BREAKING THE MINSKY–PAPERT BARRIER FOR CONSTANT-DEPTH CIRCUITS*

ALEXANDER A. SHERSTOV†

Abstract. The *threshold degree* of a Boolean function f is the minimum degree of a real polynomial p that represents f in sign: $f(x) \equiv \operatorname{sgn} p(x)$. In a seminal 1969 monograph, Minsky and Papert constructed a polynomial-size constant-depth $\{\wedge, \vee\}$-circuit in n variables with threshold degree $\Omega(n^{1/3})$. This lower bound underlies some of today's strongest results on constant-depth circuits. It has since been an open problem (O'Donnell and Servedio, STOC 2003) to improve Minsky and Papert's bound to $n^{\Omega(1)+1/3}$.

We give a detailed solution to this problem. For any fixed $k \geqslant 1$, we construct an $\{\wedge, \vee\}$-formula of size n and depth k with threshold degree $\Omega(n^{(k-1)/(2k-1)})$. This lower bound nearly matches a known $O(\sqrt{n})$ upper bound for arbitrary formulas, and is exactly tight for "regular" formulas. Our result proves a conjecture due to O'Donnell and Servedio (STOC 2003) and a different conjecture due to Bun and Thaler (2013). Applications to communication complexity and computational learning are given.

Key words. Polynomial representations of Boolean functions, polynomial threshold functions, threshold degree, computational learning theory, communication complexity theory, polynomial approximation theory

AMS subject classifications. 03D15, 68Q15, 68Q17

1. Introduction. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a given Boolean function. A real polynomial p is said to *represent f in sign* if

$$\operatorname{sgn} p(x) = \begin{cases} -1 & \text{if } f(x) = 0, \\ +1 & \text{if } f(x) = 1, \end{cases}$$

for every input $x \in \{0,1\}^n$. The main complexity measure of interest is the degree of p. The minimum degree of a sign-representing polynomial for f is called the *threshold degree* of f, denoted $\deg_\pm(f)$. This notion was introduced in 1969 in the seminal work of Minsky and Papert [34], who proved that the parity function on n variables has threshold degree n and examined the threshold degree of several other functions. Sign-representing polynomials quickly found a variety of applications in theoretical computer science, the first of which were size-depth trade-offs [37, 53] and lower bounds [29, 30] for various types of threshold circuits, oracle separations [4] for PP, and the famous proof that PP is closed under intersection [8].

Sign-representing polynomials have been especially useful in the study of constant-depth circuits, leading to algorithmic and complexity-theoretic breakthroughs in the area. One such example is the fastest known algorithm for learning DNF formulas, due to Klivans and Servedio [25], with running time $\exp\{\tilde{O}(n^{1/3})\}$. The authors of [25] obtained their algorithm by proving an upper bound of $O(n^{1/3}\log n)$ on the threshold degree of polynomial-size DNF formulas, essentially matching a classic lower bound due to Minsky and Papert [34]. Another success story is the fastest known algorithm for learning read-once formulas, due to Ambainis et al. [3], with running time $\exp\{\tilde{O}(\sqrt{n})\}$. That algorithm, too, follows from an upper bound of $O(\sqrt{n})$

*Submitted to the editors April 8, 2015. An extended abstract of this paper appeared in *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, pages 223–232, 2014.

Funding: The author was supported in part by NSF CAREER award CCF-1149018 and an Alfred P. Sloan Foundation Research Fellowship.

†University of California, Los Angeles, CA 90095 (sherstov@cs.ucla.edu).

on the threshold degree of read-once formulas, obtained in a series of breakthrough papers [36, 15, 3, 32] by learning theorists and quantum researchers.

Sign-representing polynomials have been equally influential in the *complexity-theoretic* study of constant-depth circuits. Recall that AC^0 denotes the class of $\{\wedge, \vee, \neg\}$-circuits of constant depth and polynomial size. Aspnes et al. [4] used the notion of threshold degree and its relaxations to give an ingenious new proof that AC^0 circuits cannot compute or even approximate the parity function. Another contribution [42, 43] in which threshold degree played a central role is the first construction of an AC^0 circuit with exponentially small discrepancy and hence maximum communication complexity in nearly every model. This discrepancy result was used in [42] to show the optimality of Allender's classic simulation of AC^0 functions by majority circuits, solving the open problem [29] on the relation between these two circuit classes. Subsequent work generalized the threshold degree method of [42, 43] to communication models with three or more parties, resolving well-known questions [33, 14, 6, 50, 49] in communication complexity and circuit complexity. Yet another example of the use of threshold degree in complexity theory is the first exponential lower bound on the sign-rank of AC^0 circuits [40], posed as a challenge by Babai et al. [5] twenty-two years earlier.

1.1. Our results. In light of these algorithmic and complexity-theoretic applications, the problem of determining the threshold degree of constant-depth circuits has attracted considerable attention. Forty-five years ago, Minsky and Papert [34] proved an $\Omega(n^{1/3})$ lower bound on the threshold degree of the constant-depth circuit

$$f(x) = \bigwedge_{i=1}^{n^{1/3}} \bigvee_{j=1}^{n^{2/3}} x_{ij}.$$

The only subsequent progress was a lower bound of $\Omega(n^{1/3} \log^k n)$ for an arbitrary constant k, due to O'Donnell and Servedio [36]. In other words, it has been open since 1969 to obtain a polynomial improvement on Minsky and Papert's lower bound. We give a detailed solution to this problem. Our main result is as follows:

THEOREM 1.1. *Let $k \geqslant 1$ be any fixed integer. Define $f\colon \{0,1\}^n \to \{0,1\}$ by*

$$f = \mathrm{NOR}_{n^{\frac{1}{2k-1}}} \circ \underbrace{\mathrm{NOR}_{n^{\frac{2}{2k-1}}} \circ \cdots \circ \mathrm{NOR}_{n^{\frac{2}{2k-1}}}}_{k-1}.$$

Then

$$\deg_{\pm}(f) = \Omega\left(n^{\frac{k-1}{2k-1}}\right).$$

As usual, the symbol $\circ$ denotes function composition. Thus, the function f above is a depth-k tree of NOR gates, with top fan-in $n^{1/(2k-1)}$ and all other fan-ins $n^{2/(2k-1)}$. Recall that by De Morgan's law, a tree of NOR gates is equivalent to a tree of alternating AND and OR gates of the same depth and size. For typesetting convenience, we work with NOR trees throughout this manuscript.

Several remarks are in order. For depth $k = 2$, Theorem 1.1 gives a new and entirely different proof of Minsky and Papert's classic $\Omega(n^{1/3})$ lower bound. For depth $k = 3$, Theorem 1.1 proves a conjecture of O'Donnell and Servedio [36] who proposed the function $\mathrm{AND}_{n^{1/5}} \circ \mathrm{OR}_{n^{2/5}} \circ \mathrm{AND}_{n^{2/5}}$ as a candidate for threshold degree $\Omega(n^{2/5})$. Finally, the lower bound of Theorem 1.1 is essentially optimal. As k grows, the bound approaches $\Omega(\sqrt{n})$, nearly matching a well-known $O(\sqrt{n})$ upper bound on the threshold degree of arbitrary read-once Boolean formulas [32]. Moreover, we show that for any fixed depth k, the lower bound of Theorem 1.1 is tight for a large class Boolean formulas:

THEOREM 1.2. *Let $k \geqslant 1$ be any fixed integer. Define $f\colon \{0,1\}^n \to \{0,1\}$ by*

$$f = \mathrm{NOR}_{n_1} \circ \mathrm{NOR}_{n_2} \circ \cdots \circ \mathrm{NOR}_{n_k},$$

where $n_1, n_2, \ldots, n_k$ are arbitrary integers with $n_1 n_2 \cdots n_k = n$. Then

$$\deg_\pm(f) = O\left(n^{\frac{k-1}{2k-1}} \log n\right).$$

Our techniques allow us to prove another conjecture on the threshold degree of constant-depth circuits. The *element distinctness* function $\mathrm{ED}_n\colon \{0,1\}^{n\lceil \log n\rceil} \to \{0,1\}$ is given by

$$\mathrm{ED}_n(x) = \bigwedge_{\substack{i,j=1,2,\ldots,n:\\ i\neq j}} \bigvee_{k=1}^{\lceil \log n\rceil} x_{i,k} \oplus x_{j,k}.$$

Viewing the arguments to ED_n as $\lceil \log n\rceil$-bit integers, the function evaluates to true if and only if these n integers are pairwise distinct. A moment's reflection reveals that ED_n is a CNF formula of polynomial size. Bun and Thaler [13] proposed the composed function $\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}$ as another candidate for threshold degree $\Omega(n^{2/5})$, a conjecture that we prove in this paper:

THEOREM 1.3. *Consider the depth-3 polynomial-size $\{\wedge, \vee\}$-circuit f given by*

$$f = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}.$$

Then

$$\deg_\pm(f) \geqslant \Omega(n^{2/5}).$$

The lower bound in this theorem is optimal up to a logarithmic factor. This function is quite different from the corresponding construction of Theorem 1.1 for depth $k = 3$. Remarkably, the threshold degree in both cases turns out to be the same up to a logarithmic factor: $\Omega(n^{2/5})$ versus $\Omega(n/\log n)^{2/5}$, where n denotes the total number of variables.

1.2. Further applications. Lower bounds on the threshold degree translate in a black-box manner into various lower bounds in computational learning theory and communication complexity. We focus on two illustrative applications in these research areas. By the *pattern matrix method* [42, 43, 50, 49], Theorem 1.1 gives an improved construction of a constant-depth circuit with exponentially small discrepancy:

THEOREM 1.4. *For every $\epsilon > 0$, there is an (explicitly given) two-party communication problem $f\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, representable by a read-once $\{\wedge, \vee\}$-formula of constant depth, with discrepancy*

$$\mathrm{disc}(f) \leqslant \exp\left(-\Omega\left(n^{\frac{1}{2}-\epsilon}\right)\right).$$

The best previous upper bound was $\exp(-\Omega(n/\log n)^{2/5})$, due to Bun and Thaler [13], preceded by an upper bound of $\exp(-\Omega(n^{1/3}))$ due to Buhrman et al. [11] and Sherstov [42, 43]. By the results of [50, 49], Theorem 1.4 generalizes to three or more parties.

As a second application, we consider the notions of *threshold weight* and *threshold density*, defined for a given Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ as the minimum size of a majority-of-parity and threshold-of-parity circuit for f, respectively. Both quantities play a prominent role in computational learning theory. By the black-box reduction in [29], Theorem 1.1 in this paper implies:

THEOREM 1.5. *For every $\epsilon > 0$, there is an (explicitly given) read-once $\{\wedge,\vee\}$-formula $f\colon \{0,1\}^n \to \{0,1\}$ of constant depth with threshold weight and threshold density*

$$\exp\left(\Omega\left(n^{\frac{1}{2}-\epsilon}\right)\right).$$

Prior to Theorem 1.5, the best lower bounds for circuits of constant depth were as follows: $\exp(\Omega(n/\log n)^{2/5})$ for threshold weight, due to Bun and Thaler [13], and $\exp(\Omega(n^{1/3}))$ for threshold density, due to Krause and Pudlák [29]. We defer a detailed exposition of these and other applications to Section 8.

1.3. Proof overview. Sign-representation is a particularly powerful analytic model, which explains the difficulty of proving lower bounds on the threshold degree. A much weaker model is that of *uniform approximation*, whereby a real polynomial represents a Boolean function f if it approximates f pointwise within $1/3$, ranging in $[-1/3,1/3]$ on $f^{-1}(0)$ and in $[2/3,4/3]$ on $f^{-1}(1)$. Central to our proof is a hybrid model, best thought of as *one-sided approximation* [16, 12, 45, 13], in which the representing polynomial ranges in $[-1/3,1/3]$ on $f^{-1}(0)$ and in $[2/3,+\infty)$ on $f^{-1}(1)$. The complexity measure of a Boolean function f in each of these cases is the minimum degree of a real polynomial that represents f: the threshold degree, approximate degree, and one-sided approximate degree of f, respectively.

We obtain our results by proving the following more general statement.

THEOREM 1.6. *Let f be an arbitrary Boolean function, with one-sided approximate degree d. Then for all integers $n,k \geqslant 0$,*

$$\deg_{\pm}(\mathrm{NOR}_{cn} \circ \underbrace{\mathrm{NOR}_{cn^2} \circ \cdots \circ \mathrm{NOR}_{cn^2}}_{k} \circ f) \geqslant n^k \min\{n,d\}, \tag{1.1}$$

where $c \geqslant 1$ is an absolute constant.

Theorem 1.6 gives the best possible lower bound on the threshold degree of the composition (1.1) in terms of the one-sided approximate degree of f. We consider this result to be of independent interest. It allows one to start with a function f that has high one-sided approximate degree—a weak notion of hardness—and transform it into a vastly harder function, with high threshold degree. We deduce our lower bounds in Theorems 1.1 and 1.3 from Theorem 1.6 by letting f be either the NOR function or the element distinctness function, for both of which the one-sided approximate degree is known.

We give *three* different proofs of Theorem 1.6, one for arbitrary k and two simpler ones for the special case $k = 0$. We describe all three below. While the main result of this paper (Theorem 1.1) requires the full power of Theorem 1.6 for arbitrary k, the case $k = 0$ is already sufficient to prove an $\Omega(n^{2/5})$ lower bound on the threshold degree of constant-depth circuits.

Proof for arbitrary *k*. The search for a sign-representing polynomial for a given Boolean function f can be formulated as a linear program. By strong duality, the *nonexistence* of a sign-representing polynomial is therefore equivalent to the *existence* of a certain dual object. This dual point of view has been influential in past research [36, 42, 46, 44, 12, 45] and plays a central role in our paper as well. Put another way, we prove Theorem 1.6 constructively, by exhibiting a feasible object in the dual space. This object must be a nonzero function that agrees with f in sign and is additionally orthogonal to low-degree polynomials.

