Extremal properties of polynomial threshold functions

Ryan O'Donnell*
Department of Mathematics
MIT
Cambridge, MA 02139
odonnell@theory.lcs.mit.edu

Rocco A. Servedio†
Department of Computer Science
Columbia University
New York, NY 10025
rocco@cs.columbia.edu

Abstract

In this paper we give new extremal bounds on polynomial threshold function (PTF) representations of Boolean functions. Our results include the following:

- *Almost every Boolean function has PTF degree at most* $\frac{n}{2} + O(\sqrt{n \log n})$. *Together with results of Anthony and Alon, this establishes a conjecture of Wang and Williams [26] and Aspnes, Beigel, Furst, and Rudich [3] up to lower order terms.*
- *Every Boolean function has PTF density at most* $(1 - \frac{1}{O(n)})2^n$. *This improves a result of Gotsman [12].*
- *Every Boolean function has weak PTF density at most* $o(1)2^n$. *This gives a negative answer to a question posed by Saks [23].*
- *PTF degree* $\lfloor \log_2 m \rfloor + 1$ *is necessary and sufficient for Boolean functions with sparsity* m. *This answers a question of Beigel [5].*

1. Introduction

A broad research goal in computational complexity is to understand the properties of various representation schemes for Boolean functions. Many representation schemes have been studied, such as DNF and CNF formulas, decision trees, branching programs, the Fourier representation (i.e. polynomials over the reals), polynomials over GF_2, monotone span programs, and so on.

In this paper we consider Boolean functions represented as *polynomial threshold functions*. Given a Boolean function $f : \{+1, -1\}^n \to \{+1, -1\}$, a polynomial threshold function (PTF) for f is a n-variable real polynomial p such that $\text{sgn}(p(x)) = f(x)$ for all $x \in \{+1, -1\}^n$. (Alternatively, we sometimes say that such a polynomial p *sign-represents* f.)

Polynomial threshold functions play an important role in theoretical computer science. They are very useful in structural complexity theory; the Beigel *et al.* [6] proof that PP is closed under intersection uses clever constructions of polynomial threshold functions, and many oracle results have been obtained using PTFs, e.g. [3, 4, 11, 25]. Polynomial threshold functions can be viewed as threshold-of-parity circuits and as such have been studied by researchers in circuit complexity [8, 9] and learning theory [16]. More recently, upper bounds on polynomial threshold function degree have been used to obtain learning algorithms for various classes of Boolean circuits [18, 17, 21]. Finally, polynomial threshold functions are an inherently interesting intermediate model of computation between purely algebraic models such as Fourier or GF_2 polynomials and purely combinatorial models such as decision trees or logic circuits. See Saks [23] for an extensive survey on polynomial threshold functions.

The two most basic complexity measures for a polynomial threshold function are its degree and its density (number of nonzero monomials). The *threshold degree* of a Boolean function f is the minimum degree over all polynomials p which sign-represent f, and the *threshold density* of f is the minimum density over all polynomials p which sign-represent f. Note that without loss of generality we may take any sign-representing polynomial to be multilinear, and hence every Boolean function has threshold degree at most n and threshold density at most 2^n.

Aspnes *et al.* [3] introduced a useful variant on polynomial threshold representations, namely, weak polynomial threshold representations. Given a Boolean function $f : \{+1, -1\}^n \to \{+1, -1\}$ we say the n-variable polynomial p is a *weak polynomial threshold representation* of f (alternatively, p *weakly sign-represents* f) if $p(x)$ is not identically 0 on $\{+1, -1\}^n$ and $\text{sgn}(p(x)) = f(x)$ for all $x \in \{+1, -1\}^n$ such that $p(x) \neq 0$. The "Theorem of

*Supported by NSF grant 99-12342.

†Supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship and by NSF grant CCR-98-77049. Work was performed while at the Division of Engineering and Applied Sciences, Harvard University.

	Strong degree		Weak degree	
	lower bound	upper bound	lower bound	upper bound
every function	n	n	n	n
almost every function	$\frac{n}{2}$	$\boldsymbol{\frac{n}{2} + O(\sqrt{n \log n})}$	$\boldsymbol{\frac{n}{2} - O(\sqrt{n \log n})}$	$\frac{n}{2}$

Table 1. Best bounds to date on strong and weak threshold degrees of n-variable Boolean functions. Lower bounds for "every function" mean that some function has this as a lower bound. Boldface entries are new bounds proved in this paper.

	Strong density		Weak density	
	lower bound	upper bound	lower bound	upper bound
every function	$.11\ 2^n$	$\boldsymbol{(1 - \frac{1}{O(n)})2^n}$	$\boldsymbol{2^{n/2}}$	$\boldsymbol{o(1)2^n}$
almost every function	$.11\ 2^n$	$\boldsymbol{(1 - \frac{1}{O(n)})2^n}$ $(\dagger)$	$\boldsymbol{\frac{1}{2\sqrt{n}}2^{n/2}}$	$\boldsymbol{\frac{2}{n}2^n}$

Table 2. Best bounds to date on strong and weak threshold densities of n-variable Boolean functions. Lower bounds for "every function" mean that some function has this as a lower bound. Boldface entries are new bounds proved in this paper. For $(\dagger)$, we in fact show that *every* set of $(1 - \frac{1}{O(n)})2^n$ monomials can serve as a PTF support for almost every Boolean function.

the Alternative" [3] shows that weak polynomial threshold representations are intimately connected to the usual threshold representations (see Theorem 5), and thus the study of *weak threshold degree* and *weak threshold density*, defined in analogy with threshold degree and threshold density, is of interest.

1.1. Previous Work

Prior to our work many authors have studied extremal properties of polynomial threshold functions. Here we touch briefly on the most relevant previous results (see Saks [23] for a detailed treatment).

In a famous result Minsky and Papert [20] proved upper and lower bounds of n for the threshold degree of the n-variable parity function. Aspnes *et al.* [3] proved upper and lower bounds of n for the weak threshold degree of parity as well. Both Aspnes *et al.* and Wang and Williams [26] conjectured that almost every n-variable Boolean function has threshold degree exactly $n/2$. Toward this conjecture, Anthony [2] and Alon [1] used a counting argument to show that almost every Boolean function has threshold degree at least $n/2$. For the upper bound Razborov and Rudich [22] used a counting argument to show that almost every Boolean function has threshold degree at most $\frac{19}{20}n$, and Alon [1] used results of Gotsman [12] to show that almost every Boolean function has threshold degree at most $.89n$.

