Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 75 (2016)

Improved Bounds on the Sign-Rank of AC^0

Mark Bun* Justin Thaler†

Abstract

The *sign-rank* of a matrix A with entries in $\{-1,+1\}$ is the least rank of a real matrix B with $A_{ij} \cdot B_{ij} > 0$ for all i, j. Razborov and Sherstov (2008) gave the first exponential lower bounds on the sign-rank of a function in AC^0, answering an old question of Babai, Frankl, and Simon (1986). Specifically, they exhibited a matrix $A = [F(x,y)]_{x,y}$ for a specific function $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ in AC^0, such that A has sign-rank $\exp(\Omega(n^{1/3}))$.

We prove a generalization of Razborov and Sherstov's result, yielding exponential sign-rank lower bounds for a non-trivial class of functions (that includes the function used by Razborov and Sherstov). As a corollary of our general result, we improve Razborov and Sherstov's lower bound on the sign-rank of AC^0 from $\exp(\Omega(n^{1/3}))$ to $\exp(\tilde{\Omega}(n^{2/5}))$. We also describe several applications to communication complexity, learning theory, and circuit complexity.

1 Introduction

The *sign-rank* of a matrix A with entries in $\{-1,+1\}$ is the least rank of a real matrix B with $A_{ij} \cdot B_{ij} > 0$ for all i, j. This fundamental matrix-theoretic complexity measure has diverse applications in theoretical computer science. For example:

- Upper bounds on sign-rank underly many state of the art learning algorithms, including the fastest known algorithms for PAC learning DNF and read-once formulas. Algorithms based on sign-rank are additionally robust to random classification noise, a property not satisfied by the handful of known PAC learning algorithms that cannot be captured in the sign-rank framework (all of which are based on Gaussian Elimination) [14].

- In communication complexity, sign-rank is known to characterize *unbounded error communication*. Introduced by Paturi and Simon [19] and captured by the communication complexity class $\mathbf{UPP^{cc}}$, this is a powerful communication model that lies at the frontier of our understanding. It is essentially the most powerful communication model against which we know how to prove lower bounds. In fact, the only known communication models that $\mathbf{UPP^{cc}}$ cannot efficiently simulate are the communication analogues of the polynomial hierarchy introduced by Babai, Frankl, and Simon [6]. We direct the interested reader to the recent paper of Göös et al. [13] for a detailed overview of communication complexity classes and their relationships.

*John A. Paulson School of Engineering and Applied Sciences, Harvard University. Supported by an NDSEG Fellowship and NSF grant CNS-1237235. Part of this work was done while the author was visiting Yale University.
†Yahoo Research.

ISSN 1433-8092

- In circuit complexity, sign-rank lower bounds on a matrix $A = [F(x,y)]_{x,y}$ imply lower bounds on the size of threshold-of-majority circuits computing F.

Despite the importance of these applications, our understanding of sign-rank remains rather limited, and it is possible to summarize relevant prior work in a single paragraph. Alon et al. [2] proved lower bounds on the sign-rank of random matrices. The first nontrivial lower bounds for explicit matrix families was obtained in a breakthrough work of Forster [10], who proved strong lower bounds on the sign-rank of Hadamard matrices, and more generally of any sign matrix with small spectral norm. Several subsequent works improved and generalized Forster's method [3, 11, 12, 17]. Nearly tight estimates of the sign-rank were obtained by Sherstov in [24] for all symmetric predicates, i.e., matrices of the form $[D(\sum_i x_i \vee y_i)]_{x,y}$ where $D : \{0,1,\ldots,n\} \to \{0,1\}$ is a given predicate and x, y range over $\{0,1\}^n$. Razborov and Sherstov [20] answered an old question of Babai, Frankl, and Simon [6] by giving the first exponential sign-rank lower bounds on a function in AC^0. Specifically, they gave a matrix $A = [F(x,y)]_{x,y}$ for a function $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ in AC^0, such that A has sign-rank $\exp(\Omega(n^{1/3}))$.

Our work strengthens and generalizes the results of Razborov and Sherstov on the sign-rank of AC^0.

1.1 Our Results

The *threshold degree* of a function $h\colon \{-1,1\}^n \to \{-1,1\}$, denoted $\deg_\pm(h)$, is the least degree of a real polynomial that agrees in sign with h at all inputs. Minsky and Papert [18] famously showed that the threshold degree of the DNF formula $\mathrm{MP}_n(x) = \vee_{i=1}^{n^{1/3}} \wedge_{j=1}^{n^{2/3}} x_{ij}$ — now known as the Minsky-Papert DNF — is $\Omega(n^{1/3})$. This is the same function that Razborov and Sherstov used to prove their sign-rank lower bounds, and their analysis is highly tailored to the Minsky-Papert DNF. We generalize their result as follows.

For any $d > 0$, we identify a class $\mathcal{C}_d$ of functions $f : \{-1,1\}^k \to \{-1,1\}$ such that any $f \in \mathcal{C}_d$ can be transformed into a function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$, where $n = O(dk)$, for which $A = [F(x,y)]_{x,y}$ has sign rank $\exp(\Omega(d))$. Crucially, this transformation is simple in the sense that if f is computed by a polynomial-size circuit of depth t, then F is computed by a polynomial-size circuit of depth at most $t+1$ (and in some cases, F may be shallower).

In particular, the k-variate AND_k function is in $\mathcal{C}_d$ for some $d = \Omega(k^{1/2})$. Our transformation of AND_k into a function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ for $n = O(k^{3/2})$ recovers Razborov and Sherstov's function, with the same sign-rank bound of $\exp(\Omega(k^{1/2})) = \exp(\Omega(n^{1/3}))$. We also identify a k-variate AC^0 function that is in $\mathcal{C}_d$ for some $d = \tilde{\Omega}(k^{2/3})$, which in turn yields new sign-rank lower bounds for AC^0.

The precise definition of $\mathcal{C}_d$ is rather technical, so for expository purposes, we restrict ourselves to an informal statement of this result in this introduction. We define $\mathcal{C}_d$ formally in Section 2.2.

Informal description of the class $\mathcal{C}_d$. Our class $\mathcal{C}_d$ consists of all functions of the form $f : \{-1,1\}^k \to \{-1,1\}$, where f satisfies the following (informally stated) property: there exists a "small" set $S \subseteq f^{-1}(+1)$ such that f cannot be uniformly approximated to error $1/2$ by degree d polynomials, even under the promise that the input x is in $f^{-1}(-1) \cup S$. The precise definition of $\mathcal{C}_d$ is based on a *dual* (in the sense of linear programming duality) interpretation of this property.

Transforming functions in $\mathcal{C}_d$ to functions with high sign-rank. For $g : \{-1,1\}^m \to \{-1,1\}$ and $f : \{-1,1\}^k \to \{-1,1\}$, the notation $g \circ f = g(f,\ldots,f)$ denotes the function on $n = mk$ bits

obtained by block-composing g with f. Let OR_m and AND_m denote the logical OR and AND functions on m bits, respectively. Let C be a sufficiently large universal constant. Given a function $f : \{-1,1\}^k \to \{-1,1\}$, let $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ be defined by

$$F = \mathrm{OR}_{2d} \circ f \circ \mathrm{AND}_C \circ \mathrm{OR}_2,$$

and hence $n = 2Cdk = O(dk)$.

Theorem 1.1 (Informal)**.** For any $f \in \mathcal{C}_d$, the matrix $A = [F(x,y)]_{x,y}$ has sign-rank $\exp(\Omega(d))$.

Examples of functions in $\mathcal{C}_d$. We consider two prominent examples of functions in $\mathcal{C}_d$. As mentioned above, the first is the function $\mathrm{AND}_k : \{-1,1\}^k \to \{-1,1\}$, which we show is in $\mathcal{C}_d$ for $d = \Omega(k^{1/2})$. Hence, we recover a new proof of Razborov and Sherstov's lower bound.

Corollary 1.1. Let $\mathrm{MP}_n = \mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ be the Minsky-Papert DNF. Then $A = [\mathrm{MP}_n(x \vee y)]_{x,y}$ has sign-rank $\exp(\Omega(n^{1/3}))$.

Let $\mathrm{ED}_k \colon \{-1,1\}^k \to \{-1,1\}$ denote the well-known Element Distinctness function (defined in Section 2.6). As we will show, the function ED_k is in $\mathcal{C}_d$ for some $d = \tilde{\Omega}(k^{2/3})$. Hence, we obtain the following corollary, which improves Razborov and Sherstov's lower bound on the sign-rank of AC^0 from $\exp(\Omega(n^{1/3}))$ to $\exp\left(\tilde{\Omega}(n^{2/5})\right)$.

Corollary 1.2. Let $F_n^{\mathrm{ED}} = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}} \circ \mathrm{AND}_C \circ \mathrm{OR}_2$. Then $A = [F_n^{\mathrm{ED}}(x,y)]_{x,y}$ has sign-rank $\exp(\tilde{\Omega}(n^{2/5}))$.

As discussed in Section 2.6, the function F_n^{ED} is computed by a depth-3 AC^0 circuit with logarithmic bottom fan-in.

1.2 Applications

We describe applications of Corollary 1.2 to communication complexity, learning theory, and circuit complexity in detail in Appendix A. Here, we briefly describe these applications.

- Razborov and Sherstov's result yielded a function in the communication complexity class $\mathbf{PH}^{\mathbf{cc}}$ (the communication analog of the polynomial hierarchy) that requires unbounded error communication complexity $\Omega(n^{1/3})$. This was the first separation between the communication complexity classes $\mathbf{PH}^{\mathbf{cc}}$ and $\mathbf{UPP}^{\mathbf{cc}}$, answering a longstanding open problem of Babai, Frankl, and Simon [6]. We improve this separation, giving a function in the communication complexity class $\mathbf{PH}^{\mathbf{cc}}$ (indeed, in $\mathbf{\Sigma_2^{cc}}$) that requires unbounded error communication complexity $\tilde{\Omega}(n^{2/5})$.

- Razborov and Sherstov's result implied that learning algorithms in the sign-rank framework cannot PAC learn DNF formulae in time less than $\exp(O(n^{1/3}))$. This essentially matches the $\exp(\tilde{O}(n^{1/3}))$ runtime of the sign-rank based algorithm of Klivans and Servedio [15]. It is reasonable to ask whether the sign-rank framework can be used to learn depth-3 (or deeper) AC^0 circuits in the same $\exp(\tilde{O}(n^{1/3}))$ time bound. Our results rule this out, showing that sign-rank based learning algorithms require time $\exp(\tilde{\Omega}(n^{2/5}))$ to learn depth-3 AC^0 circuits, even when the bottom fan-in is $O(\log n)$.

- Razborov and Sherstov's result implied an exponential (specifically, $\exp(\Omega(n^{1/3}))$) lower bound on the size of threshold-of-majority circuits computing a function in AC^0. We improve their lower bound to $\exp\left(\tilde{\Omega}(n^{2/5})\right)$.

1.3 Our Techniques

It is well-known that the threshold degree of any function $h\colon \{-1,1\}^n \to \{-1,1\}$ is characterized by an (exponentially large) linear program. Using this formulation, if $\deg_{\pm}(h) = d$, then strong LP duality guarantees the existence of a dual solution μ that *witnesses* the fact that $\deg_{\pm}(h) \geq d$. Specifically, μ takes the form of a distribution on $\{-1,1\}^n$ such that h is uncorrelated under μ with all polynomials of degree at most d, i.e., $\sum_{x\in\{-1,1\}^n} \mu(x)\cdot h(x)\cdot p(x) = 0$ for all polynomials p of degree at most d. Razborov and Sherstov refer to μ as a *d-orthogonalizing distribution* for h (see Section 2.5 below for details).

In order to establish sign-rank lower bounds for the matrix $A = [(h\circ \mathrm{AND}_C)(x\vee y)]_{x,y}$, Razborov and Sherstov extended a lemma of Forster to show that it is enough to give an orthogonalizing distribution μ for h that additionally satisfies a *smoothness* property (cf. Theorem 4.1 in Section 4 for details). Specifically, a d-orthogonalizing distribution for h is said to be smooth if $\mu(x) = \exp(-O(d))$ for all but an $\exp(-\Omega(d))$ fraction of inputs $x \in \{-1,1\}^n$. Intuitively, this means that μ is smooth if it places "noticeable" mass on "almost all" inputs.

Razborov and Sherstov proved (non-constructively) that there exists a smooth d-orthogonalizing distribution for the Minsky-Papert DNF, for $d = n^{1/3}$. To generalize their result, for any $d > 0$ and any function $f \in \mathcal{C}_d$, we explicitly construct a smooth d-orthogonalizing distribution for the function $\mathrm{OR}_d \circ f$. Our construction combines new ideas with insights of Razborov and Sherstov, and ideas from prior works by the authors and Sherstov [9, 27] that constructed (non-smooth) orthogonalizing distributions for functions of the form $\mathrm{OR}\circ f$.

