Bounded Independence Fools Halfspaces

Ilias Diakonikolas[*] , Parikshit Gopalan[†], Ragesh Jaiswal[‡], Rocco A. Servedio[§] , and Emanuele Viola[¶]
[*‡§] *Department of Computer Science, Columbia University, New York, NY 10027.*
Email: {ilias, rjaiswal, rocco}@cs.columbia.edu
[†] *Microsoft Research, Silicon Valley, CA 94043. Email: parik@microsoft.com*
[‡] *College of Computer and Information Science, Northeastern University, Boston, MA 02115.*
Email: viola@ccs.neu.edu

Abstract— We show that any distribution on $\{-1,+1\}^n$ that is k-wise independent fools any halfspace (a.k.a. threshold) $h : \{-1,+1\}^n \to \{-1,+1\}$, i.e., any function of the form $h(x) = \mathrm{sign}(\sum_{i=1}^{n} w_i x_i - \theta)$ where the $w_1, \dots, w_n, \theta$ are arbitrary real numbers, with error ϵ for $k = O(\epsilon^{-2}\log^2(1/\epsilon))$. Our result is tight up to $\log(1/\epsilon)$ factors. Using standard constructions of k-wise independent distributions, we obtain the first explicit pseudorandom generators $G : \{-1,+1\}^s \to \{-1,+1\}^n$ that fool halfspaces. Specifically, we fool halfspaces with error ϵ and seed length $s = k \cdot \log n = O(\log n \cdot \epsilon^{-2}\log^2(1/\epsilon))$.

Our approach combines classical tools from real approximation theory with structural results on halfspaces by Servedio (Comput. Complexity 2007).

Keywords-halfspaces; pseudorandomness; k-wise independent distributions

1. Introduction

Halfspaces, or threshold functions, are a central class of Boolean functions $h : \{-1,+1\}^n \to \{-1,+1\}$ of the form:

$$h(x) = \mathrm{sign}(w_1 x_1 + \cdots + w_n x_n - \theta),$$

where the weights $w_1, \dots, w_n$ and the threshold θ are arbitrary real numbers. These functions have been studied extensively in a variety of contexts. In computer science, the work on halfspaces dates back to the study of switching functions, see for instance the books [15], [25], [32], [51], [38]. In computational complexity, much effort has been put into understanding constant-depth circuits of halfspaces. On the one hand this has resulted in surprising inclusions (such as the simulation of depth-d circuits of halfspaces by depth-$(d+1)$ circuits of majority gates [20], [21]), but on the other hand many seemingly basic questions remain unsolved: for instance it is conceivable that every function in NP is computable by a polynomial-size depth-2 circuit of halfspaces [23], [30], [31], [19]. In learning theory, the problem of learning an unknown halfspace has arguably been the most influential problem in the development of the field, with algorithms such as Perceptron, Weighted Majority, Boosting, and Support Vector Machines emerging from this study. Halfspaces (with non-negative weights) have also been studied extensively in game theory and social choice theory, where they are referred to as “weighted majority games” and have been analyzed as models for voting, see e.g., [45], [26], [17], [52].

In this work we make progress on a natural complexity-theoretic question about halfspaces. We construct the first explicit pseudorandom generators $G : \{-1,+1\}^s \to \{-1,+1\}^n$ with short seed length s that fool any halfspace $h : \{-1,+1\}^n \to \{-1,+1\}$, i.e. satisfy

$$|\,\mathbf{E}_{x\in\{-1,+1\}^s}[h(G(x))] - \mathbf{E}_{x\in\{-1,+1\}^n}[h(x)]| \le \epsilon,$$

for a small ϵ. We actually prove that the class of distributions known as k-wise independent has this “fooling” property for a suitable k; as pointed out below, a generator can then be obtained using any of the standard explicit constructions of such distributions.

Definition 1.1. A distribution $\mathcal{D}$ on $\{-1,+1\}^n$ is *k-wise independent* if the projection of $\mathcal{D}$ on any k indices is uniformly distributed over $\{-1,+1\}^k$.

Theorem 1.2 (Main)**.** Let $\mathcal{D}$ be a k-wise independent distribution on $\{-1,+1\}^n$, and let $h : \{-1,+1\}^n \to \{-1,+1\}$ be a halfspace. Then $\mathcal{D}$ fools h with error ϵ, i.e.,

$$|\,\mathbf{E}_{x\leftarrow\mathcal{D}}[h(x)] - \mathbf{E}_{x\leftarrow\mathcal{U}}[h(x)]| \le \epsilon, \text{provided } k \ge \frac{C}{\epsilon^2}\log^2(\frac{1}{\epsilon}),$$

where C is an absolute constant and $\mathcal{U}$ is the uniform distribution over $\{-1,+1\}^n$.

Our Theorem 1.2 is tight up to $\log(1/\epsilon)$ factors, as can be seen by considering the halfspace $h(x) := \mathrm{sign}(\sum_{i\le k+1} x_i)$ and the k-wise independent distribution $(x_1, x_2, \dots, x_k, \prod_{i\le k} x_i, x_{k+2}, \dots, x_n)$ where the variables x_i are independent and uniform in $\{-1,+1\}$.

Standard explicit constructions of k-wise independent distributions over $\{-1,+1\}^n$ have seed length $O(k \cdot \log n)$ [14], [3], which is optimal up to constant factors [13]. Plugging these in Theorem 1.2, we obtain explicit pseudorandom generators $G : \{-1,+1\}^s \to \{-1,+1\}^n$ that fool any

[*] Research supported by NSF grants CCF-0728736, CCF-0525260, and by an Alexander S. Onassis Foundation Fellowship.

[‡]Research supported by DARPA award HR0011-08-1-0069.

[§] Supported in part by NSF grants CCF-0347282, CCF-0523664 and CNS-0716245, and by DARPA award HR0011-08-1-0069.

[¶]Supported by NSF grant CCF-0845003. This work was partially done while the author was a postdoctoral fellow at Columbia University, supported by grants NSF award CCF-0347282 and NSF award CCF-0523664.

halfspace $h : \{-1,+1\}^n \to \{-1,+1\}$ with error ϵ and have seed length $s = O(\log n \cdot \epsilon^{-2} \log^2(\epsilon^{-1}))$.

Discussion and comparison with previous explicit generators. The literature is rich with explicit generators for various classes, such as small constant-depth circuits with various gates [2], [42], [35], [54], [5], [10], low-degree polynomials [39], [4], [9], [34], [53], and one-way small-space algorithms [40]. Many of these classes (such as low-degree polynomials and AC^0 circuits) provably cannot implement halfspaces, and it is not known how to implement an arbitrary halfspace in any of these classes, so none of these results gives Theorem 1.2. However, some of these results [40], [35], [54] give generators for the *restricted class* of halfspaces given by $h(x) = \mathrm{sign}(\sum_{i=1}^n w_i x_i - \theta)$ where the weights are integers of magnitude at most $\mathrm{poly}(n)$. While it is well known that every halfspace has a representation with integer weights, it is not possible to represent an arbitrary halfspace with $\mathrm{poly}(n)$ integer weights. Indeed, an easy counting argument (see e.g. [36], [24]) shows that if the weights are required to be integers then almost all halfspaces require weights of magnitude $2^{\Omega(n)}$; in fact some halfspaces require weights of magnitude $2^{\Theta(n \log n)}$ [24]. Our result is for the entire class of halfspaces with no restriction on the weights, and much of the richness of halfspaces only comes in this setting; for example, the "odd-max-bit" function [6], the "universal halfspace" [20], and other important halfspaces [24] all require exponentially large integer weights. Moreover, even for the restricted class of halfspaces where the weights are integers of magnitude at most $\mathrm{poly}(n)$, previous techniques [40] give seed length $s = O(\log^2 n)$ at best, while we achieve $s = O(\log n)$ for constant error. Also note that, while halfspaces can be approximated by ones with small integer weights [50], this approximation is not immediately useful for generators as it only holds for the uniform distribution, not the pseudorandom one.

