the error made by the best classifier in $\mathcal{C}$, that is,

$$\mathsf{opt} = \inf_{g\in\mathcal{C}} \Pr_{(x,y)\sim\mathcal{D}'}[g(x) \neq y] \,.$$

Kalai, Klivans, Mansour and Servedio [18] showed that the existence of low-degree real valued polynomial ℓ_2-approximators to a class of functions, implies agnostic learning algorithms for the class. In an earlier result, Klivans, O'Donnell and Servedio [24] gave a precise relationship between polynomial approximation and noise sensitivity, essentially showing that small noise sensitivity bounds imply good low-degree polynomial ℓ_2-approximators.

Combining these two results, it follows that bounding the noise sensitivity (either Boolean or Gaussian) of a concept class $\mathcal{C}$ yields an agnostic learning algorithm for $\mathcal{C}$ (with respect to the appropriate distribution). Thus, using our bounds on noise sensitivity of PTFs, we obtain corresponding learning algorithms for PTFs with respect to the uniform distribution over the hypercube.

Theorem 2.2. *The concept class of degree-d PTFs is agnostically learnable to within ε with respect to the uniform distribution on $\{-1,1\}^n$ in time $n^{1/\varepsilon^{O(d)}}$.*

These are the first polynomial-time algorithms for agnostically learning constant-degree PTFs with respect to the uniform distribution on the hypercube (to within any constant error parameter). Previously, Klivans et al. [25] had shown that quadratic (degree-2) PTFs corresponding to spheres are agnostically learnable with respect to spherical Gaussians on $\mathbb{R}^n$. Our bounds on the Gaussian noise sensitivity of ellipsoids imply that this result can be extended to all ellipsoids with respect to (not necessarily spherical) Gaussian distributions thus resolving an open problem of Klivans et al. [25].

It is implicit from a recent paper of Blais, O'Donnell and Wimmer [7] that bounding the Boolean noise sensitivity for a concept class $\mathcal{C}$ yields non-trivial learning algorithms for a very broad class of discrete and continuous product distributions. We believe this is additional motivation for obtaining bounds on a function's Boolean noise sensitivity.

3 Organization

The rest of the paper is organized as follows. We introduce the necessary notation and preliminaries in Section 4. We then present the structural results on random restrictions of PTFs (Lemmas 5.1 and 5.2) in Section 5. In Section 6 we present our analysis of Gaussian noise sensitivity, followed by the analysis of Boolean noise sensitivity in Section 7. We remark that the analysis of the Gaussian noise sensitivity is simpler than the Boolean noise sensitivity analysis, since the Boolean case, in some sense, reduces to the "regular" or Gaussian case. We then present our bounds on average sensitivity of PTFs in Section 8.

4 Notation and preliminaries

We will consider functions/polynomials over n variables $X_1,\dots,X_n$. Corresponding to any set $I \subseteq [n]$ (possibly multi-set), there is a monomial X^I defined as $X^I = \prod_{i\in I} X_i$. The degree of the monomial X^I is the size of the set I, denoted by $|I|$. Note that if I is a "regular" set (opposed to a multi-set), then the monomial X^I is linear in each of the participating variables $X_i, i \in I$.

A polynomial of degree d is a linear combination of monomials of degree at most d, that is,

$$P(X_1,\dots,X_n) = \sum_{I\subseteq[n],|I|\le d} a_I X^I .$$

The a_I are called the coefficients of the polynomial P. By convention, we set $a_I = 0$ for all other I. If the above summation is only over sets I and not multi-sets, then the polynomial is said to be *multilinear*. Observe that while working over the hypercube, it suffices to consider only multilinear polynomials. We use the following notations throughout.

1. Unless otherwise stated, we work with a PTF f of degree d and a degree-d polynomial $P(X) = \sum_I a_I X^I$ with zero constant term (i. e., $a_\emptyset = 0$) such that

   $$f(X_1,\dots,X_n) = \mathsf{sign}(P(X_1,\dots,X_n) - \theta) .$$

   In case of ambiguity, we will refer to the coefficients a_I as $a_I(P)$.

2. For a polynomial P as above and an underlying distribution over $X = (X_1,\dots,X_n)$, the ℓ_2-norm of the polynomial over X is defined by $\|P\|^2 = \mathbb{E}\left[P(X)^2\right]$. Note that if P is a multilinear polynomial and the distribution is either the multivariate Gaussian $\mathcal{N}^n$ or the uniform distribution over the hypercube, then $\|P\|^2 = \sum_I a_I^2$.

3. For $i \in [n]$, $x^i = (x_1,\dots,x_i) \in \{1,-1\}^i$, $f_{x^i} : \{1,-1\}^{n-i} \to \{1,-1\}$ is defined by

   $$f_{x^i}(X_{i+1},\dots,X_n) = \mathsf{sign}(P(x_1,\dots,x_i,X_{i+1},\dots,X_n) - \theta) .$$

4. For $i \in [n]$, $P_{|i}(X_1,\dots,X_i) = \sum_{I\subseteq[i]} a_I X^I$ is the *restriction* of P to the variables $X_1,\dots,X_i$.

5. For a multi-set S, $x \in_u S$ denotes an uniformly chosen element from S.

6. For clarity, we suppress the exact dependence of the constants on the degree d; a more careful examination of our proofs shows that all constants depending on the degree d are at worst $2^{O(d)}$.

Definition 4.1. A partial assignment $x^i = (x_1, \ldots, x_i)$ is *ε-determining* for f, if there exists $b \in \{1, -1\}$ such that

$$\Pr_{(X_{i+1},\ldots,X_n)\in_u\{1,-1\}^{n-i}}\left[f_{x^i}(X_{i+1}, \ldots, X_n) \neq b \right] \leq \varepsilon\,.$$

We now define *regular polynomials* which play an important role in all our results. Intuitively, a polynomial is regular if no variable has high influence. For a polynomial Q, the *weight* of the i^{th} coordinate is defined by $w_i^2(Q) = \sum_{I \ni i} a_I^2$. For $i \in [n]$, let $\sigma_i(Q)^2 = \sum_{j \geq i} w_j^2(Q)$.

Definition 4.2 (regular polynomials)**.** A multilinear polynomial Q is *ε-regular* if

$$\sum_i w_i^4(P) \leq \varepsilon^2 \left(\sum_i w_i^2(P) \right)^2 = \varepsilon^2 \sigma_1^4(P)\,.$$

A PTF $f(x) = \mathsf{sign}(Q(x) - \theta)$ is ε-regular if Q is ε-regular.

We also assume without loss of generality that the variables are ordered such that

$$w_1(P) \geq w_2(P) \geq \cdots \geq w_n(P)\,.$$

We repeatedly use three powerful tools: $(2,4)$-hypercontractivity (cf. [32, Chapter 9, Bonami Lemma]), the invariance principle of Mossel et al. [30] and the anti-concentration bounds of Carbery and Wright [8]. We state the relevant results below.

Lemma 4.3 ($(2,4)$-hypercontractivity)**.** *If Q, R are degree-d multilinear polynomials, then for $X \in_u \{1,-1\}^n$, $\mathbb{E}_X\left[Q^2 \cdot R^2\right] \leq 9^d \cdot \mathbb{E}_X\left[Q^2\right] \cdot \mathbb{E}_X\left[R^2\right]$. In particular,*

$$\mathbb{E}\left[Q^4\right] \leq 9^d \cdot \mathbb{E}\left[Q^2\right]^2\,.$$

The following anti-concentration bound is a special case of Theorem 8 of [8]. (In their notation, set $q = 2d$ and the log-concave distribution μ to be $\mathcal{N}^n$.)

Theorem 4.4 (Carbery-Wright anti-concentration bound)**.** *There exists an absolute constant C such that for any polynomial Q of degree at most d with $\|Q\| = 1$ and any interval $I \subseteq \mathbb{R}$ of length α,*

$$\Pr_{X \leftarrow \mathcal{N}^n}\left[Q(X) \in I\right] \leq Cd\,\alpha^{1/d}\,.$$

The following result due to Mossel et al. [30] generalizes the classical quantitative central limit theorem for sums of independent variables, the Berry-Esséen Theorem, to low-degree polynomials over independent variables.