2 Preliminaries

2.1 Notation

We work with Boolean functions $f : \{-1,1\}^k \to \{-1,1\}$, where -1 corresponds to logical TRUE and $+1$ corresponds to logical FALSE. For $x \in \{-1,1\}^k$, let $|x| = \#\{i : x_i = -1\}$ denote the Hamming weight of x. Note that $|x|$ is computed by the linear function $|x| = \frac{k}{2} - \frac{1}{2}\sum_{i=1}^k x_i$.

2.2 Symmetrization

Definition 2.1. Let $T : \{-1,1\}^k \to D$, where D is a finite subset of $\mathbb{R}^n$ for some $n \in \mathbb{N}$. The map T is *degree non-increasing* if for every polynomial $p : \{-1,1\}^k \to \mathbb{R}$, there exists a polynomial $q : D \to \mathbb{R}$ with $\deg q \leq \deg p$ such that

$$q(T(x)) = \mathop{\mathbb{E}}_{y \text{ s.t. } T(y)=T(x)} [p(y)]$$

for every $x \in \{-1,1\}^k$. We say that a degree non-increasing map T *symmetrizes* a function $f : \{-1,1\}^k \to \mathbb{R}$ if $f(x) = f(y)$ whenever $T(x) = T(y)$, and in this case we say that T is a *symmetrization* for f.

The canonical example of a degree non-increasing map is that which computes the Hamming weight.

Lemma 2.1 (Minsky and Papert [18])**.** The map $T : \{-1,1\}^k \to \{0,1,\ldots,k\}$ defined by $T(x) = |x|$ is degree non-increasing. Hence, T is a symmetrization for any symmetric Boolean function.

For any function $\psi\colon \{-1,1\}^k \to \mathbb{R}$, a symmetrization $T\colon \{-1,1\}^k \to D$ for ψ induces a symmetrized function $\widetilde{\psi}\colon D \to \mathbb{R}$ defined via $\widetilde{\psi}(z) := \mathbb{E}_{x \in T^{-1}(z)} \psi(x)$. (If $T^{-1}(z)$ is empty, then we define $\widetilde{\psi}(z) = 0$). It will also be convenient to define an "unnormalized" version $\hat{\psi}$ of $\widetilde{\psi}$, defined via $\hat{\psi}(z) := \sum_{x \in T^{-1}(z)} \psi(x)$. Observe that if μ is a distribution on $\{-1,1\}^k$, then $\hat{\mu}$ is a distribution on D.

Similarly, let $T\colon \{-1,1\}^k \to D$ be a degree non-increasing map. A function $\hat{\psi} : D \to \mathbb{R}$ naturally induces an un-symmetrized function $\psi : \{-1,1\}^k \to \mathbb{R}$ by setting $\psi(x) = \frac{1}{|T^{-1}(z)|}\hat{\psi}(z)$ where $z = T(x)$. That is, ψ spreads the mass of $\hat{\psi}(z)$ out evenly over points $x \in T^{-1}(z)$. Observe that, for any $\hat{\psi}$ and any degree non-increasing map T, the induced function ψ is symmetrized by T.

We will often pass back and forth between a function ψ on $\{-1,1\}^k$ and its symmetrized versions $\widetilde{\psi}$ and $\hat{\psi}$ on D, when the underlying symmetrization $T\colon \{-1,1\}^k \to D$ is understood.

2.3 Norms and Inner Products

For a function $\psi : \{-1,1\}^k \to \mathbb{R}$, define the ℓ_1 norm of ψ by $\|\psi\|_1 = \sum_{x \in \{-1,1\}^k} |\psi(x)|$. For functions $\psi, \varphi : \{-1,1\}^k \to \mathbb{R}$, denote the inner product $\langle \psi, \varphi \rangle = \sum_{x \in \{-1,1\}^k} \psi(x)\varphi(x)$. We say a function $\psi : \{-1,1\}^k \to \mathbb{R}$ has *pure high degree d* if $\langle \psi, p \rangle = 0$ for every polynomial $p : \{-1,1\}^k \to \mathbb{R}$ of degree less than d.

2.4 Dual Objects and the Class $\mathcal{C}_d$

Central to our work is the following definition of a "dual object." We show that whenever a Boolean function f can be associated with such a dual object, then f can be transformed into a function F such that $[F(x,y)]_{x,y}$ has high sign-rank.

Definition 2.2. Let $f : \{-1,1\}^k \to \{-1,1\}$, and let $T : \{-1,1\}^k \to D$ be a (degree non-increasing) symmetrization for f. Let $\hat{\psi}\colon D \to \mathbb{R}$ be any function, and let ψ be the associated function on $\{-1,1\}^k$ induced by T. We say that $\hat{\psi}$ is a (d, ε, η)-*dual object* for f (with respect to T) if:

- $\langle \psi, f \rangle \geq \varepsilon$ (1)
- $\|\psi\|_1 = 1$ (2)
- $\langle \psi, p \rangle = 0$ for every polynomial $p : \{-1,1\}^k \to \mathbb{R}$ with $\deg p < d$ (3)
- $f(x) = -1 \implies \psi(x) < 0$ (4)
- $\hat{\psi}(z_+) \geq \eta$ for some $z_+ \in D$ satisfying $\widetilde{f}(z_+) = 1$ (5)

Definition 2.2 is motivated by a recent line of work establishing lower bounds for polynomial approximations via linear programming duality. We direct the reader to [7–9, 21, 22, 25–27, 29] for thorough discussions of this technique and its applications to longstanding open questions in complexity theory. In short, one can use linear programming duality to show that the existence of a $(d, 2\eta, \eta)$-dual object for a function f is implied by the *non-existence* of a degree d polynomial that approximates f in a certain precise sense. In a bit more detail (and still simplifying a little), a $(d, 2\eta, \eta)$-dual object for f always exists if f cannot be uniformly approximated to error 2η by any

degree d polynomial, even under the promise that the input x is in $f^{-1}(-1) \cup T^{-1}(z_+)$. We will not use this *primal* interpretation of dual objects in our analysis, but we spell out this implication in Appendix B for completeness and intuition.

Motivated by the study of uniform approximation of Boolean functions by polynomials, several works [7, 8, 28] have constructed dual objects directly. In particular, work of Špalek [28] and the authors [7] explicitly constructed an appropriate dual object for the AND function.

Lemma 2.2 (cf. [7, 28])**.** Let $T : \{-1,1\}^k \to \{0,1,\ldots,k\}$ be the degree non-increasing map $T(x) = |x|$ that computes the Hamming weight. The function AND_k has a $(d, 1/2, 1/4)$-dual object with respect to T for $d = \Omega(\sqrt{k})$.

We are now ready to define the class $\mathcal{C}_d$ of functions to which our techniques can be applied to yield sign-rank lower bounds.

Definition 2.3. Let $f : \{-1,1\}^k \to \{-1,1\}$ be a Boolean function, and let $d > 0$. Then f is in the class $\mathcal{C}_d$ if there exists a symmetrization $T : \{-1,1\}^k \to D$ for f such that:

- there exists a $(d, 1/2, 1/4)$-dual object for f with respect to T, and
- the function f evaluates to TRUE (i.e. $f(x) = -1$) for at most a 2^{-d} fraction of inputs $x \in \{-1,1\}^k$.

2.5 Orthogonalizing Distributions

As indicated in Section 1.3, our analysis will make essential use of orthogonalizing distributions, which represent a dual formulation of the notion of threshold degree.

Definition 2.4. A distribution $\mu : \{-1,1\}^n \to [0,1]$ is *d-orthogonalizing* for a function $h : \{-1,1\}^n \to \{-1,1\}$ if

$$\mathop{\mathbb{E}}_{x\sim\mu}[h(x)p(x)] = 0$$

for every polynomial $p : \{-1,1\}^n \to \mathbb{R}$ with $\deg p < d$. In other words, μ is d-orthogonalizing for h if the function $\mu(x)h(x)$ has pure high degree d.

2.6 The Element Distinctness Function

The Boolean function $\mathrm{ED}_k : \{-1,1\}^k \to \{-1,1\}$ is defined as follows. For simplicity, assume that $k = K \log_2 K$, where K is a power of 2. The function interprets its input x as blocks $x_1, \ldots, x_K$, where each $x_i \in \{-1,1\}^{\log_2 K}$. Each x_i is interpreted as the binary representation of $g_x(i)$ for a function $g_x : [K] \to [K]$. $\mathrm{ED}_k(x)$ is defined to equal -1 iff the function g_x is 1-to-1.

Observe that ED_k is symmetric with respect to permutations of the domain and range of g_x. That is, if $x, y \in \{-1,1\}^k$ are such that there exist permutations π, σ of $[K]$ with $g_x = \pi \circ g_y \circ \sigma$, then $\mathrm{ED}_k(x) = \mathrm{ED}_k(y)$.

In Appendix C, we show that these symmetries imply the existence of a symmetrization T for ED_k and an associated dual object.

Lemma 2.3. There exists a symmetrization $T : \{-1,1\}^k \to [K]^K$ for the Element Distinctness function $\mathrm{ED}_k : \{-1,1\}^k \to \{-1,1\}$ such that ED_k has a $(d, 1/2, 1/4)$-dual object (with respect to the map T), for some $d = \Omega(K^{2/3}/\log K)$.

Remark 1. In fact, an explicit dual object for ED_k was constructed in our prior work [8]. For completeness, in Appendix C we give an alternative *primal*-based proof of the existence of a dual object for ED_k. The proof is based on Aaronson and Shi's influential lower bound of $\Omega(K^{2/3})$ on the *approximate degree*[1] of ED_k.

The Element Distinctness function is computed by a natural CNF formula:

$$\mathrm{ED}_k(x_1,\ldots,x_K) = \bigwedge_{r=1}^{K} \bigwedge_{i\neq j} ((x_i \neq r) \vee (x_j \neq r)).$$

Notice that the fan-in of each bottom OR gate is only $2K \leq 2\log_2 k$. Recall (cf. Corollary 1.2) that our aim is to prove a sign-rank lower bound for the function $F_n^{\mathrm{ED}}(x,y) = (\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}} \circ \mathrm{AND}_C)(x \vee y)$. Using the CNF for Element Distinctness described above, the function F_n^{ED} is naturally computed by an AC^0 circuit Γ of depth 5, with an OR gate at the top. However, as we now explain, F_n^{ED} is actually computable by a *depth-3* AC^0 circuit with logarithmic bottom fan-in.

Number the layers of Γ from 1 to 5, with layer 1 corresponding to the OR gate at the top. Since each OR gate at layer 3 of Γ has fan-in $O(\log n)$ (and the gates at layers 4 and 5 have constant fan-in), the sub-circuits rooted at each gate at layer 3 of Γ are functions of only $O(\log n)$ bits of x. Since any function on $O(\log n)$ inputs can be computed by a $\mathrm{poly}(n)$ size CNF with logarithmic bottom fan-in, we can replace each sub-tree rooted at layer 3 of Γ with such a CNF, to obtain a circuit Γ' of depth 4, in which layers 2 and 3 of Γ' both consist of AND gates. Collapsing layers 3 and 4 into a single layer yields a polynomial size depth 3 circuit with logarithmic bottom fan-in that computes F_n^{ED}.

3 Constructing a Smooth Orthogonalizing Distribution

Sherstov [27] showed that whenever f has a $(d_1, 1/2, 0)$-dual object[2], the function $h_m := \mathrm{OR}_m \circ f$ has a d-orthogonalizing distribution for $d = \min\{m, d_1\}$. The goal of this section, and the main technical contribution of the paper, is to prove that whenever f has a $(d_1, 1/2, \eta)$-dual object for $\eta > 0$, the function h_m has a d-orthogonalizing distribution that places significant mass on each input $x \in h_m^{-1}(1)$. More precisely, we show:

Theorem 3.1. Suppose that $f : \{-1,1\}^k \to \{-1,1\}$ has a $(d_1, 1/2, \eta)$-dual object, and let $h_m = \mathrm{OR}_m \circ f$. Then there exists a d-orthogonalizing distribution $\mu : \{-1,1\}^{mk} \to [0,1]$ for h_m such that $\mu(x) \geq 4^{-(m+d+1)} \eta^{-m/2} 2^{-mk}$ for every $x \in h_m^{-1}(1)$, where $d = \min\{m/2, d_1\}$.

Combining this theorem with Lemmas 2.2 and 2.3 yields smooth orthogonalizing distributions for the functions $\mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ and $\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}$.