After our work, R. Meka and D. Zuckerman (personal communication, 2009) show how to use [40] to fool every halfspace with seed-length $O(\log^2(n/\epsilon))$, and they also construct a generator with seed-length $O(\log(n)\log(1/\epsilon))$.

Other related results. Several recent papers have studied the power of k-wise independent distributions. An exciting recent result of Braverman [10], which builds on an earlier breakthrough of Bazzi [5] (simplified by Razborov [48]), shows that $\mathrm{polylog}(n)$-wise independent distributions fool small constant-depth circuits, settling a conjecture of Linial and Nisan [33]. Benjamini *et al.* [7] showed that any $O(1/\epsilon^2)$-wise independent distribution $\mathcal{D}$ on $\{-1,+1\}^n$ satisfies $|\Pr_{x \leftarrow \mathcal{D}}[\sum_i x_i \geq 0] - 1/2| \leq \epsilon$, i.e., such distributions fool the majority function. (We discuss [7] in more detail shortly. Here we note that their result does not seem immediately relevant for constructing generators, because to fool the majority function with error 0 one can just output 1^n with probability $1/2$ and $(-1)^n$ with probability $1/2$.)

The problem of constructing generators for halfspaces has been considered by several authors in the recent literature. Rabani and Shpilka give an explicit construction of an ϵ-net, or ϵ-hitting set, for halfspaces [47]: a set of size $\mathrm{poly}(n, 1/\epsilon)$ which is guaranteed to contain at least one point where $h(x) = +1$ and at least one point where $h(x) = -1$ for any halfspace h which takes on both values with probability at least ϵ under the uniform distribution. However, their construction does not offer any guarantees about the distribution of these values. [47] pose as a research goal "to build methodically a theory of generators for geometric functions" such as halfspaces.

The problem of generators for halfspaces also arose in recent work by Gopalan and Radhakrishnan [22] on finding duplicates in a data stream. They required a generator that allows one to estimate the influence of a variable in a halfspace, a problem which is in fact equivalent to constructing a generator for a related halfspace. They observe that Nisan's space generator [40] suffices for the halfspaces arising in their context, and raise the problem of constructing generators for general halfspaces. Our result does not improve theirs, but it makes the analysis simpler by showing that one can use $\tilde{O}(\epsilon^{-2})$-wise independence to estimate the influence to within an additive ϵ.

1.1. Techniques

Our proof combines tools from real approximation theory with structural results regarding halfspaces. An important notion is that of an ϵ-*regular* halfspace; which is a halfspace $h(x) = \mathrm{sign}(\sum_i w_i x_i - \theta)$ where no more than an ϵ-fraction of the 2-norm of its coefficient vector $(w_1, \ldots, w_n)$ comes from any single coefficient w_i. We first show that k-wise independence fools all ϵ-regular halfspaces, and then use this to prove that k-wise independence fools all halfspaces. Our proof can be broken into three steps.

Step 1: Fooling regular halfspaces. Our starting point is Bazzi's observation [5, Theorem 4.2] (also in [7]), that to establish that every k-wise independent distribution on $\{-1,+1\}^n$ fools a Boolean function $f : \{-1,+1\}^n \to \{-1,+1\}$ with error ϵ, it is sufficient to exhibit two "sandwiching" polynomials $q_\ell, q_u : \{-1,+1\}^n \to \{-1,+1\}$ of degree at most k such that:

- $q_u(x) \geq f(x) \geq q_\ell(x)$ for all $x \in \{-1,+1\}^n$; and
- $\mathbf{E}_{\mathcal{U}}[q_u(x) - f(x)], \mathbf{E}_{\mathcal{U}}[f(x) - q_\ell(x)] \leq \epsilon$.

Using only classical tools from real approximation theory, we give a proof of the existence of univariate polynomials of degree $K(\epsilon) := \tilde{O}(1/\epsilon^2)$ which, roughly speaking, provide a good sandwich approximator to the *univariate* function $\mathrm{sign}(t)$ *under the normal distribution on* $\mathbf{R}$. This is useful because of the following simple but crucial insight: for any regular halfspace $h(x) = \mathrm{sign}(w \cdot x - \theta)$, the argument $w \cdot x - \theta$ is well-approximated by a normal random variable (a precise error-estimate is given by the Berry-Esséen theorem).

For any ϵ-regular halfspace, we can plug $w \cdot x - \theta$ into our univariate polynomials, and obtain low-degree sandwich polynomials for h, establishing that $K(\epsilon)$-wise independence fools all ϵ-regular halfspaces. The construction of these polynomials is the most technical part of this paper.

Of course, there are halfspaces $\mathrm{sign}(w \cdot x - \theta)$ that are far from being ϵ-regular and have $w \cdot x - \theta$ distributed very unlike a Gaussian. To tackle general halfspaces, we use the notion of the *ϵ-critical index of a halfspace*, which was (implicitly) introduced in [50] and has since played a useful role in several recent results on halfspaces [43], [37], [16]. Briefly, assuming that the weights $w_1, \dots, w_n$ are sorted by absolute value, the ϵ-critical index is the first index ℓ so that the weight vector $(w_\ell, w_{\ell+1}, \dots, w_n)$ is ϵ-regular. The previous Step 1 handled halfspaces that are regular, corresponding to $\ell = 1$. We now proceed by analyzing two cases, based on whether or not $1 < \ell < L(\epsilon)$, or $\ell \geq L(\epsilon)$, for $L(\epsilon) := \tilde{O}(1/\epsilon^2)$. In both cases, it is convenient to think of the variables as partitioned into a "head" part consisting the first $L(\epsilon)$ variables and corresponding to the largest weights, and of a "tail" part consisting of the rest.

Step 2: Fooling halfspaces with small critical index ($\ell < L(\epsilon)$). We argue that for every setting of the head variables, the ϵ-regularity of the tail is sufficient to ensure that the overall halfspace gives the right bias. More precisely, assume that $\mathcal{D}$ is $(K(\epsilon) + L(\epsilon))$-wise independent, and note that each setting of the ℓ head variables gives an ϵ-regular halfspace $\mathrm{sign}(w \cdot x - \theta')$ over the tail variables (with the constant θ' depending on the values of the head variables). Since the marginal distribution on the tail variables is $K(\epsilon)$-wise independent for every setting of the head variables, the distribution $\mathcal{D}$ fools all such halfspaces.

Step 3: Fooling halfspaces with large critical index ($\ell \geq L(\epsilon)$). In this case, we show that the setting of the head variables alone is very likely to determine the value of the halfspace *by a large margin*. More precisely, we show that a uniform random assignment to the head variables is very likely to yield a halfspace $\mathrm{sign}(w_T \cdot x_T - \theta')$ over the tail variables T where $|\theta'| > \|w_T\|_2/\epsilon$. As long as the tail variables are pairwise independent, Chebyshev's inequality implies that the value $w_T \cdot x_T$ will be sharply concentrated within $[-\|w_T\|_2, +\|w_T\|_2]$. So, for most settings of the head variables, we get something very close to a constant function over the tail variables. Since a $(L(\epsilon) + 2)$-wise independent distribution gives uniform randomness for the head variables and pairwise independence for the tail variables, bounded independence fools these halfspaces as well.

The idea behind the proof of the large margin property is that up to the critical index ℓ – which in this case is large ($\ell \geq L(\epsilon)$) – the weights $(w_1, \dots, w_{\ell-1})$ must be decreasing fairly rapidly; this implies strong anti-concentration for the distribution of θ', which yields large margin with good probability.

The amount of independence required for all three steps to work is $\max\{K(\epsilon), K(\epsilon) + L(\epsilon), L(\epsilon) + 2\} = \tilde{O}(1/\epsilon^2)$.

Univariate approximations to the sign **function.** As mentioned above, our approach relies on the existence of low-degree univariate sandwich approximators to the sign function under the normal distribution on $\mathbf{R}$. Low-degree approximations to the sign function have been studied in both computer science and mathematics (see for instance [44], [18], [29] and the references therein). However it appears that these results do not fit all our requirements. Below we discuss how our approach relates to the work of Benjamini *et al.* [7] and Eremenko and Yuditskii [18].

