JACKSON'S INEQUALITY ON THE HYPERCUBE

PAATA IVANISVILI, ROMAN VERSHYNIN, AND XINYUAN XIE

Abstract. We investigate the best constant $J(n,d)$ such that Jackson's inequality

$$\inf_{\deg(g)\leq d} \|f-g\|_{\infty} \leq J(n,d)s(f),$$

holds for all functions f on the hypercube $\{0,1\}^n$, where $s(f)$ denotes the sensitivity of f. We show that the quantity $J(n, 0.499n)$ is bounded below by an absolute positive constant, independent of n. This complements Wagner's theorem, which establishes that $J(n,d) \leq 1$. As a first application we show that reverse Bernstein inequality fails in the tail space $L^1_{\geq 0.499n}$ improving over previously known counterexamples in $L^1_{\geq C\log\log(n)}$. As a second application, we show that there exists a function $f:\{0,1\}^n \to [-1,1]$ whose sensitivity $s(f)$ remains constant, independent of n, while the approximate degree grows linearly with n. This result implies that the sensitivity theorem $s(f) \geq \Omega(\deg(f)^C)$ fails in the strongest sense for bounded real-valued functions even when $\deg(f)$ is relaxed to the approximate degree. We also show that in the regime $d=(1-\delta)n$, the bound

$$J(n,d) \leq C\min\{\delta, \max\{\delta^2, n^{-2/3}\}\}$$

holds. Moreover, when restricted to symmetric real-valued functions, we obtain $J_{\mathrm{symmetric}}(n,d) \leq C/d$ and the decay $1/d$ is sharp. Finally, we present results for a subspace approximation problem: we show that there exists a subspace E of dimension 2^{n-1} such that $\inf_{g\in E} \|f-g\|_{\infty} \leq s(f)/n$ holds for all f.

2020 Mathematics Subject Classification. 41A10; 41A17; 68R05; 94C10; 42C10; 60C05.

Key words. Approximation on the hypercube, Jackson's inequality, Boolean functions, sensitivity, approximate degree, random boolean functions

1. Introduction

1.1. **Approximation of real-valued functions on the hypercube.**

Let $n \geq 1$ be an integer, and let $\{0,1\}^n$ be the n-dimensional hypercube. Set $[n] := \{1,\dots,n\}$. Any function f on the hypercube admits Fourier–Walsh series representation

$$f(x) = \sum_{S\subset[n]} \widehat{f}(S) W_S(x) \quad \text{where} \quad W_S(x) = (-1)^{\sum_{j\in S} x_j}.$$

Here $x=(x_1,\dots,x_n)\in\{0,1\}^n$, $\widehat{f}(S) = \mathbb{E} f(X) W_S(X)$, and $X \sim \mathrm{unif}(\{0,1\}^n)$. The *degree* of f is the minimal number $\deg(f)$ such that $\widehat{f}(S)=0$ for all $|S|>\deg(f)$ where $|S|$ denotes the cardinality of the subset $S\subset\{1,\dots,n\}$. Clearly $\deg(f)$ conincides with the degree of a multilinear polynomial $g:\mathbb{R}^n\to\mathbb{R}$ such that $g=f$ on $\{0,1\}^n$. For any $p\geq 1$ let $\|f\|_p = (\mathbb{E}|f(X)|^p)^{1/p}$, where $X\sim \mathrm{Unif}(\{0,1\}^n)$. If $p=\infty$ we simply set $\|f\|_\infty := \max_{x\in\{0,1\}^n} |f(x)|$.

For any $f:\{0,1\}^n\to\mathbb{R}$, and any d, $0\leq d\leq n$, we are interested in the best uniform polynomial approximation of f on the hypercube, which is measured by

$$E_d^n(f) \stackrel{\text{def}}{=} \inf_{\deg(g)\leq d} \|f-g\|_\infty.$$

A recent breakthrough result due to Hao Huang [Hun19], resolving the long standing *sensitivity conjecture*, states that all Boolean functions $f:\{0,1\}^n\to\{0,1\}$ satisfy

$$s(f) \geq c_0 \deg(f)^c, \tag{1}$$

where $c_0, c > 0$ are universal constants. In an equivalent form, Huang's result can be stated as a *Jackson's inequality on the hypercube*: all Boolean functions $f : \{0,1\}^n \to \{0,1\}$ satisfy

$$E_d^n(f) \leq \frac{c_1}{d^{c_2}} s(f), \tag{2}$$

where $s(f)$ denotes the sensitivity of f, i.e.

$$s(f) := \max_{x \in \{0,1\}^n} \sum_{j=1}^{n} |f(x) - f(x^j)|, \quad x^j := (x_1, \dots, x_{j-1}, 1 - x_j, x_{j+1}, \dots, x_n)$$

and where $c_1, c_2 > 0$ are universal constants, see Subsecton 2.4.

A related result is Pisier's inequality due to Wagner [Wag97] which unlike (2) holds for all real-valued functions but with a weaker upper bound. More precisely we have

$$E_0^n(f) = \inf_{c \in \mathbb{R}} \|f - c\|_\infty \leq s(f) \quad \text{for all} \quad f : \{0,1\}^n \to \mathbb{R}. \tag{3}$$

It is natural to wonder how the *approximaton error* $E_d^n(f)/s(f)$ behaves for an arbitrary real valued $f : \{0,1\}^n \to \mathbb{R}$ and any degree $d \in [0, n]$.

1.2. **Approximation by low-degree polynomials.** Our first result shows that if f is *symmetric*, that is if the value $f(x_1, \dots, x_n)$ is independent of a permutation of the variables $x_1, \dots, x_n$, then the approximation error is $O(1/d)$.

Proposition 1 (Approximability of symmetric functions by low-degree polynomials)**.** *There exists a universal constant $C > 0$ such that for any symmetric function $f : \{0,1\}^n \to \mathbb{R}$ and any $d \in (0, n]$, we have*

$$E_d^n(f) \leq \frac{C}{d} s(f). \tag{4}$$

Moreover, this bound is optimal. For every $n \in \mathbb{N}$ and any $d \in (0, n]$, there exists a symmetric function $f : \{0,1\}^n \to \mathbb{R}$ with $s(f) > 0$ such that

$$E_d^n(f) \geq \frac{1}{8d} s(f). \tag{5}$$

However, an estimate of the type (2) fails in the strongest possible sense as soon as we drop the assumption of f being symmetric or Boolean:

Theorem 1 (Inapproximability of general functions by low-degree polynomials)**.** *For any positive $c_1 < 1/2$, there exists $c_2 > 0$ such that for all $n \geq 1$ there exists a function $f : \{0,1\}^n \to \mathbb{R}$ with $s(f) > 0$ satisfying*

$$E_{c_1 n}^n(f) \geq c_2 s(f). \tag{6}$$

The original formulation of Huang's result (1), namely

$$s(f) \geq \sqrt{\deg(f)} \tag{7}$$

is valid for all Boolean functions, but due to inhomogeneity such estimate can not generalize to real valued functions $f : \{0,1\}^n \to [-1,1]$. A simple counterexample is $f(x) = \frac{1}{n}(-1)^{x_1 + \dots + x_n}$ where $s(f) = 2$ while $\deg(f) = n$.