Corollary 3.1. There exists a d-orthogonalizing distribution μ for $h = \mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ such that $\mu(x) \geq 2^{-O(d)} 2^{-n}$ on each $x \in h^{-1}(1)$, for some $d = \Omega(n^{1/3})$.

Corollary 3.2. There exists a d-orthogonalizing distribution μ for $h = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}$ such that $\mu(x) \geq 2^{-O(d)} 2^{-n}$ on each $x \in h^{-1}(1)$, for some $d = \tilde{\Omega}(n^{2/5})$.

[1] The approximate degree of a Boolean function f is the minimum degree of a real polynomial for which $|p(x) - f(x)| \leq 1/3$ for all Boolean inputs x.

[2] The existence of a $(d_1, 1/2, 0)$-dual object for f is in fact a dual formulation of the property that f has *one-sided approximate degree* at least d_1. See [7, 27] for the definition of one-sided approximate degree.

3.1 Proof of Theorem 3.1

3.1.1 Notation

Let $f : \{-1,1\}^k \to \{-1,1\}$ be as in the statement of Theorem 3.1, and let $T : \{-1,1\}^k \to D$ be the symmetrization for f associated with the assumed $(d_1, 1/2, \eta)$-dual object for f. Define $T^m \colon \{-1,1\}^{mk} \to D^m$ by $T^m(x_1, \ldots, x_m) := (T(x_1), \ldots, T(x_m))$. Since T is degree non-increasing, it is easy to see that T^m is also degree non-increasing. Moreover, T^m is a symmetrization for h_m. The map T^m induces a symmetrized version $\widetilde{h}_m \colon D^M \to \mathbb{R}$ of h_m given by $\widetilde{h}_m = \mathrm{OR}_m \circ \widetilde{f}$.

3.1.2 Proof Outline

Let $Z^+ := \widetilde{h}_m^{-1}(1) \subseteq D^m$. At a high level, our proof will produce, for every $z \in Z^+$, a d-orthogonalizing distribution μ_z that is targeted to z, in the sense that

$$\hat{\mu}_z(z) \geq 2^{-O(m+d)} \cdot \eta^{-O(m)}.$$

Since the property of d-orthogonalization is preserved under averaging, the distribution $\mu = \frac{1}{|Z^+|} \sum_{z \in Z^+} \mu_z$ remains d-orthogonalizing, and places the required amount of probability mass on each input $x \in T^{-1}(Z^+) = h_m^{-1}(1)$. The goal therefore becomes to construct these targeted distributions μ_z. We do this in two stages.

Stage 1. In the first stage (see Claim 1 below), we construct distributions μ_z for every z belonging to a highly structured subset $G \subset Z^+$ that we now describe. Let $c \in \widetilde{f}^{-1}(1)$ denote the point on which the dual object $\hat{\psi}$ for f has $\hat{\psi}(c) \geq \eta$ (cf. Condition (5) within Definition 2.2). The set G consists of inputs in Z^+ for which $c \in D$ is repeated many times (specifically, at least $m/2$ times).

Stage 2. In the second stage (see Claim 2 below), we show that given the family of distributions $\{\mu_z : z \in G\}$ constructed in Stage 1, we can construct appropriate distributions μ_z for z belonging to the entire set Z^+.

Both stages can be viewed as generalized dual counterparts to analogous statements in the work of Razborov and Sherstov (cf. [20, Lemma 3.4] and [20, Theorem 3.6] respectively). Taking a dual perspective allows us to identify general properties (Definition 2.2) of a dual object for f that enable the construction of a smooth orthogonalizing distribution. This results in a much more general and modular framework for proving the existence of these distributions. Our framework also has the advantage of constructing smooth orthogonalizing distributions explicitly.

3.1.3 Proof Details

We begin with a relatively simple lemma that shows that the function $\mathrm{OR}_{m/2} \circ f$ has a d-orthogonalizing distribution μ such that $\hat{\mu}$ places a lot of probability mass on a particular highly structured input, where $d = \min\{d_1, m/2\}$. This distribution is an important building block in the proof of Claim 1 below.

Lemma 3.3. Let $\ell = m/2$, and let f, T, and T^ℓ be as above. Consider the function $h_\ell : \{-1,1\}^{k\ell} \to \{-1,1\}$ defined by $h_\ell(x_1, \ldots, x_\ell) = \mathrm{OR}_\ell(f(x_1), \ldots, f(x_\ell))$. There exists a function $\psi : \{-1,1\}^{k\ell} \to [0,1]$ symmetrized by T^ℓ with the following properties.

- ψ agrees in sign with h_ℓ. That is, $\psi(x) \cdot h_\ell(x) \geq 0$ for all $x \in \{-1,1\}^{k\ell}$ (6)

- $\|\psi\|_1 = 1$ (7)
- ψ has pure high degree at least $d = \min\{\ell, d_1\}$ (8)
- There exists a $c \in D$ such that $\widetilde{f}(c) = 1$ and $\hat{\psi}(\underbrace{c, \ldots, c}_{\ell \text{ times}}) \geq \eta^{-\ell}/2$ (9)

We remark that Conditions (6)-(8) are equivalent to requiring that $\mu := \psi \cdot h_\ell$ is a d-orthogonalizing distribution for h_ℓ, where $d = \min\{\ell, d_1\}$.

Proof. Sherstov [27] showed that when the function f has a $(d_1, 1/2, 0)$-dual witness, then there is a function ψ satisfying Conditions (6)-(8). For completeness, we verify these properties in Appendix D.1. Here, we show that if, in addition, f has a $(d_1, 1/2, \eta)$-dual witness with $\eta > 0$, then Sherstov's construction yields a function $\hat{\psi}$ that also satisfies Condition (9).

Construction of ψ. Let $\hat{\psi}_1 : D \to \mathbb{R}$ be a $(d_1, 1/2, \eta)$-dual object for f (with respect to the symmetrization T), and decompose $\hat{\psi}_1$ as the difference of non-negative functions $\hat{\nu}_+ - \hat{\nu}_-$. Let

$$\hat{\psi}_0 = \begin{cases} \hat{\nu}_- = \max\{-\hat{\psi}_1, 0\} & \text{on } \widetilde{f}^{-1}(1) \\ \left(\frac{1}{\langle \psi_1, f \rangle} - 1\right) \hat{\psi}_1 & \text{on } \widetilde{f}^{-1}(-1). \end{cases}$$

Recall that $1/2 \leq \langle \psi_1, f \rangle \leq 1$ (cf. Condition (1) in Definition 2.2)). Hence, the factor $\left(\frac{1}{\langle \psi_1, f \rangle} - 1\right)$ in the second case of the definition of $\hat{\psi}_0$ satisfies:

$$0 \leq \left(\frac{1}{\langle \psi_1, f \rangle} - 1\right) \leq 1. \tag{10}$$

Moreover, Condition (5) of Definition 2.2 guarantees that $\hat{\psi}_1 = \hat{\nu}_-$ on $\widetilde{f}^{-1}(-1)$, so we can rewrite this definition as

$$\hat{\psi}_0 = \begin{cases} \hat{\nu}_- = \max\{-\hat{\psi}_1, 0\} & \text{on } \widetilde{f}^{-1}(1) \\ \left(\frac{1}{\langle \psi_1, f \rangle} - 1\right) \hat{\nu}_- & \text{on } \widetilde{f}^{-1}(-1). \end{cases}$$

Sherstov constructs a function $\hat{\Psi} : D^\ell \to \mathbb{R}$ defined by

$$\hat{\Psi}(z) = \prod_{i=1}^{\ell} \hat{\nu}_+(z_i) - \prod_{i=1}^{\ell} \hat{\nu}_-(z_i) + \prod_{i=1}^{\ell} \hat{\psi}_0(z_i).$$

The desired function is then given by $\hat{\psi}(z) = \hat{\Psi}(z)/\|\Psi\|_1$.

Analysis of ψ. For completeness, Appendix D.1 proves that under this definition, ψ satisfies Conditions (6)-(8). We now turn to verifying that $\hat{\psi}$ satisfies Condition (9).

Let $c \in D$ be such that $\widetilde{f}(c) = 1$ and $\hat{\psi}_1(c) \geq \eta$ (cf. Condition (5) of Definition 2.2). Since $\widetilde{f}(c) = 1$, we have $\hat{\psi}_0(c) = \hat{\nu}_-(c) = 0$, so $\hat{\Psi}(c, \ldots, c) = \hat{\psi}_1(c)^\ell \geq \eta^\ell$.

To establish that we indeed have $\hat{\psi}(c, \ldots, c) \geq \eta^\ell/2$, we now verify that $\|\Psi\|_1 \leq 2$. To do so, first consider $z \in \widetilde{h}_\ell^{-1}(-1)$. Then there exists an index i for which $\widetilde{f}(z_i) = -1$. By Condition (4) of Definition 2.2 of a dual object, this means that $\hat{\nu}_+(z_i) = 0$. Hence,

$$|\hat{\Psi}(z)| = \left| -\prod_{i=1}^{\ell} \hat{\nu}_-(z_i) + \prod_{i=1}^{\ell} \hat{\psi}_0(z_i) \right| \leq 2 \prod_{i=1}^{\ell} \hat{\nu}_-(z_i),$$

since Expression (10) implies that $|\hat{\psi}_0| \le \hat{\nu}_-$ pointwise.

If instead $z \in \widetilde{h}_\ell^{-1}(1)$, then $\widetilde{f}(z_i) = 1$ for every i. Hence, $\hat{\psi}_0(z_i) = \hat{\nu}_-(z_i)$ for every i, and

$$\hat{\Psi}(z) = \prod_{i=1}^{\ell} \hat{\nu}_+(z_i).$$

Thus we may compute

$$\begin{aligned}
\|\Psi\|_1 &= \sum_{z \in D^\ell} |\hat{\Psi}(z)| \\
&\le \sum_{z \in \widetilde{h}_\ell^{-1}(1)} \prod_{i=1}^{\ell} \hat{\nu}_+(z_i) + 2 \sum_{z \in \widetilde{h}_\ell^{-1}(-1)} \prod_{i=1}^{\ell} \hat{\nu}_-(z_i) \\
&\le 2 \sum_{z \in D^\ell} \prod_{i=1}^{\ell} |\hat{\psi}_1(z_i)| = 2 \prod_{i=1}^{\ell} \sum_{z_i \in D} |\hat{\psi}_1(z_i)| = 2.
\end{aligned}$$

Thus, the normalized function $\hat{\psi} = \hat{\Psi}/\|\Psi\|_1$ satisfies $\hat{\psi}(c, \dots, c) \ge \eta^\ell/2$. □

To complete Stage 1 of our proof, we show that for every input $w \in D^m$ that is close in Hamming distance to the special point $(\underbrace{c, \dots, c}_{m \text{ times}})$, there is an orthogonalizing distribution for h_m that places substantial weight on w. Let $G \subset Z^+ = \widetilde{h}_m^{-1}(1)$ denote the set of inputs in $\widetilde{h}_m^{-1}(1)$ that take the value c on at least $m/2$ coordinates. That is,

$$G = \{z \in Z^+ : \exists i_1, \dots, i_{m/2} \text{ s.t. } z_{i_1} = \cdots = z_{i_{m/2}} = c\}.$$

Claim 1. Let G be as above. For every $w = (w_1, \dots, w_m) \in G$, there exists a d-orthogonalizing distribution $\nu_w : \{-1,1\}^{km} \to [0,1]$ for h_m such that ν_w is symmetrized by T^m and $\hat{\nu}_w(w) \ge \eta^{m/2}/2$.

Proof. Let $I = \{i_1, \dots, i_{m/2}\}$ denote the first $m/2$ coordinates on which w takes the value c. Define the distribution $\hat{\nu}_w$ by

$$\hat{\nu}_w(z) = \begin{cases} |\hat{\psi}(z_{i_1}, \dots, z_{i_{m/2}})| & \text{if } z_i = w_i \text{ for all } i \notin I \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{\psi}$ is the function from Lemma 3.3 for $\ell = m/2$. It is immediate from the definition that $\hat{\nu}_w$ is a distribution on D^m, and hence ν_w is a distribution on $\{-1,1\}^{km}$. Moreover, $\hat{\nu}_w(w) \ge \eta^{m/2}/2$.

To show that ν_w is d-orthogonalizing, let $p_1, \dots, p_m$ be polynomials over $\{-1,1\}^k$ whose degrees sum to at most $d-1$. Let $q_1, \dots, q_m \colon D \to \mathbb{R}$ denote polynomials satisfying the property that for all i and all z in the image of T, $q_i(z_i) := \mathbb{E}_{x \in T^{-1}(z_i)}[p_i(z_i)]$. (Since T is degree non-increasing, there exist such q_i's whose degrees sum to at most $d-1$, but we will not make use of this property in this proof).