The key challenge is ensuring the agreement in sign between the dual object and the Boolean function f. This contrasts with simpler settings such as uniform approximation, where the dual object is allowed to disagree with f on a small fraction of inputs. The vast majority of methods developed to date, including most recently the paper of Bun and Thaler [13], only work for uniform approximation.

We pursue a different approach. At a high level, the proof proceeds by induction on circuit depth. For each depth, we do more than rule out a sign-representing polynomial—rather, we construct a pair of highly structured dual objects that imply high threshold degree and additionally allow for induction. A recurring technique in this paper is the construction of dual objects with desired analytic or metric properties by taking *convex combinations* of dual objects that almost have the desired properties. The technical part of the paper includes intuitive descriptions at each level of granularity.

Proofs for $k = 0$. This case corresponds to compositions of the form $\mathrm{NOR}_n \circ f$, where f is an arbitrary Boolean function. Equivalently, we may speak of $\mathrm{OR}_n \circ f$ since threshold degree is invariant under negation. We are able to fully characterize the threshold degree of any such composition.

To build intuition for our result, suppose that

$$\left\| f - \frac{p}{q} \right\|_\infty < \frac{1}{2n},$$

where p and q are polynomials. Then $\mathrm{OR}_n \circ f$ is sign-represented by

$$\sum_{i=1}^{n} \frac{p(x_i)}{q(x_i)} - \frac{1}{2}.$$

To obtain a sign-representing polynomial for $\mathrm{OR}_n \circ f$, it suffices to multiply through by the positive quantity $\prod q(x_i)^2$. In summary, the threshold degree of $\mathrm{OR}_n \circ f$ is at most $\deg p + 2n \deg q$. This construction is due to Beigel et al. [8], who used it in an ingenious way to prove the closure of PP under intersection. In previous work [46], we showed that this construction is optimal for $n = 2$, i.e., the threshold degree of $\mathrm{OR}_2 \circ f$ equals (up to a small multiplicative constant) the least degree of a rational function that approximates f pointwise. However, no characterization was known for growing n.

Observe that the above construction works even if p/q approximates f in a *one-sided* manner. In fact, we prove that this modified construction achieves the smallest possible degree. Our proof works by manipulating a feasible solution to the dual of the one-sided rational approximation problem for f, in order to construct a feasible solution to the dual of the sign-representation problem for $\mathrm{OR}_n \circ f$. The proof in this paper is unrelated to the earlier work [46] for $n = 2$. As a corollary to the newly obtained characterization of the threshold degree of $\mathrm{OR}_n \circ f$, we recover the special case of Theorem 1.6 for $k = 0$.

We give yet another proof of Theorem 1.6 for $k = 0$ by combining our techniques with a construction due to Bun and Thaler [13]. Specifically, the authors of [13] proved that $\mathrm{OR}_n \circ f$ cannot be approximated uniformly within $\frac{1}{2} - \exp(-\Omega(n))$ by a polynomial of degree less than the one-sided approximate degree of f, a form of hardness amplification for uniform approximation. In and of itself, that result does not imply anything about the threshold degree of $\mathrm{OR}_n \circ f$. Indeed, there are examples of functions [38, 39, 48] with threshold degree 1 that cannot be approximated uniformly within $\frac{1}{2} - \exp(-\Omega(n))$ by a polynomial of degree cn for some constant $c > 0$. Nevertheless, we are able to adapt the techniques of this work to the setting of Bun and Thaler [13] and thereby obtain another proof of Theorem 1.6 for $k = 0$.

2. Preliminaries. We use the term *Euclidean space* to refer to $\mathbb{R}^n$ for some positive integer n. Throughout this paper, Boolean functions are mappings $X \to \{0,1\}$ for some finite subset X of Euclidean space, most often $X = \{0,1\}^n$. For Boolean functions $f\colon \{0,1\}^n \to \{0,1\}$ and $g\colon X \to \{0,1\}$, we let $f \circ g$ denote the componentwise composition of f with g, i.e., the Boolean function on X^n that sends $(x_1, x_2, \ldots, x_n) \mapsto f(g(x_1), g(x_2), \ldots, g(x_n))$. By associativity, this definition extends unambiguously to compositions $f_1 \circ f_2 \circ \cdots \circ f_k$ of three or more functions.

For a bit string $x \in \{0,1\}^n$, we let $|x| = x_1 + x_2 + \cdots + x_n$ denote the Hamming weight of x. The *kth level* of the Boolean hypercube $\{0,1\}^n$ is the subset $\{x \in \{0,1\}^n : |x| = k\}$. The notation $\log x$ refers to the logarithm of x to base 2. The negation of a Boolean function $f\colon X \to \{0,1\}$ is denoted $\neg f$ and defined as usual by $(\neg f)(x) = \neg f(x)$. The functions $\mathrm{AND}_n, \mathrm{OR}_n, \mathrm{NOR}_n\colon \{0,1\}^n \to \{0,1\}$ have their standard definitions:

$$\mathrm{AND}_n(x) = \bigwedge_{i=1}^n x_i, \qquad \mathrm{OR}_n(x) = \bigvee_{i=1}^n x_i, \qquad \mathrm{NOR}_n = \neg\mathrm{OR}_n.$$

The *element distinctness* function $\mathrm{ED}_n\colon (\{0,1\}^{\lceil \log n\rceil})^n \to \{0,1\}$ is given by

$$\mathrm{ED}_n(x) = \bigwedge_{\substack{i,j=1,2,\ldots,n:\\ i\neq j}} \ \bigvee_{k=1}^{\lceil \log n\rceil} x_{i,k} \oplus x_{j,k}.$$

Viewing the arguments to ED_n as $\lceil \log n\rceil$-bit integers, the function evaluates to true if and only if these n integers are pairwise distinct. The sign function is denoted

$$\operatorname{sgn} t = \begin{cases} -1 & \text{if } t < 0,\\ 0 & \text{if } t = 0,\\ 1 & \text{if } t > 0.\end{cases}$$

For a multivariate real polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$, we let $\deg p$ denote the total degree of p, i.e., the largest degree of any monomial of p. We use the terms *degree* and *total degree* interchangeably in this paper. The following simple but fundamental fact, due to Minsky and Papert [34], allows one to transform a multivariate real polynomial on $\{0,1\}^n$ into a related univariate real polynomial on $\{0,1,2,\ldots,n\}$ without an increase in degree.

PROPOSITION 2.1 (Minsky and Papert). *Let $p\colon \{0,1\}^n \to \mathbb{R}$ be an arbitrary polynomial. Then the mapping*

$$m \mapsto \mathop{\mathbf{E}}_{\substack{x\in\{0,1\}^n\\ |x|=m}} p(x) \qquad\qquad (m = 0,1,2,\ldots,n)$$

is a univariate real polynomial of degree at most $\deg p$.

We adopt the convention that $0^0 = 1$, justified by continuity.

2.1. Norms and products. For a finite set X, we let $\mathbb{R}^X$ denote the linear space of functions $f\colon X \to \mathbb{R}$. This space is equipped with the usual norms and inner product:

$$\|f\|_\infty = \max_{x\in X} |f(x)|,$$
$$\|f\|_1 = \sum_{x\in X} |f(x)|,$$
$$\langle f, g\rangle = \sum_{x\in X} f(x)g(x).$$

The *tensor product* of $f \in \mathbb{R}^X$ and $g \in \mathbb{R}^Y$ is the real function $f \otimes g \in \mathbb{R}^{X\times Y}$ defined by $(f \otimes g)(x, y) = f(x)g(y)$. The tensor product $f \otimes f \otimes \cdots \otimes f$ (n times) is abbreviated $f^{\otimes n}$. The pointwise product of $f, g \in \mathbb{R}^X$ is denoted $f \cdot g \in \mathbb{R}^X$ and is given by $(f \cdot g)(x) = f(x)g(x)$. Note that as functions, $f \cdot g$ is a restriction of $f \otimes g$. The *support* of a function $f\colon X \to \mathbb{R}$ is denoted $\operatorname{supp} f = \{x \in X : f(x) \neq 0\}$. A *convex combination* of $f_1, f_2, \ldots, f_k \in \mathbb{R}^X$ is any function of the form $\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_k f_k$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are nonnegative and sum to 1. The *convex hull* of $F \subseteq \mathbb{R}^X$, denoted $\operatorname{conv} F$, is the set of all convex combinations of functions in F.

For $f\colon X \to \mathbb{R}$, the symbols $|f|$ and $\operatorname{sgn} f$ have their usual meanings as the real functions given by $|f|(x) = |f(x)|$ and $(\operatorname{sgn} f)(x) = \operatorname{sgn} f(x)$. In the context of functions, the relational operators $\leqslant, =,$ and $\geqslant$ and arithmetic operations are applied pointwise. For example, the phrase "$f \geqslant 2|g|$ on X" means that $f(x) \geqslant 2|g(x)|$ for every $x \in X$.

Throughout this manuscript, we view probability distributions as real functions, which allows us to use the various notational devices introduced above. In particular, for probability distributions μ and λ, the symbol $\operatorname{supp} \mu$ denotes the support of μ, and $\mu \otimes \lambda$ denotes the probability distribution given by $(\mu \otimes \lambda)(x, y) = \mu(x)\lambda(y)$. If μ is a probability distribution on X, we consider μ to be defined on any superset of X with the understanding that $\mu = 0$ outside X.

2.2. Approximation by polynomials. Let $f\colon X \to \{0, 1\}$ be given, for a finite subset $X \subset \mathbb{R}^n$. The *ϵ-approximate degree* of f, denoted $\deg_\epsilon(f)$, is the least degree of a real polynomial p such that $\|f - p\|_\infty \leqslant \epsilon$. We refer to any such polynomial for f as a *uniform approximant* with error ϵ. Define

$$E(f, d) = \min_{p:\deg p\leqslant d} \|f - p\|_\infty,$$

where the minimum is over polynomials of degree at most d. In words, $E(f, d)$ is the least error to which f can be approximated by a real polynomial of degree no greater than d. In this notation, $\deg_\epsilon(f) = \min\{d : E(f, d) \leqslant \epsilon\}$. In the study of Boolean functions, the standard setting of the error parameter is $\epsilon = 1/3$.

Observe that $\deg_{1/2}(f) = 0$ for every Boolean function f, the approximant in question being the constant polynomial $1/2$. While the $1/2$-approximate degree of a Boolean function is always a trivial concept, the *limit* of the ϵ-approximate degree as $\epsilon \nearrow 1/2$ turns out to be a fundamental and mathematically rich notion. It is known as the *threshold degree* of f, denoted

$$\deg_\pm(f) = \lim_{\epsilon\nearrow 1/2} \deg_\epsilon(f).$$

It is a simple but instructive exercise to verify that $\deg_\pm(f)$ is precisely the least degree of a real polynomial p that represents f in sign:

$$\operatorname{sgn} p(x) = \begin{cases} -1 & \text{if } f(x) = 0, \\ +1 & \text{if } f(x) = 1. \end{cases}$$

Clearly,

$$\deg_{\pm}(f) \leqslant \deg_{\epsilon}(f), \qquad\qquad 0 \leqslant \epsilon < \frac{1}{2}.$$

Key to our work is a hybrid notion of approximation whereby a Boolean function f is approximated uniformly on $f^{-1}(0)$ and represented in sign on $f^{-1}(1)$. Formally, the *one-sided ϵ-approximate degree* of f, denoted $\deg_{\epsilon}^{+}(f)$, is the least degree of a real polynomial p such that

$$\begin{aligned} f(x) - \epsilon \leqslant p(x) &\leqslant f(x) + \epsilon, & x &\in f^{-1}(0), \\ f(x) - \epsilon &\leqslant p(x), & x &\in f^{-1}(1). \end{aligned}$$

We refer to any such polynomial for f as a *one-sided approximant* with error ϵ. Again, the canonical setting of the error parameter is $\epsilon = 1/3$. Threshold degree and ϵ-approximate degree are invariant under function negation:

$$\deg_{\pm}(f) = \deg_{\pm}(\neg f), \tag{2.1}$$

$$\deg_{\epsilon}(f) = \deg_{\epsilon}(\neg f) \tag{2.2}$$

for every Boolean function f and every ϵ. In contrast, the gap between the one-sided approximate degree of a Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$ versus its negation $\neg f$ can be as large as 1 versus $\Omega(\sqrt{n})$, achieved for $f = \mathrm{OR}_n$.

We will need tight bounds on the one-sided approximate degree of several functions. The following theorem, due to Nisan and Szegedy [35], was one of the first results in this line of work.

Theorem 2.2 (Nisan and Szegedy).

$$\deg_{1/3}(\mathrm{NOR}_n) = \Theta(\sqrt{n}),$$
$$\deg_{1/3}^{+}(\mathrm{NOR}_n) = \Theta(\sqrt{n}).$$

The following result, obtained recently by Bun and Thaler [13, Appendix A], generalizes earlier work [1, 2] on the approximate degree of element distinctness to the one-sided case.

Theorem 2.3 (Bun and Thaler).

$$\deg_{1/3}^{+}(\mathrm{ED}_n) = \Omega(n^{2/3}).$$

2.3. Dual characterizations. Each of the approximation-theoretic notions reviewed in the previous section has a dual characterization, obtained by an appeal to linear programming duality. For threshold degree, we have:

Theorem 2.4. *Let $f\colon X \to \{0,1\}$ be given. Then $\deg_{\pm}(f) \geqslant d$ if and only if there exists $\psi\colon X \to \mathbb{R}$ such that*

(i) $\psi(x) \geqslant 0$ *whenever* $f(x) = 1$,

(ii) $\psi(x) \leqslant 0$ *whenever* $f(x) = 0$,

(iii) $\langle \psi, p \rangle = 0$ *for every polynomial* p *of degree less than* d*, and*
(iv) $\psi \not\equiv 0$.