Theorem 4.5 (Mossel et al.)**.** *There exists a universal constant C such that the following holds. For any ε-regular multilinear polynomial P of degree at most d with $\|P\| = 1$ and $t \in \mathbb{R}$,*

$$\left| \Pr_{X\in_u\{1,-1\}^n}\left[P(X) < t\right] - \Pr_{Y \leftarrow \mathcal{N}^n}\left[P(Y) < t\right] \right| \leq C^d \varepsilon^{2/(4d+1)}.$$

The result stated in [30] uses $\max_i w_i^2(P)$ as the notion of regularity instead of $\sum_i w_i^4(P)$ as we do. However, their proof extends straightforwardly to the above.

5 Random restrictions of PTFs

We now establish our structural results on random restrictions of low-degree PTFs. The use of *critical indices* ($K(P,\varepsilon)$) in our analysis is motivated by the results of Servedio [34] and Diakonikolas et al. [9] who obtain similar results for LTFs. At a high level, we show the following.

Given any $\varepsilon > 0$, define the ε-critical index of a multilinear polynomial P, $K = K(P,\varepsilon)$, to be the least index i such that $w_j^2(P) \le \varepsilon^2 \sigma_{i+1}^2(P)$ for all $j > i$. We consider two cases depending on how large $K(P,\varepsilon)$ is and roughly, show the following (here $c, \alpha > 0$ are some universal constants).

1. $K \le 1/\varepsilon^{cd}$. In this case we show that for $x^K = (x_1,\dots,x_K) \in_u \{1,-1\}^K$, the PTF f_{x^K} is ε-regular with probability at least α.

2. $K > 1/\varepsilon^{cd}$. In this case we show that with probability at least α, the value of the threshold function is determined by the *top* $L = 1/\varepsilon^{cd}$ variables.

More concretely, we show the following.

Lemma 5.1. *For every integer d, there exist constants $a_d \in \mathbb{R}$, $\gamma_d > 0$ such that for any multilinear polynomial P of degree at most d and $K = K(P,\varepsilon)$ as defined above, the following holds. The polynomial*

$$P_{x^K}(Y_{k+1},\dots,Y_n) \stackrel{def}{=} P(x_1,\dots,x_K,Y_{K+1},\dots,Y_n)$$

in variables $Y_{K+1},\dots,Y_n$ obtained by randomly choosing $x^K = (x_1,\dots,x_K) \in_u \{1,-1\}^K$ is $a_d\varepsilon$-regular with probability at least γ_d.

Lemma 5.2. *For every d, there exist constants $b_d, c_d \in \mathbb{R}$, $\delta_d > 0$, such that for any multilinear polynomial P of degree at most d the following holds. If $K(P,\varepsilon) \ge c_d \log(1/\varepsilon)/\varepsilon^2 = L$, then a random partial assignment $(x_1,\dots,x_L) \in_u \{1,-1\}^L$ is $b_d\varepsilon$-determining for P with probability at least δ_d.*

To prove the above structural properties we need the following simple lemmas.

Lemma 5.3 ([1, Lemma 3.2])**.** *Let A be a real valued random variable satisfying $\mathbb{E}[A] = 0$, $\mathbb{E}\left[A^2\right] = \sigma^2$ and $\mathbb{E}\left[A^4\right] \le b\sigma^4$. Then,*

$$\Pr\left[A \ge \sigma/4\sqrt{b}\right] \ge 1/4^{4/3}b\,.$$

Lemma 5.4. *For $d > 0$ there exist constants $\alpha_d, \beta_d > 0$ such that for any degree at most d polynomial Q, and $X \in_u \{1,-1\}^n$,*

$$\Pr\left[Q(X) \ge \mathbb{E}[Q] + \alpha_d \sigma(Q)\right] \ge \beta_d\,,$$

where $\sigma^2(Q)$ is the variance of $Q(X) = \|Q\|^2 - (\mathbb{E}_X[Q])^2$. In particular, $\Pr[Q(X) \ge \mathbb{E}[Q]] \ge \beta_d$.

Proof. Let random variable $A = Q(X) - \mathbb{E}_X[Q(X)]$. Then, $\mathbb{E}[A] = 0$, $\mathbb{E}\left[A^2\right] = \sigma^2(Q)$ and by $(2,4)$-hypercontractivity, $\mathbb{E}\left[A^4\right] \le 9^d\,\mathbb{E}\left[A^2\right] = 9^d\sigma^4(Q)$. The claim now follows from Lemma 5.3. $\square$

5.1 Proof of Lemma 5.1

Let $X \equiv (X_1, \ldots, X_K)$. We prove the lemma as follows: (1) Bound the expectation of $\sum_{j>K} w_j^4(P_X)$ using hypercontractivity and use Markov's inequality to show that with high probability $\sum_{j>K} w_j^4(P_X)$ is small. (2) Use the fact that $\sigma_{K+1}^2(P_X) = \sum_{j>K} w_j^2(P_X)$ is a polynomial of degree at most $2d$ in X and Lemma 5.4 to lower bound the probability that $\sigma_{K+1}^2(P_X)$ is large. Let

$$\begin{aligned} P_X(Y_{K+1}, \ldots, Y_n) &= P(X_1, \ldots, X_K, Y_{K+1}, \ldots, Y_n) \\ &= R(X_1, \ldots, X_K) + \sum_{J \subseteq [K+1,n], 0<|J| \leq d} Q_J(X_1, \ldots, X_K) \prod_{j \in J} Y_j \,. \end{aligned}$$

We now bound $\mathbb{E}\left[\sum_{j>K} w_j^4(P_X)\right]$. Fix a $j > K$ and observe that $w_j^2(P_X) = \sum_{J \ni j} Q_J^2(X)$. Thus,

$$\mathop{\mathbb{E}}_X \left[w_j^2(P_X)\right] = \sum_{J \ni j} \mathop{\mathbb{E}}_X \left[Q_J^2(X)\right] = \sum_{J \ni j} \|Q_J\|^2 = w_j^2(P)\,. \tag{5.1}$$

Further, by $(2,4)$-hypercontractivity (Lemma 4.3),

$$\mathop{\mathbb{E}}_X \left[w_j^4(P_X)\right] \leq 9^d \left(\mathop{\mathbb{E}}_X \left[w_j^2(P_X)\right]\right)^2 = 9^d\, w_j^4(P)\,.$$

Hence,

$$\mathbb{E}\left[\sum_{j>K} w_j^4(P_X)\right] \leq 9^d \sum_{j>K} w_j^4(P)\,.$$

Now, from the definition of $K(P,\varepsilon)$, $w_j^2(P) \leq \varepsilon^2 \sigma_{K+1}^2(P)$ for all $j > K$. Thus,

$$\sum_{j>K} w_j^4(P) \leq \varepsilon^2 \sigma_{K+1}^2(P) \sum_{j>K} w_j^2(P) = \varepsilon^2 \sigma_{K+1}^4(P)\,.$$

Combining the above inequalities and applying Markov's inequality we get

$$\Pr_X \left[\sum_{j>K} w_j^4(P_X) \geq \gamma 9^d \varepsilon^2 \sigma_{K+1}^4(P)\right] \leq 1/\gamma. \tag{5.2}$$

Observe that $Q(X) = \sum_{j>K} w_j^2(P_X)$ is a polynomial of degree at most $2d$ in $X_1, \ldots, X_k$ and by (5.1),

$$\mathbb{E}[Q] = \sum_{j>K} w_j^2(P) = \sigma_{K+1}^2(P)\,.$$

Thus, by applying Lemma 5.4 to Q,

$$\Pr\left[\sum_{j>K} w_j^2(P_X) \geq \sigma_{K+1}^2(P)\right] \geq \beta_{2d}\,.$$

Setting $\gamma = 2/\beta_{2d}$ in (5.2) and using the above equation, we get

$$\Pr_X \left[\sum_{j>K} w_j^4(P_X) \leq a_d^2 \varepsilon^2 \left(\sum_{j>K} w_j^2(P_X)\right)^2\right] \geq \beta_{2d}/2\,,$$

where $a_d^2 = 2 \cdot 9^d / \beta_{2d}$. Thus, the polynomial $P_X(Y_{K+1}, \ldots, Y_n)$ is $(a_d \varepsilon)$-regular with probability at least $\gamma_d = \beta_{2d}/2$. $\square$

5.2 Proof of Lemma 5.2

We use the following simple lemma.