Observe that:

$$\begin{aligned}
\sum_{x=(x_1,\ldots,x_m)\in\{-1,1\}^{km}} \nu_w(x)h_m(x)\prod_{i=1}^{m} p_i(x_i) &= \sum_{z=(z_1,\ldots,z_m)\in D^m} \hat{\nu}_w(z)\widetilde{h}_m(z)\prod_{i=1}^{m} q_i(z_i) \\
&= \sum_{\substack{z=(z_1,\ldots,z_m)\in D^m \text{ s.t.} \\ \forall i\notin I\ z_i=w_i}} |\hat{\psi}(z_{i_1},\ldots z_{i_{m/2}})|\cdot \widetilde{h}_m(z)\prod_{i=1}^{m} q_i(z_i) \\
&= \sum_{\substack{z=(z_1,\ldots,z_m)\in D^m \text{ s.t.} \\ \forall i\notin I\ z_i=w_i}} \hat{\psi}(z_{i_1},\ldots z_{i_{m/2}})\prod_{i=1}^{m} q_i(z_i) \\
&= \left(\prod_{i\notin I} q_i(w_i)\right) \sum_{z=(z_1,\ldots,z_{m/2})\in D^{m/2}} \hat{\psi}(z)\prod_{i\in I} q_i(z_i) \\
&= \left(\prod_{i\notin I} q_i(w_i)\right) \sum_{x=(x_1,\ldots,x_{m/2})\in\{-1,1\}^{k\cdot m/2}} \psi(x)\prod_{i\in I} p_i(x_i) \\
&= 0.
\end{aligned}$$

Here, the second equality holds by definition of $\hat{\nu}_w$. To see that the third equality holds, recall that ψ sign-represents $h_{m/2}$, and observe that $h_m(z) = h_{m/2}(z_{i_1},\ldots,z_{i_{m/2}})$ whenever $z_i = w_i$ for all $i \in I$ (this holds because $\widetilde{f}(w_i) = 1$ for all i). The final equality holds because ψ has pure high degree at least d, and $\prod_{i\in I} p_i(x)$ is a polynomial of total degree at most $d-1$. □

We now proceed to Stage 2 of our proof, in which we use the distributions constructed in Claim 1 to give orthogonalizing distributions that place significant weight on *any* input $x \in h_m^{-1}(1)$.

Claim 2. Let G be as before, and suppose that for every $w \in G$ there exists a d-orthogonalizing distribution $\nu_w : \{-1,1\}^{km} \to [0,1]$ for h_m that is symmetrized by T^m, and satisfies $\hat{\nu}_w(w) \geq \delta$. Then for every $v \in (Z^+ \setminus G)$, there exists a d-orthogonalizing distribution ρ_v that is symmetrized by T^m, and $\hat{\rho}_v(v) \geq \delta/4^{m+d}$.

The main technical ingredient in the proof of Claim 2 is the construction of a function $\varphi : \{0,1\}^m \to \mathbb{R}$ of pure high degree d for which $\varphi(1^m)$ is "large". This can be viewed as a dual formulation of a bound on the growth of low-degree polynomials. The construction of φ appears as part of the proof of such a bound in [20].

Remark 2. We choose to state Lemma 3.4 below for a function $\varphi : \{0,1\}^m \to \mathbb{R}$, rather than applying our usual convention of working with functions over $\{-1,1\}^m$, because it makes various statements in the proof of Claim 2 cleaner. To clarify the terminology below, we say a function $\varphi : \{0,1\}^m \to \mathbb{R}$ has *pure high degree* d if $\sum_{x\in\{0,1\}^m} \varphi(x)\cdot p(x) = 0$ for every polynomial $p : \{0,1\}^m \to \mathbb{R}$ of degree less than d. The Hamming weight function $|\cdot| : \{0,1\}^m \to [m]$ counts the number of 1's in its input, i.e. $|s| = s_1 + s_2 + \cdots + s_m$.

Lemma 3.4 (cf. [20, Proof of Lemma 3.2])**.** Let d be an integer with $0 \leq d \leq m-1$. Then there exists a function $\varphi : \{0,1\}^m \to \mathbb{R}$ such that

- $\varphi(1^m) = 1$ (11)

- $\varphi(x) = 0$ for all $d \leq |x| < m$ (12)
- φ has pure high degree at least d (13)
- $\sum_{|x| \leq d} |\varphi(x)| \leq 2^d \binom{m}{d}$ (14)

Proof of Claim 2. Fix $v \in (Z^+ \setminus G)$. Define an auxiliary function $\hat{\varphi}_v : D^m \to [0,1]$ as follows. For any $z = (z_1, \ldots, z_m)$, let

$$\hat{\varphi}_v(z) := \sum_{\substack{s \in \{0,1\}^m \text{ s.t.} \\ \forall i \ z_i = (1-s_i)\cdot c + s_i \cdot v_i}} \varphi(s),$$

where φ is as in Lemma 3.4, with d set as in the conclusion of Claim 1 (observe that if there is some z_i such that $z_i \neq c$ and $z_i \neq v_i$, then $\hat{\varphi}_v(z) = 0$).

Letting φ_v denote the function on $\{-1,1\}^{km}$ induced from $\hat{\varphi}_v$ by T^m, we record some properties of φ_v and $\hat{\varphi}_v$.

- $\hat{\varphi}_v(v) = \varphi(1^m) = 1$ (15)
- $\operatorname{supp} \hat{\varphi}_v \subset G \cup \{v\}$ (16)
- φ_v has pure high degree at least d (17)
- $\|\varphi_v\|_1 \leq 2^d \binom{m}{d} + 1$ (18)
- $\hat{\varphi}_v$ is supported on at most $\frac{1}{2}2^m + 1$ points in D^m (19)

Verifying Conditions (15)-(19). Conditions (15), (16), and (19) are immediate from the definition of $\hat{\varphi}_v$, combined with Conditions (11) and (12) of Lemma 3.4. For Condition (17), it is enough to show that if $p_1, \ldots, p_m$ are polynomials over $\{-1,1\}^k$ whose degrees sum to less than d, then $\sum_{x=(x_1,\ldots,x_m)\in\{-1,1\}^{km}} \varphi_v(x) \prod_{i=1}^m p_i(x_i) = 0$. To establish this, let $q_1, \ldots, q_m \colon D \to \mathbb{R}$ denote polynomials satisfying $\deg(q_i) \leq \deg(p_i)$, and such that for all i and all z_i in the image of T, $q_i(z_i) := \mathbb{E}_{x \in T^{-1}(z_i)}[p_i(z_i)]$. Such polynomials are guaranteed to exist, since T is degree non-increasing. Then:

$$\begin{aligned}
\sum_{x=(x_1,\ldots,x_m)\in\{-1,1\}^{km}} \varphi_v(x) \prod_{i=1}^m p_i(x_i) &= \sum_{z=(z_1,\ldots,z_m)\in D^m} \hat{\varphi}_v(z) \prod_{i=1}^m q_i(z_i) \\
&= \sum_{z=(z_1,\ldots,z_m)\in D^m} \left(\sum_{\substack{s \in \{0,1\}^m \text{ s.t.} \\ \forall i \ z_i = (1-s_i)\cdot c + s_i \cdot v_i}} \varphi(s) \right) \prod_{i=1}^m q_i(z_i) \\
&= \sum_{s \in \{0,1\}^m} \varphi(s) \prod_{i=1}^m q_i((1-s_i)\cdot c + s_i \cdot v_i) \\
&= 0,
\end{aligned}$$

To see that the final equality holds, recall that that degrees of the polynomials q_i sum to strictly less than d. Hence, $p(s_1, \ldots, s_m) := \prod_{i=1}^{m} q_i((1 - s_i) \cdot c + s_i \cdot v_i)$ is a polynomial of degree strictly less than d over $\{-1, 1\}^m$. The final equality then follows from the fact that φ has pure high degree at least d.

To establish Condition (18), we check that

$$\sum_{z \in D^m, z \neq v} |\hat{\varphi}_v(z)| \leq \sum_{s \in \{0,1\}^m, s \neq 1^m} |\varphi(s)| \leq 2^d \binom{m}{d},$$

where the final inequality holds by Condition (14).

Construction and analysis of ρ_v. Up to normalization, the function $\varphi_v \cdot h_m$ has all of the properties that we need to establish Claim 2, except that there are locations where it may be negative. We obtain our desired orthogonalizing distribution ρ_v by adding correction terms to $\hat{\varphi}_v$ in the locations where $\hat{\varphi}_v$ may disagree with $\widetilde{h}_m$ in sign. These correction terms are derived from the distributions $\hat{\nu}_w$ whose existence are hypothesized in the statement of Claim 2. We start by defining

$$\hat{P}_v(z) = \frac{\delta}{2^d \binom{m}{d} + 1} \widetilde{h}_m(z) \hat{\varphi}_v(z) + \sum_{w \in (\operatorname{supp} \hat{\varphi}_v \setminus \{v\})} \hat{\nu}_w(z). \tag{20}$$

Observe that each w appearing in the sum on the right hand side of Eq. (20) is in the set G, owing to Condition (16). This guarantees that each term $\hat{\nu}_w$ in the sum is well-defined.

Now we check that $\hat{P}_v$ is nonnegative. Since each term $\hat{\nu}_w$ appearing in the sum on the right hand side of Eq. (20) is a distribution (and hence non-negative), it suffices to check that $\hat{P}_v(z) \geq 0$ for each point $z \in \operatorname{supp} \hat{\varphi}_v$. On each such point with $z \neq v$, Condition (18) guarantees that $\frac{\delta}{2^d \binom{m}{d} + 1} \widetilde{h}_m(z) \hat{\varphi}_v(z) \geq -\delta$. Moreover, the contribution of the sum is at least $\hat{\nu}_z(z) \geq \delta$ by hypothesis. Hence, $\hat{P}_v$ is a non-negative function.

Next, we check that normalizing $\hat{P}_v$ yields a distribution $\hat{\rho}_v := \hat{P}_v / \|P_v\|_1$ for which $\hat{\rho}_v(v) \geq \delta / 4^{m+d}$ as required. By construction, $\hat{P}_v(v) = \delta / \left(2^d \binom{m}{d} + 1\right)$. Moreover, Conditions (15), (18), and (19) together show that $\|\hat{P}_v\|_1 \leq \delta + \frac{1}{2} 2^m \leq 2^m$. Hence, $\hat{P}_v(v) \geq \delta / \left(2^m \cdot \left(2^d \binom{m}{d} + 1\right)\right) \geq \delta / \left(2^{m+d+1} \binom{m}{d}\right) \geq \delta / 2^{2m+d+1} \geq \delta / 4^{m+d}$.

Finally, we must check that $\rho_v = P_v / \|P_v\|_1$ is d-orthogonalizing for h_m. To see this, observe that $P_v \cdot h_m$ is a linear combination of the functions φ_v and $\nu_w \cdot h_m$ for $w \in (\operatorname{supp} \hat{\varphi}_v \setminus \{v\})$. Moreover, each of these functions has pure high degree at least d (φ_v does so by Condition (17), while $\nu_w \cdot h_m$ does by the fact that ν_w is d-orthogonalizing for h_m). By linearity, it follows that $P_v \cdot h_m$ has pure high degree at least d, so ρ_v is d-orthogonalizing for h_m as desired.

This completes the proof of Claim 2. □

We are now ready to combine the claims above to prove Theorem 3.1.

Proof of Theorem 3.1. By Claim 1, for every $w \in G$ there exists a d-orthogonalizing distribution $\nu_w : \{-1, 1\}^{km} \to [0, 1]$ for h_m that is symmetrized by T^m, with $\hat{\nu}_w(w) \geq \eta^{m/2}/2$. Thus, by Claim 2, it is also true that for every $v \in (Z^+ \setminus G)$, there is a d-orthogonalizing distribution $\rho_v : \{-1, 1\}^{km} \to [0, 1]$ that is symmetrized by T^m, with $\hat{\rho}_v(v) \geq \eta^{m/2} 4^{-(m+d+1)}$.

Now for each element $z \in Z^+$, define its weight to be $W_z = |(T^m)^{-1}(z)|$. Consider the following distribution:

$$\hat{\mu}(z) = \left(\sum_{z \in Z^+} W_z \right)^{-1} \cdot \left(\sum_{w \in G} W_w \cdot \hat{\nu}_w(z) + \sum_{v \in (Z^+ \setminus G)} W_v \cdot \hat{\rho}_v(z) \right).$$

The (un-symmetrized) distribution $\mu : (\{-1,1\}^k)^m \to [0,1]$ satisfies $\mu(x) \geq \eta^{m/2} 4^{-(m+d+1)} 2^{-km}$ for every point $x \in (T^m)^{-1}(Z^+) = h_m^{-1}(1)$. Moreover, μ remains a d-orthogonalizing distribution for h_m, as it is a convex combination of d-orthogonalizing distributions for h_m. □

4 Sign Rank Lower Bounds for $\mathbf{AC}^0$

We now use the machinery developed by Razborov and Sherstov to translate our construction of a smooth orthogonalizing distribution into a sign-rank lower bound.