A convenient shorthand for (i) and (ii), which we use often, is $(-1)^{1-f}\psi \geqslant 0$. We refer the reader to [4, 36, 46] for a proof of Theorem 2.4. Analogously, approximate degree has the following dual characterization [43, 52]:

THEOREM 2.5. *Let* $f: X \to \{0,1\}$ *be given. Then* $\deg_\epsilon(f) \geqslant d$ *if and only if there exists* $\psi: X \to \mathbb{R}$ *such that*
(i) $\langle f, \psi \rangle > \epsilon \|\psi\|_1$,
(ii) $\langle \psi, p \rangle = 0$ *for every polynomial* p *of degree less than* d.

Finally, the dual characterization of one-sided approximate degree is as follows [13].

THEOREM 2.6. *Let* $f: X \to \{0,1\}$ *be given. Then* $\deg^+_\epsilon(f) \geqslant d$ *if and only if there exists* $\psi: X \to \mathbb{R}$ *such that*
(i) $\langle f, \psi \rangle > \epsilon \|\psi\|_1$,
(ii) $\langle \psi, p \rangle = 0$ *for every polynomial* p *of degree less than* d*, and*
(iii) $\psi(x) \geqslant 0$ *whenever* $f(x) = 1$.

The dual objects that arise in Theorems 2.4 to 2.6 share the following metric properties.

PROPOSITION 2.7. *Let* $\psi: X \to \mathbb{R}$ *be given with* $\langle \psi, 1 \rangle = 0$*. Then*
(i) $\sum_{x:\psi(x)>0} |\psi(x)| = \|\psi\|_1/2$,
(ii) $\|\psi\|_\infty \leqslant \|\psi\|_1/2$,
(iii) $\langle f, \psi \rangle \leqslant \|\psi\|_1/2$ *for every Boolean function* $f: X \to \{0,1\}$.

Proof. (i) We have

$$\sum_{x:\psi(x)>0} |\psi(x)| = \frac{\langle |\psi| + \psi, 1 \rangle}{2} = \frac{\langle |\psi|, 1 \rangle}{2} = \frac{\|\psi\|_1}{2}.$$

(ii) For every $x^* \in X$,

$$0 = |\langle \psi, 1 \rangle| \geqslant |\psi(x^*)| - \sum_{x \neq x^*} |\psi(x)| = 2|\psi(x^*)| - \|\psi\|_1.$$

(iii) Immediate from (i) since f ranges in $\{0,1\}$. ◻

Most proofs in this paper involve explicit constructions of dual objects ψ as in Theorems 2.4 to 2.6. A common step in such constructions is verifying that a candidate object ψ is orthogonal to polynomials of low degree. We will make frequent use of the following observation.

PROPOSITION 2.8. *Let* n, k, d *be nonnegative integers, where* $n \geqslant 1$*. Let* $\psi: X \to \mathbb{R}$ *be a function on a finite subset* X *of Euclidean space such that*

$$\langle \psi, p \rangle = 0$$

for every polynomial p *of degree less than* d*. Let* $g: X^n \to \mathbb{R}$ *be given by*

$$g(x_1, x_2, \ldots, x_n) = \sum_{i_1 < i_2 < \cdots < i_k} g_{i_1, i_2, \ldots, i_k}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}),$$

for some functions $g_{i_1, i_2, \ldots, i_k}: X^k \to \mathbb{R}$*. Then*

$$\langle \psi^{\otimes n} \cdot g, P \rangle = 0$$

for every polynomial P *of degree less than* $(n-k)d$.

Proof. By linearity, it suffices to prove the proposition for factored polynomials $P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^n p_i(x_i)$. By hypothesis, the degrees of $p_1, p_2, \ldots, p_n$ sum to less than $(n-k)d$. In particular, every subset $S \subseteq \{1, 2, \ldots, n\}$ of cardinality $|S| \geqslant n-k$ obeys $\min_{i \in S} \deg p_i < d$ and therefore

$$\prod_{i \in S} \langle \psi, p_i \rangle = 0.$$

We conclude that

$$\begin{aligned}\langle \psi^{\otimes n} \cdot g, P \rangle &= \sum_{i_1 < i_2 < \cdots < i_k} \langle \psi^{\otimes k} \cdot g_{i_1, i_2, \ldots, i_k}, p_{i_1} \otimes p_{i_2} \otimes \cdots \otimes p_{i_k} \rangle \prod_{i \notin \{i_1, i_2, \ldots, i_k\}} \langle \psi, p_i \rangle \\ &= 0. \end{aligned}$$ ◻

We will also need an explicit dual object for the NOR function, in the sense of Theorem 2.6. There are known constructions of such objects, due to Špalek [54] and Bun and Thaler [12], but we require additional properties not ensured by previous work.

THEOREM 2.9. *Let ϵ be given, $0 < \epsilon < 1$. Then for some $\delta = \delta(\epsilon) > 0$ and every $n \geqslant 2$, there exists an (explicitly given) function $\omega \colon \{0, 1, 2, \ldots, n\} \to \mathbb{R}$ such that*

$$\begin{aligned} &\omega(0) > \frac{1-\epsilon}{2} \cdot \|\omega\|_1, \\ &(-1)^{n+t} \omega(t) \geqslant \frac{\epsilon}{4t^2} \cdot \|\omega\|_1 && (t = 1, 2, \ldots, n), \\ &\deg p < \sqrt{\delta n} \implies \langle \omega, p \rangle = 0. \end{aligned}$$

The proof of this result is an adaptation of previous analyses [54, 12] and can be found in Appendix A.

2.4. Robust polynomials. A natural approach to approximating a composed function $f \circ g$ is to approximate f and g separately and compose the resulting approximants. For this approach to work, the approximating polynomial for f needs to be robust to noise in the inputs, i.e., it needs to approximate f not only on the Boolean hypercube but also on any perturbation of a Boolean vector. The following result from [47] gives an efficient procedure for making any polynomial robust to noise.

THEOREM 2.10 (Sherstov). *Let $p \colon \{0,1\}^n \to [-1, 1]$ be a given polynomial. Then for every $\delta > 0$, there is a polynomial $p_{\mathrm{robust}} \colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(\deg p + \log \frac{1}{\delta})$ such that*

$$|p(x) - p_{\mathrm{robust}}(x + \epsilon)| < \delta$$

for every $x \in \{0,1\}^n$ and $\epsilon \in [-1/3, 1/3]^n$.

Note that the degree of the robust polynomial grows additively rather than multiplicatively with the error parameter δ. This fact will play a crucial role in the next section, where we prove our upper bound on the threshold degree of constant-depth circuits. It follows from the above result that the approximate degree is always well-behaved under function composition [47]:

COROLLARY 2.11 (Sherstov). *Let $f \colon \{0,1\}^n \to \{0,1\}$ and $g \colon X \to \{0,1\}$ be given. Then*

$$\deg_{1/3}(f \circ g) \leqslant c \deg_{1/3}(f) \deg_{1/3}(g)$$

for some absolute constant $c > 0$ independent of f, g, n.

3. The upper bound. Consider an AND-OR tree of depth k on n variables, in which the fan-in may vary from level to level but is the same for all gates at any given level. O'Donnell and Servedio [36] made the following ingenious observation in a footnote of their paper: either the product of the odd-level fan-ins is at most $\sqrt{n}$ or the product of the even-level fan-ins is at most $\sqrt{n}$, which means that the standard arithmetization of the AND-OR tree gives a sign-representing polynomial of degree at most $\sqrt{n}$.

While O'Donnell and Servedio's construction falls short of achieving our desired bound of $O(n^{\frac{k-1}{2k-1}}\log n)$, the trick of odd- versus even-level fan-ins plays an essential role in our proof. The other key ingredient is work on robust approximation [47], which allows one to make a polynomial robust to noise with essentially no overhead in degree. We start by calculating the parameters in the above construction.

LEMMA 3.1 (cf. O'Donnell and Servedio). *Let* $f = \mathrm{NOR}_{n_k} \circ \mathrm{NOR}_{n_{k-1}} \circ \cdots \circ \mathrm{NOR}_{n_1}$, *where* $n_1 n_2 \cdots n_k = n$. *Then*

$$E(f, n_2 n_4 n_6 \cdots) \leqslant \frac{1}{2} - \frac{1}{2n^{n_2 n_4 n_6 \cdots}}. \tag{3.1}$$

Proof. By working with the negation of f if necessary, we may assume that

$$f(x) = \underbrace{\cdots \bigvee_{i_3=1}^{n_3} \bigwedge_{i_2=1}^{n_2} \bigvee_{i_1=1}^{n_1}}_{k} x_{i_1,i_2,\ldots,i_k}.$$

Consider the polynomial

$$p(x) = \cdots \sum_{i_3=1}^{n_3} \prod_{i_2=1}^{n_2} \sum_{i_1=1}^{n_1} x_{i_1,i_2,\ldots,i_k}.$$

It is clear that $\deg p = n_2 n_4 n_6 \cdots$. Moreover, $f(x) = 0$ forces $p(x) = 0$, whereas $f(x) = 1$ forces

$$1 \leqslant p(x) \leqslant ((n_1^{n_2} n_3)^{n_4}\, n_5)^{n_6} \ldots \leqslant n^{n_2 n_4 n_6 \cdots}.$$

Now (3.1) is immediate, the approximant in question being

$$\frac{1}{2} + \frac{1}{n^{n_2 n_4 n_6 \cdots}} \left(p(x) - \frac{1}{2} \right). \qquad \square$$

Equation (3.1) shows that O'Donnell and Servedio's approach gives a uniform approximant with reasonable accuracy, rather than just a sign-representing polynomial. Combining this fact with results on robust approximation, we obtain a *robust* sign-representing polynomial for the AND-OR tree:

COROLLARY 3.2. *Let* $f = \mathrm{NOR}_{n_k} \circ \mathrm{NOR}_{n_{k-1}} \circ \cdots \circ \mathrm{NOR}_{n_1}$, *where* $n_1 n_2 \cdots n_k = n$. *Then there is a polynomial* $p_{\text{robust}} \colon \mathbb{R}^n \to \mathbb{R}$ *such that*

$$\deg p_{\text{robust}} \leqslant (n_2 n_4 n_6 \cdots) \cdot c \log n \tag{3.2}$$

for some absolute constant $c > 0$, *and*

$$|f(x) - p_{\text{robust}}(x + \epsilon)| \leqslant \frac{1}{2} - \frac{1}{4n^{n_2 n_4 n_6 \cdots}} \tag{3.3}$$

for every $x \in \{0,1\}^n$ *and* $\epsilon \in [-1/3, 1/3]^n$.

Proof. By Lemma 3.1, there is a polynomial $p\colon \{0,1\}^n \to [-2,2]$ of degree at most $n_2 n_4 n_6 \cdots$ such that

$$\|f-p\|_\infty \leqslant \frac{1}{2} - \frac{1}{2n^{n_2n_4n_6\cdots}}.$$

Invoking Theorem 2.10 with $\delta = 1/8n^{n_2n_4n_6\cdots}$ gives a polynomial $p_{\mathrm{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(\deg p + \log\frac{1}{\delta})$ such that

$$|p(x) - p_{\mathrm{robust}}(x+\epsilon)| \leqslant \frac{1}{4n^{n_2n_4n_6\cdots}}$$

for every $x \in \{0,1\}^n$ and $\epsilon \in [-1/3, 1/3]^n$. Now (3.2) and (3.3) are immediate. ∎

We are now in a position to describe the final construction. We start by splitting the NOR tree at some level into a top part and a bottom part. Next, we construct a robust sign-representing polynomial for the top part, and a uniform approximant with error 1/3 for the bottom part. Finally, we compose the resulting polynomials to obtain a sign-representing polynomial for the original tree. This approach is made precise in the following theorem.

THEOREM 3.3. *Let* $f = \mathrm{NOR}_{n_k} \circ \mathrm{NOR}_{n_{k-1}} \circ \cdots \circ \mathrm{NOR}_{n_1}$, *where* $n_1 n_2 \cdots n_k = n$. *Then*

$$\deg_\pm(f) \leqslant c^k \min_{i=0,1,\ldots,k-1} \{\sqrt{n_1 n_2 \cdots n_i}\, n_{i+2} n_{i+4} n_{i+6} \cdots\} \log n, \tag{3.4}$$

for some absolute constant $c \geqslant 1$.

Proof. Fix i arbitrarily and write $f = f' \circ f''$, where

$$\begin{aligned} f' &= \mathrm{NOR}_{n_k} \circ \mathrm{NOR}_{n_{k-1}} \circ \cdots \circ \mathrm{NOR}_{n_{i+1}}, \\ f'' &= \mathrm{NOR}_{n_i} \circ \mathrm{NOR}_{n_{i-1}} \circ \cdots \circ \mathrm{NOR}_{n_1}. \end{aligned}$$

Corollary 3.2 provides a polynomial p'_{robust} of degree at most $(n_{i+2} n_{i+4} n_{i+6} \cdots) \cdot c' \log n$ for some absolute constant $c' \geqslant 1$ such that

$$|f'(x) - p'_{\mathrm{robust}}(x+\epsilon)| < \frac{1}{2} \tag{3.5}$$

for every $x \in \{0,1\}^{n_{i+1}n_{i+2}\cdots n_k}$ and $\epsilon \in [-1/3, 1/3]^{n_{i+1}n_{i+2}\cdots n_k}$.