For threshold density even less was known. Saks [23] noted that results of Cover [10] imply that almost every Boolean function has threshold density at least $(.11)2^n$. Gotsman [12] proved that every Boolean function has threshold density at most $2^n - 2^{n/2}$. Aspnes *et al.* proved that every Boolean function has weak threshold density at most $\frac{1}{2}2^n$. Saks [23] has asked whether almost every Boolean function (i) has threshold density at most $(1-\epsilon)2^n$ for some $\epsilon > 0$, (ii) has weak threshold density at most $(\frac{1}{2} - \epsilon)2^n$ for some $\epsilon > 0$.

1.2. Our Results

We give many new extremal results on the degree and density of polynomial threshold functions. These results, which are summarized in Tables 1 and 2, improve on previous bounds and answer several of the questions described above. In addition to the results shown in Tables 1 and 2, we also prove a tight bound on the threshold degree of sparse Boolean functions, answering a question posed by Beigel [5].

1.3. Organization of the paper

In Section 2 we give some necessary background on strong and weak threshold representations, tail bounds, and

Fourier analysis. Section 3 gives our new upper bound on threshold density for all Boolean functions. Our results on threshold degree and threshold density for almost all Boolean functions are in Section 4. In Section 5 we give new upper and lower bounds on weak threshold density for all and almost all Boolean functions. Finally, we prove a tight bound on the threshold degree of sparse Boolean functions in Section 6. We close in Section 7 with suggestions for future work and a conjecture.

2. Preliminaries

Definition 1 *A real polynomial* $p(x_1, \ldots, x_n)$ strongly (sign-)represents *a Boolean function* $f : \{+1,-1\}^n \rightarrow \{+1,-1\}$ *if* $p(x) \neq 0$ *for all* $x \in \{+1,-1\}^n$ *and sgn*$(p(x)) = f(x)$ *for all* $x \in \{+1,-1\}^n$.

Definition 2 *A real polynomial* $p(x_1, \ldots, x_n)$ weakly (sign-)represents *a Boolean function* $f : \{+1,-1\}^n \rightarrow \{+1,-1\}$ *if* $p(x) \neq 0$ *for at least one* $x \in \{+1,-1\}^n$ *and sgn*$(p(x)) = f(x)$ *for all* $x \in \{+1,-1\}^n$ *such that* $p(x) \neq 0$.

Since $x^2 = 1$ for $x \in \{+1,-1\}$, without loss of generality any sign-representing polynomial p can be taken to be multilinear. Hence any boolean function f on n bits has threshold degree at most n and threshold density at most 2^n. We write $\mathcal{M}$ to denote the set of all 2^n multilinear monomials over $x_1, \ldots, x_n$.

Definition 3 *Given a Boolean function f, we say the* strong *(respectively,* weak*)* degree *of f is the minimum degree over all polynomials which strongly (respectively, weakly) sign-represent f. Similarly, we say the* strong *(respectively,* weak*)* density *of f is the minimum density (number of nonzero coefficients) over all polynomials which strongly (respectively, weakly) sign-represent f.*

Definition 4 *The* support *of a polynomial threshold function sgn*$(p(x))$ *is the set of monomials which have nonzero coefficients in p.*

We will use the so-called "Theorem of the Alternative" of Aspnes *et al.* [3] which relates weak and strong representations. This theorem follows immediately from the theorems of the alternative used for proving linear programming duality (e.g., Farkas's Lemma, the Stiemke Transposition Theorem). See [3, 21, 23] for more details.

Theorem 5 *Let $\mathcal{S}$ be any set of monomials over $x_1, \ldots, x_n$ and let f be any Boolean function. Then exactly one of the following statements is true:*

1. *f has a strong representation with support in $\mathcal{S}$;*
2. *f has a weak representation with suport in $\mathcal{M} - \mathcal{S}$.*

Finally, some standard tail bounds will be useful:

Chernoff Bound: *Let $X_1, \ldots, X_n$ be independent 0-1 random variables with* $\Pr[X_i = 1] = p \leq \frac{1}{2}$. *Let* $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. *Then for all* $0 < \delta < p(1-p)$, *we have*

$$\Pr[|\overline{X} - p| \geq \delta] \leq 2\exp\left(-\frac{\delta^2 n}{2p(1-p)}\right).$$

Hoeffding Bound [15]: *Let $F_1, \ldots, F_k$ be independent random variables with common mean μ and bounded deviance from the mean, $|F_i - \mu| \leq M$. Let* $\sigma^2 = \frac{1}{k}\sum_{i=1}^k$ *Var*$[F_i]$, *and let* $\overline{F} = \frac{1}{k}\sum_{i=1}^k F_i$. *Then for each $0 < t < M$, we have:*

$$\Pr[|\overline{F} - \mu| \geq t] \leq$$
$$2\exp\left((-kt/M)[(1+\sigma^2/Mt)\ln(1+Mt/\sigma^2) - 1]\right).$$

This inequality also holds in the scenario where $F_1, \ldots, F_k$ are chosen without replacement from a fixed population $\{c_1, \ldots, c_N\}$, and μ and σ^2 denote $\frac{1}{N}\sum_{i=1}^N c_i$ and $\frac{1}{N}\sum_{i=1}^N (c_i - \mu)^2$, respectively.

2.1. Fourier background

We view Boolean functions as maps $\{+1,-1\}^n \rightarrow \{+1,-1\}$. We consider the vector space V of all real-valued functions on $\{+1,-1\}^n$ endowed with inner product $\langle \cdot, \cdot \rangle$ defined by

$$\langle f, g \rangle = \mathbf{E}[f(x)g(x)],$$

where the expectation is over a uniform choice of $x \in \{+1,-1\}^n$. For $S \subseteq [n]$ we write x_S to denote $\prod_{i \in S} x_i$. As is well known, the collection of functions $\{x_S\}_{S \subseteq [n]}$ forms an orthonormal basis for V. We denote $\langle f(x), x_S \rangle$ by $\hat{f}(S)$ and hence for any function f,

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) x_S.$$

This is known as the *Fourier representation* of f. Thus the Fourier coefficient $\hat{f}(S)$ is precisely the coefficient of x_S in the (unique) multilinear polynomial for f.