One may wonder if there is a chance to have an estimate of the form (7) for all $f : \{0,1\}^n \to [-1,1]$ if we replace $\deg(f)$ by the *approximate degree* $\widetilde{\deg}(f)$, which is defined as the smallest number ℓ such that $E_\ell^n(f) \leq 1/3$, where $1/3$ here is by convention and can be replaced by any number in $(0,1)$. Notice that the approximate degree of the counterexample above is zero for $n \geq 3$. The proof of Theorem 1 implies that the answer to this question is negative.

Corollary 2 (No sensitivity bound for real-valued functions via approximate degree)**.** *Let h be any function on the real line, which increases to infinity as $x \to \infty$. The inequality*

$$s(f) \geq h\left(\widetilde{\deg}(f)\right)$$

does not hold for all functions $f : \{0,1\}^n \to [-1,1]$ and all $n \geq 1$.

1.3. Failure of reverse Bernstein inequality in the tail space $L^1_{\geq 0.499n}$.

For any $d \in [0, n]$ and any $p \geq 1$, let $L^p_{\geq d}$ denote the d'th tail space equipped with the norm $\|\cdot\|_p$ and consisting of all functions f on the hypercube that can be represented as $f(x) = \sum_{|S|\geq d} a_S W_S(x)$. Set $\Delta f(x) = \sum_{j=1}^n (f(x) - f(x^j))$. Theorem 1 gives the following

Corollary 3. *For any $c_1 \in (0, 1/2)$, there exists $c_2 > 0$ that depends only on c_1 and such that for any $n \geq 1$ there exists a function $f \in L^1_{\geq c_1 n}$, $f \not\equiv 0$, with the property*

$$\|f\|_1 \geq c_2 \|\Delta f\|_1.$$

In [MN14] Mendel and Naor showed that if the reverse Bernstein inequality $\|f\|_p \leq \frac{C_p}{n}\|\Delta f\|_p$ were to hold in the tail space $L^p_{\geq n/10}$ for vector-valued functions then this would simplify their construction of vector-valued expanders (see Remark 7.5 in [MN14]). They also proved (see [MN14, equation 145]) that the inequality fails in the tail space $L^1_{\geq c \log\log n}$. Corollary 3 shows that one can find a counterexample to the reverse Bernstein inequality even in the smaller space $L^1_{\geq cn}$. Before we conclude this section, we point out that for any $p \in (1, \infty)$, the reverse Bernstein inequality in the tail space $L^p_{\geq d}$ remains a major open problem, see [HMO17, EI20, EI23] for some partial results.

1.4. Approximation by low-dimensional subspaces.

In fact, the existence of a *poorly* approximable function f in Theorem 1 is not due to the nature of the space of polynomials of degree at most d. It is just a consequence of the dimension of this space, which equals $\binom{n}{\leq d} = \sum_{j=0}^d \binom{n}{j}$. The proof of Theorem 1 actually shows that no subspace of this dimension can be used to approximate f. In the language of approximation theory, we find a lower bound on the *Kolmogorov width* of the set of functions of given sensitivity:

Theorem 4 (Inapproximability by low-dimensional subspaces). *For any $c_1 \in (0, 1/2)$, there exists small $c_2 > 0$ and large $n_0 > 0$ such that for any $n \geq n_0$ and any subspace E of dimension at most $\binom{n}{\leq c_1 n}$ there exists a function $f : \{0,1\}^n \to \mathbb{R}$ with $s(f) > 0$ satisfying*

$$\inf_{g \in E} \|f - g\|_\infty \geq c_2 s(f).$$

It is an interesting problem to understand what happens in Theorem 1 in the critical regime where $c_1 = 1/2$. Notice that if n is even, the dimension of the space of all polynomials of degree at most $n/2$ is $\binom{n}{\leq n/2} = 2^{n-1}$. In the next theorem we show that there exists a subspace of dimension 2^{n-1} in the space of all functions on $\{0,1\}^n$ such that the approximation error is at most $s(f)/n$. In particular, it shows that Theorem 4 is sharp, i.e., one cannot relax the assumption $c_1 < 1/2$.

Theorem 5 (Approximability by a subspace of half-dimension). *There exists a subspace E of dimension 2^{n-1} such that*

$$\inf_{g \in E} \|f - g\|_\infty < \frac{s(f)}{n}$$

holds for all $f : \{0,1\}^n \to \mathbb{R}$.

1.5. Approximation by high-degree polynomials.

Let us revisit the high-degree approximation. How well can we approximate a general function $f : \{0,1\}^n \to \mathbb{R}$ by a polynomial of degree $d = (1-\delta)n$ where $\delta \in [0,1]$ is fixed? In other words, what is the smallest mulitplier $J(n,d)$ that makes Jackson-type inequality

$$E_d^n(f) \leq J(n,d) s(f)$$

hold true in this regime? Pisier's inequlity (3) with Theorem 1 show that if $\delta \in (1/2, 1]$ then

$$c_1 \leq J(n,d) \leq 1$$

where $c_1 = c_1(\delta) > 0$ depends only on δ. At the opposite end of the spectrum, setting $\delta = 0$ makes $J(n,d) = 0$, since the function f itself can be expressed as a polynomial of degree $d = n$. This makes us wonder whether and how $J(n,d)$ decreases to zero if we let δ decrease to 0. For example, is it true that $J(n, 3n/4) \to 0$ as $n \to \infty$? The following result gives a bound of this kind. For $x = (x_1,\dots,x_n)$ and $y = (y_1,\dots,y_n)$ in $\{0,1\}^n$ set

$$x \oplus y = (x_1 + y_1 \bmod(2), \dots, x_n + y_n \bmod(2)).$$

Pick any polynomial h of degree at most d on the real line, which satisfies $\mathbb{E}h(X) = 1$ where $X \sim \mathrm{Bin}(n,1/2)$, and let $H(x_1,\dots,x_n) = h(x_1 + \dots + x_n)$.

Theorem 6. *We have*

$$E_d^n(f) \le \max_{x \in \{0,1\}^n} |f(x) - \mathbb{E}_y f(y) H(y \oplus x)| \le 3\frac{s(f)}{n} \mathbb{E}X|h(X)|, \tag{8}$$

where $y \sim \mathrm{Unif}(\{0,1\}^n)$, *and* $X \sim \mathrm{Bin}(n,1/2)$.

Combining (8) with a *quadrature argument* in Section 2.9, we obtain

Corollary 7. *For any* $f : \{0,1\}^n \to \mathbb{R}$, *and any* $d \in [0,n]$, *we have*

$$E_d^n(f) \le 3\frac{k_{n,\lfloor d/2 \rfloor + 1}}{n} s(f) \tag{9}$$

where $k_{n,\ell}$ *is the smallest positive root of degree* ℓ *Kravchuk polynomial, i.e., polynomials orthogonal with respect to the measure* $\mathrm{Bin}(n,1/2)$.

Alternatively, combining (8) with a specific choice of h improves the bound (9) in the regime $\delta \le \max\{\delta^2, n^{-2/3}\}$, where $\delta = 1 - d/n$.

Corollary 8. *For any* $f : \{0,1\}^n \to \mathbb{R}$, *and any* d, $0 \le d \le n$, *we have*

$$E_d^n(f) \le 3\left(1 - \frac{d}{n}\right) s(f).$$

Corollaries 7 and 8, together with additional results on the smallest positive roots of Kravchuk polynomials, provide us with the following bound.