On the other hand, Theorem 2.2 states that $\deg_{1/3}(\mathrm{NOR}_m) = O(\sqrt{m})$, whence by Corollary 2.11 the 1/3-approximate degree of f'' does not exceed $(c'')^i \sqrt{n_1 n_2 \cdots n_i}$ for some absolute constant $c'' \geqslant 1$. Fix a polynomial p'' of that degree, with

$$\|f'' - p''\|_\infty \leqslant \frac{1}{3}. \tag{3.6}$$

By (3.5) and (3.6),

$$\|f' \circ f'' - p'_{\mathrm{robust}} \circ p''\|_\infty < \frac{1}{2}.$$

In summary, the threshold degree of $f = f' \circ f''$ is at most the product of the degrees of p'_{robust} and p'', whence (3.4). ∎

We have arrived at the main result of this section, which settles Theorem 1.2 from the Introduction.

THEOREM 3.4. *Let* $f = \mathrm{NOR}_{n_k} \circ \mathrm{NOR}_{n_{k-1}} \circ \cdots \circ \mathrm{NOR}_{n_1}$, *where* $n_1 n_2 \cdots n_k = n$. *Then*

$$\deg_{\pm}(f) \leqslant c^k \cdot n^{\frac{k-1}{2k-1}} \log n$$

for some absolute constant $c \geqslant 1$.

Proof. The idea is to carefully optimize the choice of i in the previous theorem, by replacing the minimum with a geometric mean. Specifically, let $c \geqslant 1$ be the absolute constant from Theorem 3.3. Then

$$\begin{aligned}
\frac{\deg_{\pm}(f)}{c^k \log n} &\leqslant \min_{i=0,1,\ldots,k-1} \{\sqrt{n_1 n_2 \cdots n_i}\, n_{i+2} n_{i+4} n_{i+6} \cdots\} \\
&\leqslant (n_2 n_4 n_6 \cdots)^{\frac{1}{2k-1}} \prod_{i=1}^{k-1} (\sqrt{n_1 n_2 \cdots n_i}\, n_{i+2} n_{i+4} n_{i+6} \cdots)^{\frac{2}{2k-1}},
\end{aligned}$$

where the second inequality is obtained by replacing the minimum with a weighted geometric mean of the quantities involved. Raising both sides to the power $2k-1$ and simplifying,

$$\begin{aligned}
\left(\frac{\deg_{\pm}(f)}{c^k \log n}\right)^{2k-1} &\leqslant (n_2 n_4 n_6 \cdots) \left(\prod_{i=1}^{k-1} n_1 n_2 \cdots n_i\right) \left(\prod_{i=1}^{k-1} n_{i+2}^2 n_{i+4}^2 n_{i+6}^2 \cdots\right) \\
&= \left(\prod_{j=1}^{k} n_j^{j-1-2\lfloor \frac{j-1}{2} \rfloor}\right) \left(\prod_{j=1}^{k} n_j^{k-j}\right) \left(\prod_{j=1}^{k} n_j^{2\lfloor \frac{j-1}{2} \rfloor}\right) \\
&= \prod_{j=1}^{k} n_j^{k-1} \\
&= n^{k-1}.
\end{aligned}$$

□

4. The lower bound. We prove our lower bound on the threshold degree of constant-depth circuits by induction on circuit depth. The notion of a *dual pair*, defined next, plays a central role in this inductive argument.

DEFINITION 4.1. *Let* $f \colon X \to \{0,1\}$ *be given. A* (d_0, d_1, ϵ)-dual pair *for* f *is any pair of functions* $\psi_0, \psi_1 \colon X \to \mathbb{R}$ *such that:*

(i) $\langle f, \psi_1 \rangle > \frac{1-\epsilon}{2} \|\psi_1\|_1$,

(ii) $\psi_1(x) \geqslant 0$ *whenever* $f(x) = 1$,

(iii) $\langle \psi_1, p \rangle = 0$ *for every polynomial* p *of degree less than* d_1,

(iv) $\langle \psi_0, p \rangle = 0$ *for every polynomial* p *of degree less than* d_0,

(v)
$$\psi_0(x) \begin{cases} = \max\{\psi_1(x), 0\} & \text{if } f(x) = 0, \\ \in [-\epsilon|\psi_1(x)|, \epsilon|\psi_1(x)|] & \text{if } f(x) = 1. \end{cases}$$

In the final property, the absolute value $|\psi_1(x)|$ can be replaced with $\psi_1(x)$ in view of part (ii). This definition is monotonic in ϵ, in the sense that a (d_0, d_1, ϵ)-dual pair is a (d_0, d_1, ϵ')-dual pair for every $\epsilon' > \epsilon$. In our applications, we will always take $\epsilon = 1/3$.

Properties (i)–(iii) can be summarized by saying that the function f of interest has one-sided $\frac{1-\epsilon}{2}$-approximate degree at least d_1. The dual object ψ_1 witnesses this fact, in the sense of linear programming duality (Theorem 2.6). The key difficulty is that ψ_1 need not always agree in sign with f: while such agreement is assured on $f^{-1}(1)$, there may well be inputs in $f^{-1}(0)$ on which ψ_1 is positive. The role of the accompanying object ψ_0 is to eliminate those errors without introducing new ones. For this to work efficiently, ψ_0 needs to be orthogonal to polynomials of sufficiently high degree d_0. The challenge in our proof is to inductively construct new dual pairs from old ones, while ensuring sufficiently rapid growth of d_0, d_1.

The following lemma shows how we obtain our first dual pair. It corresponds to the base case of the inductive argument.

LEMMA 4.2. *Let $f\colon X \to \{0,1\}$ be a given Boolean function, $\deg_\epsilon^+(f) > 0$. Then f has a $(1, \deg_\epsilon^+(f), \frac{1}{2\epsilon} - 1)$-dual pair.*

Proof. Abbreviate $d = \deg_\epsilon^+(f)$. By Theorem 2.6, there exists $\psi_1\colon X \to \mathbb{R}$ such that

$$\langle f, \psi_1 \rangle > \epsilon \|\psi_1\|_1 \tag{4.1}$$

as well as

$$f(x) = 1 \implies \psi_1(x) \geqslant 0, \tag{4.2}$$

$$\deg p < d \implies \langle \psi_1, p \rangle = 0. \tag{4.3}$$

Define $\psi_0\colon X \to \mathbb{R}$ by

$$\psi_0(x) = \begin{cases} \max\{\psi_1(x), 0\} & \text{if } f(x) = 0, \\[2ex] \left(1 - \dfrac{\|\psi_1\|_1}{2\langle f, \psi_1\rangle}\right)\psi_1(x) & \text{if } f(x) = 1. \end{cases}$$

With properties (4.1)–(4.3) already established, the proof will be complete once we show that

$$\left| 1 - \frac{\|\psi_1\|_1}{2\langle f, \psi_1\rangle} \right| \leqslant \frac{1}{2\epsilon} - 1, \tag{4.4}$$

$$\langle \psi_0, 1 \rangle = 0. \tag{4.5}$$

By (4.3) and Proposition 2.7 (i), (iii),

$$\sum_{x:\psi_1(x)>0} \psi_1(x) = \frac{\|\psi_1\|_1}{2}, \tag{4.6}$$

$$\langle f, \psi_1 \rangle \leqslant \frac{\|\psi_1\|_1}{2}. \tag{4.7}$$

Now the upper bound (4.4) is immediate from (4.1) and (4.7). The remaining property (4.5) can be verified as follows:

$$\begin{aligned} \langle \psi_0, 1 \rangle &= \sum_{x: f(x)=0} \psi_0(x) + \sum_{x: f(x)=1} \psi_0(x) \\ &= \sum_{x:\psi_1(x)>0} (1 - f(x))\psi_1(x) + \sum_{x \in X} f(x) \left(1 - \frac{\|\psi_1\|_1}{2\langle f, \psi_1\rangle}\right) \psi_1(x) \\ &= \underbrace{\sum_{x:\psi_1(x)>0} \psi_1(x)}_{=\|\psi_1\|_1/2} - \underbrace{\sum_{x:\psi_1(x)>0} f(x)\psi_1(x)}_{=\langle f, \psi_1\rangle} + \left(1 - \frac{\|\psi_1\|_1}{2\langle f, \psi_1\rangle}\right) \langle f, \psi_1 \rangle, \end{aligned}$$

where the final calculations use (4.2) and (4.6). ∎

The inductive step in our proof is realized by the following "amplification theorem," which transforms a dual pair for a given function f into a dual pair for the composed function $\mathrm{NOR}(f, f, \ldots, f)$.

THEOREM 4.3. *Let $\epsilon, \delta \in (0,1)$ be arbitrary. Let $f\colon X \to \{0,1\}$ be any function that has a (d_0, d_1, ϵ)-dual pair, where $d_0, d_1 \geqslant 1$. Then the function*

$$F = \mathrm{NOR}_{cn} \circ f$$

has a $(\min\{nd_0, d_1\}, \min\{nd_0, \sqrt{n} d_1\}, \delta)$-dual pair, where $c = c(\epsilon, \delta) > 0$ is a constant independent of f, n, d_0, d_1.

The proof of Theorem 4.3 is lengthy and technical, and we defer it to section 5. To complete our program, we need to bridge the notions of dual pairs and sign-representation. The following lemma does just that.

LEMMA 4.4. *Let $f\colon X \to \{0,1\}$ be any function that has a (d_0, d_1, ϵ)-dual pair for some $0 \leqslant \epsilon < 1$. Then*

$$\deg_\pm(f) \geqslant \min\{d_0, d_1\}.$$

Proof. Let (ψ_0, ψ_1) be a (d_0, d_1, ϵ)-dual pair for f. By definition,

$$\deg p < d_0 \implies \langle \psi_0, p\rangle = 0, \tag{4.8}$$

$$\deg p < d_1 \implies \langle \psi_1, p\rangle = 0, \tag{4.9}$$

$$f(x) = 1 \implies \psi_1(x) \geqslant 0, \tag{4.10}$$

$$f(x) = 1 \implies |\psi_0(x)| \leqslant \epsilon |\psi_1(x)|, \tag{4.11}$$

$$f(x) = 0 \implies \psi_0(x) = \max\{\psi_1(x), 0\}, \tag{4.12}$$

$$\langle f, \psi_1\rangle > \frac{1-\epsilon}{2}\|\psi_1\|_1. \tag{4.13}$$

Letting $\psi = \psi_1 - \psi_0$, we have

$$\deg p < \min\{d_0, d_1\} \implies \langle \psi, p\rangle = 0, \tag{4.14}$$

$$f(x) = 1 \implies \psi(x) \geqslant 0, \tag{4.15}$$

$$f(x) = 0 \implies \psi(x) \leqslant 0, \tag{4.16}$$

where the first item holds by (4.8) and (4.9), the second by (4.10) and (4.11), and the third by (4.12). Finally, we claim that

$$\psi \not\equiv 0. \tag{4.17}$$

Indeed, (4.13) implies that ψ_1 is not identically zero on $f^{-1}(1)$, whereas (4.11) ensures that $\operatorname{sgn}\psi(x) = \operatorname{sgn}\psi_1(x)$ on $f^{-1}(1)$. By (4.14)–(4.17) and Theorem 2.4, the proof is complete. ∎

Combining the above three results, we arrive at the technical centerpiece of this paper, stated previously as Theorem 1.6 in the Introduction:

THEOREM 4.5. *Let $f\colon X \to \{0,1\}$ be given. Then for all integers $n, k \geqslant 0$,*

$$\deg_\pm(\mathrm{NOR}_{cn} \circ \underbrace{\mathrm{NOR}_{cn^2} \circ \cdots \circ \mathrm{NOR}_{cn^2}}_{k} \circ f) \geqslant n^k \min\{n, \deg^+_{1/3}(f)\},$$

where $c \geqslant 1$ is an absolute constant, independent of f, n, k.

Proof. Take $c \geqslant 1$ sufficiently large, and abbreviate $m = \min\{n, \deg^+_{1/3}(f)\}$. We need only consider the case $m \geqslant 1$, the lower bound being trivial otherwise. We claim that for each $k = 0, 1, 2, \dots$, the function

$$\underbrace{\mathrm{NOR}_{cn^2} \circ \cdots \circ \mathrm{NOR}_{cn^2}}_{k} \circ f$$

has a $(\lceil mn^{k-1} \rceil, mn^k, 1/2)$-dual pair. This claim holds by induction on k, with the base case $k = 0$ settled by Lemma 4.2 and the inductive step realized by Theorem 4.3. Applying Theorem 4.3 once more shows that the function

$$\mathrm{NOR}_{cn} \circ \underbrace{\mathrm{NOR}_{cn^2} \circ \cdots \circ \mathrm{NOR}_{cn^2}}_{k} \circ f \tag{4.18}$$

has an $(mn^k, mn^k, 1/2)$-dual pair. It follows by Lemma 4.4 that the composition (4.18) has threshold degree at least mn^k, as was to be shown. ∎

Corollary 4.6. *There exists an absolute constant $c > 0$ such that for all integers $n, k \geqslant 0$,*

$$\deg_\pm(\mathrm{NOR}_n \circ \underbrace{\mathrm{NOR}_{n^2} \circ \cdots \circ \mathrm{NOR}_{n^2}}_{k}) \geqslant (cn)^k.$$

Proof. Immediate from Theorems 2.2 and 4.5. ∎

This corollary settles our main result, stated as Theorem 1.1 in the Introduction. We note that all parts of our argument (Lemmas 4.2 and 4.4 and Theorems 2.2, 4.3 and 4.5) are constructive in that they produce explicit solutions to corresponding dual linear programs. In particular, our proof produces an explicit dual object, in the sense of Theorem 2.4, that witnesses the lower bound in Corollary 4.6.