We denote by $|||f|||_p$ the quantity $\left(\sum_{S \subseteq [n]} |\hat{f}(S)|^p\right)^{1/p}$. We also write $|||f|||_\infty$ for $\max_S |\hat{f}(S)|$. An easy consequence of orthonormality of $\{x_S\}$ is Parseval's identity: for any $f : \{+1,-1\}^n \rightarrow \mathbb{R}$,

$$|||f|||_2^2 = 2^{-n} \sum_{x \in \{+1,-1\}^n} f(x)^2.$$

In particular, all Boolean functions $f : \{+1,-1\}^n \to \{+1,-1\}$ have $|||f|||_2 = 1$.

For $\mathcal{S} \subseteq 2^{[n]}$ define $f_{\mathcal{S}}(x)$ by:

$$f_{\mathcal{S}}(x) = \sum_{S \in \mathcal{S}} \hat{f}(S) x_S;$$

so $f_{\mathcal{S}}$ is obtained by zeroing the Fourier coefficients of all monomials x_T such that $T \notin \mathcal{S}$. Similarly, for $0 \le d \le n$ define $f_d(x)$ by:

$$f_d(x) = \sum_{|S| \le d} \hat{f}(S) x_S.$$

Note that $f_{\mathcal{S}}(x)$ has threshold density at most $|\mathcal{S}|$ and $f_d(x)$ has threshold degree at most d.

Finally, we will often use the following simple fact:

Fact 6 *Suppose that $\mathcal{S} \subseteq [n]$ is such that $\sum_{S \notin \mathcal{S}} |\hat{f}(S)| < 1$. Then sgn$(f_{\mathcal{S}})$ is a polynomial threshold function for f.*

3. A new upper bound for threshold density

We first study the maximum threshold density of any Boolean function. As noted earlier, for any $f : \{+1,-1\}^n \to \{+1,-1\}$ the threshold density of f is clearly at most 2^n. Gotsman [12] obtained a slightly better bound of $2^n - 2^{n/2} + 1$. The proof is straightforward: Let $\mathcal{T}$ denote the set of $2^{n/2} - 1$ monomials on which f has Fourier coefficients of smallest magnitude. Since $|||f|||_2 = 1$, we have $|||f|||_1 \le 2^{n/2}$, and hence the sum of the magnitudes of the smallest $2^{n/2}$ Fourier coefficients is at most 1. Thus $\sum_{S \in \mathcal{T}} |\hat{f}(S)| < 1$, so sgn$(f_{\mathcal{M}-\mathcal{T}})$ sign-represents f by Fact 6.

In this section we improve this upper bound to $(1 - \frac{1}{O(n)})2^n$:

Theorem 7 *Let $f : \{+1,-1\}^n \to \{+1,-1\}$ be any Boolean function. Then f has threshold density at most $(1 - \frac{1}{O(n)})2^n$.*

Proof: Let $L = |||f|||_1$. Bruck and Smolensky [9] gave a randomized construction showing that f has threshold density at most $\lceil 2nL^2 \rceil$. Hence if $L \le \frac{1}{2\sqrt{n}} 2^{n/2}$ then f has threshold density at most $\frac{1}{2}2^n$; consequently we assume $L > \frac{1}{2\sqrt{n}} 2^{n/2}$.

Let $\mathcal{T}$ be the set of monomials on which f has its Fourier coefficients of smallest magnitude, where the cutoff is selected so that:

$$\sum_{S \notin \mathcal{T}} |\hat{f}(S)| \in [n-2, n-1). \tag{1}$$

Since $\sum_{S \subseteq [n]} |\hat{f}(S)| = L$ we conclude that $|\mathcal{M} - \mathcal{T}|/2^n \le (n-1)/L < 2(n-1)\sqrt{n} 2^{-n/2}$, so

$$N := |\mathcal{T}| > \frac{1}{2} 2^n. \tag{2}$$

We now select without replacement a random subset $\mathcal{K} \subseteq \mathcal{T}$ of size $k = \frac{1}{Cn} 2^n$ (here $C > 0$ is an absolute constant to be determined later), and then form the polynomial threshold function

$$f_{\mathcal{M}-\mathcal{K}}(x) = \sum_{S \notin \mathcal{K}} \hat{f}(S) x_S.$$

We will show that for each fixed $x \in \{+1,-1\}^n$, the probability that this polynomial threshold function errs on x is at most 3^{-n}. By taking a union bound over all x's it follows that f is correctly sign-represented by some polynomial of density $(1 - \frac{1}{O(n)})2^n$.

Fix $x \in \{+1,-1\}^n$. We will prove that $|\sum_{S \in \mathcal{K}} \hat{f}(S) x_S| \ge 1$ with probability at most 3^{-n} which suffices to prove the claim. Let $(c_1, \ldots, c_N)$ denote the list of numbers $(\hat{f}(S) x_S)_{S \in \mathcal{T}}$. We have:

$$\begin{aligned}
\left| \sum_{i=1}^N c_i \right| &= \left| \sum_{S \subseteq [n]} \hat{f}(S) x_S - \sum_{S \notin \mathcal{T}} \hat{f}(S) x_S \right| \\
&\le \left| \sum_{S \subseteq [n]} \hat{f}(S) x_S \right| + \sum_{S \notin \mathcal{T}} |\hat{f}(S) x_S| \\
&< 1 + (n-1) = n.
\end{aligned}$$

Write $\mu = \frac{1}{N} \sum_{i=1}^N c_i$, so $|\mu| < n/N$. Now we bound σ^2:

$$\begin{aligned}
\sigma^2 &:= \frac{1}{N} \sum_{i=1}^N (c_i - \mu)^2 \\
&\le \frac{1}{N} \sum_{i=1}^N 2(c_i^2 + \mu^2) \\
&= 2\mu^2 + \frac{2}{N} \sum_{i=1}^N c_i^2 \\
&\le 2\mu^2 + 2/N \\
&< 3/N
\end{aligned}$$

where the next to last inequality is by Parseval's identity and the last is since $\mu^2 < (n/N)^2 < 1/2N$.

Finally, we have that $|c_i| \le \frac{1}{n-2}$ and hence $|c_i - \mu| \le \frac{2}{n}$ for all $1 \le i \le N$. The second of these inequalities follows from the first since $|\mu| < n/N < 2n2^{-n} < \frac{2}{n} - \frac{1}{n-2}$. To see the first inequality, note that otherwise we would have $|\hat{f}(S)| > 1/(n-2)$ for all $S \notin \mathcal{T}$, hence $|\mathcal{M} - \mathcal{T}| < (n-2)^2$ because $\sum_{S \notin \mathcal{T}} \hat{f}^2(S) \le 1$ by Parseval. But by (1) and Cauchy-Schwarz we have:

$$\begin{aligned}
n - 2 &\le \sum_{S \notin \mathcal{T}} |\hat{f}(S)| \\
&\le \sqrt{|\mathcal{M} - \mathcal{T}|} \sqrt{\sum_{S \notin \mathcal{T}} \hat{f}^2(S)} \\
&\le \sqrt{|\mathcal{M} - \mathcal{T}|},
\end{aligned}$$

which is a contradiction.