Benjamini *et al.* prove that $O(1/\epsilon^2)$-wise independence suffices to fool the majority function, using machinery from the theory of the classical moment problem. However, their proof seems to be tailored quite specifically to the majority function, where the moments can be understood in terms of Krawtchouk polynomials and known bounds on such polynomials can be applied, so it seems difficult to extend their approach to general halfspaces (or indeed even to slight variants of the majority function).

Bazzi's condition on the existence of sandwiching polynomials mentioned above is in fact both necessary and sufficient for all k-wise independent distributions to fool a function f. Thus the result of [7] implies the existence of $O(1/\epsilon^2)$-degree multivariate sandwich polynomials for the majority function; symmetrization then implies that there exist univariate polynomials which, roughly speaking, provide good sandwich approximation to the function $\mathrm{sign}(t)$ under the binomial distribution. This is similar in spirit to the result we establish (mentioned in Step 1 above) about univariate polynomial approximators, but there is a crucial difference: since the binomial distribution is supported only on the integers $\{-n, \dots, n\}$, it seems difficult to infer much about the behavior of the univariate polynomial on values outside of $\{-n, \dots, n\}$. Hence, it is unclear whether these polynomials can be used for general (or even regular) halfspaces.

In contrast, we work with the *best possible* pointwise approximation to the function $\mathrm{sign}(t)$ on the (piecewise) *continuous* domain $[-1, -a] \cup [a, 1]$. This uniform error bound is convenient for dealing with regular halfspaces; moreover, working with the optimal pointwise approximator allows us to exploit various properties of optimal approximators that follow from the theory of Chebyshev approximation, in a way that is crucial for us to obtain the required "univariate sandwich approximators."

We note that a recent work in approximation theory [18] analyzes the error achieved by this optimal polynomial and in particular establishes the limiting behavior of the error, using tools from complex analysis. For our purposes, though, we require the error to converge to the limit fairly rapidly and it is unclear whether the results of [18] guarantee this.

We present an error analysis which is elementary (it only uses basic approximation theory) and moreover matches the limiting bounds of [18] up to a constant factor.

Finally, we briefly discuss some other work on polynomial approximations to halfspaces, a topic that has been studied extensively, motivated by applications to complexity theory and computational learning [41], [44], [28], [27], [29]. Nisan and Szegedy showed that the n-variable OR function has a pointwise (ℓ_∞) approximation of degree $O(\sqrt{n})$ [41], and Paturi showed that such approximations to Majority require degree $\Omega(n)$. A beautiful theorem by Peres shows that halfspaces have noise sensitivity $O(\sqrt{\epsilon})$ [46], improving on an $O(\epsilon^{1/4})$ bound due to Benjamini *et al.* [8]. Klivans *et al.* used this to show that every halfspace has an ϵ-approximation in ℓ_2 of degree $O(\epsilon^{-2})$ [28]. We note that while low-degree ℓ_2 approximations do imply the existence of low-degree ℓ_1 approximations, Benjamini *et al.* [7] showed that they do not imply the existence of sandwich approximations: indeed, recursive Majorities of depth 2 have ℓ_2 approximations of degree $O(\epsilon^{-4})$ but require degree $\Omega(\sqrt{n})$ for sandwich approximations. Thus this paper's results do not follow from the $O(\epsilon^{-2})$-degree ℓ_2 approximators of [28].

Organization. In Section 2 we record some useful probabilistic facts. In Sections 3 and 4 we show how to fool regular halfspaces. First, we discuss how a certain univariate polynomial approximator to $\mathrm{sign}(t)$ yields low-degree sandwich polynomials for regular halfspaces, then in Section 3.1 we construct the required univariate polynomial, and finally in Section 4 we put everything together. Due to space restrictions, the paper does not contain details about fooling non-regular halfspaces. This can be found in the full version of the paper.

2. Probability Background

We will need the Berry-Esséen theorem, a version of the Central Limit Theorem with explicit error bounds.

Theorem 2.1. *(Berry-Esséen) Let $X_1, \ldots, X_n$ be a sequence of independent random variables satisfying $\mathbf{E}[X_i] = 0$ for all i, $\sqrt{\sum_i \mathbf{E}[X_i^2]} = \sigma$, and $\sum_i \mathbf{E}[|X_i|^3] = \rho_3$. Let $S = (X_1 + \cdots + X_n)/\sigma$ and let F denote the cumulative distribution function (cdf) of S. Then $\sup_x |F(x) - \Phi(x)| \le \rho_3/\sigma^3$, where Φ is the cdf of a standard Gaussian random variable (with mean zero and variance one).*

Corollary 2.2. *Let $x_1, \ldots, x_n$ denote independent uniformly ± 1 random signs and let $w_1, \ldots, w_n \in \mathbf{R}$. Write $\sigma = \sqrt{\sum_i w_i^2}$, and assume $|w_i|/\sigma \le \tau$ for all i. Then for any interval $[a, b] \subseteq \mathbf{R}$, $\left|\Pr[a \le w_1x_1 + \cdots + w_nx_n \le b] - \Phi([\frac{a}{\sigma}, \frac{b}{\sigma}])\right| \le 2\tau$, where $\Phi([c, d]) := \Phi(d) - \Phi(c)$. In particular,*

$$\Pr[a \le w_1x_1 + \cdots + w_nx_n \le b] \le \frac{|b-a|}{\sigma} + 2\tau.$$

Theorem 2.3 (Hoeffding)**.** *For any $w \in \mathbf{R}^n$, $\gamma > 0$, we have $\Pr_{x\leftarrow\mathcal{U}}[w \cdot x \ge \gamma\|w\|] \le e^{-\gamma^2/2}$.*

3. Fooling regular halfspaces

Throughout this paper we assume without loss of generality that halfspaces are normalized to satisfy $\sum_i w_i^2 = 1$. Such a representation can always be obtained by appropriate scaling.

Definition 3.1 (Regular Halfspace)**.** A halfspace f is said to be *ϵ-regular* if it can be expressed as $f(x) = \mathrm{sign}(w \cdot x - \theta)$ where for all $i = 1, \ldots, n$, we have $|w_i| \le \epsilon$.

An ϵ-regular halfspace $f(x) = \mathrm{sign}(w \cdot x - \theta)$ has the convenient property that the cumulative distribution function (cdf) of $w \cdot x - \theta$ is everywhere within $\pm O(\epsilon)$ of the cdf of the shifted Gaussian $N(-\theta, 1)$. This is a direct consequence of the Berry-Esséen theorem (Theorem 2.1). In this section we show how to fool regular halfspaces. Given $\epsilon > 0$, we define the following parameters:

$$a(\epsilon) := \frac{\epsilon^2}{C\log(1/\epsilon)},$$
$$K(\epsilon) := \frac{4c\log(\frac{1}{\epsilon})}{a} + 2 < \frac{5c}{a}\log(1/\epsilon) = O\left(\log^2(1/\epsilon)/\epsilon^2\right).$$

We assume without loss of generality that ϵ is a sufficiently small power of 2 (i.e., $\epsilon = 2^{-i}$ for some integer i). The positive constants C and c will be chosen later; but (with foresight), we will require that $C \gg c$.

Theorem 3.2 (Fooling ϵ-regular halfspaces)**.** *Any $K(\epsilon)$-wise independent distribution fools ϵ-regular halfspaces with error 12ϵ.*

To prove the theorem we construct certain "sandwiching" polynomials. We now define such polynomials and then explain why they are sufficient for our purposes.

Definition 3.3. Let $f : \{-1, +1\}^n \to \{-1, +1\}$ be a Boolean function. A pair of real-valued polynomials $q_\ell(x_1, \ldots, x_n)$, $q_u(x_1, \ldots, x_n)$ are said to be *ϵ-sandwich polynomials of degree k for f* if they have the following properties:

- $\deg(q_u), \deg(q_\ell) \le k$;
- $q_u(x) \ge f(x) \ge q_\ell(x)$ for all $x \in \{-1, +1\}^n$;
- $\mathbf{E}_{x\leftarrow\mathcal{U}}[q_u(x) - f(x)] \le \epsilon$ and $\mathbf{E}_{x\leftarrow\mathcal{U}}[f(x) - q_\ell(x)] \le \epsilon$.