Theorem 4.1 (Implicit in [20, Theorem 1.1])**.** Let $h\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, and suppose there exists a d-orthogonalizing distribution μ for h such that $\mu(x) \geq 2^{-cd} 2^{-n}$ for all but a 2^{-cd} fraction of inputs $x \in \{-1,1\}^n$. Then there exists a constant C (depending only on c) such that if $F(x,y) := h(\dots, \wedge_{j=1}^{C}(x_{ij} \vee y_{ij}), \dots)$, then the matrix $[F(x,y)]_{x,y}$ has sign-rank $\exp(\Omega(d))$.

We sketch the proof of Theorem 4.1 in Appendix E.

Combining Theorem 4.1 with Theorem 3.1 yields the main result of this work.

Theorem 4.2. Let $f : \{-1,1\}^k \to \{-1,1\}$ be a Boolean function in the class $\mathcal{C}_d$. Let $F : \{-1,1\}^n \to \{-1,1\}^n$ be defined by

$$F = \mathrm{OR}_{2d} \circ f \circ \mathrm{AND}_C \circ \mathrm{OR}_2,$$

where C is the universal constant of Theorem 4.1 (and hence $n = O(dk)$). The sign-rank of the matrix $[F(x,y)]_{x,y}$ is $\exp(\Omega(d))$.

Proof. Let $h_{2d} : \{-1,1\}^{2dk} \to \{-1,1\}$ denote the function $h_{2d} = \mathrm{OR}_{2d} \circ f$. By Theorem 3.1, there exists a d-orthogonalizing distribution μ for h_{2d} such that $\mu(x) \geq 2^{-9d} 2^{-2dk}$ for every $x \in h_{2d}^{-1}(1)$. Since $f \in \mathcal{C}_d$, we have by a union bound that $h_{2d}^{-1}(1)$ contains all but a $(2d) \cdot 2^{-d} \leq 2^{-d/2}$ fraction of the points in $\{-1,1\}^{2dk}$. Thus, by Theorem 4.1, there is a universal constant C for which $[F(x,y)]_{x,y}$ has sign-rank $\exp(\Omega(d))$. □

Corollary 4.1. Let $\mathrm{MP}_n = \mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ be the Minsky-Papert DNF. Then $[\mathrm{MP}_n(x \vee y)]_{x,y}$ has sign-rank $\exp(\Omega(n^{1/3}))$

Proof. The function AND_k evaluates to TRUE on exactly 1 out of 2^k inputs. Hence, by Lemma 2.2, we have $\mathrm{AND}_k \in \mathcal{C}_d$ for $d = \Omega(k^{1/2})$. Let $F = \mathrm{MP}_n \circ \mathrm{AND}_C \circ \mathrm{OR}_2$. Applying Theorem 4.2 implies that the sign-rank of $[F(x,y)]_{x,y} = \exp(\Omega(n^{1/3}))$. Merging the two adjacent layers of AND gates in the natural circuit computing F yields the desired result. □

Corollary 4.2. Let $F_n^{\mathrm{ED}} = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}} \circ \mathrm{AND}_C$. Then $[F_n^{\mathrm{ED}}(x \vee y)]_{x,y}$ has sign-rank $\exp(\tilde{\Omega}(n^{2/5}))$

Proof. Assume for simplicity that $k = K \log K$. The function ED_k evaluates to TRUE on exactly $K!$ inputs, which is an $\exp(-O(K))$ fraction of the $2^k = K^K$ total inputs. Hence, by Lemma 2.3, we have $\mathrm{ED}_k \in \mathcal{C}_d$ for $d = \Omega(K^{2/3}/\log K)$. The result follows by applying Theorem 4.2. □

Acknowledgments. We thank Lijie Chen for pointing out several small errors in a previous version of this manuscript.

References

[1] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004.

[2] Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science, Portland, Oregon, USA, 21-23 October 1985*, pages 277–280. IEEE Computer Society, 1985.

[3] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank, VC dimension and spectral gaps. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:135, 2014.

[4] Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[5] Andris Ambainis, Andrew M. Childs, Ben Reichardt, Robert Spalek, and Shengyu Zhang. Any and-or formula of size n can be evaluated in time $n^{1/2+o(1)}$ on a quantum computer. *SIAM J. Comput.*, 39(6):2513–2530, 2010.

[6] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 337–347. IEEE Computer Society, 1986.

[7] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and markov-bernstein inequalities. In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *ICALP (1)*, volume 7965 of *Lecture Notes in Computer Science*, pages 303–314. Springer, 2013.

[8] Mark Bun and Justin Thaler. Dual polynomials for collision and element distinctness. *CoRR*, abs/1503.07261, 2015.

[9] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, volume 9134 of *Lecture Notes in Computer Science*, pages 268–280. Springer, 2015. Full version available at `http://eccc.hpi-web.de/report/2013/151`.

[10] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.

[11] Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans-Ulrich Simon. Relations between communication complexity, linear arrangements, and computational complexity. In Ramesh Hariharan, Madhavan Mukund, and V. Vinay, editors, *FST TCS 2001: Foundations of Software Technology and Theoretical Computer Science, 21st Conference, Bangalore, India, December 13-15, 2001, Proceedings*, volume 2245 of *Lecture Notes in Computer Science*, pages 171–182. Springer, 2001.

[12] Jürgen Forster and Hans-Ulrich Simon. On the smallest possible dimension and the largest possible margin of linear arrangements representing given concept classes uniform distribution. In Nicolò Cesa-Bianchi, Masayuki Numao, and Rüdiger Reischuk, editors, *Algorithmic Learning Theory, 13th International Conference, ALT 2002, Lübeck, Germany, November 24-26, 2002, Proceedings*, volume 2533 of *Lecture Notes in Computer Science*, pages 128–138. Springer, 2002.

[13] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:49, 2015.

[14] Lisa Hellerstein and Rocco A. Servedio. On PAC learning algorithms for rich boolean function classes. *Theor. Comput. Sci.*, 384(1):66–76, 2007.

[15] Adam R. Klivans and Rocco A. Servedio. Learning DNF in time $2^{\tilde{o}(n^{1/3})}$. *J. Comput. Syst. Sci.*, 68(2):303–318, 2004.

[16] Samuel Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(1):29–36, 2005.

[17] Nati Linial, Shahar Mendelson, Gideon Schechtman, and Adi Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

[18] Marvin Minsky and Seymour Papert. *Perceptrons - an introduction to computational geometry.* MIT Press, 1969.

[19] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

[20] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of ac^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010.

[21] A. A. Sherstov. The power of asymmetry in constant-depth circuits. In *FOCS*, 2015.

[22] Alexander A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008.

[23] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

[24] Alexander A. Sherstov. The unbounded-error communication complexity of symmetric functions. *Combinatorica*, 31(5):583–614, 2011.

[25] Alexander A. Sherstov. Approximating the and-or tree. *Theory of Computing*, 9(20):653–663, 2013.

[26] Alexander A. Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013.

[27] Alexander A. Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 223–232. ACM, 2014.

[28] Robert Spalek. A dual polynomial for OR. *CoRR*, abs/0803.4516, 2008.

[29] Justin Thaler. Lower bounds for the approximate degree of block-composed functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:150, 2014.

[30] Leslie G. Valiant. A theory of the learnable. In Richard A. DeMillo, editor, *Proceedings of the 16th Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1984, Washington, DC, USA*, pages 436–445. ACM, 1984.

A Applications

A.1 Communication Complexity

Sign-rank completely characterizes the notion of *unbounded error communication complexity* defined by Babai et al. [6] as follows. Consider a Boolean function $f : X \times Y \to \{-1, 1\}$. Alice receives an input $x \in X$, and Bob receives an input $y \in Y$. Each has an unlimited source of private randomness, and their goal is to compute the joint function $f(x, y)$ of their inputs with minimal communication. We say a protocol computes f if for any input (x, y), the output of the protocol is correct with probability strictly greater than $1/2$. The cost of a protocol for computing f is the maximum number of bits exchanged on any input (x, y). The unbounded error communication complexity $\mathrm{UPP}(f)$ of a function f is the minimum cost of a protocol computing f. A function f is in the complexity class $\mathbf{UPP^{cc}}$ if $\mathrm{UPP}(f) = O(\log^c n)$ for some constant c.

Paturi and Simon [19] showed that $\mathrm{UPP}(f) = \log(\text{sign-rank}([f(x, y)]_{x \in X, y \in Y})) + O(1)$. Thus, our main result gives an improved lower bound on the unbounded error communication complexity of a communication problem in AC^0. The previous best lower bound was $\Omega(n^{1/3})$, obtained by Razborov and Sherstov [20].

Corollary A.1. There exists an (explicitly given) function $F : \{-1, 1\}^n \times \{-1, 1\}^n \to \{-1, 1\}$ computed by an AC^0 circuit of depth 3 with logarithmic bottom fan-in (with an OR as the top gate), for which $\mathrm{UPP}(F) = \tilde{\Omega}(n^{2/5})$.

The fact that F is computed by a depth-3 AC^0 circuit with logarithmic bottom fan-in (with an OR as the top gate) implies that $F \in \mathbf{\Sigma_2^{cc}}$, the communication complexity analogue of the second level of the polynomial hierarchy. These complexity classes were also introduced by Babai et al. [6], and we refer the reader to their work for precise definitions. Our result thus gives an improved separation between the polynomial hierarchy $\mathbf{PH^{cc}}$ (indeed, $\mathbf{\Sigma_2^{cc}}$) and $\mathbf{UPP^{cc}}$.

A.2 Learning Theory

Sign-rank upper bounds have important consequences for the design of learning algorithms in Valiant's PAC model [30]. Let $\mathcal{C}$ denote a *concept class* of Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$. The characteristic matrix of $\mathcal{C}$ is given by $M(\mathcal{C}) = [f(x)]_{f\in\mathcal{C},x\in\{-1,1\}^n}$. Upper bounds on the sign-rank of $\mathcal{C}$ underly state-of-the-art PAC learning algorithms for many concept classes, including DNF [15] and read-once formulas [5], as follows. A matrix $M(\mathcal{C})$ has sign-rank at most r if and only if there are functions $h_1, \ldots, h_r : \{-1,1\}^n \to \{-1,1\}$ such that every $f \in \mathcal{C}$ can be written as

$$f(x) = \operatorname{sgn}(w_1 h_1(x) + \cdots + w_r h_r(x))$$

for some real coefficients $w_1, \ldots, w_r$. In other words, every $f \in \mathcal{C}$ can be expressed as a halfspace over the "feature" functions $h_1, \ldots, h_r$. When these feature functions are efficiently computable, $\mathcal{C}$ may be learned in time $\operatorname{poly}(n, r)$ using a linear-programming based algorithm.

Many of the best-known PAC learning algorithms fit into this sign-rank framework. For instance, Klivans and Servedio [15] gave an upper bound of $2^{\tilde{O}(n^{1/3})}$ on the sign-rank of $\mathcal{C}$ for the class of DNF formulas. Specifically, their upper bound used feature functions $h_1, \ldots, h_r$ corresponding to monomials of degree up to $\tilde{O}(n^{1/3})$.

Sign-rank *lower bounds* in turn rule out the existence of efficient learning algorithms in this general framework. In particular, Razborov and Sherstov's lower bound on the sign-rank of $[\mathrm{MP}_n(x \vee y)]_{x,y}$ showed that Klivans and Servedio's algorithm is essentially optimal for all sign-rank based algorithms for learning DNF – even those for which the feature space is not restricted to low-degree monomials. Our improved lower bound shows that the sign-rank framework cannot be used to learn depth-3 AC^0 circuits with logarithmic bottom fan-in in time faster than $\exp(\tilde{\Omega}(n^{2/5}))$.

Corollary A.2. Let $\mathcal{C}_s$ be the concept class of depth-3 circuits $f : \{-1,1\}^n \to \{-1,1\}$ of size s and bottom fan-in $O(\log s)$. There exists an $s = \operatorname{poly}(n)$ for which $M(\mathcal{C}_s)$ has sign-rank $\exp(\tilde{\Omega}(n^{2/5}))$.