Corollary 4.7. *For every Boolean function $f\colon X \to \{0,1\}$ and every $n \geqslant 1$,*

$$\deg_\pm(\mathrm{OR}_n \circ f) \geqslant \min\{cn, \deg^+_{1/3}(f)\},$$

where $c > 0$ is an absolute constant. In particular,

$$\deg_\pm(\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}) = \Omega(n^{2/5}).$$

Proof. The first claim holds by taking $k = 0$ in Theorem 4.5 and recalling that threshold degree is invariant under function negation. The second claim is immediate from the first in light of Theorem 2.3. ∎

This settles Theorem 1.3 from the Introduction. In sections 6 and 7 we will present two alternate proofs of Corollary 4.7, completely different from the proof just given. In fact, we will fully characterize the threshold degree of $\mathrm{OR}_n \circ f$ for every f.

5. Proof of Theorem 4.3. The objective of this section is to prove Theorem 4.3 (the "amplification theorem"), which transforms a dual pair for a given Boolean function into a dual pair of higher degree for the composed function $\mathrm{NOR}(f, f, \dots, f)$. We start by reviewing the notation and hypothesis of the theorem. We then introduce auxiliary dual objects and establish their properties. In the final subsection, we put these ingredients together to obtain the desired dual pair.

5.1. Notation. We adopt verbatim the notation and hypothesis of Theorem 4.3. Specifically, f is an arbitrary Boolean function on a finite subset X of Euclidean space; the reals $0<\epsilon<1$ and $0<\delta<1$ are arbitrary parameters; and it is assumed that f has a (d_0,d_1,ϵ)-dual pair (ψ_0,ψ_1), for some positive integers d_0,d_1. By definition,

$$\deg p < d_0 \implies \langle \psi_0, p\rangle = 0, \tag{5.1}$$
$$\deg p < d_1 \implies \langle \psi_1, p\rangle = 0, \tag{5.2}$$
$$f(x)=1 \implies \psi_1(x) \geqslant 0, \tag{5.3}$$
$$f(x)=1 \implies |\psi_0(x)| \leqslant \epsilon|\psi_1(x)|, \tag{5.4}$$
$$f(x)=0 \implies \psi_0(x) = \max\{\psi_1(x),0\}, \tag{5.5}$$

and

$$\langle f,\psi_1\rangle > \frac{1-\epsilon}{2}\|\psi_1\|_1. \tag{5.6}$$

A simple but vital consequence of (5.2) is that

$$\langle \psi_1, 1\rangle = 0. \tag{5.7}$$

It follows from (5.6) that

$$\psi_1 \not\equiv 0, \tag{5.8}$$

whence by homogeneity we may assume that

$$\|\psi_1\|_1 = 1. \tag{5.9}$$

Define α by

$$\langle f,\psi_1\rangle = \frac{1-\alpha}{2}. \tag{5.10}$$

Then

$$0 \leqslant \alpha < \epsilon, \tag{5.11}$$

where the upper bound is immediate from (5.6) and (5.9), whereas the lower bound holds by (5.7), (5.9), and Proposition 2.7 (iii).

The objective of the proof is to construct a dual pair (Ψ_0,Ψ_1) with sufficiently high degrees for the Boolean function $F\colon X^N \to \{0,1\}$ given by

$$F = \mathrm{NOR}_N \circ f,$$

where $N = cn$ for some constant $c = c(\epsilon,\delta) > 0$. The construction will proceed in stages, shown schematically in Figure 5.1. The inputs to the construction, shaded in gray, are the function f, its dual pair (ψ_0,ψ_1), and the parameters n,ϵ,δ. These are combined to build more complex intermediate objects, resulting eventually in the desired dual pair (Ψ_0,Ψ_1) for F. To be precise, the intermediate objects are function *families*, indexed by nonnegative integers as in $\omega_n, L_d, \Lambda^N_{k,m}$. Throughout the proof, small letters $(f,\psi_0,\psi_1,\mu_0,\mu_1,\mu_*,p)$ are reserved for functions on X, whereas capital letters $(\Psi_0,\Psi_1,L,\Lambda,\tilde{\Lambda},P,P_0,P_1)$ refer to functions on X^N.

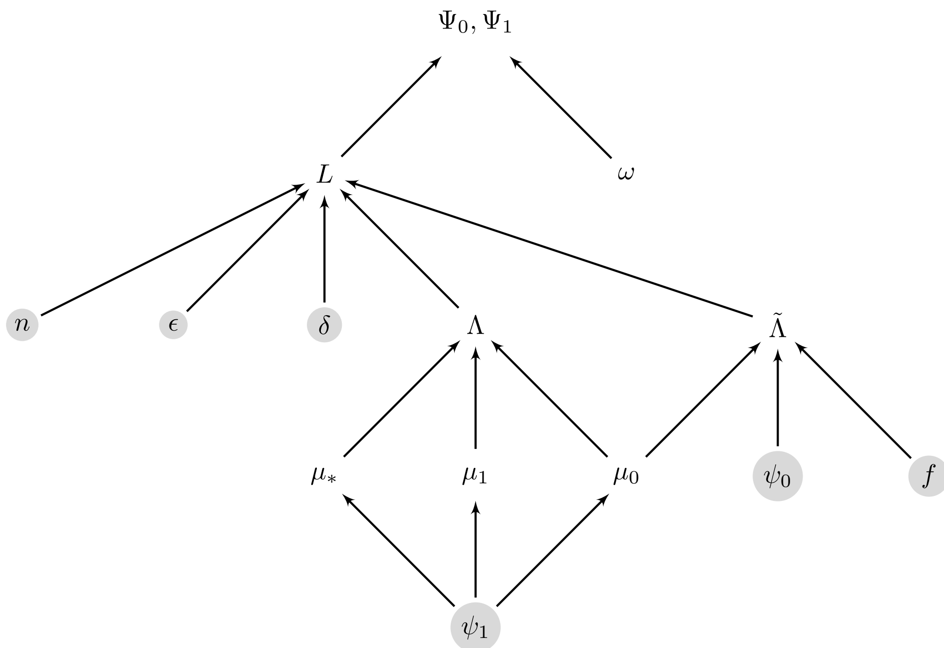

FIG. 5.1. *Construction of the dual pair* (Ψ_0, Ψ_1). *Arrows indicate dependencies.*

5.2. Fundamental distributions. We start by examining several probability distributions induced on X by the sign behavior of ψ_1. By (5.9), the function $|\psi_1|$ is a probability distribution on X, legitimizing the following definition.

DEFINITION 5.1. *Let μ_0 and μ_1 be the probability distributions induced by $|\psi_1|$ on the sets $\{x \in X : \psi_1(x) < 0\}$ and $\{x \in X : \psi_1(x) > 0\}$, respectively.*

Equations (5.7) and (5.8) guarantee that $\{x : \psi_1(x) < 0\}$ and $\{x : \psi_1(x) > 0\}$ are nonempty, so that μ_0 and μ_1 are well-defined. By (5.7),

$$\psi_1 = \frac{1}{2}\mu_1 - \frac{1}{2}\mu_0. \tag{5.12}$$

We now claim that

$$\deg p < d_1 \implies \langle \mu_0, p \rangle = \langle \mu_1, p \rangle, \tag{5.13}$$
$$f(x) = 1 \implies 2\psi_1(x) = \mu_1(x), \tag{5.14}$$
$$f(x) = 1 \implies 2|\psi_0(x)| \leqslant \epsilon\mu_1(x), \tag{5.15}$$
$$f(x) = 0 \implies 2\psi_0(x) = \mu_1(x). \tag{5.16}$$

The first item is a direct consequence of (5.2) and (5.12); the second follows from (5.3) and (5.12); the third follows from (5.4) and (5.14); and the final item holds by (5.5).

DEFINITION 5.2. *Define $\mu_* \colon X \to [0,1]$ by $\mu_*(x) = (1 - f(x))\mu_1(x)$.*

We have

$$\begin{aligned}\langle 1-f,\mu_1\rangle &= 1-\langle f,\mu_1\rangle \\ &= 1-\langle f,\mu_1-\mu_0\rangle && \text{since } \operatorname{supp}\mu_0\subseteq f^{-1}(0) \\ &= 1-2\langle f,\psi_1\rangle && \text{by (5.12)} \\ &= \alpha && \text{by (5.10)},\end{aligned}\tag{5.17}$$

whence

$$\|\mu_*\|_1=\alpha.\tag{5.18}$$

We will need the following technical result from [45, Claim 3.3], which continues to hold with μ_* replaced by any function.

LEMMA 5.3. *For every polynomial $P\colon X^N\to\mathbb{R}$ and every $k=0,1,2,\dots,N$, the mapping*

$$z\mapsto\left\langle \mu_*^{\otimes k}\otimes\bigotimes_{i=1}^{N-k}\mu_{z_i},P\right\rangle,\qquad\qquad z\in\{0,1\}^{N-k},\tag{5.19}$$

is a polynomial of degree at most $(\deg P)/d_1$.

Proof (adapted from [45]). By linearity, it suffices to consider factored polynomials of the form $P(x_1,\dots,x_N)=p_1(x_1)\cdots p_N(x_N)$. In this case (5.19) simplifies to

$$z\mapsto\prod_{i=1}^{k}\langle\mu_*,p_i\rangle\cdot\prod_{i=1}^{N-k}\langle\mu_{z_i},p_{k+i}\rangle,\qquad\qquad z\in\{0,1\}^{N-k}.\tag{5.20}$$

By (5.13), polynomials p_i of degree less than d_1 satisfy $\langle\mu_0,p_i\rangle=\langle\mu_1,p_i\rangle$ and therefore do not contribute to the degree of (5.20) as a real function on $\{0,1\}^{N-k}$. It follows that the degree of (5.20) is at most $|\{i:\deg p_i\geqslant d_1\}|\leqslant(\deg P)/d_1$. ◻

5.3. Auxiliary objects in the tensor space. The fundamental distributions μ_0 and μ_1 on X naturally give rise to the following family of functions $\Lambda^N_{k,m}\colon X^N\to[0,1]$.

DEFINITION 5.4. *For nonnegative integers k,m with $k+m\leqslant N$, define*

$$\Lambda^N_{k,m}(x_1,x_2,\dots,x_N)=\mathop{\mathbf{E}}_{S,T}\left[\prod_{i\in S}\mu_*(x_i)\cdot\prod_{i\in T}\mu_1(x_i)\cdot\prod_{i\notin S\cup T}\mu_0(x_i)\right],\tag{5.21}$$

where the expectation is with respect to a uniformly random pair of disjoint sets $S,T\subseteq\{1,2,\dots,N\}$ of size $|S|=k$ and $|T|=m$.

We proceed to examine basic analytic and metric properties of $\Lambda^N_{k,m}$.

LEMMA 5.5.
- (i) $\operatorname{supp}\Lambda^N_{k,0}\subseteq F^{-1}(1)$,
- (ii) $\langle\Lambda^N_{k,m},1\rangle=\|\Lambda^N_{k,m}\|_1=\alpha^k$,
- (iii) $\Lambda^N_{k,m}=\Lambda^N_{k',m'}$ *on $F^{-1}(1)$ whenever $k+m=k'+m'$*,
- (iv) $\langle F,\Lambda^N_{k,m}\rangle=\alpha^{k+m}$,
- (v) $\Lambda^N_{k,m}(x)\neq 0$ *only if $|\{i:\psi_1(x_i)>0\}|=k+m$.*

Proof. (i) Immediate from the fact that $\operatorname{supp}\mu_0 \subseteq f^{-1}(0)$ and $\operatorname{supp}\mu_* \subseteq f^{-1}(0)$.

(ii) The first equality holds because $\Lambda^N_{k,m}$ is nonnegative, whereas the second is immediate from the fact that the nonnegative functions μ_0, μ_1, μ_* satisfy $\|\mu_0\|_1 = \|\mu_1\|_1 = 1$ by definition and $\|\mu_*\|_1 = \alpha$ by (5.18).

(iii) Recall that $\mu_* = \mu_1$ on $f^{-1}(0)$. Since $F^{-1}(1) = f^{-1}(0)^N$, the claim follows.

(iv) We have

$$\begin{aligned} \langle F, \Lambda^N_{k,m}\rangle &= \langle F, \Lambda^N_{k+m,0}\rangle && \text{by (iii)} \\ &= \langle 1, \Lambda^N_{k+m,0}\rangle && \text{by (i)} \\ &= \alpha^{k+m} && \text{by (ii).} \end{aligned}$$

(v) Immediate from the fact that $\operatorname{supp}\mu^* \subseteq \operatorname{supp}\mu_1 = \{x \in X : \psi_1(x) > 0\}$ and $\operatorname{supp}\mu_0 = \{x \in X : \psi_1(x) < 0\}$. ∎

LEMMA 5.6. *For any polynomial $P\colon X^N \to \mathbb{R}$, the mapping*

$$m \mapsto \langle \Lambda^N_{k,m}, P\rangle \qquad (m = 0, 1, 2, \ldots, N-k) \tag{5.22}$$

is a univariate polynomial of degree at most $(\deg P)/d_1$.

Proof. For $S \subseteq \{1, 2, \ldots, N\}$ with $|S| = k$, define

$$\Lambda^N_{S,m}(x) = \mathop{\mathbf{E}}_{T}\left[\prod_{i\in T}\mu_1(x_i) \cdot \prod_{i\notin S\cup T}\mu_0(x_i)\right]\prod_{i\in S}\mu_*(x_i),$$

where the expectation is over a uniformly random subset $T \subseteq \{1, 2, \ldots, N\} \setminus S$ of cardinality $|T| = m$. It is clear that $\Lambda^N_{k,m} = \mathbf{E}_{|S|=k}\,\Lambda^N_{S,m}$, and therefore (5.22) is a convex combination of mappings

$$m \mapsto \langle \Lambda^N_{S,m}, P\rangle \qquad (m = 0, 1, 2, \ldots, N-k) \tag{5.23}$$

as S ranges over k-element subsets. As a result, the proof will be complete once we show that (5.23) is a polynomial of degree at most $(\deg P)/d_1$.