Lemma 5.5. *For $1 \leq i < j < K(P,\varepsilon)$, $\sigma_j^2(P) \leq (1-\varepsilon^2)^{j-i}\sigma_i^2(P)$.*

Proof. For $1 \leq i < K(P,\varepsilon)$, we have

$$\sigma_i^2(P) = w_i^2(P) + \sigma_{i+1}^2(P) \geq \varepsilon^2\sigma_i^2(P) + \sigma_{i+1}^2(P)\,.$$

Thus, $\sigma_{i+1}^2(P) \leq (1-\varepsilon^2)\sigma_i^2(P)$. The lemma follows. □

Proof of Lemma 5.2. Suppose that $K(P,\varepsilon) \geq L = c_d \log(1/\varepsilon)/\varepsilon^2$ for a constant c to be chosen later and let

$$Q(X_1,\ldots,X_n) = P(X_1,\ldots,X_n) - P_{|L}(X_1,\ldots,X_L)\,.$$

The proof proceeds as follows. We first show that $\|Q\|$ is significantly smaller than $\|P_{|L}\|$. We then use Lemma 5.4 applied to $P_{|L}-\theta$ and Markov's inequality applied to $|Q(X)|$ to show that $|P_{|L}(X_1,\ldots,X_L)-\theta|$ is larger than $|Q(X)|$, so that $Q(X)$ cannot flip the sign of $P_{|L}(X_1,\ldots,X_L)-\theta$, with at least a constant probability. We first bound $\|Q\|$.

Claim 5.6. *For a suitably large constant c_d, $\|Q\| \leq \sqrt{\varepsilon}\,\alpha_d\,\|P_{|L}\|$.*

Proof. Let α_d, β_d be the constants from Lemma 5.4. By definition $\|Q\|^2 = \sum_{I:I\not\subseteq[L]} a_I^2 \leq \sigma_L^2(P)$. Now,

$$\begin{aligned}\sigma_1^2(P) &= \sum_{j<L} w_j^2(P) + \sigma_L^2(P) \leq d \sum_{I:I\cap[L]\neq\emptyset} a_I^2 + \sigma_L^2(P) \leq d \sum_{I:\emptyset\neq I\subseteq[L]} a_I^2 + d \sum_{I:I\not\subseteq[L]} a_I^2 + \sigma_L^2(P) \\ &\leq d \sum_{I:\emptyset\neq I\subseteq[L]} a_I^2 + d\sum_{j>L} w_j^2(P) + \sigma_L^2(P) \leq d \sum_{I:\emptyset\neq I\subseteq[L]} a_I^2 + (d+1)\,\sigma_L^2(P)\,.\end{aligned}$$

Further, by Lemma 5.5, $\sigma_L^2(P) \leq (1-\varepsilon^2)^{L-1}\sigma_1^2(P)$. Combining the above inequalities we get,

$$\sigma_L^2(P) \leq O_d\left((1-\varepsilon^2)^{L-1}\right) \sum_{I:\emptyset\neq I\subseteq[L]} a_I^2 = O_d\left((1-\varepsilon^2)^{L-1}\right)\sigma^2(P)\,. \tag{5.3}$$

Choosing $L = c_d\log(1/\varepsilon)/\varepsilon^2$ for large enough c_d, we get the claim. □

By Claim 5.6 and Markov's inequality,

$$\Pr_{x\in_u\{1,-1\}^n}\left[|Q(x_1,\ldots,x_n)| \geq \alpha_d\,\|P_{|L}\|\right] \leq \Pr_{x\in_u\{1,-1\}^n}\left[|Q(x_1,\ldots,x_n)| \geq \|Q\|/\sqrt{\varepsilon}\right] \leq \varepsilon. \tag{5.4}$$

Let $S \subseteq \{1,-1\}^L$ be the set of all *bad* $x^L \in \{1,-1\}^L$ such that,

$$\Pr_{(X_{L+1},\ldots,X_n)\in_u\{1,-1\}^n}\left[|Q(x_1,\ldots,x_L,X_{L+1},\ldots,X_n)| \geq \alpha_d\,\|P_{|L}\|\right] \geq 2\varepsilon/\beta_d\,.$$

Then, from (5.4) and the above equation, $\Pr_{x^L\in_u\{1,-1\}^L}\left[x^L\in S\right] \leq \beta_d/2$. Now, let $T \subseteq \{1,-1\}^L$ be such that for $x^L \in T$, $|P_{|L}(x_1,\ldots,x_L)-\theta| \geq \alpha_d\,\|P_L\|$ and $x^L \notin S$. Observe that all $x^L \in T$ are $(2\varepsilon/\beta_d)$-determining and by Lemma 5.4 and the above equations,

$$\Pr_{x^L\in_u\{1,-1\}^L}\left[x^L\in T\right] \geq \Pr_{x^L\in_u\{1,-1\}^L}\left[\left|P_{|L}(x_1,\ldots,x_L)-\theta\right| \geq \alpha_d\,\|P_L\|\right] - \Pr_{x^L\in_u\{1,-1\}^L}\left[x^L\in S\right] \geq \beta_d/2\,.$$

The lemma now follows. □

6 Gaussian noise sensitivity of PTFs

In this section we bound the Gaussian noise sensitivity of PTFs and thus prove Theorem 1.4. The proof is simpler than the Boolean case and only makes use of an anti-concentration bound for polynomials in Gaussian space.

Although Theorem 1.4 was stated only for multilinear polynomials and ellipsoids, we give a proof below that works for all degree-d polynomials using ideas from Diakonikolas et al. [11], who were the first to prove a bound on the Gaussian noise sensitivity of general degree-d polynomials (see remarks after the statement of Claim 6.1).

Proof of Theorem 1.4. Let f be the degree-d PTF and P the corresponding degree-d polynomial such that $f(x) = \mathsf{sign}(P(x))$. We may assume without loss of generality that P is normalized, i. e., $\|P\|^2 = \mathbb{E}\left[P^2(X)\right] = 1$.

The proof is based on the Carbery-Wright anti-concentration bound (Theorem 4.4) for degree-d PTFs. Let $X, Y \sim \mathcal{N}^n$ and

$$Z \stackrel{\text{def}}{=} (1-\delta)X + \sqrt{1-(1-\delta)^2}\,Y = (1-\delta)X + \sqrt{2\delta-\delta^2}\,Y\,.$$

Let $\rho = \sqrt{2\delta-\delta^2}$. Define the perturbation polynomial

$$Q(X,Y) = P(Z) - P(X) = P((1-\delta)X + \rho Y) - P(X)\,.$$

Now, for $\gamma > 0$ to be chosen later,

$$\begin{aligned}
\Pr\left[\mathsf{sign}(P(X)) \neq \mathsf{sign}(P(Z))\right] &= \Pr\left[\mathsf{sign}(P(X)) \neq \mathsf{sign}(P(X)+Q(X,Y))\right] \\
&\leq \Pr\left[|P(X)| < |Q(X,Y)|\right] \\
&\leq \Pr\left[|P(X)| < \gamma\right] + \Pr\left[|Q(X,Y) > \gamma\right] \\
&\leq C_d \gamma^{1/d} + \Pr\left[|Q(X,Y)| > \gamma\right],
\end{aligned}$$

where the last inequality follows from the anti-concentration bound (Theorem 4.4). In Claim 6.1, we show that the norm $\|Q\|$ of the perturbation polynomial is at most $c_d\sqrt{\delta}$ for some constant c_d (dependent on d). We can now apply Markov's inequality to bound the second quantity as follows.

$$\Pr\left[|Q(X,Y)| > \gamma\right] \leq \|Q\|^2/\gamma^2 \leq c_d\,\delta/\gamma^2.$$

Thus,

$$\mathbb{GNS}_\delta(f) \leq C_d \gamma^{1/d} + \frac{c_d \delta}{\gamma^2}.$$

The theorem follows by setting $\gamma = \delta^{d/(2d+1)}$ in which case we get $\mathbb{GNS}_\delta(f) = O_d(\delta^{1/(2d+1)})$. $\square$

We note that we can get a slightly stronger bound of

$$O_d\left(\delta^{1/2d}\sqrt{\log(1/\delta)}\right)$$

if we used a stronger tail bound instead of Markov's in the above argument.

Claim 6.1. *There exists a constant c_d such that $\|Q\| \leq c_d\sqrt{\delta}$.*

An earlier version of this paper had an error in the proof of this claim. As pointed out to us by the authors of [11], that proof worked only for multilinear polynomials and ellipsoids. Diakonikolas et al. [11] proved the claim for general degree-d polynomials. For the sake of completeness, we give a simplified presentation of their proof (that works for all degree-d polynomials) in Section 9.