Suppose that we select k of the c_i's at random, without replacement. Let X denote the sum of the selected numbers. Our goal is to show that $|X| \geq 1$ with probability at most 3^{-n}. By Hoeffding's bound, with $t = \frac{1}{2k}$ and $M = \frac{2}{n}$, we have:

$$\Pr[|X/k - \mu| \geq t] \leq \\ 2\exp\left((-kt/M)[(1+\sigma^2/Mt)\ln(1+Mt/\sigma^2)-1]\right)$$

which implies

$$\Pr[|X - k\mu| \geq 1/2] \leq \\ 2\exp\left((-n/4)[(1+3nk/N)\ln(1+N/3nk)-1]\right)$$

which implies

$$\Pr[|X| \geq 1/2 + |k\mu|] \leq \\ 2\exp\left((-n/4)[(1+3nk/N)\ln(1+N/3nk)-1]\right)$$

which by (2) implies

$$\Pr[|X| \geq 1/2 + 2/C] \leq \\ 2\exp\left((-n/4)[(1+6/C)\ln(1+C/6)-1]\right)$$

which in turn implies

$$\Pr[|X| \geq 1] \leq 2\exp(-2n)$$

by taking C to be a large enough constant, since $(1+x)\ln(1+1/x) - 1 \to \infty$ as $x \to 0^+$. Hence $\Pr[|X| \geq 1] \leq 3^{-n}$ as desired, and the proof is complete. □

4. Upper bounds on density and degree for almost all functions

In the previous section we showed that every Boolean function has threshold density at most $(1 - \frac{1}{O(n)})2^n$. In this section we show that every subset of $(1-\frac{1}{O(n)})2^n$ monomials can serve as a polynomial threshold support for almost every Boolean function. More precisely, we prove:

Theorem 8 *Let $\mathcal{S} \subseteq 2^{[n]}$ be any collection of subsets of $[n]$ such that $|\mathcal{S}| \geq (1-\frac{1}{6n})2^n$. Then for all but a $1/2^n$ fraction of Boolean functions f on n bits, there is a polynomial p whose support is contained in $\mathcal{S}$ such that p sign-represents f.*

An interesting special case of Theorem 8 occurs when we take $\mathcal{S}$ to be the $(1 - \frac{1}{6n})2^n$ smallest subsets of $2^{[n]}$. By the Chernoff bound we then have that $|S| \leq \frac{n}{2} + O(\sqrt{n \log n})$ for all $S \in \mathcal{S}$. We thus obtain the following corollary:

Corollary 9 *Almost all Boolean functions have threshold degree at most $\frac{n}{2} + O(\sqrt{n \log n})$.*

As noted earlier, Anthony and also Alon have used a counting argument to show that almost every Boolean function has threshold degree at least $\lfloor n/2 \rfloor$. Together with this lower bound, our upper bound answers in the affirmative a conjecture of Wang and Williams [26] and Aspnes *et al.* [3] up to lower order terms. (They conjectured that almost all Boolean functions have threshold degree exactly $n/2$.) We note here that Corollary 9 has also been independently proved by Samorodnitsky [24].

Using the Theorem of the Alternative, Aspnes *et al.* gave a simple proof that for any n-bit Boolean function f, the sum of the strong degree of f and the weak degree of $f \cdot \mathsf{PARITY}_n$ is exactly n (Lemma 2.5 of [3]). Hence Corollary 9 also implies that almost all Boolean functions have weak degree at least $n/2 - O(\sqrt{n \log n})$.

4.1. Proof of Theorem 8

Let $f : \{+1,-1\}^n \to \{+1,-1\}$ denote a randomly chosen Boolean function. In the sequel, all probabilities are taken over this choice of f. To motivate our proof of Theorem 8, we sketch Alon and Gotsman's simpler proof (see [12, 23]) of the weaker upper bound $2^n - \frac{1}{2\sqrt{n}}2^{n/2}$ in Appendix A.

Alon and Gotsman's argument uses a "worst-case" assumption about the magnitude of the sum of the omitted Fourier coefficients. If the Fourier coefficients of the random function f were not just binomially distributed but were *independent* random variables, then we could use standard tail inequalities on sums of independent random variables to obtain a stronger bound. However the Fourier coefficients are not at all independent, so this direct approach does not seem to work.

We get around this by showing that in fact the error term $\sum_{|S|>d} \hat{f}(S)x_S$ can be viewed as a sum of independent random variables. These new independent variables no longer correspond to the individual Fourier coefficients $\hat{f}(S)$, and thus we cannot use the arguments of Alon and Gotsman to bound their deviation. However, as shown below, we can exactly characterize the variance of the sum of these new random variables, and this enables us to push the argument through.

We now proceed with the proof. For $z \in \{+1,-1\}^n$ let $\delta_z : \{+1,-1\}^n \to \mathbb{R}$ be the function

$$\delta_z(x) = \begin{cases} 1 & \text{if } x = z \\ 0 & \text{otherwise.} \end{cases}$$

The Fourier representation of δ_z is easily seen to be

$$\begin{aligned} \delta_z(x) &= \frac{(1+z_1x_1)(1+z_2x_2)\cdots(1+z_nx_n)}{2^n} \\ &= \frac{1}{2^n}\sum_{S\subseteq[n]} z_S x_S. \end{aligned}$$

Consequently any function $f : \{+1,-1\}^n \to \mathbb{R}$ may be written as:

$$f(x) = \sum_{z \in \{+1,-1\}^n} f(z) \frac{1}{2^n} \sum_{S \subseteq [n]} z_S x_S.$$

For any $\mathcal{S} \subseteq 2^{[n]}$ we thus have

$$f_{\mathcal{S}}(x) = \frac{1}{2^n} \sum_{z \in \{+1,-1\}^n} f(z) \sum_{S \in \mathcal{S}} z_S x_S. \qquad (3)$$

Let $\delta_{\mathcal{S},z}(x) = \sum_{S \in \mathcal{S}} z_S x_S$. It is clear that $\delta_{\mathcal{S},x}(x) = |\mathcal{S}|$ for any $x \in \{+1,-1\}^n$. We now claim:

Lemma 10 *For any* $x \in \{+1,-1\}^n$, *we have* $\sum_{z \neq x} \delta_{\mathcal{S},z}(x)^2 = 2^n|\mathcal{S}| - |\mathcal{S}|^2$.