A.3 Circuit Complexity

A threshold gate is a Boolean function $T(x_1, \ldots, x_k) = \operatorname{sgn}(a_1 x_1 + \cdots + a_k x_k - \theta)$ parameterized by integer weights $a_1, \ldots, a_k, \theta$. Forster et al. [11] showed that sign-rank lower bounds imply lower bounds on the weights of depth-2 threshold circuits. In particular, if $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ is computed by a depth-2 threshold circuit of size s, with the (absolute values) of weights at the bottom layer summing to w, then $\text{sign-rank}([F(x,y)]_{x,y}) \leq O(snw)$.

Corollary A.3. There exists a depth-3 AC^0 circuit F such that every depth-2 threshold circuit computing F requires the sum of the weights of the bottom gates to be at least $\exp(\tilde{\Omega}(n^{2/5}))$.

In particular, every threshold-of-majority circuit computing F requires size $\exp(\tilde{\Omega}(n^{2/5}))$. This strengthens the lower bound of $\exp(\Omega(n^{1/3}))$ due to Razborov and Sherstov [20].

B A Primal Formulation of Dual Objects

In this section, we argue that the *non-existence* of a certain kind of low-degree polynomial approximation to f implies the existence of a suitable dual object for f. Throughout this discussion, we fix a degree bound $d \geq 1$.

For a subset $S \subseteq \{1, \ldots, k\}$ and $x \in \{-1, 1\}^k$, let $\chi_S(x) = \prod_{i \in S} x_i$. Given a Boolean function f, let $V \subseteq f^{-1}(1)$ be any subset of inputs on which f evaluates to 1. Let $p(x) = \sum_{|S| \leq d} c_S \chi_S(x)$ be a polynomial of degree d, where the coefficients c_S are real numbers.

Definition B.1. We say that p is an *ε-error one-sided approximation to f under the promise that the input x is in $V \cup f^{-1}(-1)$* if p satisfies the following three properties:

- $p(x) \leq -1 + \varepsilon$ for all $x \in f^{-1}(-1)$.
- $|p(x) - 1| \leq \varepsilon$ for all $x \in V$.
- $|p(x)| \leq 1 + \varepsilon$ for all $x \in \{-1, 1\}^k \setminus (V \cup f^{-1}(-1))$.

The non-existence of an ε-error one-sided approximation to f under the promise that the input x is in $T^{-1}(z_+) \cup f^{-1}(-1)$, for some $z_+ \in \widetilde{f}^{-1}(1)$ "almost" implies the existence of an appropriate dual object for f. In order to make a precise statement, however, we actually require the non-existence of certain *weak* type of a polynomial approximation to f stated below. Our definition of a weak approximation captures polynomials p that may not actually approximate f pointwise, yet do approximate f when averaged over small sets of the form $T^{-1}(z)$ for $z \in D$.

Definition B.2. Let $T : \{-1, 1\}^k \to D$ be a symmetrization for a Boolean function f. Let $V = T^{-1}(\tilde{V}) \subseteq \{-1, 1\}^k$ for some $\tilde{V} \subseteq \tilde{f}^{-1}(1)$. We say that $p : \{-1, 1\}^k \to \mathbb{R}$ is a *weak ε-error one-sided approximation to f under the promise that the input x is in $V \cup f^{-1}(-1)$ (with respect to T)* if the following holds. Define $q : \{-1, 1\}^k \to \mathbb{R}$ by $q(x) = \mathbb{E}_{y:T(y)=T(x)}[p(y)]$. Then q satisfies the following three properties:

- $q(x) \leq -1 + \varepsilon$ for all $x \in f^{-1}(-1)$.
- $|q(x) - 1| \leq \varepsilon$ for all $x \in V$.
- $|q(x)| \leq 1 + \varepsilon$ for all $x \in \{-1, 1\}^k \setminus (V \cup f^{-1}(-1))$.

The goal of this section is to prove the following general criterion for the existence of a dual object.

Theorem B.1. Let T be a symmetrization for f. Let $V = T^{-1}(z_+)$ for some $z_+ \in \tilde{f}^{-1}(1)$. If there does not exist a weak (2η)-error, degree-d one-sided approximation to f under the promise that the input is in $V \cup f^{-1}(-1)$, then f has a $(d, 2\eta, \eta)$-dual object with respect to T.

B.1 Linear Programming Formulation of Definition B.2

The existence of a p satisfying Definition B.2 is equivalent to the existence of a solution of value at least ε to the following linear program, which we refer to as LP1 (LP1 is not written in standard form, but it can easily be transformed into one that is).

$$
\begin{array}{lll}
\min & \varepsilon & \\
\text{such that} & & \\
& \sum_{|S|\le d} c_S \cdot \mathbb{E}_{x\in T^{-1}(z)}[\chi_S(x)] \le -1+\varepsilon & \text{for each } z \in \widetilde{f}^{-1}(-1) \\
& \left|1 - \sum_{|S|\le d} c_S \cdot \mathbb{E}_{x\in T^{-1}(z)}[\chi_S(x)]\right| \le \varepsilon & \text{for each } z \in \tilde{V} \\
& \left|\sum_{|S|\le d} c_S \cdot \mathbb{E}_{x\in T^{-1}(z)}[\chi_S(x)]\right| \le 1+\varepsilon & \text{for each } z \in D \setminus (\tilde{V} \cup \widetilde{f}^{-1}(-1)) \\
& c_S \in \mathbb{R} & \text{for each } |S| \le d \\
& \varepsilon \ge 0 &
\end{array}
$$

The dual LP is as follows (we refer to this linear program as LP2), with variables given by $\hat{\psi}(z)$ for each $z \in D$ and with $\psi(x) := \hat{\psi}(T(x))$.

$$
\begin{array}{ll}
\max & \sum_{x\in V\cup f^{-1}(-1)} \psi(x) f(x) - \sum_{x\in\{-1,1\}^n\setminus(V\cup f^{-1}(-1))} |\psi(x)| \\
\text{such that} & \sum_{x\in\{-1,1\}^k} |\psi(x)| = 1 \\
& \sum_{x\in\{-1,1\}^k} \psi(x)\chi_S(x) = 0 \text{ for each } |S| \le d \\
& \psi(x) \le 0 \text{ for each } x \in f^{-1}(-1) \\
& \psi(x) = \hat{\psi}(T(x)) \in \mathbb{R} \text{ for each } x \in \{-1,1\}^k
\end{array}
$$

B.2 Proof of Theorem B.1

Let T be a symmetrization for f, let $\eta > 0$, let z_+ be an arbitrary element of $\tilde{f}^{-1}(1)$, and let $V = T^{-1}(z_+)$. Suppose that the value of LP1 is less than 2η. That is, suppose there does not exist any weak (2η)-error one-sided approximation to f of degree at most d under the promise that the input x is in $V \cup f^{-1}(-1)$. Then strong LP-duality implies the existence of a solution ψ to LP2 of value at least 2η. Moreover, this solution is symmetrized by T since $\psi(x) = \hat{\psi}(T(x))$ for some $\hat{\psi} : D \to \mathbb{R}$.

It is immediate from LP2 that ψ satisfies Conditions (1)-(4) of Definition 2.2 with $\varepsilon = 2\eta$. We now turn to showing that ψ also satisfies Condition (5) of Definition 2.2.

Let

$$B = \{x \in \{-1,1\}^k \setminus (V \cup f^{-1}(-1))\} \cup \{x \in V : \psi(x) < 0\}.$$

Since $\sum_{x\in\{-1,1\}^k} |\psi(x)| = 1$ (cf. Constraint 1 of LP2), and $V \subseteq f^{-1}(1)$, it holds that the value of ψ in LP2 is

$$\sum_{x\in V\cup f^{-1}(-1)} \psi(x) f(x) - \sum_{x\in\{-1,1\}^k\setminus(V\cup f^{-1}(-1))} |\psi(x)| \le 1 - 2\cdot\sum_{x\in B} |\psi(x)|.$$

Since ψ has value at least 2η, this implies that

$$2\eta \ge 1 - 2\cdot\sum_{x\in B} |\psi(x)| \Longrightarrow \sum_{x\in B} |\psi(x)| \le 1/2 - \eta.$$

This in turn implies that

$$\sum_{x\in f^{-1}(-1)} |\psi(x)| + \sum_{x\in V:\psi(x)\ge 0} \psi(x) \ge 1/2 + \eta. \tag{21}$$

Now, since $d \geq 1$ and ψ has pure high degree at least d (cf. Constraint 2 of LP2), it holds in particular that

$$\sum_{x \in \{-1,1\}^k} \psi(x) = 0. \tag{22}$$

Combining Eq. (22) with the fact that $\psi(x) < 0$ for all $x \in f^{-1}(-1)$ (cf. Constraint 3 of LP2) and the fact that $\sum_{x \in \{-1,1\}^k} |\psi(x)| = 1$ (cf. Constraint 1 of LP2), we conclude that

$$\sum_{x \in V : \psi(x) < 0} |\psi(x)| + \sum_{x \in f^{-1}(-1)} |\psi(x)| \leq 1/2. \tag{23}$$

Combining Inequalities (21) and (23), we conclude that

$$\begin{aligned}
&\sum_{x \in V} \psi(x) = \sum_{x \in V : \psi(x) \geq 0} \psi(x) - \sum_{x \in V : \psi(x) < 0} |\psi(x)| \\
&= \left(\sum_{x \in f^{-1}(-1)} |\psi(x)| + \sum_{x \in V : \psi(x) \geq 0} \psi(x) \right) - \left(\sum_{x \in f^{-1}(-1)} |\psi(x)| + \sum_{x \in V : \psi(x) < 0} |\psi(x)| \right) \\
&\geq 1/2 + \eta - 1/2 = \eta.
\end{aligned}$$

It follows that ψ satisfies Condition (5). This completes the proof of Theorem B.1. □

C A Dual Object for Element Distinctness

In our prior work [8], we constructed an explicit dual witness for the high *one-sided approximate degree* of the Element Distinctness function.[3] That is, we constructed an object satisfying Conditions (1)-(4) of Definition 2.2 with respect to a specific symmetrization T for ED_k, for $\varepsilon = 1/2$. Additionally, it can be shown that this dual witness satisfies Condition (5) of Definition 2.2 (with $\eta = 1/4$) with respect to T as well. For both completeness and intuition, in this appendix we detail an alternative proof of Lemma 2.3. This proof uses the primal formulation of dual objects described in Appendix B, and builds directly on results of Aaronson and Shi [1] and Ambainis [4].

See Section 2.6 for the definition of ED_k, and relevant notation.

Lemma 2.3. There exists a symmetrization $T : \{-1,1\}^k \to [K]^K$ for the Element Distinctness function $\mathrm{ED}_k : \{-1,1\}^k \to \{-1,1\}$ such that ED_k has a $(d, 1/2, 1/4)$-dual object (with respect to the map T), for some $d = \Omega(K^{2/3}/\log K)$.

C.1 A Symmetrization T for Element Distinctness

As stated in Lemma C.1 below, Ambainis showed that the Element Distinctness function is symmetric with respect to a certain degree non-increasing symmetrization. Ambainis' results are described in a slightly different setting than our own, so we provide a proof that explains how to translate his result into Lemma C.1 below.

[3] The result of [8] is stated for the "large-range" version of the Element Distinctness function, but the construction can be modified to work for the "small-range" case considered in this work.

Lemma C.1 (Ambainis [4])**.** Define a map $T^* : \{-1,1\}^k \to [K]^K$ as follows. For $x \in \{-1,1\}^k$, write $T^*(x) = (T_1^*(x), \ldots, T_K^*(x))$, where the component $T_j^*(x)$ counts the number of elements of $[K]$ that are mapped to j under the function $g_x : [K] \to [K]$. That is, $T_j^*(x) = |g_x^{-1}(j)|$. The map T^* is degree non-increasing.

Proof. Ambainis' [4] analysis considers polynomials defined over a different domain than ED_k. Specifically, whereas the domain of ED_k is $\{-1,1\}^k$ for $k = K \log K$, and ED_k interprets its input x as blocks $x_1, \ldots, x_K$ where each $x_i \in \{-1,1\}^{\log_2 K}$, Ambainis considers polynomials defined over the domain $\{-1,1\}^{K^2}$, where an input $y = (y_{1,1}, \ldots, y_{K,K}) \in \{-1,1\}^{K^2}$ is interpreted as corresponding to an input $x \in \{-1,1\}^k$ in the following manner: $y_{i,j} = -1 \iff x_i = j$.

Note that not all inputs $y \in \{-1,1\}^{K^2}$ correspond to some $x \in \{-1,1\}^k$. Specifically, y corresponds to some $x \in \{-1,1\}^k$ if and only if for all $i = 1, \ldots, K$, there exists exactly one value of j for which $y_{i,j} = -1$. In the remainder of this proof, we call any such y *valid*, and denote the set of all valid y's by Y. For any $x \in \{-1,1\}^k$, let us denote the corresponding y by $y(x)$. Similarly, for any valid y, let us denote the corresponding $x \in \{-1,1\}^k$ by $x(y)$.