7 Noise sensitivity of PTFs

We now bound the noise sensitivity of PTFs and prove Theorem 1.3. We do so by first bounding the noise sensitivity of regular PTFs and then use the results of the previous section to reduce the general case to the regular case.

7.1 Noise sensitivity of regular PTFs

At a high level, we bound the noise sensitivity of regular PTFs as follows: (1) Reduce the problem to that of proving certain anti-concentration bounds for regular PTFs over the hypercube. (2) Use the invariance principle of Mossel et al. [30] to reduce proving anti-concentration bounds over the hypercube to that of proving anti-concentration bounds over Gaussian distributions. (3) Use the Carbery-Wright anti-concentration bounds [8] for polynomials over log-concave distributions.

For the rest of this section, we fix a degree-d multilinear polynomial P and a corresponding degree-d PTF f. Recall that it suffices to consider multilinear polynomials as we are working over the hypercube. We first reduce bounding noise sensitivity to proving anti-concentration bounds.

Lemma 7.1. *For $0 < \rho < 1$ and $\delta > 0$,*

$$\mathbb{NS}_\rho(f) \leq (d+1)\,\delta + \Pr_{x \in \{1,-1\}^n}\left[\,|P(x)-\theta| \leq 2\sqrt{\rho}/\delta\,\right].$$

Proof. Let S be a random subset $S \subseteq [n]$ where each $i \in [n]$ is in S independently with probability ρ. From the definition of noise sensitivity it easily follows that

$$\begin{aligned}
\mathbb{NS}_\rho(f) &= \Pr_{X \in_u \{1,-1\}^n, S}\left[\mathsf{sign}\,(P(X)-\theta) \neq \mathsf{sign}\left(P(X) - 2\sum_{I:|I\cap S| \text{ is odd}} a_I X^I - \theta\right)\right] \\
&\leq \Pr_{X \in_u \{1,-1\}^n, S}\left[|P(x)-\theta| \leq 2\left|\sum_{I:|I\cap S| \text{ is odd}} a_I X^I\right|\right] \\
&\leq \Pr_{X \in_u \{1,-1\}^n, S}\left[\left|\sum_{I:|I\cap S| \text{ is odd}} a_I X^I\right| \geq \sqrt{\rho}/\delta\right] + \Pr_{X \in_u \{1,-1\}^n}\left[\,|P(X)-\theta\,| \leq 2\sqrt{\rho}/\delta\,\right]. \qquad (7.1)
\end{aligned}$$

Define a non-negative random variable P_S as follows: $P_S^2 = \sum_{I:|I\cap S| \text{ is odd}} a_I^2$. We can then bound the first quantity in the above expression using P_S as follows:

$$\Pr_{X\in_u\{1,-1\}^n,S}\left[\left|\sum_{I:|I\cap S| \text{ is odd}} a_I X^I\right| \geq \sqrt{\rho}/\delta\right] \leq \Pr_{X\in_u\{1,-1\}^n,S}\left[\left|\sum_{I:|I\cap S| \text{ is odd}} a_I X^I\right| \geq P_S/\sqrt{\delta}\right] + \Pr_S\left[P_S \geq \sqrt{\rho}/\sqrt{\delta}\right]. \quad (7.2)$$

Since $\mathbb{E}_X(\sum_{I|I\cap S| \text{ is odd}} a_I X^I)^2 = P_S^2$, by Markov's inequality, we have

$$\Pr_{x\in_u\{1,-1\}^n}\left[\left|\sum_{I:|I\cap S| \text{ is odd}} a_I X^I\right| \geq P_S/\sqrt{\delta}\right] \leq \delta. \quad (7.3)$$

Now, note that $P_S^2 \leq \sum_{i\in S} w_i^2(P)$. Thus,

$$\mathop{\mathbb{E}}_S\left[P_S^2\right] \leq \mathop{\mathbb{E}}_S\left[\sum_{i\in S} w_i^2(P)\right] = \rho\sum_i w_i^2(P) \leq d\,\rho\,.$$

Hence, by Markov's inequality, $\Pr_S[P_S \geq \sqrt{\rho}/\sqrt{\delta}] \leq d\,\delta$. The lemma now follows by combining equations (7.1), (7.2), (7.3) and the above equation. □

We now prove an anti-concentration bound for regular PTFs.

Lemma 7.2. *If P is ε-regular, then for any interval $I \subseteq \mathbb{R}$ of length at most α,*

$$\Pr_{X\in_u\{1,-1\}^n}\left[P(X) \in I\right] = O_d(\alpha^{1/d} + \varepsilon^{2/(4d+1)})\,.$$

Proof. Let $Z_1 = P(X), Z_2 = P(Y)$ for $X \in_u \{1,-1\}^n, Y \leftarrow \mathcal{N}^n$. Then, since P is ε-regular, by Theorem 4.5, for all $t \in \mathbb{R}$,

$$|\Pr[Z_1 > t] - \Pr[Z_2 > t]| = O_d(\varepsilon^{2/(4d+1)})\,.$$

Now, by the above equation and Theorem 4.4 applied to the random variable Y for interval I,

$$\Pr[Z_1 \in I] = \Pr[Z_2 \in I] + O_d(\varepsilon^{2/(4d+1)}) = O_d(\alpha^{1/d} + \varepsilon^{2/(4d+1)})\,. \qquad \square$$

We can now obtain a bound on noise sensitivity of regular PTFs.

Theorem 7.3. *If f is an ε-regular PTF of degree d, then $\mathbb{NS}_\varepsilon(f) \leq O_d\left(\varepsilon^{1/(2d+2)}\right)$.*

Proof. Let $\delta > 0$ to be chosen later. Then, by Lemma 7.1 and Lemma 7.2 above,

$$\mathbb{NS}_\varepsilon(f) = O_d(\delta + \varepsilon^{2/(4d+1)} + \varepsilon^{1/2d}/\delta^{1/d})\,.$$

Choosing $\delta = \varepsilon^{1/(2d+2)}$ we get $\mathbb{NS}_\varepsilon(f) = O_d(\varepsilon^{1/(2d+2)})$. □

7.2 Noise sensitivity of arbitrary PTFs

We prove Theorem 1.3 by recursively applying the following lemma.

Lemma 7.4. *For every d there exist universal constants $c_d, \Delta_d \in \mathbb{N}, \alpha_d \in (0,1)$ such that for $M = \min(K(P,\varepsilon), c_d \log(1/\varepsilon)/\varepsilon^2)$ and $X^M = (X_1, \dots, X_M) \in_u \{1,-1\}^M$,*

$$\Pr_{X^M}\left[\mathbb{NS}_\varepsilon(f_{X^M}) \le \Delta_d \varepsilon^{1/(2d+2)}\right] \ge \alpha_d\,. \tag{7.4}$$

Proof. Let $a_d, b_d, c_d, \gamma_d, \delta_d$ be the constants from Lemmas 5.1 and 5.2. Let $\alpha_d = \min(\gamma_d, \delta_d)$. We consider two cases.

Case (i): $M = K(P,\varepsilon)$. Then, by Lemma 5.1 and Theorem 7.3, for $X^K \in_u \{1,-1\}^K$, with probability at least α_d, $\mathbb{NS}_\varepsilon(f_{x^K}) \le \Delta_d \varepsilon^{1/(2d+2)}$ for some constant Δ_d.