Proof:

$$\begin{aligned}
\sum_{z \neq x} \delta_{\mathcal{S},z}(x)^2 &= \sum_{z \in \{+1,-1\}^n} \delta_{\mathcal{S},z}(x)^2 - \delta_{\mathcal{S},x}(x)^2 \\
&= \sum_{z \in \{+1,-1\}^n} \delta_{\mathcal{S},z}(x)^2 - |\mathcal{S}|^2 \\
&= \sum_{z \in \{+1,-1\}^n} \delta_{\mathcal{S},x}(z)^2 - |\mathcal{S}|^2 \qquad (4) \\
&= 2^n \sum_{S \subseteq [n]} \hat{\delta}_{\mathcal{S},x}(S)^2 - |\mathcal{S}|^2 \qquad (5) \\
&= 2^n|\mathcal{S}| - |\mathcal{S}|^2, \qquad (6)
\end{aligned}$$

where (4) is because $\delta_{\mathcal{S},z}(x) = \delta_{\mathcal{S},x}(z)$, (5) is Parseval's identity, and (6) follows because $\delta_{\mathcal{S},x}$ has exactly $|\mathcal{S}|$ nonzero Fourier coefficients, each of magnitude exactly 1. $\square$

To prove Theorem 8, fix any $\mathcal{S} \subseteq 2^{[n]}$ with $|\mathcal{S}| \geq (1 - \frac{1}{6n})2^n$. Fix any $x \in \{+1,-1\}^n$. We will show that for a random Boolean function f, with probability at least $1 - 1/4^n$ we have $\text{sgn}(f_{\mathcal{S}}(x)) = f(x)$. If this is the case, then for a random Boolean function f we have that $\text{sgn}(f_{\mathcal{S}}(x)) = f(x)$ for all x with probability at most $1/2^n$ and the theorem is proved.

We have

$$\begin{aligned}
\text{sgn}(f_{\mathcal{S}}(x)) &= \text{sgn}\left(\sum_{z \in \{+1,-1\}^n} f(z)\delta_{\mathcal{S},z}(x)\right) \\
&= \text{sgn}\left(f(x)|\mathcal{S}| + \sum_{z \neq x} f(z)\delta_{\mathcal{S},z}(x)\right). \qquad (7)
\end{aligned}$$

Since each $f(z)$ is an independent random ± 1 value, we may view the sum over $z \neq x$ in (7) as a sum of $2^n - 1$ independent random variables, where the z-th random variable takes values $\pm\delta_{\mathcal{S},z}(x)$ each with probability $1/2$. From Lemma 10 we know that the sum of the squares of $\delta_{\mathcal{S},z}(x)$ is precisely $2^n|\mathcal{S}| - |\mathcal{S}|^2$, and hence the variance of the sum of these $2^n - 1$ random variables is precisely $\sigma^2 = \frac{2^n|\mathcal{S}| - |\mathcal{S}|^2}{2^n - 1}$. We can bound each random variable's deviance from the mean 0 by noting that $|\delta_{\mathcal{S},z}(x)| \leq 2^n - |\mathcal{S}|$ for all $z \neq x$ (this holds since by adding $\sum_{S \notin \mathcal{S}} z_S x_S$ to $\delta_{\mathcal{S},z}(x)$ we would get $\sum_{S \subseteq [n]} z_S x_S$ which is 0). Hence by Hoeffding's bound, with $k = 2^n - 1$, $t = \frac{|\mathcal{S}|}{2^n - 1}$, and $M = 2^n - |\mathcal{S}|$, we have:

$$\begin{aligned}
&\Pr[|\tfrac{1}{k} \sum_{z \neq x} f(z)\delta_{\mathcal{S},z}(x)| \geq t] \leq \\
&\qquad 2\exp\left((-kt/M)[(1 + \sigma^2/Mt)\ln(1 + Mt/\sigma^2) - 1]\right)
\end{aligned}$$

which implies

$$\begin{aligned}
&\Pr[|\sum_{z \neq x} f(z)\delta_{\mathcal{S},z}(x)| \geq |\mathcal{S}|] \\
&\quad \leq 2\exp\left(-\left(\frac{|\mathcal{S}|}{2^n - |\mathcal{S}|}\right)[(1+1)\ln(1+1) - 1]\right) \\
&\quad \leq 2\exp(-2n) \qquad (8) \\
&\quad < 1/4^n,
\end{aligned}$$

where the last line uses $|\mathcal{S}| \geq (1 - \frac{1}{6n})2^n$. But when $|\sum_{z \neq x} f(z)\delta_{\mathcal{S},z}(x)| < |\mathcal{S}|$, the right-hand side of (7) is just $\text{sgn}(f(x)|\mathcal{S}|) = f(x)$, and the theorem is proved. (Theorem 8) ■

5. Weak threshold density

In this section we give an upper bound on weak threshold density which holds for all Boolean functions and a stronger upper bound which holds for almost all Boolean functions. These bounds give a negative answer to a question of Saks. We also give a lower bound on weak threshold density which holds for almost all Boolean functions and a stronger lower bound which holds for a particular Boolean function. To the best of our knowledge these are the only lower bounds known for weak threshold density.

5.1. Upper bounds for weak threshold density

Since any strong representation of a Boolean function f is also a weak representation, Theorem 5 implies that for any function f and any set $\mathcal{S} \subseteq \mathcal{M}$ of monomials either f has a weak representation with support contained in $\mathcal{S}$ or f has a weak representation with support contained in $\mathcal{M} - \mathcal{S}$ (or both). Taking $\mathcal{S}$ to be any set of $\frac{1}{2}2^n$ monomials, it follows that the weak density of every Boolean function is at most $\frac{1}{2}2^n$.