The following fact proved via LP-duality relates sandwiching polynomials to fooling [5]. We only use the "if" direction of this lemma, which follows easily by linearity of expectation.

Lemma 3.4 (Bazzi)**.** *Let $f : \{-1, +1\}^n \to \{-1, +1\}$ be a Boolean function. Every k-wise independent distribution ϵ-fools f if and only if there exist ϵ-sandwich polynomials of degree k for f.*

The crux of our construction of sandwiching polynomials for regular halfspaces is good univariate approximations to the sign function:

Theorem 3.5. Let $0 < \epsilon < 0.1$ and let a and K be as defined above. There is a univariate polynomial $P(t)$ such that $\deg(P) \le K$ with the following properties:

(1) $P(t) \ge \text{sign}(t) \ge -P(-t)$ for all $t \in \mathbf{R}$;
(2) $P(t) \in [\text{sign}(t), \text{sign}(t) + \epsilon]$ for $t \in [-1/2, -2a] \bigcup [0, 1/2]$;
(3) $P(t) \in [-1, 1+\epsilon]$ for $t \in (-2a, 0)$;
(4) $|P(t)| \le 2 \cdot (4t)^K$ for all $|t| \ge 1/2$.

Property (1) says that $P(t)$ is an upper sanwdich to the sign function. By property (2), P gives a pointwise approximation with error ϵ in the interval $[-1/2, 1/2]$, except for the interval $[-2a, 0]$ where it has error at most $2 + \epsilon$ by property (3). For $t \ge \frac{1}{2}$, property (4) bounds how rapidly $P(t)$ grows. For a qualitative depiction of P we refer the reader to Figure 1 (this figure is not an actual plot, it is intended to illustrate the behavior of P on various intervals; also the parameter $1/2$ is replaced by $1 - a \ge 1/2$ for later needs). Before constructing P, we outline the proof of Theorem 3.2 using the polynomial P; the full proof is in Section 4.

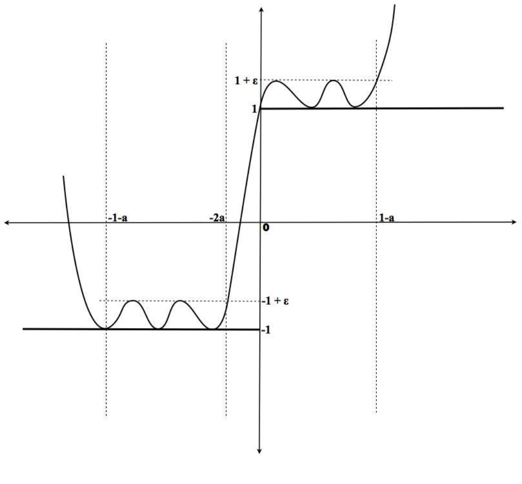

Figure 1. Qualitative plot of polynomial P.

Overview of the proof of Theorem 3.2.: Let $h(x) = \text{sign}(w \cdot x - \theta)$ be an ϵ-regular halfspace, and assume that $|\theta|$ is small (the case where $|\theta|$ is large is simpler). Let us define

$$t := \frac{w \cdot x - \theta}{Z}$$

where we choose the scaling factor Z to be $\tilde{\Theta}(\epsilon^{-1})$. We use $q_u(x) = P(t)$ and $q_\ell(x) = -P(-t)$ as the upper and lower sandwich polynomials respectively. The sandwiching property is easy to verify, the crux is to bound $\mathbf{E}_x[q_u(x) - h(x)]$. We do this by case analysis.

(1) If t lies in the interval $[-2a, 0]$ then, although the error $q_u(x) - h(x)$ may be large, by our choice of Z it must be the case that $w \cdot x$ lands in an interval of length $O(\epsilon)$. By the anti-concentration of $w \cdot x$ (which is a consequence of the ϵ-regularity of w), this only happens with probability $O(\epsilon)$. Thus the contribution to $\mathbf{E}_x[q_u(x) - h(x)]$ from this event is $O(\epsilon)$.

(2) In the event that t lies in $[-1/2, 1/2] \setminus [-2a, 0]$, the pointwise error $q_u(x) - h(x)$ is at most ϵ because, by Property (2), P gives a good pointwise approximation to the sign function in this range. So this event contributes at most $O(\epsilon)$ to $\mathbf{E}_x[q_u(x) - h(x)]$.

(3) Finally, the event that the input t has absolute value bigger than $1/2$ corresponds to the event that $|w \cdot x - \theta| \ge Z/2$. Since $\sum_i w_i^2 = 1$, $|\theta|$ is small, and Z is $\tilde{\Theta}(\epsilon^{-1})$, we can bound this probability using the Hoeffding bound. In this event, the pointwise error is large but we can bound it from above using Property (4). Our choice of parameters ensures that the Hoeffding bound dominates the growth of the polynomial P, so that the contribution to $\mathbf{E}_x[q_u(x) - h(x)]$ is again at most $O(\epsilon)$.

Thus, overall $\mathbf{E}_x[q_u(x) - h(x)] = O(\epsilon)$. One can similarly bound the error of q_ℓ.

3.1. Constructing P

This section contains our proof of Theorem 3.5. The key step is to exhibit a low-degree univariate polynomial that approximates $\text{sign}(t)$ well when $|t| \in [a, 1]$ and is well-behaved even for larger values of $|t|$ to be compatible with the sandwich condition. We phrase this as a problem in univariate approximation. The solution we use is a low-degree polynomial $p(t)$ which is an optimal pointwise approximator to $\text{sign}(t)$ on $[-1, -a] \cup [a, 1]$. Such an optimal polynomial exists and we prove that it is well-behaved for large $|t|$, using ideas from classical approximation theory. However, it seems difficult to construct this polynomial explicitly and bound its error.

Recent work by [18] analyzes the error achieved by such a polynomial and in particular establishes the limiting behavior of the error function. For our purposes, though, we require the error to converge to the limit fairly rapidly and it is unclear whether the results of [18] guarantee this.

Instead, we bound the error by constructing a small error approximator $q(t)$ using Jackson's theorem together with standard amplification ideas. While $q(t)$ might not be well-behaved for large values of t, we only use it to bound from above the error of $p(t)$ on $[-1, -a] \cup [a, 1]$. Our approach has the advantage of being fairly elementary (using only standard ingredients from basic approximation theory) and matches the limiting bounds of [18] up to a constant factor.

For a bounded continuous function $f : [-1, 1] \to \mathbf{R}$, we define its *modulus of continuity* $\omega_f(\delta)$ as

$$\omega_f(\delta) := \sup\{|f(x) - f(y)| : x, y \in [-1, 1]; |x - y| \le \delta\}.$$

A classical result of Dunham Jackson from the early twentieth century bounds the error of the best degree-ℓ approximation to f.

Theorem 3.6. (Jackson's Theorem) *[11, Page 104], [12]. For f as above and any integer $\ell \geq 1$, there exists a polynomial $J(t)$ with $\deg(J) \leq \ell$ so that*

$$\max_{t\in[-1,1]} |J(t) - f(t)| \leq 6\omega_f\left(\frac{1}{\ell}\right).$$

Recall the parameter $a = \frac{\epsilon^2}{C\log(1/\epsilon)}$. We now define $m := \frac{c\log(1/\epsilon)}{a}$. It will be crucial for us that m is even (see in particular the last paragraph in the proof of Theorem 3.10.); for this condition to be satisfied, it is of course enough that c is even. (We also note that the parameters K and m are such that $K = 4m + 2$.)