Case (ii): $M = c_d \log(1/\varepsilon)/\varepsilon^2$. Then, by Lemma 5.2, $X^M \in_u \{1,-1\}^M$ is $b_d\varepsilon$-determining with probability at least α_d. Further, if X^M is $b_d\varepsilon$-determining, with f_{X^M} biased towards $b \in \{1,-1\}$, then

$$\mathbb{NS}_\varepsilon(f_{X^M}) \;=\; \Pr_{\substack{Z_1 \in_u \{1,-1\}^{n-M},\\ Z_2 \in_\varepsilon Z_1}}\left[f_{X^M}(Z_1) \ne f_{X^M}(Z_2)\right] \;\le\; 2 \Pr_{Z \in_u \{1,-1\}^{n-M}}\left[f_{X^M}(Z) \ne b\right] \;\le\; 2b_d\varepsilon\,,$$

where "$Z_2 \in_\varepsilon Z_1$" is short-hand to denote that Z_2 is an ε-perturbation of Z_1. The lemma now follows. □

Proof of Theorem 1.3. Let c_d, Δ_d, α_d be as in the above lemma and let

$$L = c_d \log(1/\varepsilon)/\varepsilon^2 \qquad \text{and} \qquad t = \log_{1-\alpha_d}(1/\varepsilon)\,.$$

We will show that for

$$\delta = \varepsilon^{1/(2d+2)}/(Lt) = O_d(\varepsilon^{(4d+5)/(2d+2)}/\log^2(1/\varepsilon))\,,$$

we have

$$\mathbb{NS}_\delta(f) = O_d(\,\varepsilon^{1/(2d+2)}\,)\,.$$

Observe that for this setting, $\delta \le \varepsilon$. For $S \subseteq [n]$ and $x \in \{1,-1\}^n$ let $P_{x,S} : \{1,-1\}^{\bar{S}} \to \mathbb{R}$ be the polynomial of degree at most d defined by $P_{x,S}(X_{\bar{S}}) = P(x_{|S}, X_{\bar{S}})$. Fix $x = (x_1, \dots, x_n) \in \{1,-1\}^n$ and define $S_{x,i} \subseteq [n]$ for $i \ge 1$, recursively as follows. $S_{x,1}$ is the set of $M_1 \le L$ largest weight coordinates in P given by applying Lemma 7.4 to P. For $i \ge 1$, let $S^{x,i} = S_{x,1} \cup S_{x,2} \cup \ldots \cup S_{x,i}$.

For $i > 1$, let $S_{x,i+1}$ be the set of $M_{i+1} \le L$ largest weight coordinates in $P_{x,S^{x,i}}$ given by applying Lemma 7.4 to the polynomial $P_{x,S^{x,i}}$. Define $f_{x,i}$ by $f_{x,i}(\cdot) \equiv \mathsf{sign}(P_{x,S^{x,i}}(\cdot) - \theta)$. Note that the definition of $f_{x,i}$ only depends on x_j for $j \in S^{x,i}$ and that $|S^{x,i}| \le L \cdot i$.

Call $x \in \{1,-1\}^n$ (ε, f)-good if there exists an i, $1 \le i \le t$ such that $\mathbb{NS}_\varepsilon(f_{x,i}) \le \Delta_d\, \varepsilon^{1/(2d+2)}$ and let t_x be such an i for a (ε, f)-good x. Then, from the definition of $f_{x,i}$ and Lemma 7.4,

$$\Pr_{x \in_u \{1,-1\}^n}\left[x \text{ is } (\varepsilon, f)\text{-good}\right] \ge 1 - \varepsilon\,. \tag{7.5}$$

Let $y \in_\delta x$ be a δ-perturbation of $x \in_u \{1,-1\}^n$. Then, since $|S^{x,t_x}| \le Lt$,

$$\Pr_{x,y}\left[x_{|S^{x,t_x}} \neq y_{|S^{x,t_x}}\right] \le Lt\,\delta = \varepsilon^{1/(2d+2)}. \tag{7.6}$$

Also note that for any $i \ge 1$, conditioned on an assignment for the values in $x_{|S^{x,i}}$ and $x_{|S^{x,i}} = y_{|S^{x,i}}$,

$$\Pr_{x,y}[f(x) \neq f(y)] = \mathbb{NS}_\delta(f_{x,i}) \le \mathbb{NS}_\varepsilon(f_{x,i}).$$

Thus, conditioned on x being (ε, f)-good and $x_{|S^{x,t_x}} = y_{|S^{x,t_x}}$,

$$\Pr_{x,y}[f(x) \neq f(y)] \le \mathbb{NS}_\varepsilon(f_{x,t_x}) \le \Delta_d\, \varepsilon^{1/(2d+2)}. \tag{7.7}$$

Combining (7.5), (7.6), (7.7), we get

$$\mathbb{NS}_\delta(f) \le \varepsilon + Lt\,\delta + \Delta_d \varepsilon^{1/(2d+2)} = O_d\left(\varepsilon^{1/(2d+2)}\right).$$

Since $\delta = O_d\left(\varepsilon^{\frac{4d+5}{2d+2}}/\log^2(1/\varepsilon)\right)$ and the above is applicable for all $\varepsilon > 0$, we get that for all $\rho > 0$,

$$\mathbb{NS}_\rho(f) = O_d\left(\log(1/\rho)\rho^{1/(4d+5)}\right) = O_d\left(\rho^{1/(4d+6)}\right). \qquad \square$$

8 Average sensitivity of PTFs

In this section we bound the average sensitivity of PTFs on the Boolean hypercube, proving Theorem 1.6. We first prove a lemma bounding the average sensitivity of a Boolean function in terms of its noise sensitivity. Theorem 1.6 follows immediately from Theorem 1.3 and the following lemma:

Lemma 8.1 (noise sensitivity to average sensitivity)**.** *For any Boolean function $f : \{1,-1\}^n \to \{1,-1\}$,*

$$\mathbb{AS}(f) \le ne\, \mathbb{NS}_{(1/n)}(f).$$

Proof. Let $\delta = 1/n$. Let $X \in_u \{1,-1\}^n$ and let $S \subseteq [n]$ be a random set with each element $i \in [n]$ present in S independently with probability δ. Let $X(S)$ be the vector obtained by flipping the coordinates of X in S. Then, $\mathbb{NS}(f) = \Pr_{X,S}[f(X) \neq f(X(S))]$. Observe that for $i \in [n]$,

$$\Pr[S = \{i\}] = \delta(1-\delta)^{n-1} = (1/n)(1-1/n)^{n-1} > 1/(ne).$$

Therefore,

$$\begin{aligned}
\mathbb{NS}_\delta(f) &= \Pr_{X,S}[f(X) \neq f(X(S))] \\
&= \sum_i \Pr_S[S = \{i\}] \cdot \Pr_X[f(X) \neq f(X(S)) \,|\, S = i] + \Pr_S[|S| \neq 1] \cdot \Pr_{X,S}[f(X) \neq f(X(S)) \,|\, |S| \neq 1] \\
&> \sum_i \frac{1}{ne} \Pr_X[f(X) \neq f(X(\{i\}))] = \frac{1}{ne} \mathbb{AS}(f).
\end{aligned} \qquad \square$$

We now give a bound of $O(n^{1-2^{-d}})$ on the average sensitivity using a different, combinatorial, argument (not using the noise sensitivity bounds).

Theorem 8.2. *For any degree-d PTF $f : \{1,-1\}^n \to \{1,-1\}$, $\mathbb{AS}(f) \le 3n^{1-2^{-d}}$.*

We first show the theorem using Lemma 1.7.

Proof. Let $P(x) = x_i P_i(x_{-i}) + Q_i(x_{-i})$, where $P_i(\,), Q_i(\,)$ are degree-$(d-1)$ and degree-d polynomials, respectively, that do not depend on x_i. Define $f_i(x_{-i}) = \mathsf{sign}(P_i(x_{-i}))$ and $g_i(x) = f(x) f_i(x_{-i})$. Then,

$$\begin{aligned}\mathbb{I}_i(f) &= \Pr_{X \in_u \{1,-1\}^n}\left[f(X) \ne f(X^{(i)})\right] = \Pr_{X \in_u \{1,-1\}^n}\left[f(X) f_i(X_{-i}) \ne f(X^{(i)}) f_i(X_{-i})\right] \\ &= \Pr_{X \in_u \{1,-1\}^n}\left[f(X) f_i(X_{-i}) \ne f(X^{(i)}) f_i((X^{(i)})_{-i})\right] = \Pr_{X \in_u \{1,-1\}^n}\left[g_i(X) \ne g_i(X^{(i)})\right] \\ &= \mathbb{I}_i(g_i)\,.\end{aligned}$$

Observe that g_i is monotone increasing in x_i as the coefficient of x_i in g_i written as a polynomial is $f_i(x_{-i}) \cdot P_i(x_{-i}) \ge 0$. Hence, for $i \in [n]$, $\mathbb{I}_i(g_i) = \mathbb{E}_X\left[X_i g_i(X)\right]$. Thus,

$$\mathbb{AS}(f) = \sum_i \mathbb{I}_i(f) = \sum_i \mathbb{I}_i(g_i) = \sum_i \mathop{\mathbb{E}}_X\left[X_i g_i(X)\right] = \sum_i \mathop{\mathbb{E}}_X\left[X_i f(X) f_i(X_{-i})\right] = \mathop{\mathbb{E}}_X\left[f(X) \sum_i X_i f_i(X_{-i})\right].$$

Since $|f(x)| \le 1$ for all x, we have

$$\mathbb{AS}(f) \le \mathop{\mathbb{E}}_X\left[\left|\sum_i X_i f_i(X_{-i})\right|\right]. \tag{8.1}$$