Saks has asked the following question (Question 2.28.2 of [23]): is it true that for all $\epsilon > 0$ almost all Boolean functions have weak density at least $(\frac{1}{2}-\epsilon)2^n$? Our next two theorems show that the answer is "no" in a rather strong sense:

Theorem 11 *Almost all Boolean functions have weak density at most $\frac{2}{n}2^n$.*

Theorem 12 *All Boolean functions have weak density $o(1)2^n$.*

The intuition behind the proof of Theorem 11 is straightforward: with high probability a random Boolean function f has some small subcube on which f is "simple." We take advantage of this simplicity to construct a low-density polynomial p which weakly represents f on this subcube. Multiplying p by another polynomial which is 0 off of the subcube, we obtain a weak representative for f. More precisely, we use the following lemma:

Lemma 13 *Let τ be a restriction which fixes $n-k$ variables from $x_1,\ldots,x_n$ and keeps k variables free. Let D denote the weak density of $f|_\tau$. Then the weak density of f is at most $2^{n-k}D$.*

Proof: Without loss of generality we can suppose that τ is the restriction which maps variables $x_1,\ldots,x_{n-k}$ to 1 and leaves the remaining k variables free. Let p be a polynomial over $x_{n-k+1},\ldots,x_n$ which weakly represents $f|_\tau$ and has D nonzero monomials. Then the polynomial

$$\begin{aligned} P(x_1,\ldots,x_n) &= (x_1+1)(x_2+1)\cdots(x_{n-k}+1)\cdot \\ &\quad p(x_{n-k+1},\ldots,x_n) \end{aligned}$$

has density $2^{n-k}D$. To see that P weakly represents f, note that on any input $x=1^{n-k}y$ we have $P(x)=2^{n-k}p(x)$, while on any other input we have $P(x)=0$. Since p is a weak representative of $f|_\tau$ it must be somewhere nonzero, so the same is true for P. □

Proof of Theorem 11: Let f be a random Boolean function. Consider the 2^{n-k} disjoint k-dimensional subcubes of $\{+1,-1\}$ corresponding to restrictions τ which fix variables $x_1,\ldots,x_{n-k}$. For any such restriction τ we have

$$\Pr[f|_\tau \text{ is not identically } 1] = 1-\frac{1}{2^{2^k}}$$

and hence

$$\Pr[f|_\tau \not\equiv 1 \text{ for all such } \tau] = \left(1-\frac{1}{2^{2^k}}\right)^{2^{n-k}}.$$

Taking $k=\log n-1$, the above probability is $(1-2^{-n/2})^{2^{n-\log n+1}} < e^{-2^{n/2+1}/n}$. Thus with overwhelmingly high probability there is some restriction τ fixing $n-\log n+1$ variables such that $f|_\tau$ is identically 1, and hence the weak density of $f|_\tau$ is 1. Now use Lemma 13. □

Using Lemma 13 it is easy to prove an upper bound of $\frac{1}{2}2^n$ on the weak density of all Boolean functions without using Theorem 5. For any Boolean function f on n variables, the polynomial

$$(x_1+1)(x_2+1)\cdots(x_{n-1}+1)y$$

is easily seen to be a weak representative of f which has density $\frac{1}{2}2^n$, where $y\in\{-1,1,-x_n,x_n\}$ is suitably chosen depending on the two values of $f(1^{n-1},1)$ and $f(1^{n-1},-1)$.

By looking at subcubes of dimension greater than 1 it is possible to improve this bound. A straightforward case analysis shows the following:

Fact 14 *Every Boolean function on 3 variables has weak density at most 3.*

Together with Lemma 13, this yields

Corollary 15 *Every Boolean function has weak density at most $\frac{3}{8}2^n$.*

While Corollary 15 already gives a strong negative answer to the question of Saks, we can obtain the stronger upper bound of Theorem 12 by using more powerful tools from Ramsey theory. A *k-dimensional affine subspace* of a vector space V is a translate of a k-dimensional vector subspace of V. The following is a special case of the Affine Ramsey Theorem of Graham *et al.* [13, 14]:

Theorem 16 *Let A be a finite field. For all $r,k\geq 1$ there exists n such that if the points of A^n are r-colored, then some k-dimensional affine subspace of A^n has all of its points the same color.*

Taking $r=2$ and $A=GF_2$, we can rephrase this as:

Corollary 17 *There is a function $g(n)=\omega(1)$ such that for any Boolean function $f:(GF_2)^n\to\{-1,1\}$, there is some $g(n)$-dimensional affine subspace of $(GF_2)^n$ on which f is constant.*

Proof of Theorem 12: Let f be any Boolean function on n variables and let W' be the affine subspace whose existence is asserted by Corollary 17. Any $g(n)$-dimensional vector subspace W of $(GF_2)^n$ is the set of solutions to some system of $n-g(n)$ homogeneous linear equations, i.e.,

$$W=\{x\in(GF_2)^n : Ax=(0)^{n-g(n)}\}$$

where A is an $(n-g(n))\times n$ matrix over GF_2. Thus the $g(n)$-dimensional affine subspace W' is the set of solutions to some system of $n - g(n)$ not necessarily homogeneous linear equations, i.e.,

$$W' = \{x \in (GF_2)^n : Ax = b\}$$

for some $b \in (GF_2)^{n-g(n)}$. If we identify GF_2 with the set $\{+1, -1\}$, then this system of equations becomes:

$$\begin{aligned}
\prod_{j:A_{1,j}=1} x_j &= b_1, \\
\prod_{j:A_{2,j}=1} x_j &= b_2, \\
&\vdots \\
\prod_{j:A_{n-g(n),j}=1} x_j &= b_{n-g(n)}.
\end{aligned}$$

Without loss of generality we may suppose that $f(x) = 1$ for all $x \in W'$. It is easy to see that the points of $\{+1, -1\}^n$ on which the polynomial

$$\prod_{i=1}^{n-g(n)} \left(b_i \left(\prod_{j:A_{i,j}=1} x_i \right) + 1 \right)$$

is nonzero are exactly the points in W', and that moreover this polynomial always takes value exactly $2^{n-g(n)}$ on W'. Thus this polynomial is a weak representative for f of density $2^{n-g(n)} = o(1)2^n$, and Theorem 12 is proved. □

5.2. Lower bounds for weak threshold density

Here we give our lower bounds for weak threshold density. The first lower bound holds for almost every Boolean function:

Theorem 18 *Almost all Boolean functions have weak threshold density at least* $\frac{1}{2\sqrt{n}} 2^{n/2}$.