Lemma 3.7. *For a, m as above, there is a polynomial $q(t)$ of degree at most $2m$ such that*

$$\max_{|t|\in[a,1]} |q(t) - \operatorname{sign}(t)| \leq \epsilon^2.$$

Proof: Define the piecewise linear continuous function $f(t)$ as

$$f(t) = \begin{cases} \operatorname{sign}(t) & a \leq |t| \leq 1 \\ t/a & |t| \leq a. \end{cases}$$

Thus $f(t)$ increases linearly from -1 to 1 in the range $[-a, a]$. A simple calculation yields that $\omega_f(\frac{1}{\ell}) = 1/(a\ell)$. Taking $\ell \geq 25/a$, Jackson's theorem gives a polynomial $J(t)$ of degree at most ℓ such that

$$\max_{a\leq|t|\leq 1} |J(t) - \operatorname{sign}(t)| \leq \max_{t\in[-1,1]} |J(t) - f(t)| \leq \frac{6}{a\ell} < \frac{1}{4}.$$

Our goal is to bring the error down to ϵ^2. Rather than using Jackson's theorem for this (which would require degree $\tilde{O}(\epsilon^{-4})$), we use the degree-k amplifying polynomial

$$A_k(u) := \sum_{j\geq\frac{k}{2}} \binom{k}{j} \left(\frac{1+u}{2}\right)^j \left(\frac{1-u}{2}\right)^{k-j}. \qquad (1)$$

This polynomial has the following properties (easily proved via elementary calculation and also following from the Chernoff bound):

Claim 3.8. *The polynomial $A_k(u)$ satisfies:*

1) *If $u \in [3/5, 1]$, then $2A_k(u) - 1 \in [1 - 2e^{-k/6}, 1]$.*
2) *If $u \in [-1, -3/5]$, then $2A_k(u) - 1 \in [-1, -1 + 2e^{-k/6}]$.*

We define the polynomial

$$q(t) := 2A_k\left(\frac{4}{5}J(t)\right) - 1$$

where $k = 15\log(1/\epsilon)$. Scaling $J(t)$ by $\frac{4}{5}$ ensures that the argument to A_k lies in the range $[-1, -3/5] \cup [3/5, 1]$ whenever $|t| \in [a, 1]$. Applying Claim 3.8 with $k = 15\log(1/\epsilon)$ gives

$$\max_{|t|\in[a,1]} |q(t) - \operatorname{sign}(t)| < 2e^{-k/6} < \epsilon^2.$$

Finally, by selecting c large enough, we have

$$\begin{aligned} \deg(q) &\leq \deg(J)\deg(A_k) \\ &\leq \frac{25}{a} \cdot 15\log(1/\epsilon) < \frac{2c}{a}\log(1/\epsilon) = 2m. \end{aligned}$$

■

We use Chebyshev's classical theorem on (weighted) real polynomial approximation.

Theorem 3.9. (Chebyshev's Theorem) *[1, Page 55]. Let $f : [a, b] \to \mathbf{R}$ be a continuous function. Let $s : [a, b] \to \mathbf{R}$ be a continuous function that does not vanish on $[a, b]$. The polynomial $r(z)$ of degree m that minimizes*

$$M(m) = \max_{z\in[a,b]} |f(z) - s(z)r(z)|$$

is unique, and it is characterized by the property that there exist $m + 2$ points $a \leq z_0 < z_1 \cdots < z_{m+1} \leq b$ such that for each z_i

$$M(m) = |f(z_i) - s(z_i)r(z_i)|$$

and the sign of the error at the z_i's alternates.

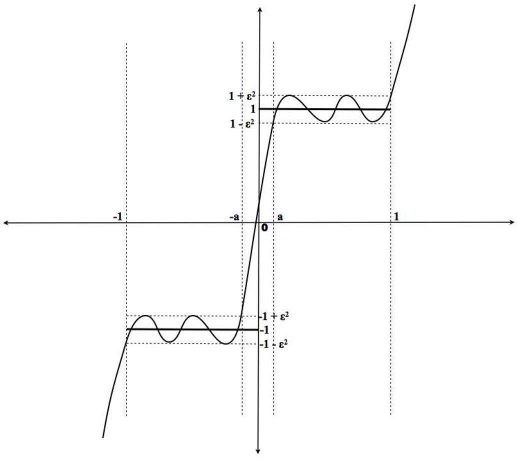

Figure 2. Qualitative representation of polynomial p.

We now present the "well-behaved" polynomial $p(t)$ mentioned at the beginning of this section. To help the reader visualize $p(t)$, we provide a schematic representation in Figure 2. (As before, this figure is not an actual plot, but rather is intended to illustrate the behavior of p on various intervals.)

Theorem 3.10. *Let a and m be as previously specified. There is a univariate polynomial $p(t)$ where $\deg(p) \leq 2m + 1$ such that:*

1) *$p(t) \in [\operatorname{sign}(t) - \epsilon^2, \operatorname{sign}(t) + \epsilon^2]$ for all $|t| \in [a, 1]$;*
2) *$p(t) \in [-(1 + \epsilon^2), 1 + \epsilon^2]$ for all $t \in [-a, a]$;*
3) *$p(t)$ is monotonically increasing on the intervals $(-\infty, -1]$ and $[1, \infty)$.*

Proof: Intuitively, the polynomial p is the "best possible" approximator to the function sign. However, some care is required because the function sign is not continuous. We present an analysis that assumes no background in approximation theory.

Invoking Theorem 3.9, let $r(z)$ be the polynomial of degree m that minimizes

$$\max_{z\in[a^2,1]} |\sqrt{z}r(z)-1|.$$

Define $p(t) := t\cdot r(t^2)$.

Bounding the error of $p(t)$ for $|t|\in[a,1]$: A polynomial $p^*(t)$ is *odd* if the coefficients of the even powers of t are 0; it can be written as $p^*(t)=t\cdot r^*(t^2)$. Note that

$$\begin{aligned} \max_{|t|\in[a,1]} |p^*(t)-\mathrm{sign}(t)| &= \max_{|t|\in[a,1]} |t\cdot r^*(t^2)-\mathrm{sign}(t)| \\ &= \max_{z\in[a^2,1]} |\sqrt{z}\cdot r^*(z)-1|. \end{aligned}$$

By Theorem 3.6 there exists a polynomial $p^*(t)$ of degree $2m \le 2m+1$ such that

$$\max_{|t|\in[a,1]} |p^*(t)-\mathrm{sign}(t)| \le \epsilon^2.$$

We can assume that $p^*(t)$ is odd, for else we can replace it by the odd polynomial $(p^*(t)-p^*(-t))/2$ whose error is no worse. Therefore we can write $p^*(t)=t\cdot r^*(t^2)$. Using Equation 2, the definition of r, and the property of p^* above, we can now bound the error of p as follows:

$$\begin{aligned} &\max_{|t|\in[a,1]} |p(t)-\mathrm{sign}(t)| = \max_{z\in[a^2,1]} |\sqrt{z}\cdot r(z)-1| \\ &\le \max_{z\in[a^2,1]} |\sqrt{z}\cdot r^*(z)-1| \le \max_{|t|\in[a,1]} |p^*(t)-\mathrm{sign}(t)| \le \epsilon^2. \end{aligned}$$

This concludes the proof of Property (1).

Other properties of p: By Theorem 3.9 we find that there is a sequence of points

$$a^2 \le z_0 < z_1 \ldots < z_{m+1} \le 1$$

so that the error $\sqrt{z}r(z)-1$ achieves its maximum magnitude exactly at the points z_i, and the sign of the error alternates. Set $t_i=\sqrt{z_i}>0$ so that $a\le t_0<t_1\ldots<t_{m+1}\le 1$. Let $\phi(t)$ be the error function $\phi(t)=p(t)-\mathrm{sign}(t)$. Note that for $t\ge a$, we have

$$\begin{aligned} &\phi(t) = p(t)-1, \text{ and} \\ &\phi(-t) = p(-t)-(-1) = -p(t)+1 = -\phi(t). \end{aligned}$$

In particular, for each t_i we have $|\phi(t_i)|=|\phi(-t_i)|$.