We now use induction and Lemma 1.7. For an LTF f, f_i as defined above are constants. Therefore, by equation (8.1) and the Cauchy-Schwarz inequality,

$$\mathbb{AS}(f) \le \mathop{\mathbb{E}}_X\left[\left|\sum_i X_i f_i(X_{-i})\right|\right] = \mathop{\mathbb{E}}_X\left[\left|\sum_i X_i\right|\right] \le \sqrt{n}\,.$$

Suppose the theorem is true for degree-d PTFs and let f be a degree-$(d+1)$ PTF and let f_i be as defined before. Then, by equation (8.1) and Lemma 1.7

$$\mathbb{AS}(f)^2 \le 2\sum_i \mathbb{AS}(f_i) + n \le \sum_i 6n^{1-2^{-d}} + n \le 7n^{2-2^{-d}}.$$

Therefore, $\mathbb{AS}(f) \le 3n^{1-2^{-(d+1)}}$. The theorem follows by induction. $\square$

Proof of Lemma 1.7. For brevity, let $f_i(x) = f_i(x_{-i})$. By Cauchy-Schwarz, for any random variable Z we have $\mathbb{E}[|Z|]^2 \leq \mathbb{E}\left[Z^2\right]$. Thus,

$$\begin{aligned}\mathop{\mathbb{E}}_X\left[\left|\sum_i X_i f_i(X_{-i})\right|\right]^2 &\leq \mathop{\mathbb{E}}_X\left[\left(\sum_i X_i f_i(X_{-i})\right)^2\right] \\ &= \mathop{\mathbb{E}}_X\left[\sum_{i,j} X_i X_j f_i(X) f_j(X)\right] \\ &= n + \sum_{i\neq j} \mathop{\mathbb{E}}_X\left[X_i X_j f_i(X) f_j(X)\right]. \end{aligned} \tag{8.2}$$

For $i \neq j \in [n]$, let $x_{-ij} = (x_k : k \in [n], k \neq i, j)$ and let

$$S_i^j = \{x \in \{1,-1\}^n : f_i(x) \neq f_i(x \oplus e_j)\}.$$

Note that $\mathbb{I}_j(f_i) = \Pr_X\left[X \in S_i^j\right]$. Now,

$$\mathop{\mathbb{E}}_X\left[X_i X_j f_i(X) f_j(X)\right] = \sum_{x \in S_i^j \cup S_j^i} \mu(x) x_i x_j f_i(x) f_j(x) + \sum_{x \notin S_i^j \cup S_j^i} \mu(x) x_i x_j f_i(x) f_j(x), \tag{8.3}$$

where $\mu(x) = 1/2^n$ is the probability of choosing x under the uniform distribution. We bound the first term in the above expression by the average sensitivity of the f_i and show that the second term vanishes. Observe that

$$\sum_{x \in S_i^j \cup S_j^i} \mu(x) x_i x_j f_i(x) f_j(x) \leq \mu(S_i^j \cup S_j^i) \leq \mu(S_i^j) + \mu(S_j^i) = \mathbb{I}_j(f_i) + \mathbb{I}_i(f_j). \tag{8.4}$$

Note that for $x \notin S_i^j \cup S_j^i$, $f_i(x), f_j(x)$ are both independent of the values of x_i, x_j. For such x (abusing notation) let $f_i(x_{-ij}) = f_i(x)$, $f_j(x_{-ij}) = f_j(x)$ and let

$$T_{ij} = \{(x_k : k \neq i, j) : x \notin S_i^j \cup S_j^i\}.$$

Then, since for $x \notin S_i^j \cup S_j^i$, $f_i(x), f_j(x)$ depend only on x_{-ij}, we get that $x \notin S_i^j \cup S_j^i$ if and only if $x_{-ij} \notin T_{ij}$. Therefore,

$$\begin{aligned}\sum_{x \notin S_i^j \cup S_j^i} \mu(x) x_i x_j f_i(x) f_j(x) &= \sum_{x \notin S_i^j \cup S_j^i} \mu(x_{-ij})\, \mu(x_i)\, \mu(x_j)\, f_i(x_{-ij})\, f_j(x_{-ij})\, x_i x_j \\ &= \sum_{x_{-ij} \notin T_{ij}} \mu(x_{-ij}) f_i(x_{-ij})\, f_j(x_{-ij}) \mathop{\mathbb{E}}_{x_i, x_j}[x_i x_j] = 0. \end{aligned} \tag{8.5}$$

From equations (8.2), (8.3), (8.4),(8.5) we have,

$$\mathop{\mathbb{E}}_X\left[\left|\sum_i X_i f_i(X_{-i})\right|\right]^2 \leq n + \sum_{i \neq j}(\mathbb{I}_j(f_i) + \mathbb{I}_i(f_j)) = n + 2\sum_i \sum_{j:j\neq i} \mathbb{I}_j(f_i) = n + 2\sum_i \mathbb{AS}(f_i). \qquad \square$$

Remark 8.3. The bound of Lemma 1.7 is tight up to a constant factor if we only have bounds on the average sensitivity of the f_i to go with. For example, consider f_i defined as follows. Divide $[n]$ into $m = \sqrt{n}$ blocks $B_1, \ldots, B_m$ of size m each and for $1 \le j \le m$, $i \in B_j$, let $f_i = \prod_{k \in B_j : k \neq i} x_k$. Then, the left hand side of the lemma is $\Theta(n^{3/2})$ and $\mathbb{AS}(f_i) = m - 1 = \Theta(\sqrt{n})$ for all i.

9 Bounding the perturbation polynomial in the Gaussian setting

In this section we prove the bound on the perturbation polynomial (Claim 6.1). For this, we need some background on Hermite polynomials, which we provide in Section 9.1.

9.1 Background on Hermite polynomials

The univariate Hermite polynomials are defined as follows:

$$H_k(x) = \frac{(-1)^k}{\sqrt{k!}} \mathrm{e}^{x^2/2} \frac{d^k}{dx^k} \mathrm{e}^{-x^2/2}.$$

The univariate Hermite polynomials satisfy $H_k'(x) = \sqrt{k} H_{k-1}(x)$.

The multivariate Hermite polynomials in n variables $(x_1, \ldots, x_n)$ are defined as follows. Let $S \subseteq [n]$ be a multiset. It will be convenient to denote a multiset S by a sequence of n indices as $S = (s_1, \ldots, s_n)$ where each s_i denotes the multiplicity of element $i \in [n]$ in the set S. Note, by this notation, $|S| = \sum s_i$.

$$H_S(x_1, \ldots, x_n) = \prod_{i=1}^{n} H_{s_i}(x_i).$$

The Hermite polynomials are especially useful while working over the (multivariate) normal distribution due to the following orthonormality conditions.

$$\mathop{\mathbb{E}}_{X \leftarrow \mathcal{N}^n} [H_S(X) H_T(X)] = \begin{cases} 1 & \text{if } S = T, \\ 0 & \text{otherwise.} \end{cases}$$

This implies that $\|P\|^2 = \mathbb{E}_{X \leftarrow \mathcal{N}^n}[P^2(X)] = \sum \hat{P}_S^2$.

We will need to work with the Taylor series expansion for the Hermite polynomials. For this, we first observe that the partial derivatives of the multivariate Hermite polynomials can be calculated as follows

$$(\partial H_S)_i(x_1, \ldots, x_n) = \sqrt{s_i} H_{s_i - 1}(x_i) \prod_{j \neq i} H_{s_j}(x_j) = \sqrt{s_i} H_{S \setminus \{i\}}(x_1, \ldots, x_n).$$

Furthermore, the iterative partial derivatives can be calculated as follows. Let $R = (r_1, \ldots, r_n) \subseteq S$ be any multiset.

$$(\partial H_S)_R = \sqrt{\prod_{i=1}^{n} \frac{s_i!}{(s_i - r_i)!}} \cdot H_{S \setminus R}.$$

This in particular gives the follow Taylor series expansion for $H_S(z) = H_S(z_1,\ldots,z_n)$ about the point $x = (x_1,\ldots,x_n)$ for multisets S. Let $|S| = d$. Since H_S depends on at most d variables, we can assume without loss of generality that H_S is defined on the first d variables, i.e., $S \subseteq [d]$ and $H_S(x) = H_S(x_1,\ldots,x_d)$.

$$\begin{aligned} H_S(z) &= H_S(x) + \sum_{k=1}^{d} \sum_{R:|R|=k} \frac{1}{\prod_{i=1}^{d} r_i!} (\partial H_S)_R(x) \cdot \left(\prod_{i=1}^{d} (z_i - x_i)^{r_i} \right) \\ &= H_S(x) + \sum_{k=1}^{d} \sum_{R:|R|=k} \frac{1}{\prod_{i=1}^{d} r_i!} \sqrt{\prod_{i=1}^{d} \frac{s_i!}{(s_i - r_i)!}} \cdot H_{S\setminus R}(x) \cdot \left(\prod_{i=1}^{d} (z_i - x_i)^{r_i} \right). \end{aligned} \tag{9.1}$$

The multivariate Hermite polynomials up to degree d form a basis for the set of all multivariate polynomials of degree d. In particular, given any degree-d polynomial $P(x_1,\ldots,x_n)$, we can write it as a linear combination of Hermite polynomials as follows

$$P(x_1,\ldots,x_n) = \sum_{S\subset[n]:|S|\le d} \hat{P}_S H_S(x_1,\ldots,x_n).$$

The values $\hat{P}_S$ are called the Hermite coefficients of P.