Proof: Recall the proof of Theorem 8; in particular, equation (8). If we consider sets $\mathcal{S}$ of size $(1 - \epsilon)2^n$, then the probability that f has no PTF over $\mathcal{S}$ is bounded by $2\exp(-.38/\epsilon)$. There are exactly $\binom{2^n}{\epsilon 2^n}$ such sets $\mathcal{S}$. Hence if we select ϵ such that $\binom{2^n}{\epsilon 2^n} \cdot 2\exp(-.38/\epsilon)$ is at most $1/2^n$, then a union bound tells us that almost every Boolean function can be sign-represented using *any* set of $(1 - \epsilon)2^n$ monomials. In this case Theorem 5 implies that for almost every Boolean function, no set of $\epsilon 2^n$ monomials can serve as the support of a weak sign-representation. Taking $\epsilon = \frac{1}{2\sqrt{n}} 2^{-n/2}$, it is easily shown that $\binom{2^n}{\epsilon 2^n} \cdot 2\exp(-.38/\epsilon) < 1/2^n$, and the theorem is proved. □

We can give a slightly better bound for an explicit Boolean function. For $n = 2k$ let IP denote the "inner product mod 2" function, i.e. $IP(x_1, \ldots, x_k, y_1, \ldots, y_k) = \oplus_{i=1}^k (x_i \wedge y_i)$ where $\oplus$ denotes exclusive-OR (parity) and $\wedge$ denotes AND.

Theorem 19 *IP has weak density at least* $2^{n/2}$.

Proof: It is known [8, 19] that IP is a *bent* function, i.e. a function for which $|\hat{f}(S)| = \frac{1}{2^{n/2}}$ for all $S \subseteq [n]$. Consequently, for any set $\mathcal{S}$ of $2^n - 2^{n/2} + 1$ monomials, the function sgn$(f_{\mathcal{S}}(x))$ is a strong representative of f by Fact 6. By Theorem 5 this means that for any set $\mathcal{T}$ of $2^{n/2} - 1$ monomials, it is not the case that f has a weak representative whose support is contained in $\mathcal{T}$. Hence the weak degree of f is at least $2^{n/2}$. □

6. Threshold degree of sparse functions

The following question was posed by Richard Beigel [5]: are sparse sets easy for low-degree polynomial threshold functions? More concretely, let $f : \{+1, -1\}^n \to \{+1, -1\}$ be a Boolean function such that $|f^{-1}(1)| = m \ll 2^n$, so f is the characteristic function of a sparse subset of the Boolean cube. What is the maximum polynomial threshold function degree for such an f? The following theorem gives a complete answer for all values of m:

Theorem 20 *For $1 \le m \le \frac{1}{2}2^n$, let $\mathcal{F}_m$ be the set of all Boolean functions $f : \{+1, -1\}^n \to \{+1, -1\}$ such that $m = \min\{|f^{-1}(1)|, |f^{-1}(-1)|\}$. Then the maximum threshold degree over all $f \in \mathcal{F}_m$ is exactly $\lfloor \log m \rfloor + 1$.*

Proof: We assume without loss of generality that $1 \le |f^{-1}(1)| = m \le \frac{1}{2}2^n$. For the lower bound, let f be any function which is such that if the last $n - (\lfloor \log m \rfloor + 1)$ inputs are fixed to 1 then f computes parity on the first $\lfloor \log m \rfloor + 1$ inputs. (Note that this uses up $2^{\lfloor \log m \rfloor} \le m$ of the ones in f's output; any remaining ones can be located arbitrarily). Since any polynomial threshold function which computes parity on k variables must have degree at least k, it follows that any polynomial threshold function for f must have degree at least $\lfloor \log m \rfloor + 1$.

For the upper bound, we begin by constructing an m-leaf decision tree over variables $x_1, \ldots, x_n$ such that each string in $f^{-1}(1)$ arrives at a different leaf. Such a tree can be constructed by a greedy algorithm: initially all strings in $f^{-1}(1)$ are at the root of the tree. Let x_i be any variable such that there are two strings in $f^{-1}(1)$ which disagree on x_i (such a variable must exist as long as $|f^{-1}(1)| \ge 2$). Label the root with x_i. The strings $\{x : x \in f^{-1}(1), x_i = -1\}$ go to the left child and the strings $\{x : x \in f^{-1}(1), x_i = 1\}$ go to the right child. Now recurse on each child. At the end of this process we have an m-leaf tree in which each (unlabeled) leaf has a unique string in $f^{-1}(1)$ which reaches that leaf.

Let ℓ be a leaf in this tree and let z be the element of $f^{-1}(1)$ which reaches that leaf. We label ℓ with the degree-1 polynomial threshold function sgn$(p(x))$ where $p(x) = x_1 z_1 + \cdots + x_n z_n - n + \frac{1}{2}$. Note that $p(z) = \frac{1}{2}$, and $p(x) \leq -\frac{1}{2}$ for all binary inputs $x \neq z$. Thus we now have an m-leaf decision tree T in which internal nodes are labeled with variables and leaves are labeled with degree-1 polynomial threshold functions, such that T computes exactly f.

The rest of our proof follows the proof of Theorem 2 in [18]. Recall that the rank of a decision tree T is defined inductively as follows:

- If T is a single leaf then $rank(T) = 0$.
- If T has subtrees T_0 and T_1 then $rank(T)$ equals $\max(rank(T_0), rank(T_1))$ if $rank(T_0) \neq rank(T_1)$ and equals $rank(T_0) + 1$ if $rank(T_0) = rank(T_1)$.

It follows from this definition that the rank of an m-leaf tree is at most $\lfloor \log m \rfloor$. Now we use the fact (see [7]) that a rank-r decision tree with functions $f_1, f_2, \ldots, f_m$ at the leaves is equivalent to some r-decision list, i.e., to a function "if $C_1(x)$ then output $f_1(x)$ else if $C_2(x)$ then output $f_2(x)$ else $\ldots$ else output $f_m(x)$" where each C_i is a conjunction on at most r variables. Thus, our decision tree T is equivalent to such a decision list, where $r = \lfloor \log m \rfloor$ and each f_i is a degree-1 polynomial threshold function sgn(p_i) as described above.