Now consider the interval $[a,1]$, on which $\phi(t)=p(t)-1$. Note that $\phi'(t)$ is well defined and equals $p'(t)$ at any point in $(a,1)$. The points $t_1,\ldots,t_m$ lie in $(a,1)$ and they are local maxima/minima, since $\phi(t)$ cannot increase in magnitude in the neighborhood of t_i. Thus $\phi'(t_i)=p'(t_i)=0$ for each $i\in[m]$. Similarly, we can show that $\phi'(-t_i)=p'(-t_i)=0$ for $i\in[m]$. But $\deg(p')$ is at most $2m$, and so we have located all its roots. As we now show, this allows us to determine the sign of p in the intervals $[-\infty,-1],[-a,a]$ and $[1,\infty]$.

Note that $p(t_1)$ is close to 1 whereas $p(-t_1)$ is close to -1, and thus p increases monotonically in the interval $(-t_1,t_1)$ which includes $[-a,a]$. This gives Property (2). Also t_1 is a local maximum for p, which shows that the t_i's are maxima when i is odd, and minima when i is even. Thus, since m is even, $p(t_m)$ is a local minimum, so $p(t)$ increase monotonically in the range (t_m,∞), which includes $[1,\infty)$. Since $p(t)$ is odd, this also implies that $p(t)$ is monotonically increasing in the range $(-\infty,-t_m)$ which contains $(-\infty,-1]$. This gives Property (3). ■

Using the polynomial $p(t)$, we construct the polynomial $P(t)$ which is a good "upper" approximator to $\mathrm{sign}(t)$ (i.e. $P(t)\ge\mathrm{sign}(t)$ for all t), completing the proof of Theorem 3.5.

Proof of Theorem 3.5: Let p denote the polynomial of degree $2m+1$ from Theorem 3.10. Consider the following polynomial:

$$P(t) = \frac{1}{2}(1+\epsilon^2+p(t+a))^2-1.$$

Note that $\deg(P)=2\deg(p)\le K$. We now consider the behavior of P on the relevant intervals. We repeatedly use the inequality $\frac{1}{2}(2+2\epsilon^2)^2-1=1+4\epsilon^2+2\epsilon^4\le 1+\epsilon$ which holds since $\epsilon<\frac{1}{10}$. Note that $P(t)\ge -1$ holds for all t. We now analyze the behavior of $P(t)$ interval by interval:

(a) $t\in[-1-a,-2a]$. Here $p(t+a)\in[-1-\epsilon^2,-1+\epsilon^2]$, hence $P(t)\in[-1,-1+\epsilon]$.
(b) $t\in(-2a,0)$. Here $p(t+a)\in[-1-\epsilon^2,1+\epsilon^2]$, hence $P(t)\in[-1,1+\epsilon]$.
(c) $t\in[0,1-a]$. Here $p(t+a)\in[1-\epsilon^2,1+\epsilon^2]$, hence $P(t)\in[1,1+\epsilon]$.
(d) $t\in(1-a,\infty]$. Here $p(t+a)\ge 1-\epsilon^2$, hence $P(t)\ge 1$.

This shows that $P(t)\ge\mathrm{sign}(t)$ for all $t\in\mathbf{R}$. Thus we also have

$$P(-t)\ge\mathrm{sign}(-t) \;\Rightarrow\; \mathrm{sign}(t)\ge -P(-t)$$

which establishes Property (1). Properties (2) and (3) follow immediately from (a), (b) and (c) above.

For Property (4), we use the following standard fact from approximation theory.

Fact 3.11. *[11, Page 61], [49]. Let $a(t)$ be a polynomial of degree at most d for which $|a(t)|\le b$ in the interval $[-1,1]$. Then $|a(t)|\le b|2t|^d$ for all $|t|\ge 1$.*

Taking $a(t)$ to be $P(t/2)$, properties (2) and (3) give us that $|P(t/2)|\le 2$ for $t\in[-1,1]$. So the fact gives $|P(t/2)|<2|2t|^{4m+2}$ for $|t|\ge 1$, i.e. $|P(t)|<2|4t|^{4m+2}$ for $|t|\ge 1/2$. Theorem 3.5 is proved. ■

4. Proof of Theorem 3.2

In this section we prove Theorem 3.2: any $K(\epsilon)$-wise independent distribution fools ϵ-regular halfspaces with error 12ϵ. In light of Lemma 3.4, it is sufficient to exhibit sandwiching polynomials. For this, we use our univariate polynomial approximator P from the previous section.

Let $h(x) = \mathrm{sign}(w \cdot x - \theta)$ be an ϵ-regular halfspace (and recall $\sum_i w_i^2 = 1$.) Let

$$Z := \frac{\epsilon}{2a} = \frac{C \log(1/\epsilon)}{2\epsilon}.$$

We break the analysis into the following two cases, based on the magnitude of the threshold θ.

4.1. $|\theta|$ is small ($|\theta| \leq Z/4$)

The sandwich polynomials we use are:

$$q_u(x) := P\left(\frac{w \cdot x - \theta}{Z}\right), \quad q_l(x) := -P\left(\frac{\theta - w \cdot x}{Z}\right). \tag{2}$$

First, observe that for every $x \in \{-1, +1\}^n$ we have

$$q_u(x) \geq h(x) \geq q_l(x).$$

This is because from Theorem 3.5 with $t = (w \cdot x - \theta)/Z$ we get

$$q_u(x) \geq \mathrm{sign}\left(\frac{w \cdot x - \theta}{Z}\right) = \mathrm{sign}(w \cdot x - \theta) = h(x) \geq q_l(x).$$

In the rest of this section we bound the error of the approximation.

Lemma 4.1. $\mathbf{E}_x[q_u(x) - h(x)] < 10\epsilon$.

Proof: Define the random variable $H(x) = (w \cdot x - \theta)/Z$. We prove the desired upper bound by partitioning the space into three events and bounding the contribution from each:

1) S_1 is the event that $H(x) \in [-\epsilon/Z, 0]$.
2) S_2 is the event that $|H(x)| \leq 1/2$, but S_1 does not happen.
3) S_3 is the event that $|H(x)| > 1/2$.

We have

$$\mathbf{E}_x[q_u(x) - h(x)] = \sum_{i=1}^{3} \Pr_x[S_i]\, \mathbf{E}_x[q_u(x) - h(x) | S_i].$$

Case 1: In this case, the pointwise error is moderate – at most $(2 + \epsilon)$ – and we use gaussian anti-concentration to argue that the event has small probability mass. The event $H(x) \in [-\epsilon/Z, 0]$ implies that

$$\begin{aligned} \frac{w \cdot x - \theta}{Z} \in [-2a, 0] \;&\Rightarrow\; q_u(x) \leq 1 + \epsilon \\ &\Rightarrow\; q_u(x) - h(x) \leq 2 + \epsilon, \end{aligned}$$

using Item (3) in Theorem 3.5.