By symmetry, we may assume that $S = \{1, 2, \ldots, k\}$. By Lemma 5.3, the function $\phi\colon \{0,1\}^{N-k} \to \mathbb{R}$ given by

$$\phi(z) = \left\langle \mu_*^{\otimes k} \otimes \bigotimes_{i=1}^{N-k} \mu_{z_i}, P \right\rangle$$

has degree at most $(\deg P)/d_1$. Therefore by Proposition 2.1,

$$m \mapsto \mathop{\mathbf{E}}_{\substack{z\in\{0,1\}^{N-k}\\ |z|=m}} \phi(z) \qquad (m = 0, 1, 2, \ldots, N-k) \tag{5.24}$$

is a univariate polynomial of degree at most $(\deg P)/d_1$. It remains to note that the right-hand side of (5.24) is precisely $\langle \Lambda^N_{S,m}, P\rangle$. ∎

We now define a real function $\tilde{\Lambda}^{N,r}_k$ that approximates $\Lambda^N_{k,0}$ pointwise and is additionally orthogonal to low-degree polynomials. The parameter r controls the accuracy of the approximation.

DEFINITION 5.7. *For integers k, r with $0 \leqslant k \leqslant N$ and $0 \leqslant r < k$, define $\tilde{\Lambda}_k^{N,r} \colon X^N \to \mathbb{R}$ by*

(5.25)
$$\tilde{\Lambda}_k^{N,r}(x) = \frac{2^k}{(k-r-1)!} \mathop{\mathbf{E}}_{|S|=k} \left[\prod_{i\in S} \psi_0(x_i) \cdot \prod_{i=1}^{k-r-1} \left(i - \sum_{j\in S} f(x_j) \right) \cdot \prod_{i\notin S} \mu_0(x_i) \right],$$

where the expectation is over a uniformly random $S \subseteq \{1, 2, \ldots, N\}$ of size $|S| = k$.

LEMMA 5.8.

(i) $\langle \tilde{\Lambda}_k^{N,r}, P \rangle = 0$ for every polynomial P of degree less than $(r+1)d_0$,

(ii) $\tilde{\Lambda}_k^{N,r}(x) \neq 0$ only if $|\{i : \psi_1(x_i) > 0\}| = k$,

(iii) $\tilde{\Lambda}_k^{N,r} = \Lambda_{0,k}^N$ on $F^{-1}(1)$,

(iv) $|\tilde{\Lambda}_k^{N,r}| \leqslant \epsilon^{k-r} \binom{k}{r} \Lambda_{0,k}^N$ on $F^{-1}(0)$.

Proof. (i) For $t = 0, 1, 2, \ldots$, it follows from (5.1) that ${\psi_0}^{\otimes t}$ is orthogonal to every polynomial of degree less than td_0. Now Proposition 2.8 implies that the function

$$x \mapsto \prod_{i\in S} \psi_0(x_i) \cdot \prod_{i=1}^{k-r-1} \left(i - \sum_{j\in S} f(x_j) \right) \cdot \prod_{i\notin S} \mu_0(x_i),$$

where $S \subseteq \{1, 2, \ldots, N\}$ is a given subset, is orthogonal to every polynomial of degree less than $(|S| - (k-r-1))d_0$. Since $\tilde{\Lambda}_k^{N,r}$ is a linear combination of such functions with $|S| = k$, the claim follows.

(ii) We have $\operatorname{supp} \psi_0 \subseteq \{x \in X : \psi_1(x) > 0\}$ by (5.3)–(5.5), and $\operatorname{supp} \mu_0 = \{x \in X : \psi_1(x) < 0\}$ by definition. The claim is now immediate from the defining equation, (5.25).

(iii) For every $x \in F^{-1}(1)$, we have $f(x_i) = 0$ and $2\psi_0(x_i) = \mu_1(x_i)$ for every i, where the former holds by definition and the latter by (5.16). Making these substitutions in (5.25),

$$\tilde{\Lambda}_k^{N,r}(x) = \mathop{\mathbf{E}}_{|S|=k} \left[\prod_{i\in S} \mu_1(x_i) \cdot \prod_{i\notin S} \mu_0(x_i) \right] = \Lambda_{0,k}^N(x).$$

(iv) Fix any x with $F(x) = 0$. We claim that for every subset $S \subseteq \{1, 2, \ldots, N\}$ of size $|S| = k$,

(5.26)
$$\frac{2^k}{(k-r-1)!} \prod_{i\in S} |\psi_0(x_i)| \cdot \prod_{i=1}^{k-r-1} \left| i - \sum_{j\in S} f(x_j) \right| \cdot \prod_{i\notin S} \mu_0(x_i) \leqslant \epsilon^{k-r} \binom{k}{r} \prod_{i\in S} \mu_1(x_i) \cdot \prod_{i\notin S} \mu_0(x_i).$$

To see this, consider the nonempty set $A = \{i : f(x_i) = 1\}$. There are three possibilities.

- If $A \not\subseteq S$, then both sides of (5.26) vanish because μ_0 is supported on $f^{-1}(0)$.
- If $A \subseteq S$ and $1 \leqslant |A| \leqslant k-r-1$, then $\prod_{i=1}^{k-r-1} |i - \sum_{j\in S} f(x_j)| = 0$ and the left-hand side of (5.26) vanishes.

- If $A \subseteq S$ and $k-r \leqslant |A| \leqslant k$, then the left-hand side of (5.26) simplifies to

$$\begin{aligned}
\binom{|A|-1}{k-r-1} & \prod_{i\in S}|2\psi_0(x_i)| \cdot \prod_{i\notin S}\mu_0(x_i) \\
& \leqslant \binom{k}{r}\prod_{i\in S}|2\psi_0(x_i)| \cdot \prod_{i\notin S}\mu_0(x_i) \\
& = \binom{k}{r}\prod_{i\in A}|2\psi_0(x_i)| \cdot \prod_{i\in S\setminus A}|2\psi_0(x_i)| \cdot \prod_{i\notin S}\mu_0(x_i) \\
& = \binom{k}{r}\prod_{i\in A}|2\psi_0(x_i)| \cdot \prod_{i\in S\setminus A}\mu_1(x_i) \cdot \prod_{i\notin S}\mu_0(x_i) && \text{by (5.16)} \\
& \leqslant \binom{k}{r}\epsilon^{k-r}\prod_{i\in A}\mu_1(x_i) \cdot \prod_{i\in S\setminus A}\mu_1(x_i) \cdot \prod_{i\notin S}\mu_0(x_i) && \text{by (5.15).}
\end{aligned}$$

This completes the proof of (5.26). One now obtains $|\tilde\Lambda_k^{N,r}(x)| \leqslant \epsilon^{k-r}\binom{k}{r}\Lambda_{0,k}^N(x)$ by passing to expectations on both sides of (5.26) with respect to a uniformly random subset S of cardinality k. ∎

5.4. Simulating symmetric structure. The next step in our construction is a family of real functions $L_1, L_2, \ldots, L_m, \ldots$ with pairwise disjoint support whose role is to mimic the levels of the Boolean hypercube, in the sense that inner product with L_m roughly corresponds to the averaging operation on the mth level of the hypercube. In this way, we are able to simulate symmetric structure in a context with little actual symmetry.

Let $c' = c'(\delta) > 0$ be a sufficiently large *even* integer. Then for each $k = 0, 1, 2, \ldots, n$, Theorem 2.9 gives an explicit function $\omega_k\colon \{0,1,2,\ldots,c'n-k\} \to \mathbb{R}$ such that

$$\|\omega_k\|_1 = 1, \tag{5.27}$$

$$\omega_k(0) > \frac{1}{2} - \frac{\delta}{12}, \tag{5.28}$$

$$|\omega_k(t)| \geqslant \frac{\delta}{24t^2} \qquad (t \geqslant 1), \tag{5.29}$$

$$\operatorname{sgn}\omega_k(t) = \begin{cases} 1 & \text{if } t = 0, \\ (-1)^{k+t} & \text{otherwise,} \end{cases} \tag{5.30}$$

$$\deg p < \sqrt{n} \implies \langle \omega_k, p\rangle = 0. \tag{5.31}$$

By Proposition 2.7 (ii),

$$\|\omega_k\|_\infty \leqslant \frac{1}{2}. \tag{5.32}$$

We will work with the following integer parameters:

$$c'' = \min\left\{c \geqslant 2 : \epsilon^{c-1}2^{cH(1/c)} < \frac{\delta}{20}\right\}, \tag{5.33}$$

$$N = c'c''n, \tag{5.34}$$

where H is the binary entropy function. Observe that $c'' = c''(\epsilon, \delta) > 0$ is a constant.

DEFINITION 5.9. *Define* $L_1, L_2, \ldots, L_{c'n} \colon X^N \to \mathbb{R}$ *by*

$$L_m = \sum_{k=0}^{m-1} \left(\frac{4}{\delta}\right)^k \omega_k(m-k)\Lambda^N_{c''k,c''(m-k)} \qquad (m \leqslant n),$$

$$L_m = \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \omega_k(m-k)\left(\Lambda^N_{c''k,c''(m-k)} - \tilde{\Lambda}^{N,n}_{c''m}\right) \qquad (m \geqslant n+1).$$

The following lemma collects key properties of this function family.

LEMMA 5.10.
(i) $L_m(x) \neq 0$ *only if* $|\{i : \psi_1(x_i) > 0\}| = c''m$,
(ii) $L_m = 0$ *on* $F^{-1}(1)$ *for every* $m \geqslant n+1$,
(iii) $(-1)^m L_m \geqslant 0$,
(iv) $L_m = \sum_{k=0}^{m-1} (4/\delta)^k \omega_k(m-k)\Lambda^N_{c''m,0}$ *on* $F^{-1}(1)$ *for every* $m = 1, 2, \ldots, n$,
(v) $\|L_m\|_1 = \sum_{k=0}^{\min\{m-1,n\}} (4\alpha^{c''}/\delta)^k |\omega_k(m-k)|$.

Proof. (i) Immediate from Lemma 5.5 (v) and Lemma 5.8 (ii).

(ii) On $F^{-1}(1)$, we have the following identity for every k:

$$\Lambda^N_{c''k,c''(m-k)} - \tilde{\Lambda}^{N,n}_{c''m} = \Lambda^N_{0,c''m} - \tilde{\Lambda}^{N,n}_{c''m} = 0,$$

where the first step uses Lemma 5.5 (iii), and the second Lemma 5.8 (iii). The claim is now immediate from the defining equation of L_m for $m \geqslant n+1$.

(iii) For $m = 1, 2, \ldots, n$, the claim follows directly from (5.30) and the nonnegativity of $\Lambda^N_{c''k,c''(m-k)}$. Consider now L_m for $m \geqslant n+1$ and fix an arbitrary point $x \in \operatorname{supp} L_m$. Then $F(x) = 0$ by (ii). As a result,

$$|\tilde{\Lambda}^{N,n}_{c''m}(x)| \leqslant \epsilon^{c''m-n} \binom{c''m}{n} \Lambda^N_{0,c''m}(x) \tag{5.35}$$

by Lemma 5.8 (iv). In light of (5.30), the defining equation of L_m for $m \geqslant n+1$ gives

$$\begin{aligned}(-1)^m L_m(x) &= \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k |\omega_k(m-k)| \left(\Lambda^N_{c''k,c''(m-k)}(x) - \tilde{\Lambda}^{N,n}_{c''m}(x)\right) \\ &= \sum_{k=1}^{n} \left(\frac{4}{\delta}\right)^k |\omega_k(m-k)| \Lambda^N_{c''k,c''(m-k)}(x) \\ &\qquad + \left\{ |\omega_0(m)| \Lambda^N_{0,c''m}(x) - \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k |\omega_k(m-k)| \tilde{\Lambda}^{N,n}_{c''m}(x) \right\}.\end{aligned}$$

Using the estimates (5.29), (5.32), and (5.35), we arrive at

$$\begin{aligned}(-1)^m L_m(x) \geqslant &\sum_{k=1}^{n} \left(\frac{4}{\delta}\right)^k |\omega_k(m-k)| \Lambda^N_{c''k,c''(m-k)}(x) \\ &\qquad + \left\{ \frac{\delta}{24m^2} - \left(\frac{4}{\delta}\right)^n \cdot \epsilon^{c''m-n} \binom{c''m}{n} \right\} \Lambda^N_{0,c''m}(x).\end{aligned}$$

The terms in the summation are nonnegative. Thus, the proof will be complete once we show that the expression in braces is nonnegative as well, which is accomplished by the following calculation:

$$\begin{aligned}
&\left(\frac{4}{\delta}\right)^{n}\cdot\epsilon^{c''m-n}\binom{c''m}{n}\\
&\qquad\leqslant\left(\frac{4}{\delta}\right)^{m-1}\cdot\epsilon^{c''m-m}\binom{c''m}{m} &&\text{since } m\geqslant n+1 \text{ and } c''\geqslant 2\\
&\qquad\leqslant\frac{\delta}{4}\left(\frac{4}{\delta}\cdot\epsilon^{c''-1}2^{c''H(1/c'')}\right)^{m}\\
&\qquad\leqslant\frac{\delta}{4\cdot 5^m} &&\text{by (5.33)}\\
&\qquad\leqslant\frac{\delta}{24m^2} &&\text{since } m\geqslant 2.
\end{aligned}$$

(iv) Immediate from Lemma 5.5 (iii).