Corollary 9 (Approximation by high-degree polynomials)**.** *For any* $f : \{0,1\}^n \to \mathbb{R}$, *and any* $d \in [0,n]$ *we have*

$$E_d^n(f) \le C \min\{\delta, \max\{\delta^2, n^{-2/3}\}\} s(f), \tag{10}$$

where $C > 0$ *is an absolute constant, and* $\delta = 1 - d/n$.

1.6. **Approximation of random Boolean functions.** Our results so far have been about *arbitrary* real-valued functions on the hypercube. Next we consider the same approximation problem for a *random* Boolean function. We say that $f : \{0,1\}^n \to \{-1,1\}$ is a random Boolean function if $\{f(x)\}_{x \in \{0,1\}^n}$ are independent identically distributed symmetric ± 1 Bernoulli random variables. The following result is due to [OS08] which shows that $f_{\le d}$, where d is slighly greater than $n/2$, is a good approximation in L^∞ to f with high probability. Set

$$f_{>d}(x) := \sum_{|S| > d} \widehat{f}(S) W_S(x),$$

and denote $f_{\le d} := f - f_{>d}$. It is clear that $\inf_{\deg(g) \le d} \|f - g\|_2$ is achieved on $g = f_{\le d}$.

Theorem 10 (Approximability of random Boolean functions by polynomials of degree $> n/2$)**.** *If* $d \ge n/2 + C\sqrt{n \log n}$, *then a random Boolean function* $f : \{0,1\}^n \to \{-1,1\}$ *can be uniformly* $1/n^{10}$*-approximated by a polynomial of degree* d. *Specifically, we have*

$$\|f_{>d}\|_\infty \le \frac{1}{n^{10}} \tag{11}$$

with probability at least $1 - 2^{-n}$.

The original proof of Theorem 10 in [OS08] did not use $f_{\leq d}$ as an approximating polynomial, however, their argument still implies the bound (11). In Subsection 2.12 we will present a direct and simpler proof of Theorem 10.

It is an interesting and open porblem, Wang–Williams conjecture, see also [BV19, Conjecture 8.2], what should be the correct term instead of $C\sqrt{n\log(n)}$ in Theorem 10. In general the degree $n/2$ is needed in Theorem 10, i.e., half of the Boolean functions cannot be uniformly well approximated by polynomials of degree at most $n/2$. This result is due to [An95]. Since the proof is short we decided to present it in Subsection 2.13 for completeness.

Theorem 11 (Degree $n/2$ is a sharp threshold)**.** *At least* 50% *of all Boolean functions* $f : \{0,1\}^n \to \{-1,1\}$ *satisfy*

$$\inf_{\deg(g)\leq n/2} \|f-g\|_\infty \geq 1.$$

Equivalently, at least 50% *of all Boolean functions are not sign-representable by any PTF of degree* $\leq n/2$*, where PTF stands for polynomial threshold function, i.e., sign of degree at most* d *real-valued polynomial on the hypercube.*

2. Proofs

2.1. **Proof of Proposition 1.** First we verify the upper bound (4) in Proposition 1. Since f is symmetric, the value of f at any point $x = (x_1,\ldots,x_n) \in \{0,1\}^n$ is uniquely determined by the Hamming weight $x_1+\cdots+x_n$. In other words, for any symmetric function $f : \{0,1\}^n \to \mathbb{R}$ there exists a function $\varphi : \{0,\ldots,n\} \to \mathbb{R}$ such that

$$f(x_1,\ldots,x_n) = \varphi(x_1+\cdots+x_n) \quad \text{for all } x \in \{0,1\}^n. \tag{12}$$

Conversely, for any function $\varphi : \{0,\ldots,n\} \to \mathbb{R}$, the function $f : \{0,1\}^n \to \mathbb{R}$ defined by (12) is symmetric. This observation allows us to use symmetry to greatly simplify the computations of the approximation error and sensitivity.

Lemma 1 (Approximation error for symmetric functions)**.** *For any symmetric* $f : \{0,1\}^n \to \mathbb{R}$ *we have*

$$\inf_{\deg(g)\leq d} \|f-g\|_\infty = \inf_{\substack{\deg(g)\leq d\\ g \text{ is symmetric}}} \|f-g\|_\infty = \inf_{\deg(h)\leq d} \max_{k\in\{0,\ldots,n\}} |\varphi(k)-h(k)|,$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ *is a function that satisfies* (12).

Proof. To check the first equality, recall that by definition, the left hand-side term is less or equal than the right hand-side term. For the reverse direction, for any g, we can obtain a symmetric function with smaller error

$$\tilde{g}(x) := \frac{1}{n!}\sum_{\sigma\in S_n} g(x_{\sigma(1)},\ldots,x_{\sigma(n)}),$$

where S_n is the symmetry group of $[n]$. Since

$$\|f(x)-\tilde{g}(x)\|_\infty \leq \frac{1}{n!}\sum_{\sigma\in S_n} \|f(x)-g(\sigma(x))\|_\infty = \|f(x)-g(x)\|_\infty,$$

where the last equality follows from the fact that f is symmetric and $f(x) = f(x_{\sigma(1)},\ldots,x_{\sigma(n)})$.

For the right hand-side equality, we again notice that any symmetric polynomial g on the hypercube $\{0,1\}^n$ of degree at most d can be written as $h(x_1+\cdots+x_n)$, where h is a polynomial on the real line of degree at most d. □

Lemma 2 (Sensitivity of symmetric functions)**.** *For any symmetric* $f : \{0,1\}^n \to \mathbb{R}$ *we have*

$$\frac{n}{2}\max_{k\in[n]} |\varphi(k)-\varphi(k-1)| \leq s(f) \leq n\max_{k\in[n]} |\varphi(k)-\varphi(k-1)|,$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ *is a function that satisfies* (12).

Proof. The right hand-side inequality follows from the definition of $s(f)$. For the left hand-side inequality, let m be the number that attains the maximum *jump*, i.e.,

$$|\varphi(m)-\varphi(m-1)| = \max_{k\in[n]}|\varphi(k)-\varphi(k-1)|.$$

Pick a vector $x \in \{0,1\}^n$ with $|x| = m$. We have

$$s(f)(x) = \sum_{i=1}^{n}|f(x)-f(x^i)| \geq m\max_{k\in[n]}|\varphi(k)-\varphi(k-1)|.$$

Similarly, we obtain that

$$s(f)(x^j) \geq (n-m+1)\max_{k\in[n]}|\varphi(k)-\varphi(k-1)|,$$

where j is a coordinate at which $x_j = 1$. Thus

$$s(f) \geq \max(s(f)(x), s(f)(x^j)) \geq \frac{n}{2}\max_{k\in[n]}|\varphi(k)-\varphi(k-1)|.$$

□

Lemma 3. *For any real numbers $a_0,\ldots,a_n$, we have*

$$\inf_{\deg(h)\leq d}\max_{k\in\{0,\ldots,n\}}|a_k - h(k)| \leq \frac{Cn}{d}\max_{k\in[n]}|a_k - a_{k-1}|$$