9.2 Proof of Claim 6.1

Recall that we must prove there exists a constant c_d such that $\|Q\| \le c_d\sqrt{\delta}$.

Proof. Given any degree-d multivariate polynomial P, we can write it in the Hermite basis as $P(x) = \sum_{S:|S|\le d} \hat{P}_S H_S(x)$ and use this expansion to bound $\|Q\| = \|P(Z) - P(X)\|$ as follows.

$$\begin{aligned} \|Q\|^2 &= \mathbb{E}\left[(P(Z)-P(X))^2\right] = \mathbb{E}\left[\left(\sum_{S:|S|\le d} \hat{P}_S(H_S(Z)-H_S(X))\right)^2\right] \\ &= \sum_{S,T} \hat{P}_S\hat{P}_T\,\mathbb{E}\left[(H_S(Z)-H_S(X))\cdot(H_T(Z)-H_T(X))\right] \\ &= \sum_{S} \hat{P}_S^2\,\mathbb{E}\left[(H_S(Z)-H_S(X))^2\right] + \sum_{S\ne T} \hat{P}_S\hat{P}_T\,\mathbb{E}\left[(H_S(Z)-H_S(X))\cdot(H_T(Z)-H_T(X))\right] \\ &= \sum_{S} \hat{P}_S^2\,\mathbb{E}\left[(H_S(Z)-H_S(X))^2\right] - \sum_{S\ne T} \hat{P}_S\hat{P}_T\left(\mathbb{E}\left[H_S(Z)H_T(X)\right] + \mathbb{E}\left[H_S(X)H_T(Z)\right]\right), \end{aligned} \tag{9.2}$$

where the last step follows from the orthonormality of the Hermite polynomials.

We will now show that

$$\mathbb{E}[H_S(X)H_T(Z)] = \mathbb{E}[H_S(X_1\ldots X_n)H_T(Z_1\ldots Z_n)] = 0$$

for $S \ne T$. Since $S \ne T$ and $(X_1,\ldots,X_n),(Z_1,\ldots,Z_n)$ are product distributions, it suffices to show the following univariate case: $\mathbb{E}[H_s(X_1)H_t(Z_1)] = 0$ for $s \ne t$. We now observe that the joint distribution (X_1,Z_1) is identical to the distribution (Z_1,X_1). Hence, to calculate $\mathbb{E}[H_s(X_1)H_t(Z_1)]$ for $s \ne t$ we can assume without loss of generality that $s > t$. Now, $H_t(Z_1) = H_t((1-\delta)X_1 + \rho Y_1)$ is a bivariate degree-t polynomial and can be expanded in the Hermite basis as $\sum_{i,j=0}^{t} \alpha_{ij} H_i(X_1) H_j(Y_1)$. We thus have

$$\mathop{\mathbb{E}}_{X_1,Y_1}[H_s(X_1)H_t(Z_1)] = \sum_{i,j=0}^{t} \alpha_{ij} \mathop{\mathbb{E}}_{X_1}[H_s(X_1)H_i(X_1)] \cdot \mathop{\mathbb{E}}_{Y_1}[H_j(Y_1)] = 0$$

since $s > t \geq i$.

Plugging this into the expression for $\|Q\|$ in (9.2), we have

$$\|Q\|^2 = \sum_S \hat{P}_S^2 \, \mathbb{E}\left[(H_S(Z) - H_S(X))^2\right] .$$

Since $\|P\|^2 = \sum_S \hat{P}_S^2 = 1$, to prove the claim it suffices if we show that there exists a constant c_d such that for any multiset S, $\|H_S(Z) - H_S(X)\|^2 \leq c_d^2 \delta$. We bound the norm $\|H_S(Z) - H_S(X)\|$ using the Taylor series expansion of $H_S(Z)$ as stated in equation (9.1). Let $|S| = d$; then we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{X,Y}[|H_S(Z) - H_S(X)|] &\leq \sum_{k=1}^{d} \sum_{R:|R|=k} \frac{1}{\prod_{i=1}^{d} r_i!} \sqrt{\prod_{i=1}^{d} \frac{s_i!}{(s_i - r_i)!}} \cdot \mathbb{E}\left[\left| H_{S\setminus R}(X) \cdot \left(\prod_{i=1}^{d} (Z_i - X_i)^{r_i} \right) \right|\right] \\
&\leq \sum_{k=1}^{d} \sum_{R:|R|=k} \frac{1}{\prod_{i=1}^{d} r_i!} d^{k/2} \, \mathbb{E}\left[\left| H_{S\setminus R}(X) \cdot \left(\prod_{i=1}^{d} (Z_i - X_i)^{r_i} \right) \right|\right] \\
&\qquad [\text{ since each } s_i \leq d \text{ and } \textstyle\sum r_i = |R| = k\ ] \\
&\leq \sum_{k=1}^{d} d^{k/2} \sum_{R:|R|=k} \frac{1}{\prod_{i=1}^{d} r_i!} \sqrt{\mathbb{E}\left[H_{S\setminus R}^2(X)\right] \cdot \mathbb{E}\left[\prod_{i=1}^{d} (Z_i - X_i)^{2r_i}\right]} \\
&\qquad [\text{ by the Cauchy-Schwarz inequality }] \\
&= \sum_{k=1}^{d} d^{k/2} \sum_{R:|R|=k} \frac{1}{\prod_{i=1}^{d} r_i!} \sqrt{\prod_{i=1}^{d} \mathbb{E}\left[(Z_i - X_i)^{2r_i}\right]} \\
&\qquad [\text{ by orthonormality of } H_{S\setminus R} \text{ and independence of } (Z_i - X_i) \text{ over the values } i\ ] \\
&= \sum_{k=1}^{d} d^{k/2} \sum_{R:|R|=k} \frac{1}{\prod_{i=1}^{d} r_i!} \sqrt{\prod_{i=1}^{d} \delta^{r_i} \frac{(2r_i)!}{r_i!}} \\
&\qquad [\text{ since } Z_i - X_i \sim \mathcal{N}(0, \sqrt{2\delta}) \text{ whose } 2r\text{-th moment is } \delta^r \frac{(2r)!}{r!}\ ] \\
&= \sum_{k=1}^{d} (d\delta)^{k/2} \sum_{R:|R|=k} \sqrt{\prod_{i=1}^{d} \binom{2r_i}{r_i}} \leq \sum_{k=1}^{d} (d\delta)^{k/2} \cdot d^k \cdot 2^{k/2} = \sum_{k=1}^{d} (d^{3/2}\sqrt{2\delta})^k \\
&\leq 2d^{3/2}\sqrt{2\delta} \qquad\qquad [\text{ if } d^{3/2}\sqrt{2\delta} \leq 1/2\ ].
\end{aligned}$$

To complete the proof we need the following $(1,2)$-hypercontractivity for degree-d polynomials under the normal distribution.