(v) For $m=1,2,\ldots,n$,

$$\begin{aligned}
\|L_m\|_1 &= \langle L_m, \operatorname{sgn} L_m\rangle\\
&= (-1)^m\langle L_m, 1\rangle &&\text{by (iii)}\\
&= (-1)^m\sum_{k=0}^{m-1}\left(\frac{4}{\delta}\right)^{k}\omega_k(m-k)\langle\Lambda^{N}_{c''k,c''(m-k)},1\rangle\\
&= (-1)^m\sum_{k=0}^{m-1}\left(\frac{4\alpha^{c''}}{\delta}\right)^{k}\omega_k(m-k) &&\text{by Lemma 5.5 (ii)}\\
&= \sum_{k=0}^{m-1}\left(\frac{4\alpha^{c''}}{\delta}\right)^{k}|\omega_k(m-k)| &&\text{by (5.30).}
\end{aligned}$$

The analysis for $m\geqslant n+1$ is similar but has an additional step:

$$\begin{aligned}
\|L_m\|_1 &= \langle L_m, \operatorname{sgn} L_m\rangle\\
&= (-1)^m\langle L_m, 1\rangle &&\text{by (iii)}\\
&= (-1)^m\sum_{k=0}^{n}\left(\frac{4}{\delta}\right)^{k}\omega_k(m-k)\langle\Lambda^{N}_{c''k,c''(m-k)}-\tilde\Lambda^{N,n}_{c''m},1\rangle\\
&= (-1)^m\sum_{k=0}^{n}\left(\frac{4}{\delta}\right)^{k}\omega_k(m-k)\langle\Lambda^{N}_{c''k,c''(m-k)},1\rangle &&\text{by Lemma 5.8 (i)}\\
&= (-1)^m\sum_{k=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^{k}\omega_k(m-k) &&\text{by Lemma 5.5 (ii)}\\
&= \sum_{k=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^{k}|\omega_k(m-k)| &&\text{by (5.30).} \qquad \square
\end{aligned}$$

5.5. Constructing the dual objects. We are finally in a position to construct the claimed dual pair (Ψ_0, Ψ_1) for F. Let

$$\Psi_0 = \sum_{\substack{m=1,2,\dots,c'n:\\ m \text{ even}}} L_m + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \left(\omega_m(0) - \frac{1}{2}\right) \Lambda^N_{c''m,0}, \tag{5.36}$$

$$\Psi_1 = \sum_{m=1}^{c'n} L_m + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \omega_m(0) \Lambda^N_{c''m,0}. \tag{5.37}$$

The next two lemmas establish useful facts about these functions.

Lemma 5.11. There are $\tilde{\Lambda}_0, \tilde{\Lambda}_1 \in \operatorname{span}\{\tilde{\Lambda}^{N,n}_m : n+1 \leqslant m \leqslant N\}$ such that

$$\Psi_0 = \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \left(\left(\omega_k(0) - \frac{1}{2}\right) \Lambda^N_{c''k,0} + \sum_{\substack{m=1,2,\dots,c'n-k:\\ m\equiv k \pmod 2}} \omega_k(m) \Lambda^N_{c''k,c''m} \right) + \tilde{\Lambda}_0, \tag{5.38}$$

$$\Psi_1 = \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \sum_{m=0}^{c'n-k} \omega_k(m) \Lambda^N_{c''k,c''m} + \tilde{\Lambda}_1. \tag{5.39}$$

Proof. Substituting the defining equation for L_m in (5.36),

$$\begin{aligned}
\Psi_0 &= \sum_{\substack{m=1,2,\dots,c'n:\\ m \text{ even}}} \sum_{k=0}^{\min\{m-1,n\}} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \Lambda^N_{c''k,c''(m-k)} \\
&\qquad\qquad + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \left(\omega_m(0) - \frac{1}{2}\right) \Lambda^N_{c''m,0} + \tilde{\Lambda}_0 \\
&= \sum_{k=0}^{n} \sum_{\substack{m=1,2,\dots,c'n-k:\\ m\equiv k \pmod 2}} \left(\frac{4}{\delta}\right)^k \omega_k(m) \Lambda^N_{c''k,c''m} \\
&\qquad\qquad + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \left(\omega_m(0) - \frac{1}{2}\right) \Lambda^N_{c''m,0} + \tilde{\Lambda}_0,
\end{aligned}$$

where $\tilde{\Lambda}_0$ is as claimed in the lemma statement. Now (5.38) is immediate.

The proof for Ψ_1 is analogous. Substituting the defining equation for L_m in (5.37),

$$\begin{aligned}
\Psi_1 &= \sum_{m=1}^{c'n} \sum_{k=0}^{\min\{m-1,n\}} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \Lambda^N_{c''k,c''(m-k)} \\
&\qquad\qquad + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \omega_m(0) \Lambda^N_{c''m,0} + \tilde{\Lambda}_1 \\
&= \sum_{m=0}^{c'n} \sum_{k=0}^{\min\{m,n\}} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \Lambda^N_{c''k,c''(m-k)} + \tilde{\Lambda}_1,
\end{aligned}$$

where $\tilde{\Lambda}_1$ is as claimed in the lemma statement. The final expression is equivalent to (5.39) by basic algebra. ◻

Lemma 5.12. *On $F^{-1}(1)$, one has*

$$\Psi_0 = \sum_{\substack{m=0,1,\ldots,n:\\ m \text{ even}}} \left(-\frac{1}{2}\left(\frac{4}{\delta}\right)^m + \sum_{k=0}^{m} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \right) \Lambda^N_{c''m,0} + \sum_{\substack{m=0,1,\ldots,n:\\ m \text{ odd}}} \left(\frac{4}{\delta}\right)^m \left(\omega_m(0) - \frac{1}{2}\right) \Lambda^N_{c''m,0}, \tag{5.40}$$

$$\Psi_1 = \sum_{m=0}^{n} \left(\sum_{k=0}^{m} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \right) \Lambda^N_{c''m,0}. \tag{5.41}$$

Proof. For any input x with $F(x) = 1$,

$$\begin{aligned}
\Psi_0(x) &= \sum_{\substack{m=1,2,\ldots,c'n:\\ m \text{ even}}} L_m(x) + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \left(\omega_m(0) - \frac{1}{2}\right) \Lambda^N_{c''m,0}(x) \\
&= \sum_{\substack{m=1,2,\ldots,n:\\ m \text{ even}}} \left(\sum_{k=0}^{m-1} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \right) \Lambda^N_{c''m,0}(x) \\
&\qquad + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \left(\omega_m(0) - \frac{1}{2}\right) \Lambda^N_{c''m,0}(x),
\end{aligned}$$

where the first equality holds by definition, and the second by Lemma 5.10 (ii), (iv). This proves (5.40).

The proof of (5.41) is closely analogous. For $x \in F^{-1}(1)$,

$$\begin{aligned}
\Psi_1(x) &= \sum_{m=1}^{c'n} L_m(x) + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \omega_m(0)\Lambda^N_{c''m,0}(x) \\
&= \sum_{m=1}^{n} \left(\sum_{k=0}^{m-1} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \right) \Lambda^N_{c''m,0}(x) + \sum_{m=0}^{n} \left(\frac{4}{\delta}\right)^m \omega_m(0)\Lambda^N_{c''m,0}(x),
\end{aligned}$$

where the first equality holds by definition, and the second equality is valid by Lemma 5.10 (ii), (iv). ◻

We are now in a position to establish one by one the properties required of Ψ_0, Ψ_1 to be a dual pair for F. The five lemmas that follow, Lemmas 5.13 to 5.17, are independent and can be read in any order.

Lemma 5.13. $\langle F, \Psi_1 \rangle > \frac{1-\delta}{2} \|\Psi_1\|_1$.

Proof. We have

$$
\begin{aligned}
\langle F, \Psi_1\rangle &= \sum_{m=0}^{n}\sum_{k=0}^{m}\left(\frac{4}{\delta}\right)^k \omega_k(m-k)\langle F, \Lambda^N_{c''m,0}\rangle && \text{by (5.41)}\\
&= \sum_{m=0}^{n}\sum_{k=0}^{m}\left(\frac{4}{\delta}\right)^k \omega_k(m-k)\alpha^{c''m} && \text{by Lemma 5.5 (iv)}\\
&= \sum_{k=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^k \sum_{m=0}^{n-k}\alpha^{c''m}\omega_k(m) && \text{by basic algebra}\\
&\geqslant \sum_{k=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^k \Big(\omega_k(0)-\alpha^{c''}\|\omega_k\|_1\Big)\\
&> \sum_{k=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^k \left(\frac{1}{2}-\frac{\delta}{12}-\alpha^{c''}\right) && \text{by (5.27) and (5.28)}\\
&\geqslant \sum_{k=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^k \frac{1-\delta}{2} && \text{by (5.11) and (5.33).}
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\|\Psi_1\|_1 &\leqslant \sum_{m=1}^{c'n}\|L_m\|_1 + \sum_{m=0}^{n}\left(\frac{4}{\delta}\right)^m|\omega_m(0)|\|\Lambda^N_{c''m,0}\|_1 && \text{by (5.37)}\\
&= \sum_{m=1}^{c'n}\sum_{k=0}^{\min\{m-1,n\}}\left(\frac{4\alpha^{c''}}{\delta}\right)^k|\omega_k(m-k)|\\
&\qquad\qquad + \sum_{m=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^m|\omega_m(0)| && \text{by Lemma 5.10(v) and 5.5(ii)}\\
&= \sum_{k=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^k\sum_{m=0}^{c'n-k}|\omega_k(m)| && \text{by basic algebra}\\
&= \sum_{k=0}^{n}\left(\frac{4\alpha^{c''}}{\delta}\right)^k && \text{by (5.27).} \qquad \square
\end{aligned}
$$

LEMMA 5.14. $\Psi_1(x)\geqslant 0$ *whenever* $F(x)=1$.

Proof. By (5.41), it suffices to show that

$$
\left(\frac{4}{\delta}\right)^m\omega_m(0)\geqslant\sum_{k=0}^{m-1}\left(\frac{4}{\delta}\right)^k|\omega_k(m-k)| \qquad (m=0,1,\ldots,n). \tag{5.42}
$$

This relation follows directly from the properties of ω_k. Specifically, by (5.28) the left-hand side of (5.42) is at least $(4/\delta)^m(1-\delta/6)/2\geqslant(4/\delta)^m/3$, whereas by (5.32) the right-hand side of (5.42) is at most $\sum_{k=0}^{m-1}(4/\delta)^k/2\leqslant(4/\delta)^m/6$. $\square$

LEMMA 5.15. *Let* $P_0, P_1\colon X^N\to\mathbb{R}$ *be polynomials with*

$$
\begin{aligned}
&\deg P_0<\min\{nd_0,d_1\},\\
&\deg P_1<\min\{nd_0,\sqrt{n}d_1\}.
\end{aligned}
$$

Then

$$\langle \Psi_0, P_0 \rangle = \langle \Psi_1, P_1 \rangle = 0.$$

Proof. Lemma 5.6 ensures the existence of univariate polynomials $p_0, p_1, \ldots, p_n$ such that

$$\langle \Lambda^N_{c''k,m}, P_1 \rangle = p_k(m) \qquad (k = 0, 1, \ldots, n; \quad m = 0, 1, \ldots, N - c''k), \tag{5.43}$$

$$\deg p_k < \sqrt{n} \qquad (k = 0, 1, \ldots, n). \tag{5.44}$$

Thus,

$$\begin{aligned}
\langle \Psi_1, P_1 \rangle &= \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \sum_{m=0}^{c'n-k} \omega_k(m) \langle \Lambda^N_{c''k,c''m}, P_1 \rangle && \text{by Lemma 5.11 and Lemma 5.8 (i)} \\
&= \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \sum_{m=0}^{c'n-k} \omega_k(m) p_k(c''m) && \text{by (5.43)} \\
&= \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \cdot 0 && \text{by (5.31) and (5.44)} \\
&= 0.
\end{aligned}$$

We now prove the claim for Ψ_0. By (5.27), (5.31), and Proposition 2.7 (i),

$$\sum_{m:\omega_k(m)>0} \omega_k(m) = \frac{1}{2}$$

for every k, which in view of (5.30) is equivalent to

$$\omega_k(0) + \sum_{\substack{m=1,2,\ldots,c'n-k: \\ m \equiv k \pmod 2}} \omega_k(m) = \frac{1}{2}. \tag{5.45}$$

From this point on, the analysis is similar to the one above for Ψ_1. By Lemma 5.6, there are reals $a_0, a_1, \ldots, a_n$ (i.e., zero-degree polynomials) such that

$$\langle \Lambda^N_{c''k,m}, P_0 \rangle = a_k \qquad (k = 0, 1, \ldots, n; \quad m = 0, 1, \ldots, N - c''k). \tag{5.46}$$

By Lemma 5.11 and Lemma 5.8 (i),

$$\begin{aligned}
\langle \Psi_0, P_0\rangle &= \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \Bigg(\left(\omega_k(0) - \frac{1}{2}\right) \langle \Lambda^N_{c''k,0}, P_0\rangle + \\
&\qquad \sum_{\substack{m=1,2,\ldots,c'n-k:\\ m\equiv k \pmod 2}} \omega_k(m) \langle \Lambda^N_{c''k,c''m}, P_0\rangle \Bigg) \\
&= \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \left(\omega_k(0) - \frac{1}{2} + \sum_{\substack{m=1,2,\ldots,c'n-k:\\ m\equiv k \pmod 2}} \omega_k(m) \right) a_k && \text{by (5.46)} \\
&= \sum_{k=0}^{n} \left(\frac{4}{\delta}\right)^k \cdot 0 && \text{by (5.45)} \\
&= 0. && \square
\end{aligned}$$

LEMMA 5.16. $\Psi_0 = \max\{\Psi_1, 0\}$ *on* $F^{-1}(0)$.