Proof. For any $v:[a,b]\to\mathbb{R}$, let

$$w(v,t) = \sup_{x,y\in[a,b],|x-y|\leq t}|v(x)-v(y)|,$$

be the modulus of continuity of v. It is straightforward to verify that $t\mapsto w(v,t)$ is non-decreasing, and $t\mapsto w(v,t)$ is sub-additive, i.e., $w(v,t_1+t_2)\leq w(v,t_1)+w(v,t_2)$. Therefore, it follows that

$$w(v,ta) \leq (t+1)w(v,a) \quad \forall t>1, \forall a>0. \tag{13}$$

Jackson's theorem for algebraic polynomials [Jac12] states that $\forall v\in C([a,b])$,

$$\inf_{\deg(h)\leq d}\max_{x\in[a,b]}|v(x)-h(x)| \leq Cw(v,\frac{b-a}{2d}).$$

Take $[a,b]=[0,n]$. Set $v(t)$ to be constant a_k on each interval $[k-\frac{1}{3},k+\frac{1}{3}]$ (if $k=0$ or n, then the interval would be $[0,\frac{1}{3}]$ and $[n-\frac{1}{3},n]$ respectively), and interpolate by a linear function on the remaining parts the domain so that $v(t)$ is continous. Then, clearly

$$w(v,\frac{1}{3}) = \max_{k\in[n]}|a_k - a_{k-1}|.$$

Thus, using (13) and the fact that $d\leq n$, we have

$$w(v,\frac{n}{2d}) \leq \left(\frac{3n}{2d}+1\right)w(v,\frac{1}{3}) \leq \frac{3n}{d}\max_{k\in[n]}|a_k - a_{k-1}|.$$

Then the lemma follows from

$$\inf_{\deg(h)\leq d}\max_{k\in\{0,\ldots,n\}}|a_k-h(k)| \leq \inf_{\deg(h)\leq d}\max_{x\in[0,n]}|v(x)-h(x)| \leq \frac{Cn}{d}\max_{k\in[n]}|a_k-a_{k-1}|.$$

□

Now the upper bound in Proposition 1 can be proved as follows: we have

$$\begin{aligned}
E_d^n &= \inf_{\deg(g)\le d} \|f-g\|_\infty \\
&= \inf_{\deg(h)\le d} \max_{k\in\{0,\dots,n\}} |\varphi(k)-h(k)| \quad \text{(by Lemma 1)} \\
&\le \frac{Cn}{d} \max_{k\in[n]} |\varphi(k)-\varphi(k-1)| \quad \text{(by Lemma 3)} \\
&\le \frac{2C}{d} s(f) \quad \text{(by Lemma 2).}
\end{aligned}$$

The proof of the lower bound in Proposition 1 is based on a general lower bound of the Kolmogorov width of a spaces $C(K)$. Such a result was proved by G. G. Lorenz in [Lor60, Theorem 3] by using a "hat-packing" argument similar to the one in the proof of Theorem 1. A simplified version of Lorenz's result can be stated as follows:

Theorem 12 (Lorenz's bound)**.** *Consider a metric space (K,d) that consists of $k+1$ points. Suppose that all pairwise distances between these points are bounded below by $a>0$. Let H be a k-dimensional linear space of real-valued functions on K. Then there exists a function $\varphi : K \to \mathbb{R}$ that is 1-Lipschits, i.e. satisfying $|\varphi(x)-\varphi(y)| \le d(x,y)$ for all $x,y\in K$, which satisfies*

$$\|\varphi - h\|_\infty \ge \frac{a}{4} \quad \textit{for all } h\in H.$$

To prove the lower bound in Proposition 1, we apply Lorenz's Theorem 12 for the metric space $K=\{0,\dots,n\}$ equipped with the usual distance, and for the space H of (restrictions of) univariate polynomials on K whose degree is bounded by d. Then $k := \dim(H) = d+1$. Placing the points at equal distances, we can find $k+1$ points in K whose all pairwise distances are at least $a = \lfloor n/(k+1)\rfloor = \lfloor n/(d+2)\rfloor$. Therefore Lorenz's Theorem 12 yields the existence of a 1-Lipschitz function φ on K that satisfies

$$\inf_{\deg(h)\le d} \max_{k\in K} |\varphi(k)-h(k)| \ge \frac{a}{4} \ge \frac{n}{8d}.$$

Recalling Lemma 1 and the representation of symmetric functions in (12), we conclude that there exists a symmetric function $f:\{0,1\}^n \to \mathbb{R}$ such that

$$E_d^n(f) = \inf_{\deg(g)\le d} \|f-g\|_\infty \ge \frac{n}{8d}.$$

On the other hand, Lemma 2 and the fact that φ is 1-Lipschitz imply

$$s(f) \le n \max_{k\in[n]} |\varphi(k)-\varphi(k-1)| \le n.$$

Combining these two bounds concludes the proof of the lower bound in Proposition 1.

2.2. **Proof of Theorem 1.** The main idea of the proof is to construct a class of low sensitivity functions that is *complex* enough, so that some of them can not be approximated efficiently.

We construct such class of functions by considering the maximal number of disjoint *smooth* hamming balls (the *hat* functions defined in the latter sentence) that we can pack into the cube. More specifically, for $x\in\{0,1\}^n$, let h_x be a *hat* function centered at the point x with radius m, which is defined by assigning $\frac{m-k}{m}$ to each point with hamming distance k to the center if $k<m$, and zero otherwise. Let K be a maximal packing with distance $2m$ on the hypercube $\{0,1\}^n$, i.e., a subset of $\{0,1\}^n$ with maximal number of elements such that the hamming distance between any two point in the subset is strictly greater than $2m$. Consider the class of functions

$$F := \left\{ \sum_{x\in K} \varepsilon_x h_x \mid \varepsilon_x \in \{1,-1\}, \forall x\in K \right\}.$$

The cardinality of F is $2^{P(K,2m)}$, where $P(K,2m)$ denotes the packing number of K with distance $2m$. Notice that

$$s(f) = \frac{n}{m} \quad \text{for all} \quad f \in F.$$

Next we show that some of elements in F cannot be approximated by low degree polynomials. Fix $d \in \{1,\dots,n\}$. The polynomials of degree at most d form linear space

$$L_d := \{g : \deg(g) \le d\} \quad \text{with} \quad \dim(L_d) = \binom{n}{\le d} =: N,$$

where we used the notation $\binom{n}{\le k} := \sum_{j=0}^{k} \binom{n}{j}$.

For each $x \in \{0,1\}^n$, consider the central hyperplane

$$H_x := \{g \in L_d : g(x) = 0\}.$$

There are 2^n such hyperplanes which partition the space L_d into regions (here we think every function on the cube as a vector in $\mathbb{R}^{2^n}$, and $L_d \setminus (\cup_x h_x)$ is a union of connected components, which are called regions here). Notice that each region is the collection of polynomials $g \in L_d$ that have the same sign pattern $\mathrm{sign}(g)$, i.e., $\mathrm{sign}(g_1) = \mathrm{sign}(g_2)$ if and only if g_1, g_2 are in the same connected component. Thus, the number of all possible sign patterns $\mathrm{sign}(g)$ equals the number of regions. On the other hand, for $f \in F$, in order to have $\inf_{g \in L_d} \|f - g\|_\infty < 1 = c_2 s(f)$, one must have $\mathrm{sign}(g) = \mathrm{sign}(f)$ on K. Thus, the number of $f \in F$ that satisfy this inequality is less or equal than the number of all possible sign patterns.