Since h is ϵ-regular, from Corollary 2.2 it follows that $\Pr_x[H(x) \in [-\epsilon/Z, 0]] \leq 3\epsilon$. So,

$$\Pr_x[S_1]\, \mathbf{E}_x[q_u(x) - h(x) | S_1] \leq (2 + \epsilon) \cdot 3\epsilon < 8\epsilon.$$

Case 2: This event has high probability, but in this range we get good pointwise approximation. The event S_2 implies that

$$\begin{aligned} H(x) \in [-1/2, 1/2] \setminus [-2a, 0] \;&\Rightarrow\; q_u(x) \leq h(x) + \epsilon \\ &\Rightarrow\; q_u(x) - h(x) \leq \epsilon, \end{aligned}$$

where we used Item (2)in Theorem 3.5. So,

$$\Pr_x[S_2]\, \mathbf{E}_x[q_u(x) - h(x) | S_2] \leq 1 \cdot \epsilon \leq \epsilon.$$

Case 3: Here we trade off the large magnitude of error (Item (4) in Theorem 3.5) with the small probability of the event (bounded by the Hoeffding bound). Define the intervals

$$\begin{aligned} I_j^+ &= \left[\frac{j}{2}, \frac{(j+1)}{2}\right) \text{ for } j = 1, 2, \ldots \\ I_k^- &= \left(\frac{-(k+1)}{2}, \frac{-k}{2}\right] \text{ for } k = 1, 2, \ldots. \end{aligned}$$

We can write

$$\begin{aligned} &\Pr_x[S_3]\, \mathbf{E}_x[q_u(x) - h(x) | S_3] = \\ &\quad \sum_{j \geq 1} \Pr_x[H(x) \in I_j^+]\, \mathbf{E}_x[q_u(x) - h(x) | H(x) \in I_j^+] \\ &\quad + \sum_{k \geq 1} \Pr_x[H(x) \in I_k^-]\, \mathbf{E}_x[q_u(x) - h(x) | H(x) \in I_k^-]. \end{aligned} \tag{3}$$

Fix any integer $j \geq 1$. If $H(x) \in I_j^+$, then

$$\frac{j}{2} \leq H(x) < \frac{j+1}{2}.$$

Recalling that we have $|P(t)| \leq 2 \cdot (4t)^K$ for $t \geq 1/2$, we get that

$$q_u(x) = P(H(x)) \leq 2(2j + 2)^K.$$

Since $h(x) = 1$, we get

$$q_u(x) - h(x) = q(x) - 1 \leq 2(2j + 2)^K - 1. \tag{4}$$

Next we bound $\Pr_x[H(x) \in I_j^+]$ using the Hoeffding bound.

$$\begin{aligned} \Pr[H(x) \in I_j^+] &\leq \Pr_x\left[w \cdot x - \theta \geq \frac{jZ}{2}\right] \\ &\leq \Pr_x\left[w \cdot x \geq \frac{jZ}{4}\right] \leq e^{-j^2 Z^2/32} \end{aligned} \tag{5}$$

where the second inequality uses the fact that $|\theta| \leq Z/4$.

The analysis of the intervals I_k^- is similar (except $h(x) = -1$). For $H(x) \in I_k^-$ we get

$$\begin{aligned} |H(x)| \leq \frac{k+1}{2} \;&\Rightarrow\; q_u(x) \leq 2(k+1)^K \\ &\Rightarrow\; q_u(x) - h(x) \leq 2(2k + 2)^K + 1. \end{aligned} \tag{6}$$

Similarly, the Hoeffding bound gives

$$\Pr[H(x) \in I_k^-] \leq \Pr_x\left[w \cdot x - \theta \leq \frac{-kZ}{2}\right]$$
$$\leq \Pr_x\left[w \cdot x \leq \frac{-kZ}{4}\right] \leq e^{-k^2Z^2/32}. \quad (7)$$

Plugging equations (4), (5), (6), (7) back into (3), we get

$$\Pr_x[S_3]\,\mathbf{E}_x[q_u(x) - h(x)|S_3] \leq$$
$$\sum_{j\geq 1} \frac{2(2j+2)^K - 1}{e^{j^2Z^2/32}} + \sum_{k \geq 1} \frac{2(2k+2)^K + 1}{e^{k^2Z^2/32}}$$
$$= 4 \sum_{j\geq 1} \frac{(2j+2)^K}{e^{j^2Z^2/32}} < 4 \sum_{j\geq 1} e^{j(2K - Z^2/32)},$$

where the last inequality follows by noting that, for $j \geq 1$, $(2j+2)^K < e^{2Kj}$ and $e^{j^2Z^2/32} \geq e^{jZ^2/32}$. But now observe that

$$2K - \frac{Z^2}{32} < \frac{C\log^2(1/\epsilon)}{\epsilon^2}\left(10c - \frac{C}{128}\right).$$

For a suitable choice of $C \gg c$, we have that $10c - C/128 \leq -1$, so

$$\Pr_x[S_3]\,\mathbf{E}_x[q_u(x) - h(x)|S_3] < 4 \sum_j e^{-jC\frac{\log^2(1/\epsilon)}{\epsilon^2}} < \epsilon.$$

Thus overall, we have $\mathbf{E}_x[q_u(x) - h(x)] \leq 10\epsilon$. ∎

The lower sandwich bound follows by symmetry:

Lemma 4.2. $\mathbf{E}_x[h(x) - q_l(x)] < 10\epsilon$.

Proof: Since $q_l(x) \leq h(x)$ for every x, we also have $-h(x) \leq -q_l(x)$. Thus

$$-q_l(x) = P\left(\frac{\theta - w \cdot x}{Z}\right)$$

is an upper sandwich for the function $-h(x) = \mathrm{sign}(\theta - w \cdot x)$. As this does not change the magnitude of θ, we can apply the analysis of Lemma 4.1 to conclude that
$\mathbf{E}_x[h(x) - q_l(x)] = \mathbf{E}_x[-q_l(x) - (-h(x))] < 10\epsilon$. ∎

4.2. $|\theta|$ is large ($|\theta| > Z/4$)

We assume for simplicity that $\theta \geq Z/4$ (the case when θ is negative is handled similarly). The sandwich polynomials we use are:

$$r_u(x) = P\left(\frac{w \cdot x - Z/4}{Z}\right), \quad r_l(x) = -1. \quad (8)$$

Lemma 4.3. *$h(x) \geq r_l(x)$ for all $x \in \{-1, +1\}^n$. Further,* $\mathbf{E}_x[h(x) - r_l(x)] \leq 2\epsilon$.

Proof: Note that $\mathbf{E}_x[h(x) - r_l(x)] = 2\Pr_x[h(x) = 1]$. For large enough C we have $\Pr_x[h(x) = 1] = \Pr_x[w \cdot x \geq \theta] < e^{-Z^2/32} < \epsilon$. ∎

Lemma 4.4. *$r_u(x) \geq h(x)$ for all $x \in \{-1, +1\}^n$. Further,* $\mathbf{E}_x[r_u(x) - h(x)] \leq 12\epsilon$.

Proof: Observe that $r_u(x)$ is the upper sandwich polynomial for the halfspace $h'(x) = \mathrm{sign}(w \cdot x - Z/4)$ as specified in Section 4.1. Thus we have $r_u(x) \geq h'(x) \geq h(x)$ hence

$$\mathbf{E}_x[r_u(x) - h(x)] = \mathbf{E}_x[r_u(x) - h'(x)] + \mathbf{E}_x[h'(x) - h(x)].$$

By Lemma 4.1, $\mathbf{E}_x[r_u(x) - h'(x)] \leq 10\epsilon$ whereas by the Hoeffding bound $\mathbf{E}_x[h'(x) - h(x)] \leq 2\epsilon$ which completes the proof. ∎

REFERENCES

[1] N. Achieser, *Theory of Approximation*. New York: Frederik Ungar Publishing Co, 1956.

[2] M. Ajtai and A. Wigderson, “Deterministic simulation of probabilistic constant depth circuits,” in *Proc. 26th IEEE Symposium on Foundations of Computer Science (FOCS)*, 1985, pp. 11–19.

[3] N. Alon, L. Babai, and A. Itai, “A fast and simple randomized algorithm for the maximal independent set problem,” *Journal of Algorithms*, vol. 7, pp. 567–583, 1986.

[4] N. Alon, O. Goldreich, J. Håstad, and R. Peralta, “Simple constructions of almost k-wise independent random variables,” *Random Structures & Algorithms*, vol. 3, no. 3, pp. 289–304, 1992.

[5] L. Bazzi, “Polylogarithmic independence can fool DNF formulas,” in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007)*. IEEE Computer Society, 2007, pp. 63–73.

[6] R. Beigel, “Perceptrons, PP, and the Polynomial Hierarchy,” *Computational Complexity*, vol. 4, pp. 339–349, 1994.

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

[8] I. Benjamini, G. Kalai, and O. Schramm, “Noise sensitivity of Boolean functions and applications to percolation,” *Inst. Hautes Études Sci. Publ. Math.*, vol. 90, pp. 5–43, 1999.