Lemma 9.1 ((1,2)-hypercontractivity [16, Remark 5.13])**.** *If P is a multivariate polynomial of degree d, then*

$$\|P\| = \sqrt{\mathop{\mathbb{E}}_{X \leftarrow \mathcal{N}^n}[P(X)^2]} \le \mathrm{e}^d \cdot \mathop{\mathbb{E}}_{X \leftarrow \mathcal{N}^n}[|P(X)|] .$$

Thus, if $d^{3/2}\sqrt{2\delta} \le 1/2$, then $\mathbb{E}[|H_S(Z) - H_S(X)|] \le 2d^{3/2}\sqrt{2\delta}$. We can now use the above (1,2)-hypercontractivity and bound $\|H_S(Z) - H_S(X)\|$ as follows.

$$\|H_S(Z) - H_S(X)\|^2 \le \mathrm{e}^d\,\mathbb{E}\,[|H_S(Z) - H_S(X)|] \le 2d^{3/2}\mathrm{e}^d\sqrt{2\delta} .$$

If $d^{3/2}\sqrt{2\delta} > 1/2$, we have

$$\mathbb{E}[|H_S(Z) - H_S(X)|^2] \le 2\,\mathbb{E}[H_S^2(Z) + H_S^2(X)] \le 4 < 8d^{3/2}\sqrt{2\delta} .$$

Thus, either way, we have that there exists a constant c_d such that $\|H_S(Z) - H_S(X)\| \le c_d\sqrt{\delta}$. □

Acknowledgments

We thank Ilias Diakonikolas for pointing out an error in an earlier version of this writeup.

References

[1] NOGA ALON, GREGORY GUTIN, AND MICHAEL KRIVELEVICH: Algorithms with large domination ratio. *J. Algorithms*, 50(1):118–131, 2004. [doi:10.1016/j.jalgor.2003.09.003] 9

[2] 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] 2

[3] ROBERT BEALS, HARRY BUHRMAN, RICHARD CLEVE, MICHELE MOSCA, AND RONALD DE WOLF: Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001. Preliminary version in FOCS'98. [doi:10.1145/502090.502097] 2

[4] RICHARD BEIGEL: The polynomial method in circuit complexity. In *Proc. 8th IEEE Conf. on Structure in Complexity Theory (SCT'93)*, pp. 82–95. IEEE Comp. Soc. Press, 1993. [doi:10.1109/SCT.1993.336538] 2

[5] MICHAEL BEN-OR AND NATHAN LINIAL: Collective coin flipping, robust voting schemes and minima of Banzhaf values. In *Proc. 26th FOCS*, pp. 408–416. IEEE Comp. Soc. Press, 1985. [doi:10.1109/SFCS.1985.15] 2

[6] ITAI BENJAMINI, GIL KALAI, AND ODED SCHRAMM: Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90(1):5–43, 1999. See also at arXiv. [doi:10.1007/BF02698830] 2

[7] ERIC BLAIS, RYAN O'DONNELL, AND KARL WIMMER: Polynomial regression under arbitrary product distributions. *Machine Learning*, 80(2-3):273–294, 2010. Preliminary version in COLT'08. [doi:10.1007/s10994-010-5179-6] 6

[8] 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] 5, 8, 13

[9] 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] 4, 9

[10] 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] 1

[11] 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] 1, 3, 5, 12, 13

[12] ILIAS DIAKONIKOLAS, ROCCO A. SERVEDIO, LI-YANG TAN, AND ANDREW WAN: A regularity lemma and low-weight approximators for low-degree polynomial threshold functions. *Theory of Computing*, 10(2):27–53, 2014. Preliminary version in CCC'10. See also at arXiv. [doi:10.4086/toc.2014.v010a002] 5

[13] CRAIG GOTSMAN AND NATHAN LINIAL: Spectral properties of threshold functions. *Combinatorica*, 14(1):35–50, 1994. [doi:10.1007/BF01305949] 2, 4

[14] JOHAN HÅSTAD: Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. Preliminary version in STOC'97. [doi:10.1145/502090.502098] 2

[15] DAVID HAUSSLER: Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Inform. and Comput.*, 100(1):78–150, 1992. Preliminary version in ALT'90. [doi:10.1016/0890-5401(92)90010-D] 6

[16] SVANTE JANSON: *Gaussian Hilbert Spaces*. Cambridge Univ. Press, 1997. [doi:10.1017/CBO9780511526169] 22

[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] 2

[18] ADAM TAUMAN KALAI, ADAM R. KLIVANS, YISHAY MANSOUR, AND ROCCO A. SERVEDIO: Agnostically learning halfspaces. *SIAM J. Comput.*, 37(6):1777–1805, 2008. Preliminary version in FOCS'05. [doi:10.1137/060649057] 2, 6

[19] GIL KALAI: Noise sensitivity and chaos in social choice theory. In *Fete of Combinatorics and Computer Science*, pp. 173–212. Springer, 2010. [doi:10.1007/978-3-642-13580-4_8] 2

[20] DANIEL M. KANE: The Gaussian surface area and noise sensitivity of degree-d polynomial threshold functions. *Comput. Complexity*, 20(2):389–412, 2011. See also at arXiv. [doi:10.1007/s00037-011-0012-6] 5

[21] 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] 5

[22] MICHAEL J. KEARNS, ROBERT E. SCHAPIRE, AND LINDA SELLIE: Toward efficient agnostic learning. *Machine Learning*, 17(2-3):115–141, 1994. Preliminary version in COLT'92. [doi:10.1007/BF00993468] 6

[23] SUBHASH KHOT, GUY KINDLER, ELCHANAN MOSSEL, AND RYAN O'DONNELL: Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007. Preliminary version in FOCS'04. See also at ECCC. [doi:10.1137/S0097539705447372] 2

[24] ADAM R. KLIVANS, RYAN O'DONNELL, AND ROCCO A. SERVEDIO: Learning intersections and thresholds of halfspaces. *J. Comput. System Sci.*, 68(4):808–840, 2004. Preliminary version in FOCS'02. [doi:10.1016/j.jcss.2003.11.002] 2, 6

[25] ADAM R. KLIVANS, RYAN O'DONNELL, AND ROCCO A. SERVEDIO: Learning geometric concepts via Gaussian surface area. In *Proc. 49th FOCS*, pp. 541–550. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.64] 2, 6

[26] ADAM R. KLIVANS AND ROCCO A. SERVEDIO: Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. System Sci.*, 68(2):303–318, 2004. Preliminary version in STOC'01. [doi:10.1016/j.jcss.2003.07.007] 2

[27] 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] 2

[28] 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. 5

[29] 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] 4

[30] 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] 5, 8, 13

[31] RYAN O'DONNELL: Hardness amplification within NP. *J. Comput. System Sci.*, 69(1):68–94, 2004. Preliminary version in STOC'02. [doi:10.1016/j.jcss.2004.01.001] 2

[32] RYAN O'DONNELL: *Analysis of Boolean Functions*. Cambridge Univ. Press, 2014. See online version here. 8

[33] YUVAL PERES: Noise stability of weighted majority. Technical report, 2004. [arXiv:math/0412377] 5

[34] 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] 4, 9

[35] ALEXANDER A. SHERSTOV: Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in STOC'07. [doi:10.1137/08071421X] 2

[36] ALEXANDER A. SHERSTOV: The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. Preliminary version in STOC'08. See also at arXiv. [doi:10.1137/080733644] 2

[37] YAOYUN SHI: Lower bounds of quantum black-box complexity and degree of approximating polynomials by influence of Boolean variables. *Information Processing Letters*, 75(1-2):79–83, 2000. See also at arXiv. [doi:10.1016/S0020-0190(00)00069-7] 2

AUTHORS

Prahladh Harsha
Tata Institute of Fundamental Research
Mumbai, India
prahladh@tifr.res.in
`http://www.tcs.tifr.res.in/~prahladh/`

Adam Klivan
The University of Texas at Austin
klivans@cs.utexas.edu
`http://www.cs.utexas.edu/~klivans/`

Raghu Meka
Microsoft Research, Silicon Valley
raghu@ias.edu
`http://www.math.ias.edu/~raghu`

ABOUT THE AUTHORS

PRAHLADH HARSHA is a reader at the Tata Institute of Fundamental Research, Mumbai, India. He obtained his Ph. D. from MIT under the supervision of Madhu Sudan. Prahladh's research interests include complexity theory, probabilistically checkable proofs, and communication complexity.

ADAM KLIVANS is an Associate Professor of Computer Science at the University of Texas at Austin. His interests are in learning theory and its relationship with computational complexity. He is frequently confused with Adam Kalai.

RAGHU MEKA is a researcher at Microsoft Research, Silicon Valley, Mountain View. He received his Ph. D. from the University of Texas at Austin in 2011 under the supervision of David Zuckerman. His main interests are in complexity theory, pseudo-randomness, algorithm design, learning theory and more generally in anything to do with probability.