Schlaffli's formula in the hyperplane arrangement literature [BV19, Lemma 2.2] states that the number of sign patterns S generated by polynomials of degree at most d is at most $2\binom{2^n-1}{\le N-1}$ which, in turn, is at most

$$\binom{2^n}{\le N}$$

whenever $d \le n/2$ because $\binom{2^n}{N} = \frac{2^n}{N}\binom{2^n-1}{N-1}$, and $\frac{2^n}{N} = \frac{2^n}{\binom{2^n}{\le d}} \ge 2$.

Thus if $\inf_{g \in L_d} \|f - g\|_\infty < c_2 s(f)$ holds for all $f \in F$ then we must have

$$2^{P(K,2m)} \le \binom{2^n}{\le \binom{n}{\le d}}. \tag{14}$$

Let us show that this inequality is not possible in the regime $d = c_1 n$, where $c_1 < 1/2$ is fixed, and n is large. Indeed, consider binary entropy function $E(p) = p\log_2(1/p) + (1-p)\log_2(1/(1-p))$. This is a concave function on $[0,1]$, symmetric with respect to $p = 1/2$, and such that $E(0) = E(1) = 0$, and $E(1/2) = 1$. Let $m = c_2 n$, where $c_2 \in (0,1/4)$ is so small that $E(2c_2) + E(c_1) - 1 < 0$. Pick any $n > \frac{1}{1-E(c_1)}$. By [FlGr, Lemma 16.19] we have $\binom{n}{\le k} \le 2^{nE(k/n)}$, whenever $k/n \in (0,1/2]$. Thus

$$\binom{2^n}{\le \binom{n}{\le d}} \le 2^{2^n E\left(\frac{\binom{n}{\le c_1 n}}{2^n}\right)} \le 2^{2^n E(2^{n(E(c_1)-1)})} \tag{15}$$

where in the second inequality we used [FlGr, Lemma 16.19], the fact that $2^{n(E(c_1)-1)} \in (0,1/2)$, and $E(t)$ is nondecreasing on $(0,1/2]$.

Replacing volume by cardinality in [Ver18, Proposition 4.2.12] we have

$$2^n / \binom{n}{\le 2m} \le P(K,2m), \tag{16}$$

which is also addressed in [Ver18, Exercise 4.1.16]. Therefore validity of the inequality (14) implies

$$1 \le \binom{n}{\le 2m} E(2^{n(E(c_1)-1)}) \le 2^{nE(2c_2)} E(2^{n(E(c_1)-1)}) \le 2 \cdot 2^{n(E(2c_2)+E(c_1)-1)} n(1 - E(c_1)),$$

where in the last inequality we used the simple fact that $E(t) \leq 2t\log_2(1/t)$ for all $t \in (0,1/2]$. On the other hand the inequality $1 \leq 2 \cdot 2^{n(E(2c_2)+E(c_1)-1)} n(1-E(c_1))$ fails for n sufficiently large (say all $n \geq n_0(c_1)$) due to the assumption $E(2c_2)+E(c_1)-1 < 0$. However, notice that in the beginning of the proof we could assume without loss of generality that $n \geq n_0(c_1)$ because for small finite number of n's with $1 \leq n \leq n_0(c_1)$ we can just take any function f of degree exactly n and in this case we trivially have $\inf_{\deg(g)\leq c_1 n} \|f-g\|_\infty \geq c_2(n)s(f)$ holds for some $c_2(n) > 0$. This finishes the proof of the theorem.

2.3. **Proof of Corollary 2.** The example $f : \{0,1\}^n \to [-1,1]$ constructed in the proof of Theorem 1 for sufficienlty large n has the property $s(f) = 1/c_2$, and $E^n_{c_1 n}(f) \geq 1$. Clearly this implies $\widetilde{\deg(f)} \geq c_1 n$, and hence $s(f) \geq h(\widetilde{\deg(f)})$ is not possible as $n \to \infty$.

2.4. **Proof of equiavlence (1) and (2).** First let us show the implication (1) implies (2). Assume $s(f) \geq c_0(\deg(f))^c$ holds with some universal constants $c_0, c_1 > 0$. We claim that $E^n_d(f) \leq \frac{1}{c_0 d^c} s(f)$ holds for all Boolean f. Indeed, if $\deg(f) = 0$ then the claim is trivial. Assume $\deg(f) \geq d$, then $\frac{1}{c_0 d^c} s(f) \geq 1$. Since $E^n_d(f) \leq 1$ the claim follows.

Next let us show the implication (2) implies (1). Assume $E^n_d(f) \leq \frac{c_1}{d^{c_2}} s(f)$ holds with some universal constants $c_1, c_2 > 0$. Choose $d = \widetilde{\deg}(f) - 1$. Then we obtain

$$\frac{1}{3} \leq \frac{c_1}{(\widetilde{\deg}(f)-1)^{c_2}} s(f). \tag{17}$$

Since $\deg(f) \leq c\widetilde{\deg}(f)^8$, see [NS93], we obtain (1). In this argument without loss of generality we assume $\widetilde{\deg}(f) > 100$ this assumption avoids the issue in (17) when $\widetilde{\deg}(f) = 1$.

2.5. **Proof of Corollary 3.** Since $s(f)(x) \geq |\Delta f|(x)$ it follows that the inequality (6) in Theorem 1 holds true if we replace $s(f)$ by $\|\Delta f\|_\infty$. By the duality argument presented in [EI23, Theorem 1] (in our case we use Laplacian instead of the discrete gradient), the estimate $E^n_{c_1 n}(f) \geq c_2 \|\Delta f\|_\infty$ implies

$$\sup_{f(x)=\sum_{|S|>c_1 n} \widehat{f}(S) W_S(x)} \frac{\|f\|_1}{\|\Delta f\|_1} \geq C > 0,$$

for some absolute constant C depending on c_1.

2.6. **Proof of Theorem 4.** The *sign pattern argument* presented in the proof of Theorem 1 works verbatim for any subspace of the same dimension $\binom{n}{\leq c_1 n}$ where $n \geq n_0(c_1)$.

2.7. **Proof of Theorem 5.** Without loss of generality, we can assume that $s(f) = 1$. Every such function is *approximately harmonic*: its value at any point x is within $1/n$ from the arithmetic mean of the values at the neighbors of x. This follows from

$$|f(x) - \frac{1}{n}\sum_{i\in[n]} f(x^i)| \leq \frac{1}{n}|\sum_{i\in[n]} f(x) - f(x^i)| \leq \frac{1}{n} s(f). \tag{18}$$

The cube consists of even points and odd points (according to the number of ones). Define E as the set of all *odd-harmonic functions* – those whose value at every odd point is the arithmetic mean of its neighbors (which are all even). The dimension of E is 2^{n-1} since it is defined by 2^{n-1} linear constrains – one per every odd point.

Pick any function f with $s(f) \leq 1$ and choose $g \in E$ that coincides with f at all even points. Since $g \in E$, the value of g at each odd point x is the arithmetic mean of the values of g at the neighbors of x. By the fact that f is *approximately harmonic*, this value is within $1/n$ from f.