We now show that the degree-$(\lfloor \log m \rfloor + 1)$ polynomial threshold function sgn$(P(x))$ computes T, where $P(x)$ equals

$$A_1 \tilde{C}_1(x) p_1(x) + A_2 \tilde{C}_2(x) p_2(x) + \cdots A_m \tilde{C}_m(x) p_m(x).$$

Here $\tilde{C}_i$ is the polynomial of degree at most $\lfloor \log m \rfloor$ which outputs 1 if C_i is true and 0 if C_i is false, and $A_1 \gg A_2 \gg A_3 \gg \cdots \gg A_m > 0$ are appropriately chosen positive values. To see that this works, note that if C_i is the first conjunction in the decision list which is satisfied by x, then we have

$$P(x) = A_i p_i(x) + \sum_{j > i, C_j(x) = 1} A_j p_j(x).$$

Since $|p_i(x)| \geq \frac{1}{2}$ and $A_i \gg A_j > 0$ for $j > i$, the sign of $P(x)$ is the same as the sign of $p_i(x)$, and we are done. □

7. Conclusion

While we have made significant progress on extremal bounds for threshold degree and threshold density, there is still room for improvement. One goal is to improve the lower order term in our $n/2 + O(\sqrt{n \log n})$ upper bound for the threshold degree of almost every Boolean function. Another goal is to give tighter bounds on the maximum threshold density of Boolean functions. Saks [23] has asked whether almost all Boolean functions have threshold density at least $(1 - \epsilon)2^n$ for some $\epsilon > 0$. We conjecture that the answer is "no" in a strong sense:

Conjecture 21 *For n sufficiently large, every Boolean function $f : \{+1, -1\}^n \to \{+1, -1\}$ has threshold density at most $\frac{1}{2}2^n$.*

Finally, a large gap remains between our upper and lower bounds for weak threshold density; it would be interesting to tighten these bounds.

References

[1] N. Alon. Personal communication to M. Saks, reported in M. Saks, "Slicing the Hypercube", 1993.

[2] M. Anthony. Classification by polynomial surfaces. *Discrete Applied Mathematics*, 61:91–103, 1995.

[3] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):1–14, 1994.

[4] R. Beigel. Perceptrons, PP, and the Polynomial Hierarchy. *Computational Complexity*, 4:339–349, 1994.

[5] R. Beigel. Personal communication, 2000.

[6] R. Beigel, N. Reingold, and D. Spielman. PP is closed under intersection. *Journal of Computer and System Sciences*, 50(2):191–202, 1995.

[7] A. Blum. Rank-r decision trees are a subclass of r-decision lists. *Information Processing Letters*, 42(4):183–185, 1992.

[8] J. Bruck. Harmonic analysis of polynomial threshold functions. *SIAM Journal on Discrete Mathematics*, 3(2):168–177, 1990.

[9] J. Bruck and R. Smolensky. Polynomial threshold functions, AC^0 functions and spectral norms. *SIAM Journal on Discrete Mathematics*, 21:33–42, 1992.

[10] T. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, EC-14(3):326–334, 1965.

[11] B. Fu. Separating PH from PP by relativization. *Acta Math. Sinica*, 8(3):329–336, 1992.

[12] C. Gotsman. On boolean functions, polynomials and algebraic threshold functions. Technical Report TR-89-18, Department of Computer Science, Hebrew University, 1989.

[13] R. L. Graham, K. Leeb, and B. L. Rothschild. Ramsey's theorem for a class of categories. *Adv. Math.*, 8:417–433, 1972.

[14] R. L. Graham, B. L. Rothschild, and J. L. Spencer. *Ramsey Theory*. Wiley-Interscience, New York, 1990.

[15] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[16] J. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55:414–440, 1997.

[17] A. Klivans, R. O'Donnell, and R. Servedio. Learning intersections and thresholds of halfspaces. In *Proceedings of the 43rd Annual Symposium on Foundations of Computer Science*, 2002.

[18] A. Klivans and R. Servedio. Learning DNF in time $2^{\tilde{o}(n^{1/3})}$. In *Proceedings of the Thirty-Third Annual Symposium on Theory of Computing*, pages 258–265, 2001.

[19] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, New York, 1977.

[20] M. Minsky and S. Papert. *Perceptrons: an introduction to computational geometry*. MIT Press, Cambridge, MA, 1968.

[21] R. O'Donnell and R. Servedio. New degree bounds for polynomial threshold functions. Proceedings of the 35th ACM Symposium on Theory of Computing, 2003.

[22] A. Razborov and S. Rudich. Natural proofs. *Journal of Computer and System Sciences*, 55(1):24–35, 1997.

[23] M. Saks. *Slicing the hypercube*, pages 211–257. London Mathematical Society Lecture Note Series 187, 1993.

[24] A. Samorodnitsky. Unpublished manuscript, 2000. Personal communication, 2003.

[25] N. Vereshchagin. Lower bounds for perceptrons solving some separation problems and oracle separation of AM from PP. In *Proceedings of the Third Annual Israel Symposium on Theory of Computing and Systems*, 1995.

[26] C. Wang and A.C. Williams. The threshold order of a boolean function. *Discrete Applied Mathematics*, 31:51–69, 1991.

A. Alon and Gotsman's threshold density upper bound

Theorem 22 *(Alon and Gotsman) Let $\mathcal{S} \subseteq 2^{[n]}$ be any collection of subsets of $[n]$ such that $|\mathcal{S}| \geq 2^n - \frac{1}{2\sqrt{n}}2^{n/2}$. Then for all but a $1/2^n$ fraction of Boolean functions f on n bits, there is a polynomial p whose support is contained in $\mathcal{S}$ such that p sign-represents f.*

Proof: We first claim that for any $S \subseteq [n]$, the value $\hat{f}(S)$ is distributed as $\frac{1}{2^n}B(\pm1, 2^n)$, where $B(\pm1, 2^n)$ denotes the binomial random variable given by the sum of 2^n independent uniform ±1 values. To see this, note that:

$$\hat{f}(S) = \mathbf{E}[f(x)x_S] = \Pr[f(x) = x_S] - \Pr[f(x) \neq x_S],$$

and that for a random function f, each input $x \in \{+1, -1\}$ satisfies $f(x) = x_S$ with probability $\frac{1}{2}$.

It's now straightforward to show that $|||f|||_\infty \leq \frac{2\sqrt{n}}{2^{n/2}}$ with probability $1 - 1/2^n$; a Chernoff bound tells us that for each fixed monomial S, $\Pr[|\hat{f}(S)| > \frac{2\sqrt{n}}{2^{n/2}}] \leq 1/4^n$ and we take a union bound over all 2^n subsets S. Hence for any fixed set of monomials $\mathcal{S}$ with $|\mathcal{S}| \geq 2^n - \frac{1}{2\sqrt{n}}2^{n/2}$ we get $\sum_{S \notin \mathcal{S}} |\hat{f}(S)| < 1$ with probability $1 - 1/2^n$. Fact 6 completes the proof. □

Corollary 23 *All but a $1/2^n$ fraction of n-bit Boolean functions have threshold degree $\leq .89n$.*

Proof: In Theorem 22, take the set of $2^n - \frac{1}{2\sqrt{n}}2^{n/2}$ smallest monomials. By standard binomial tail bounds, these all have size at most $.89n$. □