[9] A. Bogdanov and E. Viola, “Pseudorandom bits for polynomials.” *SIAM J. Comp.*, to appear. In *48th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2007, pp. 41–51.

[10] M. Braverman, “Poly-logarithmic independence fools AC^0 circuits,” in *Proc. 24th Annual IEEE Conference on Computational Complexity (CCC)*, 2009, pp. 3–8.

[11] N. Carothers. A short course on approximation theory, 1998. Available at http://personal.bgsu.edu/∼carother/Approx.html.

[12] E. Cheney, *Introduction to approximation theory*. New York, New York: McGraw-Hill, 1966.

[13] B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, and R. Smolensky, “The bit extraction problem and t-resilient functions,” in *26th Annual Symposium on Foundations of Computer Science*, IEEE, 1985, pp. 396–407.

[14] B. Chor and O. Goldreich, “On the power of two-point based sampling,” *Journal of Complexity*, vol. 5, no. 1, pp. 96–106, Mar. 1989.

[15] M. Dertouzos, *Threshold logic: a synthesis approach*. Cambridge, MA: MIT Press, 1965.

[16] I. Diakonikolas and R. Servedio, “Improved approximation of linear threshold functions,” in *Proc. 24th Annual IEEE Conference on Computational Complexity (CCC)*, 2009, pp. 161–172.

[17] P. Dubey and L. Shapley, "Mathematical properties of the banzhaf power index," *Mathematics of Operations Research*, vol. 4, pp. 99–131, 1979.

[18] A. Eremenko and P. Yuditskii, "Uniform approximation of $\mathrm{sgn}(x)$ by polynomials and entire functions," *J. d'Analyse Math.*, vol. 12, pp. 313–324, 2007.

[19] J. Forster, M. Krause, S. Lokam, R. Mubarakzjanov, N. Schmitt, and H.-U. Simon, "Relations between communication complexity, linear arrangements, and computational complexity," in *FSTTCS*, 2001, pp. 171–182.

[20] M. Goldmann, J. Håstad, and A. Razborov, "Majority gates vs. general weighted threshold gates," *Computational Complexity*, vol. 2, pp. 277–300, 1992.

[21] M. Goldmann and M. Karpinski, "Simulating threshold circuits by majority circuits," *SIAM Journal on Computing*, vol. 27, no. 1, pp. 230–246, 1998.

[22] P. Gopalan and J. Radhakrishnan, "Finding duplicates in a data stream," in *Proc. 20th Annual Symposium on Discrete Algorithms (SODA)*, 2009, pp. 402–411.

[23] A. Hajnal, W. Maass, P. Pudlak, M. Szegedy, and G. Turan, "Threshold circuits of bounded depth," *Journal of Computer and System Sciences*, vol. 46, pp. 129–154, 1993.

[24] J. Håstad, "On the size of weights for threshold gates," *SIAM Journal on Discrete Mathematics*, vol. 7, no. 3, pp. 484–492, 1994.

[25] S. Hu, *Threshold Logic*. University of California Press, 1965.

[26] J. Isbell, "A Counterexample in Weighted Majority Games," *Proc. of the AMS*, vol. 20, no. 2, pp. 590–592, 1969.

[27] A. Kalai, A. Klivans, Y. Mansour, and R. Servedio, "Agnostically learning halfspaces," in *Proc. of the 46th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005, pp. 11–20.

[28] A. Klivans, R. O'Donnell, and R. Servedio, "Learning intersections and thresholds of halfspaces," in *Proc. of the 43rd Annual Symposium on Foundations of Computer Science*, 2002, pp. 177–186.

[29] A. R. Klivans and A. A. Sherstov, "A lower bound for agnostically learning disjunctions," in *Conference on Learning Theory (COLT'07)*, 2007, pp. 409–423.

[30] M. Krause, "Geometric arguments yield better bounds for threshold circuits and distributed computing," in *Proc. 6th Structure in Complexity Theory Conference*, 1991, pp. 314–322.

[31] M. Krause and S. Waack, "Variation ranks of communication matrices and lower bounds for depth two circuits having symmetric gates with unbounded fanin," in *Proc. 32nd IEEE Symposium on Foundations of Computer Science (FOCS)*, 1991, pp. 777–782.

[32] P. Lewis and C. Coates, *Threshold Logic*. New York, Wiley, 1967.

[33] N. Linial and N. Nisan, "Approximate inclusion-exclusion," *Combinatorica*, vol. 10, no. 4, pp. 349–365, 1990.

[34] S. Lovett, "Unconditional pseudorandom generators for low degree polynomials," in *40th Annual Symposium on the Theory of Computing (STOC)*. ACM, 2008, pp. 557–562.

[35] M. Luby, B. Velickovic, and A. Wigderson, "Deterministic approximate counting of depth-2 circuits," in *Proc. of the 2nd ISTCS*, 1993, pp. 18–24.

[36] W. Maass and G. Turan, "How fast can a threshold gate learn?" in *Computational Learning Theory and Natural Learning Systems: Volume I: Constraints and Prospects*. MIT Press, 1994, pp. 381–414.

[37] K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio, "Testing halfspaces," *SIAM J. Comp.*, to appear. In Proc. Symp. Discrete Algorithms (SODA) (2009), pp. 256-264. Full version available at http://www.cs.cmu.edu/˜odonnell/.

[38] S. Muroga, *Threshold logic and its applications*. New York: Wiley-Interscience, 1971.

[39] J. Naor and M. Naor, "Small-bias probability spaces: efficient constructions and applications," *SIAM J. on Comput.*, vol. 22(4), pp. 838–856, 1993, earlier version in STOC'90.

[40] N. Nisan, "Pseudorandom generators for space-bounded computations," *Combinatorica*, vol. 12, no. 4, pp. 449–461, 1992.

[41] N. Nisan and M. Szegedy, "On the degree of Boolean functions as real polynomials," in *Proc. 24th Annual Symposium on Theory of Computing*, 1992, pp. 462–467.

[42] N. Nisan, "Pseudorandom bits for constant depth circuits," *Combinatorica*, vol. 11, no. 1, pp. 63–70, 1991.

[43] R. O'Donnell and R. Servedio, "The Chow Parameters Problem," in *Proc. 40th Annual ACM Symposium on Theory of Computing (STOC)*, 2008, pp. 517–526.

[44] R. Paturi, "On the degree of polynomials that approximate symmetric Boolean functions," in *Proc. of the 24th Symposium on Theory of Computing*, 1992, pp. 468–474.

[45] L. Penrose, "The elementary statistics of majority voting," *Journal of the Royal Statistical Society*, vol. 109, no. 1, pp. 53–57, 1946.

[46] Y. Peres, "Noise stability of weighted majority," 2004. Available: http://arxiv.org/abs/math/0412377

[47] Y. Rabani and A. Shpilka, "Explicit construction of a small epsilon-net for linear threshold functions," 2009, to appear in *Proc. 40th Annual ACM Symposium on Theory of Computing (STOC)*.

[48] A. Razborov, "A simple proof of bazzi's theorem," 2008, available at http://eccc.hpi-web.de/eccc-reports/2008/TR08-081/index.html.

[49] T. J. Rivlin, *The Chebyshev Polynomials*. John Wiley and Sons, 1974.

[50] R. Servedio, "Every linear threshold function has a low-weight approximator," *Computational Complexity*, vol. 16, no. 2, pp. 180–209, 2007.

[51] Q. Sheng, *Threshold Logic*. London, New York, Academic Press, 1969.

[52] A. Taylor and W. Zwicker, "A Characterization of Weighted Voting," *Proc. of the AMS*, vol. 115, no. 4, pp. 1089–1094, 1992.

[53] E. Viola, "The sum of d small-bias generators fools polynomials of degree d." *Comp. Complexity*, to appear. In *Proc. 23nd Annual IEEE Conference on Computational Complexity (CCC)*, 2008, pp. 124–127.

[54] ——, "Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates," *SIAM Journal on Computing*, vol. 36, no. 5, pp. 1387–1403, 2007.