2.8. **Proof of Theorem 6.** The first inequality in (8) is trivial since the convolution $\mathbb{E}_y f(y)H(y\oplus x)$ is a polynomial of degree at most d. We verify the second inequality

$$\max_{x\in\{0,1\}^n} |f(x)-\mathbb{E}_y f(y)H(y\oplus x)| \le 3\frac{s(f)}{n}\mathbb{E}X|h(X)|,$$

where $y\sim \mathrm{Unif}(\{0,1\}^n)$, and $X\sim \mathrm{Bin}(n,1/2)$. Let $x^*\in\{0,1\}^n$ be such that $\max_x|f(x)-\mathbb{E}_y f(y)H(y\oplus x)| = |f(x^*)-\mathbb{E}_y f(y)H(y\oplus x^*)|$. Set $F(x)=f(x\oplus x^*)$. Notice that $s(F)=s(f)$, and

$$\max_x |f(x)-\mathbb{E}_y f(y)H(y\oplus x)| = |F(0,\dots,0)-\mathbb{E}_y F(y\oplus x^*)H(y\oplus x^*)| =$$
$$|F(0,\dots,0)-\mathbb{E}_y F(y)H(y)| \le \mathbb{E}|F(0,\dots,0)-F(y)||H(y)| =$$
$$\frac{1}{2^n}\sum_{k=0}^n |h(k)|\sum_{|y|=k}|F(0,\dots,0)-F(y)|.$$

To complete the proof of the theorem it suffices to show

Lemma 4. *For each $k=1,\dots,n$ we have*

$$\sum_{|y|=k}|F(0,\dots,0)-F(y)| \le 3s(F)\binom{n-1}{k-1}.$$

Proof. Without loss of generality assume $s(F)=1$.

Consider the case $k\le \frac{n}{2}$. We have

$$\sum_{|y|=k}|F(0,\dots,0)-F(y)| \le \frac{1}{k!}\sum_{|y|=k}\sum_{\mathrm{paths}(a_0-a_1-\dots-a_k)}|F(a_j)-F(a_{j+1})|,$$

where inner sum runs over all *monotone paths* over vertices $(0,\dots,0)=a_0,..,a_k=y$ joining the points $(0,\dots,0)$ and y. Here $\|a_{j+1}-a_j\|_{\ell_1^n}=1$ and $a_{j+1}\succ a_j$ i.e., $(a_{j+1})_k\ge (a_j)_k$ for all $k=1,\dots,n$. We claim that

$$\frac{1}{k!}\sum_{|y|=k}\sum_{\mathrm{paths}(a_0-a_1-\dots-a_k)}|F(a_j)-F(a_{j+1})| = \sum_{|x|\le k-1} p_{|x|}\, s(F)(x), \tag{19}$$

where

$$p_\ell := \frac{\ell!(n-\ell)!}{k!(n-k)!(n+1)} - \frac{(-1)^{k+\ell}}{n+1}, \quad \ell=0,\dots,k.$$

Indeed, a direct calculation shows

$$p_\ell + p_{\ell+1} = \frac{\ell!(n-\ell-1)!}{k!(n-k)!} = \frac{\ell!(k-\ell-1)!}{k!}\binom{n-\ell-1}{k-\ell-1}$$

for all $\ell=0,\dots,k-1$, and $p_k=0$. For each $y\in\{0,1\}^n$, $|y|\le k-1$, the term $|F(y)-F(y^i)|$ appears $p_\ell+p_{\ell+1}$ times in the right hand side of (19), where $|y|=\ell$ and $|y^i|=\ell+1$. Let us count the total number of paths passing through the edge $[y,y^i]$ in the left hand side of (19). We enter the vertex y in $\ell!$ ways, and exit the vertex y^i in $(k-\ell-1)!\binom{n-\ell-1}{k-\ell-1}$ ways, and this gives $\ell!(k-\ell-1)!\binom{n-\ell-1}{k-\ell-1}$ total number of paths.

Next we show

$$\sum_{\ell=0}^k |p_\ell|\binom{n}{\ell} \le 2\binom{n-1}{k-1}.$$

The inequality simplifies to

$$\sum_{\ell=0}^k \left|1-(-1)^{k+\ell}\frac{\binom{n}{\ell}}{\binom{n}{k}}\right| \le \frac{2k(n+1)}{n}.$$

The inequality follows from the fact that $\binom{n}{\ell}\le\binom{n}{k}$ for all $\ell=0,\dots,k$ and all $k\le n/2$.

Next, let us consider the case $k \geq n/2$. The main result in [Wag97] says that $|F(0,\dots,0) - F(1,\dots,1)| \leq s(F)$. Therefore, since $\sum_{|y|=k} 1 = \binom{n}{k} \leq 2\binom{n-1}{k-1}$ it suffices to show

$$\sum_{|y|=k} |F(1,\dots,1) - F(y)| \leq \binom{n-1}{k-1}.$$

On the other hand for $\tilde{F}(x) = F(x \oplus (1,\dots,1))$ we have $s(F) = s(\tilde{F})$, and

$$\sum_{|y|=k} |F(1,\dots,1) - F(y)| = \sum_{|w|=n-k} |\tilde{F}(0,\dots,0) - \tilde{F}(w)| \leq \binom{n-1}{n-k-1} \leq \binom{n-1}{k-1}.$$

□

2.9. **Proof of Corollary 7.** We need the following lemma which follows from [DIM24, Theorem 2.1 equation (2.7)].

Lemma 5. *Let N be an odd integer, and let $\gamma(t)$ denote the moment curve in $\mathbb{R}^N$ on the interval $[a,b]$, i.e., $\gamma(t) = (t,\dots,t^N), t \in [a,b]$. Then $\sum_{k=1}^{(N+3)/2} \omega_k \gamma(t_k)$ is an interior point of the convex hull of the moment curve, where $a = t_1 < \cdots < t_{(N+3)/2} = b$ and $\sum_{k=1}^{(N+3)/2} \omega_k = 1, \omega_k \in (0,1)$.*

Corollary 13. *If a probaility measure μ on $[a,b]$ is supported on finitely many but at least $(N+3)/2$ points, and the support includes the boundary points $\{a,b\}$, then $\int \gamma(t)\, d\mu(t)$ is an interior point of the convex hull of the moment curve.*

Proof. Notice that the integral is a convex combination of finitely many but more than $(N+3)/2$ points. Hence we can rewrite the convex combination as a convex combination of $(N+3)/2$ points, including $\gamma(a)$ and $\gamma(b)$, and a convex combination of the reminder. By the above Lemma 5, the first convex combination (when normilized, i.e., divided by the sum of its weights) is an interior point of the convex hull. Since a convex combination of an interior point and some other points in a convex set is again an interior point of the convex set, the value of the integral is an interior point of the convex hull of the moment curve. □

Now we turn to the proof of Corollary 7. Notice that

$$\inf_{\deg(h) \leq d, \mathbb{E}(h(X))=1} \mathbb{E}X|h(X)| \leq \inf_{\deg(Q) \leq \lfloor d/2 \rfloor, \mathbb{E}Q^2(X)=1} \mathbb{E}XQ^2(X).$$