Consider the map $T^{**} \colon Y \to [K]^K$ defined via $T^{**}(y) = T^*(x(y))$. Ambainis [4, Lemma 3.4] shows that for any polynomial $p' \colon Y \to \mathbb{R}$, there exists a polynomial $q \colon [K]^K \to \mathbb{R}$ satisfying $\deg(q) \le \deg(p')$ such that for all z in the image of T^{**},

$$q(z) = \mathbb{E}_{y \in (T^{**})^{-1}(z)}[p'(y)]. \tag{24}$$

We now argue that this implies that T^* is a degree non-increasing map. Given any polynomial $p \colon \{-1,1\}^k \to \mathbb{R}$, it is easy to turn p into a polynomial $p' \colon Y \to \mathbb{R}$ such that $\deg(p') \le \deg(p)$ and $p'(y(x)) = p(x)$ for all $x \in \{-1,1\}^k$. Indeed, this follows from the observation that for any $y \in Y$, each bit of $x(y)$ is computed by a degree 1 polynomial in y. Specifically

$$x_{i,\ell}(y) = 1 + 2 \cdot \sum_{j \in \{-1,1\}^{\log_2 K} \text{ such that } j_\ell = -1} (1 - y_{i,j})/2.$$

That T^* is degree non-increasing is an immediate consequence of [4, Lemma 3.4] combined with the observation above. In more detail, given any polynomial $p \colon \{-1,1\}^k \to \mathbb{R}$, first turn p into the polynomial $p' \colon Y \to \mathbb{R}$ described above that satisfies $p'(y(x)) = p(x)$ for all $x \in \{-1,1\}^k$, and let q be the polynomial satisfying Eq. (24). Then $\deg(q) \le \deg(p') \le \deg(p)$, and

$$q(z) = \mathbb{E}_{y \in (T^{**})^{-1}(z)}[p'(y)] = \mathbb{E}_{x \in (T^*)^{-1}(z)}[p(x)]$$

as desired. □

In order to show that ED_k has an appropriate dual object, we take Ambainis' symmetrization one step further. Let $H \colon [K]^K \to [K]^K$ be the mapping that on input z, sorts the coordinates of z in non-decreasing order. That is, $H(z) = (z_{\pi(1)}, z_{\pi(2)}, \ldots, z_{\pi(K)})$ where π is a permutation that makes $z_{\pi(1)} \le z_{\pi(2)} \le \cdots \le z_{\pi(K)}$.

Define the mapping $T : \{-1,1\}^k \to [K]^K$ by $T = H \circ T^*$. Looking ahead, this mapping is specifically constructed so that, if $\widetilde{\mathrm{ED}}_k$ represents the symmetrization of ED_k with respect to T, we have that

- There is exactly one element $z_- = (1, 1, \ldots, 1) \in [K]^K$ such that $\widetilde{\mathrm{ED}}_k(z_-) = -1$, and

- The element $z_+ = (0, 1, 1, \ldots, 1, 2) \in [K]^K$ satisfies $\widetilde{\mathrm{ED}}_k(z_+) = 1$; observe that z_+ is the image under T of all inputs that represent an "almost" 1-to-1 function, in the sense that they have exactly 1 collision. This element z_+ will be the one for which we exhibit a dual object ψ for which $\hat{\psi}(z_+)$ is large.

Lemma C.2. The map T is degree non-increasing, and in fact symmetrizes ED_k.

Proof of Lemma C.2. We first show that T is degree non-increasing. Since the map T^* is degree non-increasing, it suffices to show that if $q : [K]^K \to \mathbb{R}$ is a polynomial, then there exists a polynomial $r : [K]^K \to \mathbb{R}$ with $\deg r \leq \deg q$ such that

$$r(H(z)) = \mathop{\mathbb{E}}_{w \text{ s.t. } H(w)=C(z)} [q(z)]$$

for every $z \in [K]^K$. The polynomial

$$r(z) = \mathop{\mathbb{E}}_{\pi \in S_K} [q(z_{\pi(1)}, \ldots, z_{\pi(K)})]$$

does the trick. □

C.2 Proof of Lemma 2.3

To prove Lemma 2.3, we study the approximability of a promise variant of the ED_k function (cf. Appendix B). For notational convenience, in the remainder of this section, we will consider polynomials defined over the same domain as Ambainis [4] (cf. the proof of Lemma C.1). We also work with the "large-range" generalization of the Element Distinctness function. That is, we think of a function $g : [K] \to [R]$, for some $R \geq K$, as being represented by Boolean variables $(y_{1,1}, \ldots, y_{K,R})$ where $y_{i,j} = -1$ if $g(i) = j$, and $y_{i,j} = 1$ otherwise.

Lemma C.3. Let $R \geq \Omega(K^2)$ and let $p : \{-1, 1\}^{K \cdot R} \to \mathbb{R}$ be a polynomial such that

- $p(y) \in [-11/10, -9/10]$ if y represents a 1-to-1 function,
- $p(y) \in [9/10, 11/10]$ if y represents a function g with exactly one collision, i.e., if there exist $i_1 \neq i_2$ such that $g(i_1) = g(i_2)$ and g is 1-to-1 on $[K] \setminus \{i_1, i_2\}$,
- $p(y) \in [-11/10, 11/10]$ whenever y represents a function from $[K]$ to $[R]$.

Then $\deg p \geq \Omega(K^{2/3})$.

The proof of Lemma C.3 is by a standard reduction to the Collision lower bound. We state the Collision lower bound as Theorem C.1 below.

Theorem C.1 (Collision lower bound. Cf. [1, 4, 16].)**.** Let $q : \{-1, 1\}^{R^2} \to \mathbb{R}$ be a polynomial. If there are constants $a_1 < a_2 < b_1 < b_2$ such that

- $q(y) \in [a_1, a_2]$ whenever y represents a 1-to-1 function,
- $q(y) \in [b_1, b_2]$ whenever y represents a 2-to-1 function (i.e., a function g such that for every $i \in R$, there exists exactly one $i' \neq i$ such that $g(i) = g(i')$)

- and $q(y) \in [a_1, b_2]$ whenever y represents a function from $[R]$ to $[R]$,

then $\deg q \geq \Omega(R^{1/3})$.

Proof of Lemma C.3. Let $p : \{-1,1\}^{K \cdot R} \to \mathbb{R}$ be a polynomial satisfying the conditions of Lemma C.3. We show how to transform p into a polynomial q satisfying $\deg(q) \leq \deg(p)$, and moreover q satisfies all of the requirements necessary to invoke Theorem C.1.

Definition of q. Define $q : \{-1,1\}^{R^2} \to \mathbb{R}$ by

$$q(y_{1,1}, \ldots, y_{R,R}) = \frac{1}{\binom{R}{K}} \sum_{1 \leq i_1 < i_2 < \cdots < \ldots i_K \leq R} p(y_{i_1,1}, \ldots, y_{i_1,R}, y_{i_2,1}, \ldots, y_{i_2,R}, \ldots, y_{i_K,1}, \ldots, y_{i_K,R}).$$

That is, q is the expected value of p over a random subset of K blocks of its input. It is clear from the definition that $\deg q \leq \deg p$.

Analysis of q. Let S be any set $S = \{i_1, \ldots, i_K\} \subseteq \{1, \ldots, R\}$, and let $y|_S$ denote the restricted input $(y_{i_1,1}, \ldots, y_{i_K,R})$. Then whenever y represents a function from $[R]$ to $[R]$, the restricted input $y|_S$ represents a function from $[K]$ to $[R]$.

If y represents a 1-to-1 function, then $y|_S$ also represents a 1-to-1 function for every restriction S. Hence $q(y) \in [-11/10, -9/10]$.

On the other hand, if y represents a 2-to-1 function, then the following claim establishes that $y|_S$ represents a function with exactly one collision with constant probability.

Claim 3. Fix $M \in \mathbb{N}$. There exists $K = \Theta(\sqrt{M})$ such that

$$\frac{\binom{M}{K-1}(K-1)2^{K-2}}{\binom{2M}{K}} \geq \frac{1}{2e}.$$

Proof of Claim 3. Rewrite the left hand side as

$$\begin{aligned}
\frac{K(K-1)}{4(M-K+1)} \cdot \frac{2^K M(M-1)\ldots(M-K+1)}{(2M)(2M-1)\ldots(2M-K+1)} &= \frac{K(K-1)}{4(M-K+1)} \cdot \frac{1}{\prod_{j=0}^{K-1}\left(1 + \frac{j}{2M-2j}\right)} \\
&\geq \frac{K(K-1)}{4(M-K+1)} \cdot \frac{1}{\exp\left(\sum_{j=1}^{K-1} \frac{j}{2M-2j}\right)} \\
&\geq \frac{K(K-1)}{4(M-K+1)} \cdot \exp\left(-\frac{(K-1)(K-2)}{2(2M-2K+1)}\right) \\
&\geq \frac{1}{2e}
\end{aligned}$$

by taking $K = \lceil 2\sqrt{M} \rceil$. □

The expression on the left hand side of the claim, for $M = R/2$, represents the probability that $y|_S$ represents a function with exactly one collision when y represents a 2-to-1 function. Thus, for $K = \Theta(\sqrt{R})$,

$$\begin{aligned}
\mathbb{E}[p(y|_S)] &\geq \frac{1}{2e} \cdot (9/10) + \left(1 - \frac{1}{2e}\right) \cdot (-11/10) \\
&\geq -\frac{4}{5}.
\end{aligned}$$

This allows us to conclude that $q(y) \in [-4/5, 11/10]$.

Finally, if y represents any other function, then $q(y) \in [-11/10, 11/10]$ by averaging. To summarize,

- $q(y) \in [-11/10, -9/10]$ whenever y represents a 1-to-1 function,
- $q(y) \in [-4/5, 11/10]$ whenever y represents a 2-to-1 function,
- $q(y) \in [-11/10, 11/10]$ when y represents any function.

Thus, by Theorem C.1, $\deg p \geq \deg q \geq \Omega(R^{1/3}) = \Omega(K^{2/3})$. This completes the proof of Lemma C.3. □

In order to prove Lemma 2.3, we actually need the following technical strengthening of Lemma C.3.

Lemma C.4. Let $p : \{-1, 1\}^{K^2} \to \mathbb{R}$ be a polynomial such that:

- $p(y) \leq -1/2$ if y represents a 1-to-1 function (25)
- $p(y) \in [1/2, 3/2]$ if y represents a function g with exactly one collision, (26)
  i.e., if there exist $i_1 \neq i_2$ such that $g(i_1) = g(i_2)$ and g is 1-to-1 on $[K] \setminus \{i_1, i_2\}$.
- $p(y) \in [-3/2, 3/2]$ whenever y represents a function from $[K]$ to $[K]$. (27)

Then $\deg p \geq \Omega(K^{2/3})$.