We can obtain the exact value of the right hand-side by the following *quadrature argument*. Let $p = \lfloor d/2 \rfloor$. Consider the moment curve $\gamma(t) = (t,\dots,t^{2p+1})$ on $[0,n]$, and a random variable $X \sim \mathrm{Bin}(n,1/2)$. By Corollary 13, $\mathbb{E}\gamma(X)$ is an interior point of the convex hull of $\gamma([0,n])$. By [DIM24] there exist $0 < t_1 < \cdots < t_{p+1} < n$ such that

$$\mathbb{E}\gamma(X) = \sum_{k=1}^{p+1} \omega_k \gamma(t_k),$$

for some $\omega_k \in (0,1)$ with $\sum_{k=1}^{p+1} \omega_k = 1$. This implies that for any polynomial $P(t)$ of degree at most $2p+1$, we have that

$$\mathbb{E}P(X) = \sum_{k=1}^{p+1} \omega_k P(t_k).$$

In particular, choosing $P(t)$ to be $t^i K_{n,p+1}(t), i \in \{0,\dots,p\}$, where $K_{n,p+1}$ is the Kravchuk polynomial of degree $p+1$ with respect to $\mathrm{Bin}(n,1/2)$, we obtain

$$0 = \mathbb{E}X^i K_{n,p+1}(X) = \sum_{k=1}^{p+1} \omega_k t_k^i K_{n,p+1}(t_k),\ i \in \{0,\dots,p\}.$$

Since the matrix $\{t_k^i\}_{0\le i\le p, 1\le k\le p+1}$ is a Vandermonde matrix with nonzero determinant, the linear system above only has the trivial solution, which means that t_k are the roots of $p+1$ degree Kravchuk polynomial $K_{n,p+1}(t)$, and $t_1 = k_{n,p+1}$.

Let h be a polynomial of degree at most p with $\mathbb{E}h^2(X) \neq 0$, then

$$\frac{\mathbb{E}Xh^2(X)}{\mathbb{E}h^2(X)} = \frac{\sum_{k=1}^{p+1} \omega_k t_k h^2(t_k)}{\sum_{k=1}^{p+1} \omega_k h^2(t_k)} \geq t_1,$$

since $\frac{\mathbb{E}Xh^2(X)}{\mathbb{E}h^2(X)}$ is a convex combination of $t_k's$. Thus, we have

$$\inf_{\deg(h)\le p, \mathbb{E}h^2(X)\neq 0} \frac{\mathbb{E}Xh^2(X)}{|\mathbb{E}h^2(X)|} \geq t_1.$$

On the other hand, if we take h to be a degree p polynomial that vanishes at t_k's for $k = 2,\dots,p+1$, but does not vanish at t_1, then

$$\frac{\mathbb{E}Xh^2(X)}{\mathbb{E}h^2(X)} = t_1,$$

which shows that the inequality above is an equality finishing the proof of the corollary.

2.10. **Proof of Corollary 8.** By Theorem 6, it suffices to show that

$$\inf_{\deg(h)\le d, \mathbb{E}h(X)=1} \mathbb{E}X|h(X)| \le n-d, \quad X \sim Bin(n, 1/2).$$

Consider the following degree d polynomial

$$\tilde{h}(x) := \prod_{k=n-d+1}^{n} (k-x),$$

and set $h(x) = \tilde{h}(x)/\mathbb{E}\tilde{h}(X)$. Notice that $h \geq 0$ on $\{0,\dots,n\}$ and $\mathbb{E}h(X) = 1$. Since $h = 0$ on $\{n-d+1,\dots,n\}$ it follows that $\mathbb{E}X|h(X)| \le n-d$ which completes the proof.

2.11. **Proof of Corollary 9.** By Corollary 8, for $d = (1-\delta)n$ we have

$$\inf_{\deg(g)\le d} \|f-g\|_\infty \le 3\delta s(f)$$

which gives the first term in minimum of the right hand-side of (10). Bounding $k_{n,p}$ from above will give the second term. By [ADGP13, Theorem 5.1] (using the upper bound of $\kappa_{n,n}(p,N)$ in the reference with $p = 1/2$; in our case N is n, and n is $p := \lfloor d/2 \rfloor + 1$), we have that

$$k_{n,p} \le \frac{n}{2} - \sqrt{\frac{n-p+1}{2}} h_{p,1}, \tag{20}$$

where $h_{p,1}$ is the largest root of the Hermite polynomial. By [S75, equation (6.32.5)] we have

$$h_{n,p} = (2p+1)^{\frac{1}{2}} - 6^{-\frac{1}{2}}(2p+1)^{-\frac{1}{6}}(i_1 + \varepsilon_n),$$

where i_1 is the lowest zero of the Airy function and it satisfies $6^{-\frac{1}{2}} i_1 = 1.85575\dots$, and $\varepsilon_n \to 0$ as $n \to \infty$. Combining this result and inequality (20), we obtain that

$$k_{n,p} \le \frac{n}{2} - \sqrt{\frac{(n-p+1)(2p+1)}{2}} + c_n \sqrt{\frac{n-p+1}{2}} (2p+1)^{-1/6},$$

where $c_n \geq 0$ and $c_n \to 1.85575\dots$. We have

$$\frac{n}{2} - \sqrt{\frac{(n-p+1)(2p+1)}{2}} = \frac{n}{2} - \sqrt{\frac{(n-\lfloor d/2\rfloor)(2\lfloor d/2\rfloor+3)}{2}} \le \frac{n}{2} - \sqrt{(n-\frac{d}{2})\frac{d}{2}} = \frac{n}{2}(1-\sqrt{1-\delta^2}) \le \frac{n}{2}\delta^2,$$

where the first inequaltiy follows from $n-\lfloor d/2\rfloor \geq n-d/2$ and $2\lfloor d/2\rfloor+3\geq d$, and the second inequality follows from $\sqrt{x}\geq x$ for all $x\in[0,1]$. For the remainder term, we have that

$$c_n\sqrt{\frac{n-p+1}{2}}(2p+1)^{-1/6} = c_n\sqrt{\frac{n-\lfloor d/2\rfloor}{2}}(2\lfloor d/2\rfloor+3)^{-1/6} \leq$$

$$\frac{c_n}{\sqrt{2}}\sqrt{n-\frac{d}{2}+1}(d+3)^{-1/6} = \frac{c_n}{\sqrt{2}}\sqrt{(\frac{1+\delta}{2}+\frac{1}{n})n}[(1-\delta+\frac{3}{n})n]^{-1/6} \leq \frac{c_n}{(1-\delta)^{1/6}}n^{1/3}.$$

Combining the two estimates above, we have that

$$k_{n,p} \leq [0.5\delta^2+\frac{c_n}{(1-\delta)^{1/6}}n^{-2/3}]n.$$

Thus by the Theorem 7, we have that

$$\inf_{\deg(g)\leq d}\|f-g\|_\infty \leq 3(0.5\delta^2+\frac{c_n}{(1-\delta)^{1/6}}n^{-2/3})s(f).$$

Thus we obtain

$$\inf_{\deg(g)\leq d}\|f-g\|_\infty \leq 3\min\left\{\delta, 0.5\delta^2+\frac{c_n}{(1-\delta)^{1/6}}n^{-2/3}\right\}s(f) \leq C\min\{\delta,\max\{\delta^2,n^{-2/3}\}\}$$

finishing the proof of the corollary.