Proof. Let p be a polynomial satisfying conditions (25)-(27). Our prior work [9, Proof of Theorem 2] shows how to transform such a polynomial p into a polynomial $p' : \{-1, 1\}^{K^2} \to \mathbb{R}$ such that $\deg(p') \leq \deg(p)$ such that:

- $p'(y) \in [-3/2, -1/2]$ if y represents a 1-to-1 function (28)
- $p'(y) \in [1/2, 3/2]$ if y represents a function g with exactly one collision, (29)
  i.e., if there exist $i_1 \neq i_2$ such that $g(i_1) = g(i_2)$ and g is 1-to-1 on $[K] \setminus \{i_1, i_2\}$.
- $p'(y) \in [-3/2, 3/2]$ whenever y represents a function from $[K]$ to $[K]$. (30)

Next, Ambainis' techniques [4] for porting approximate degree lower bounds in the "large range" case $R \geq K$ to the "small range" case $R = K$ show that there exists a polynomial $p'' : \{-1, 1\}^{K \cdot R} \to \mathbb{R}$ for the value of R used in the statement of Lemma C.3, with $\deg(p'') \leq \deg(p')$, such that:

- $p''(y) \in [-3/2, -1/2]$ if y represents a 1-to-1 function (31)
- $p''(y) \in [1/2, 3/2]$ if y represents a function g with exactly one collision, (32)
  i.e., if there exist $i_1 \neq i_2$ such that $g(i_1) = g(i_2)$ and g is 1-to-1 on $[K] \setminus \{i_1, i_2\}$.
- $p''(y) \in [-3/2, 3/2]$ whenever y represents a function from $[K]$ to $[R]$. (33)

Finally, we use standard error amplification techniques to produce a low degree polynomial satisfying the conditions of Lemma C.3. Namely, let $r : \mathbb{R} \to \mathbb{R}$ be a polynomial of degree $O(1)$ such that $r([-3/2, -1/2]) \subseteq [-11/10, -9/10]$, $r([1/2, 3/2]) \subseteq [9/10, 11/10]$, and $r([-3/2, 3/2]) \subseteq [-11/10, 11/10]$. Then the polynomial $p'''(y) := r(p''(y))$ satisfies the conditions of Lemma C.3, and has $\deg(p''') = O(\deg(p))$. Thus, $\deg(p) = \Omega(K^{2/3})$, completing the proof. □

Proof of Lemma 2.3. We are at last ready to combine Lemma C.4 with Theorem B.1 of Appendix B to prove Lemma 2.3. Let $T : \{-1,1\}^k \to D$ be the degree non-increasing symmetrization defined in Section C.1. Denote by z_- the unique input $(1,1,\ldots,1)$ in D for which $\widetilde{\mathrm{ED}}_k(z_-) = -1$. Let $V = T^{-1}(z_+) \subseteq \{-1,1\}^k$, where $z_+ = (0,1,1,\ldots,1,2)$ is the input in D corresponding to functions with exactly one collision. Suppose $p : \{-1,1\}^k \to \mathbb{R}$ is a weak $(1/2)$-error one-sided approximation to f under the promise that the input x is in $V \cup f^{-1}(-1)$ (with respect to T). Then by definition, the function $p^{\mathrm{sym}} : \{-1,1\}^k \to \mathbb{R}$ defined by $p^{\mathrm{sym}}(x) = \mathbb{E}_{y:T(y)=T(x)}[p(y)]$ has the following properties:

- $p^{\mathrm{sym}}(x) \le -1/2$ for every $x \in T^{-1}(z_-)$,
- $|p^{\mathrm{sym}}(x) - 1| \le 1/2$ for every $x \in T^{-1}(z_+)$,
- $|p^{\mathrm{sym}}(x)| \le 3/2$ for all $x \in \{-1,1\}^k \setminus T^{-1}(\{z_-, z_+\})$.

Our prior work (cf. [9, Lemma 23]) showed that the function p^{sym} satisfies $\deg(p^{\mathrm{sym}}) \le (\log K) \cdot \deg(p)$. Now let $r : \{-1,1\}^{K^2} \to \mathbb{R}$ be defined by $r(y) = p^{\mathrm{sym}}(x(y))$. Since x is computed by a polynomial of degree 1, we have that $\deg(r) \le (\log K) \cdot \deg(p)$ and satisfies the conditions of Lemma C.4. We conclude that $\deg p \ge \Omega(K^{2/3}/\log K)$.

□

D Omitted Details in the Proof of Theorem 3.1

D.1 Proof of Lemma 3.3

We first recall the construction of the desired dual object $\hat{\psi}$. Let $\hat{\psi}_1 : D \to \mathbb{R}$ be a $(d_1, 1/2, \eta)$-dual object for $\widetilde{f}$, and decompose $\hat{\psi}_1$ as the difference of non-negative functions $\hat{\nu}_+ - \hat{\nu}_-$. Let

$$\hat{\psi}_0 = \begin{cases} \hat{\nu}_- = \max\{-\hat{\psi}_1, 0\} & \text{on } \widetilde{f}^{-1}(1) \\ \left(\frac{1}{\langle \psi_1, f\rangle} - 1\right)\hat{\psi}_1 = \left(\frac{1}{\langle \psi_1, f\rangle} - 1\right)\hat{\nu}_- & \text{on } \widetilde{f}^{-1}(-1). \end{cases}$$

Recall that $\langle \psi_1, f\rangle \ge 1/2$ (cf. Condition (1) in Definition 2.2), so the factor $\left(\frac{1}{\langle \psi_1, f\rangle} - 1\right)$ in the second case of the definition of $\hat{\psi}_0$ is at most 1. The constant $\left(\frac{1}{\langle \psi_1, f\rangle} - 1\right)$ is chosen so that ψ_0 is

orthogonal to the constant functions. That is,

$$\begin{aligned}
\sum_{x\in\{-1,1\}^k}\psi_0(x) &= \sum_{z\in D}\hat{\psi}_0(z) = \sum_{z\in\widetilde{f}^{-1}(1)}\hat{\psi}_0(z) + \sum_{z\in\widetilde{f}^{-1}(-1)}\hat{\psi}_0(z)\\
&= -\frac{1}{2}\sum_{z \text{ s.t. } \hat{\psi}_1(z)<0}(1+\widetilde{f}(z))\hat{\psi}_1(z) + \frac{1}{2}\sum_{z\in D}(1-\widetilde{f}(z))\left(\frac{1}{\langle\psi_1,f\rangle}-1\right)\hat{\psi}_1(z)\\
&= -\frac{1}{2}\sum_{z \text{ s.t. } \hat{\psi}_1(z)<0}\hat{\psi}_1(z) - \frac{1}{2}\sum_{z \text{ s.t. } \hat{\psi}_1(z)<0}\widetilde{f}(z)\hat{\psi}_1(z) +\\
&\qquad \frac{\left(\frac{1}{\langle\psi_1,f\rangle}-1\right)}{2}\sum_{z\in D}\hat{\psi}_1(z) - \frac{\left(\frac{1}{\langle\psi_1,f\rangle}-1\right)}{2}\sum_{z\in D}\widetilde{f}(z)\hat{\psi}_1(z)\\
&= -\frac{1}{2}\left(\langle f,\psi_1\rangle - \frac{1}{2}\right) - \frac{1}{2}\left(-\frac{1}{2}\right) + 0 - \frac{\left(\frac{1}{\langle\psi_1,f\rangle}-1\right)}{2}\langle f,\psi_1\rangle\\
&= 0,
\end{aligned}$$

where the penultimate equality follows by Conditions (1)-(3) of Definition 2.2.

Recall that we defined $\hat{\Psi}: D^\ell \to \mathbb{R}$ via

$$\hat{\Psi}(z) = \prod_{i=1}^{\ell}\hat{\nu}_+(z_i) - \prod_{i=1}^{\ell}\hat{\nu}_-(z_i) + \prod_{i=1}^{\ell}\hat{\psi}_0(z_i).$$

The desired function is then given by $\hat{\psi}(z) = \hat{\Psi}(z)/\|\Psi\|_1$.

We now proceed to verify that Ψ satisfies Conditions (6)-(8) as in [27].

First we check that Ψ sign represents h_ℓ. If $z \in \widetilde{h}_\ell^{-1}(-1)$, then there exists an i such that $\widetilde{f}(z_i) = -1$. By Condition (4), this means that $\hat{\nu}_+(z_i) = 0$. Hence,

$$\hat{\Psi}(z) = -\prod_{i=1}^{\ell}\hat{\nu}_-(z_i) + \prod_{i=1}^{\ell}\hat{\psi}_0(z_i) \le 0,$$

since $|\hat{\psi}_0| \le \hat{\nu}_-$ pointwise.

If instead $z \in \widetilde{h}_\ell^{-1}(1)$, then $\widetilde{f}(z_i) = 1$ for every i. Hence, $\hat{\psi}_0(z_i) = \hat{\nu}_-(z_i)$ for every i, and

$$\hat{\Psi}(z) = \prod_{i=1}^{\ell}\hat{\nu}_+(z_i) \ge 0.$$

Now we show that $\hat{\Psi}$ has pure high degree d. By linearity, it suffices to show that Ψ is orthogonal to all factored polynomials $p(x) = p_1(x_1)\cdot\cdots\cdot p_\ell(x_\ell)$ of total degree less than $d = \min\{\ell, d_1\}$. Since the function $\hat{\Psi}$ telescopes as

$$\hat{\Psi}(z) = \prod_{i=1}^{\ell}\hat{\psi}_0(z_i) + \sum_{i=1}^{\ell}\hat{\nu}_+(z_1)\ldots\hat{\nu}_+(z_{i-1})(\hat{\nu}_+(z_i) - \hat{\nu}_-(z_i))\hat{\nu}_-(z_{i+1})\ldots\hat{\nu}_-(z_\ell),$$

we have, by the fact that ψ_0 is orthogonal to the constant functions and that $\hat{\psi}_1 = \hat{\nu}_+ - \hat{\nu}_-$ has pure high degree d_1, that

$$\begin{aligned}\langle \Psi, p\rangle &= \prod_{i=1}^{\ell} \langle \psi_0, p_i\rangle + \sum_{i=1}^{\ell} \langle \nu_+, p_1\rangle \dots \langle \nu_+, p_{i-1}\rangle \langle \nu_+ - \nu_-, p_i\rangle \langle \nu_-, p_{i+1}\rangle \cdots \langle \nu_-, p_\ell\rangle \\ &= 0.\end{aligned}$$

E Sign-Rank Lower Bounds from Smooth Orthogonalizing Distributions

For convenience, we restate Theorem 4.1 here, before sketching its proof.

Theorem 4.1 (Implicit in [20, Theorem 1.1])**.** Let $h\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, and suppose there exists a d-orthogonalizing distribution μ for h such that $\mu(x) \geq 2^{-cd}2^{-n}$ for all but a 2^{-cd} fraction of inputs $x \in \{-1,1\}^n$. Then there exists a constant C (depending only on c) such that if $F(x,y) := h(\dots, \wedge_{j=1}^{C}(x_{ij} \vee y_{ij}), \dots)$, then the matrix $[F(x,y)]_{x,y}$ has sign-rank $\exp(\Omega(d))$.

Razborov and Sherstov's proof of Theorem 4.1 builds on Sherstov's *pattern matrix method* [23], a powerful framework that was introduced to prove lower bounds in communication complexity. In order to describe Razborov and Sherstov's proof, we start by defining pattern matrices. Let n and N be positive integers for which n divides N. Let $\mathcal{P}(N,n)$ denote the collection of subsets $S \subset [N]$ for which S contains exactly one member of each block $\{1,2,\dots,N/n\}, \{N/n+1,\dots,2N/n\}, \dots, \{(n-1)N/n+1,\dots,N\}$. For $x \in \{-1,1\}^N$ and $S \in \mathcal{P}(N,n)$, let $x|_S$ denote the restriction of x to S, i.e. $x|_S = (x_{s_1},\dots,x_{s_n})$ where $s_1 < \cdots < s_n$ are the elements of S.

Definition E.1. For $\varphi : \{-1,1\}^n \to \mathbb{R}$, the (N,n,φ)-pattern matrix M is given by

$$M = [\varphi(x|_S \oplus w)]_{x \in \{-1,1\}^N, (S,w) \in \mathcal{P}(N,n) \times \{-1,1\}^n}.$$

Note that M is a matrix with 2^N rows and $(N/n)^n 2^n$ columns.

Proof Sketch of Theorem 4.1. Let M be the (N,n,h) pattern matrix for $N = 2^{3c}n$. Let P be the (N,n,μ)-pattern matrix where μ is the guaranteed smooth orthogonalizing distribution for h. Razborov and Sherstov's extension of Forster's lower bound (cf. [20, Theorem 5.1]) establishes that

$$\text{sign-rank}(M) \geq \text{sign-rank}(M \circ P) \geq \min\left\{\frac{2^{-cd}2^{-n}\sqrt{s}}{2\|M \circ P\|}, 2^{cd}\right\},$$

where $s = 2^{N+n}(N/n)^n$ is the number of entries in $M \circ P$, and $\|\cdot\|$ denotes the spectral norm. The proof of this bound exploits only the smoothness of μ.

Razborov and Sherstov then directly invoke the pattern matrix method [23, Theorem 4.3], which uses the fact that μ is a d-orthogonalizing distribution for h to place an upper bound on $\|M \circ P\|$. In particular,

$$\|M \circ P\| \leq \sqrt{s}2^{-n}\left(\frac{N}{n}\right)^{-d/2}.$$

Thus, the sign-rank of M is at least $2^{\Omega(d)}$.

It remains to show that the pattern matrix M appears as a submatrix of $[F(x,y)]_{x,y}$ for some C depending only on c. For $x \in \{-1,1\}^N$ and $(S,w) \in \mathcal{P}(N,n) \times \{-1,1\}^n$, the relevant entry of the pattern matrix M takes the form $f(x|_S \oplus w)$. Each bit of $(x|_S \oplus w)$ can be written as

$$(x|_S \oplus w)_i = \bigwedge_{j=1}^{2^{3c}} \bigwedge_{b \in \{-1,1\}} ((x_{ij} = b) \vee (w_{ij} = b) \vee (s_i \neq j)).$$

Thus, M is a submatrix of $[F(x,y)]_{x,y}$ for $C = 2^{3c+1}$.

□

ECCC ISSN 1433-8092

http://eccc.hpi-web.de