2.12. **Proof of Theorem 10.** Let $\langle\cdot,\cdot\rangle$ denote the dot product in $\mathbb{R}^{2^n}$. The vectors in $\mathbb{R}^{2^n}$ will be indexed by the set $\{0,1\}^n$. For example, given $f:\{0,1\}^n\to\{-1,1\}$ we will write

$$\langle f, W_S\rangle = \sum_{x\in\{0,1\}^n} f(x)W_S(x).$$

Fix any $\theta\in\{0,1\}^n$. We have

$$f_{>d}(\theta) = \frac{1}{2^n}\sum_{|S|>d}\langle f, W_S\rangle W_S(\theta) = \langle f, v_\theta\rangle \quad \text{where} \quad v_\theta(x) = \frac{1}{2^n}\sum_{|S|>d} W_S(\theta)W_S(x).$$

Here we think about f as a vector in $\mathbb{R}^{2^n}$ with independent Rademacher coefficients, and v_θ as a fixed vector in $\mathbb{R}^{2^n}$. Hoeffding's inequality gives

$$\|\langle f, v_\theta\rangle\|_{\psi_2} \lesssim \|v_\theta\|_2 = \frac{1}{2^n}\left(\sum_{|S|>d}\|W_S\|_2^2\right)^{1/2} \qquad \text{(by orthogonality of monomials } W_S(x))$$

$$= \sqrt{2^{-n}\binom{n}{>d}} \quad \text{(since } \|W_S\|_2^2 = 2^n \text{ for each } S).$$

It follows from definition of ψ_2 norm [Ver18, Definition 2.5.6] that

$$\mathbb{P}\left\{|\langle f, v_\theta\rangle| > C\sqrt{2^{-n}\binom{n}{>d}n}\right\} \leq 2^{-2n}.$$

Taking a union bound yields

$$\mathbb{P}\left\{\exists\theta\in\{0,1\}^n : |\langle f, v_\theta\rangle > C\sqrt{2^{-n}\binom{n}{>d}n}\right\} \leq 2^n\cdot 2^{-2n} = 2^{-n}.$$

Thus, with probability at least $1-2^{-n}$, we have

$$\|f_{>d}\|_\infty \lesssim \sqrt{2^{-n}\binom{n}{>d}n}.$$

By Hoeffding's inequality, the quantity on the right hand side is bounded by $1/n^{10}$ whenever $d \ge n/2 + C\sqrt{n \log n}$.

2.13. **Proof of Theorem 11.** Fix $d \in \{1, \dots, n\}$. Polynomials of degree at most d form the linear space

$$L_d := \{g : \deg(g) \le d\} \quad \text{with} \quad \dim(L_d) = \binom{n}{\le d} =: m.$$

For each $x \in \{0,1\}^n$, consider the central hyperplane

$$H_x := \{g \in L_d : g(x) = 0\}.$$

There are $p := 2^n$ such hyperplanes, which partition the space L_d into regions. Each region is formed by polynomials g that have the same sign pattern $\operatorname{sign}(g)$. Thus, the number of all possible sign patterns $\operatorname{sign}(g)$ equals the number of regions. On the other hand, in order to have $\inf_{g \in L_d} \|f - g\|_\infty < 1$, one must have $\operatorname{sign}(g) = f$. Thus, the number of Boolean functions that satisfy this inequality equals the number of sign patterns, which equals the number of regions.

Schlaffli's formula [BV19, Lemma2.2] states that for any arrangement of p central hyperplanes in $\mathbb{R}^m$, the number of regions is bounded by

$$r(p,m) := 2\binom{p-1}{\le m-1}.$$

If $d \le n/2$ then

$$m = \binom{n}{\le d} \le 2^n/2 = p/2.$$

Thus $m - 1 < (p-1)/2$, hence $r(p,m) \le 2 \cdot 2^{p-1}/2 = 2^{2^n}/2$. We showed that the number of Boolean functions that satisfy $\inf_{g \in L} \|f - g\|_\infty < 1$ is at most $2^{2^n}/2$.

Acknowledgements

P.I. was supported in part by NSF CAREER-DMS-2152401. R.V. was supported from NSF DMS-1954233, NSF DMS-2027299, U.S. Army 76649CS, and NSF+Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning. We thank Alexandros Eskenazis for helpful comments that improved the exposition in this paper.

References

[An95] M. Anthony, Classification by polynomial surfaces. *Discrete Applied Mathematics*, Volume 61, Issue 2, 28 July 1995, Pages 91–103

[ADGP13] I. Area, D.K. Dimitrov, E. Godoy, V.G. Pasvhoa. Zeros of classical orthogonal polynomials of a discrete variable. *Mathematics of Computation*, 82(282):1069-95, 2013.

[BV19] P. Baldi, R. Vershynin. Polynomial threshold functions, hyperplane arrangements, and random tensors. *SIAM Journal on Mathematics of Data Science*, 1(4), 699-729, 2019.

[DIM24] J. de Dios Pont, P. Ivanisvili, J. Madrid. A new proof of the description of the convex hull of space curves with totally positive torsion. *To appear Michigan Math. Journal* (2024)

[EI20] A. Eskenazis, P. Ivanisvili. Polynomial inequalities on the Hamming cube. *Probability Theory and Related Fields*, Volume 178, pages 235–287, (2020).

[EI23] A. Eskenazis, P. Ivanisvili. Sharp growth of the Ornstein—Uhlenbeck operator on Gaussian tail spaces. *Israel Journal of Mathematics*, 253(1), 469-485, 2023.

[FlGr] J. Flum, M. Grohe. Parametrized Complexity Theory. *Springer*, 2006

[HMO17] S. Heilman, E. Mossel and K. Oleszkiewicz. Strong contraction and influences in tail spaces, *Transactions of the American Mathematical Society*, 69, 4843–4863, 2017.

[Hun19] H. Huang. Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. *Annals of Mathematics*, 190.3: 949-955, 2019.

[Jac12] D. Jackson. On approximation by trigonometric sums and polynomials. *Transactions of the American Mathematical Society*, 3, 491–515, 1912.

[Lor60] G. G. Lorentz, Lower bounds for the degree of approximation. *Transactions of the American Mathematical Society* (1960), 25-34.

[MN14] M. Mendel, A. Naor. Nonlinear spectral calculus and super-expanders. *Publications mathématiques de l'IHÉS* 119:1-95, 2014.

[NS93] N. Nisan, M. Szegedy. On the degree of boolean functions as real-valued polynomials. *Comput Complexity*, 4, 301–313, 1994.

[NS64] D. J. Newman, H. S. Shapiro. Jackson's Theorem in Higher Dimensions. *On Approximation Theory/Über Approximationstheorie: Proceedings of the Conference held in the Mathematical Research Institute at Oberwolfach*, Black Forest, August 4–10, 1963.

[OS08] R. O'Donnell, R A. Servedio. Extremal properties of polynomial threshold functions. *Journal of Computer and System Sciences*, 74(3): 298-312, 2008.

[S75] G. Szegö. Orthogonal Polynomials. *Amer. Math. Soc. Coll. Publ.*, 4th ed., Vol. 23, Providence, RI, 1975

[Ver18] R. Vershynin. High-dimensional probability: An introduction with applications in data science. *Cambridge university press*, Vol. 47, 2018.

[Wag97] R. Wagner. Notes on an inequality by Pisier for functions on the discrete cube. . *In: Milman, V.D., Schechtman, G. (eds) Geometric Aspects of Functional Analysis.* Lecture Notes in Mathematics, vol 1745. Springer, Berlin, Heidelberg, 25(1):206–214, 1997.

(P. I.) Department of Mathematics, University of California, Irvine, CA 92697, USA
Email address: pivanisv@uci.edu

(R. V.) Department of Mathematics, University of California, Irvine, CA 92697, USA
Email address: rvershyn@uci.edu

(X. X.) Department of Mathematics, University of California, Irvine, CA 92697, USA
Email address: xinyuax7@uci.edu