Proof. Recall from Lemma 5.5 (i) that for any k, the support of $\Lambda^N_{k,0}$ is contained in $F^{-1}(1)$. As a result, the defining equations (5.36) and (5.37) simplify on $F^{-1}(0)$ to

$$\Psi_0 = \sum_{\substack{m=1,2,\ldots,c'n:\\ m \text{ even}}} L_m, \qquad\qquad \Psi_1 = \sum_{m=1}^{c'n} L_m.$$

This completes the proof since by Lemma 5.10 (i), (iii), the functions in question $L_1, L_2, \ldots, L_m, \ldots$ have pairwise disjoint support, with $\operatorname{sgn} L_m = (-1)^m$ on the support of L_m. $\square$

LEMMA 5.17. $|\Psi_0| \leqslant \delta\Psi_1$ *on* $F^{-1}(1)$.

Proof. Recall from Lemma 5.5 (v) that the functions $\Lambda^N_{c''m,0}$ for $m = 0, 1, 2, \ldots, n$ have pairwise disjoint support. Therefore, the claimed result will follow immediately from Lemma 5.12 once we verify the inequality

$$\max\left\{ \left| -\frac{1}{2}\left(\frac{4}{\delta}\right)^m + \sum_{k=0}^{m} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \right|, \left(\frac{4}{\delta}\right)^m \left| \frac{1}{2} - \omega_m(0) \right| \right\} \leqslant \delta \sum_{k=0}^{m} \left(\frac{4}{\delta}\right)^k \omega_k(m-k) \tag{5.47}$$

for every $m = 0, 1, \ldots, n$. We have

$$\begin{aligned}
\left|\omega_m(0) - \frac{1}{2}\right| &\leqslant \frac{\delta}{12} && \text{by (5.28) and (5.32)}, \\
|\omega_k(m-k)| &\leqslant \frac{1}{2} && \text{by (5.32)}.
\end{aligned}$$

Thus, the left-hand side of (5.47) is at most

$$\left(\frac{4}{\delta}\right)^m \left|\frac{1}{2} - \omega_m(0)\right| + \sum_{k=0}^{m-1} \left(\frac{4}{\delta}\right)^k |\omega_k(m-k)| \leqslant \left(\frac{4}{\delta}\right)^m \left(\frac{\delta}{12} + \frac{\delta}{8-2\delta}\right)$$
$$\leqslant \left(\frac{4}{\delta}\right)^{m-1},$$

whereas the right-hand side of (5.47) is at least

$$\delta\left(\frac{4}{\delta}\right)^m \omega_m(0) - \delta \sum_{k=0}^{m-1} \left(\frac{4}{\delta}\right)^k |\omega_k(m-k)| \geqslant \delta\left(\frac{4}{\delta}\right)^m \left(\frac{1}{2} - \frac{\delta}{12} - \frac{\delta}{8-2\delta}\right)$$
$$\geqslant \left(\frac{4}{\delta}\right)^{m-1}. \qquad \square$$

Lemmas 5.13 to 5.17 establish that (Ψ_0, Ψ_1) is a $(\min\{nd_0, d_1\}, \min\{nd_0, \sqrt{n}d_1\}, \delta)$-dual pair for F. This completes the proof of Theorem 4.3.

5.6. Generalizations. The proof of Theorem 4.3 presented in this section can be generalized in several ways. As a concrete example, define a *generalized* (d_0, d_1, ϵ)-*dual pair* for $f\colon X \to \{0,1\}$ to be any pair of real functions $\psi_0, \psi_1\colon X \to \mathbb{R}$ such that

(i) $\langle f, \psi_1\rangle > \frac{1-\epsilon}{2}\|\psi_1\|_1$,
(ii) $\psi_1(x) \geqslant 0$ whenever $f(x) = 1$,
(iii) $\langle \psi_1, p\rangle = 0$ for every polynomial p of degree less than d_1,
(iv) $\langle \psi_0, p\rangle = 0$ for every polynomial p of degree less than d_0,
(v)
$$\psi_0(x) \in \begin{cases} [\psi_1(x), 2\psi_1(x)] & \text{if } f(x) = 0 \text{ and } \psi_1(x) > 0, \\ [-\epsilon|\psi_1(x)|, \epsilon|\psi_1(x)|] & \text{otherwise.} \end{cases}$$

This definition extends the notion of a (d_0, d_1, ϵ)-dual pair from section 4. Indeed, requirements (i)–(iv) are unchanged but the final requirement (v) is significantly weaker than before. It is not hard to adapt our proof of Theorem 4.3 to this alternate definition of a dual pair, for a small absolute constant $\epsilon > 0$.

6. A complete characterization of the threshold degree. In this section, we study composed functions of the form $\mathrm{OR}_n \circ f$. We fully characterize the threshold degree of any such composition in terms of an approximation-theoretic property of f. Specifically, we show that up to a logarithmic factor, the threshold degree of $\mathrm{OR}_n \circ f$ for $n \geqslant 2$ equals

$$\min_{d_0, d_1} \{nd_0 + d_1\},$$

where the minimum is over all $d_0, d_1 \geqslant 0$ such that f can be approximated in a one-sided manner to within $1/3$ by a rational function with denominator degree d_0 and numerator degree d_1. As a limiting case, we show that the threshold degree of $\mathrm{OR}_n \circ f$ for n large essentially coincides with the one-sided approximate degree of f. The work in this section gives a different proof of Corollary 4.7.

6.1. One-sided rational approximation. Analogous to the one-sided approximation of Boolean functions by *polynomials*, reviewed in section 2, the definition below formalizes one-sided approximation by *rational functions*.

Definition 6.1. *For $d_0 \geqslant 0$ and a Boolean function $f\colon X \to \{0,1\}$, define $\deg_\epsilon^+(f,d_0)$ to be the smallest $d_1 \geqslant 0$ for which there exist polynomials p_0, p_1 of degree at most d_0, d_1, respectively, with*

$$f(x)=0 \qquad\Longrightarrow\qquad \left|\frac{p_1(x)}{p_0(x)}\right| \leqslant \epsilon,$$

$$f(x)=1 \qquad\Longrightarrow\qquad \frac{p_1(x)}{p_0(x)} \geqslant 1-\epsilon.$$

Implicit in this definition is the requirement that $p_0(x) \neq 0$ for every $x \in X$. Since a polynomial can be viewed as a rational function with denominator degree 0, we have

$$\deg_\epsilon^+(f) = \deg_\epsilon^+(f,0).$$

There is a partial equivalence between one-sided and two-sided approximation by rational functions. Specifically, any one-sided rational approximant for f with denominator degree d_0 and numerator degree d_1 gives a two-sided (ℓ_∞-norm) approximant for the same function with a numerator and denominator of degree at most $2d_0+2d_1$. This equivalence has no bearing on our paper because we treat numerator degree and denominator degree as distinct complexity measures—indeed, our interest is precisely in the trade-off between them. Nevertheless, we include a proof of this interesting fact for the sake of completeness.

Proposition 6.2. *For every function $f\colon X \to \{0,1\}$ and every $0<\epsilon<1/2$,*

$$M \leqslant \min_{p,q}\left\{\deg p + \deg q : \left\|f - \frac{p}{q}\right\|_\infty \leqslant \epsilon\right\} \leqslant 4M,$$

where

$$M = \min_{d=0,1,2,\ldots}\{d + \deg_\epsilon^+(f,d)\}.$$

Proof. The lower bound is trivial since one-sided approximation is a weaker requirement than approximation in the ℓ_∞ norm. In the other direction, fix an integer $d \geqslant 0$ and polynomials p_0, p_1 of degree at most d and $\deg_\epsilon^+(f,d)$, respectively, with $|p_1/p_0| \leqslant \epsilon$ on $f^{-1}(0)$ and $p_1/p_0 \geqslant 1-\epsilon$ on $f^{-1}(1)$. Letting

$$\tilde{f} = \frac{p_1^2}{p_1^2 + \epsilon(1-\epsilon)p_0^2},$$

we have $0 \leqslant \tilde{f} \leqslant \epsilon$ on $f^{-1}(0)$ and $1-\epsilon \leqslant \tilde{f} \leqslant 1$ on $f^{-1}(1)$. ∎

Analogous to polynomial approximation, there is a generic way to rapidly reduce the error in a one-sided approximation by rational functions.

Proposition 6.3. *For any function $f\colon X \to \{0,1\}$ and any $k=1,2,3,\ldots,$*

$$\deg^+_{\frac{\epsilon^k}{\epsilon^k+(1-\epsilon)^k}}(f,kd) \leqslant k\deg_\epsilon^+(f,d).$$

Proof. Fix $d \geqslant 0$ and polynomials p_0, p_1 of degree at most d and $\deg_\epsilon^+(f,d)$, respectively, such that $|p_1/p_0| \leqslant \epsilon$ on $f^{-1}(0)$ and $p_1/p_0 \geqslant 1-\epsilon$ on $f^{-1}(1)$. Letting $q_0 = p_0^k$ and $q_1 = p_1^k/(\epsilon^k+(1-\epsilon)^k)$, we obtain

$$\left|\frac{q_1}{q_0}\right| \leqslant \frac{\epsilon^k}{\epsilon^k+(1-\epsilon)^k} \qquad \text{on } f^{-1}(0),$$

$$\frac{q_1}{q_0} \geqslant 1 - \frac{\epsilon^k}{\epsilon^k+(1-\epsilon)^k} \qquad \text{on } f^{-1}(1). \qquad ∎$$

A substantial disadvantage of one-sided approximate degree, in the setting of rational functions, is its lack of a clean and exact dual characterization. We therefore consider a closely related quantity that admits such a characterization.

DEFINITION 6.4. *For $d_0, d_1 \geqslant 0$ and a Boolean function $f\colon X \to \{0,1\}$, define $R(f, d_0, d_1)$ as the infimum over all $\epsilon > 0$ for which there exist polynomials p_0, p_1 of degree at most d_0, d_1, respectively, such that*

$$f(x) = 0 \qquad \Longrightarrow \qquad |p_1(x)| < \epsilon p_0(x), \tag{6.1}$$

$$f(x) = 1 \qquad \Longrightarrow \qquad |p_0(x)| < \epsilon p_1(x). \tag{6.2}$$

It is clear that $R(f, d_0, d_1)$ is always well-defined and ranges in $[0, 1]$. We now have two notions of error for the one-sided rational approximation of Boolean functions: one-sided approximate degree and the new quantity $R(f, d_0, d_1)$. Fortunately, the two notions are equivalent, with $\deg^+_\epsilon(f, d_0) > d_1$ roughly corresponding to $R(f, d_0, d_1) \geqslant \sqrt{\epsilon/(1-\epsilon)}$. The proposition below makes this correspondence formal.

PROPOSITION 6.5. *For $d_0, d_1 \geqslant 0$ and every Boolean function $f\colon X \to \{0,1\}$,*

$$\deg^+_\epsilon(f, d_0) > d_1 \qquad \Longrightarrow R\left(f, \frac{d_0}{2}, \frac{d_1}{2}\right) \geqslant \sqrt[4]{\frac{\epsilon}{1-\epsilon}}\ , \tag{6.3}$$

$$\deg^+_\epsilon(f, d_0) \leqslant d_1 \qquad \Longrightarrow R\left(f, 2d_0, 2d_1\right) \leqslant \frac{\epsilon}{1-\epsilon}. \tag{6.4}$$

Proof. Assume that $\deg^+_\epsilon(f, d_0) > d_1$ and fix $\delta > R(f, d_0/2, d_1/2)$ arbitrarily. Then by definition, there are polynomials p_0, p_1 of degree at most $d_0/2$ and $d_1/2$, respectively, such that $|p_1| < \delta p_0$ on $f^{-1}(0)$ and $|p_0| < \delta p_1$ on $f^{-1}(1)$. In particular, the infimum

$$\inf_{\zeta > 0} \left\{ \frac{\delta^2}{1+\delta^4} \cdot \frac{p_1^2(x)}{p_0^2(x) + \zeta} \right\}$$

has absolute value less than $\delta^4/(1+\delta^4)$ on $f^{-1}(0)$ and exceeds $1/(1+\delta^4)$ on $f^{-1}(1)$. We obtain

$$\deg^+_{\frac{\delta^4}{1+\delta^4}}(f, d_0) \leqslant d_1,$$

whence

$$\delta > \sqrt[4]{\frac{\epsilon}{1-\epsilon}}$$

by the premise of (6.3). Since $\delta > R(f, d_0/2, d_1/2)$ was chosen arbitrarily, (6.3) follows.

In the other direction, assume that $\deg^+_\epsilon(f, d_0) \leqslant d_1$. Then for every $\delta > \epsilon$, there are polynomials p_0, p_1 of degree at most d_0, d_1, respectively, such that $|p_1/p_0| < \delta$ on $f^{-1}(0)$ and $p_1/p_0 > 1-\delta$ on $f^{-1}(1)$. Letting $q_0 = p_0^2$ and $q_1 = p_1^2/(\delta - \delta^2)$, we obtain $|q_1| < q_0\delta/(1-\delta)$ on $f^{-1}(0)$ and $|q_0| < q_1\delta/(1-\delta)$ on $f^{-1}(1)$. Put another way,

$$R(f, 2d_0, 2d_1) \leqslant \frac{\delta}{1-\delta}.$$

Since the choice of $\delta > \epsilon$ was arbitrary, (6.4) follows. ∎

6.2. Passing to the dual program. One-sided rational approximation, as formalized by the quantity $R(f, d_0, d_1)$, admits the following intuitive dual characterization.

THEOREM 6.6. *Let $f\colon X \to \{0,1\}$ be a given Boolean function, $d_0, d_1 \geqslant 0$. Then for every $\epsilon > 0$, the nonexistence of polynomials p_0, p_1 such that*

(i) $|p_1| < \epsilon p_0$ *on* $f^{-1}(0)$,
(ii) $|p_0| < \epsilon p_1$ *on* $f^{-1}(1)$,
(iii) $\deg p_0 \leqslant d_0$,
(iv) $\deg p_1 \leqslant d_1$,

is equivalent to the existence of $\psi_0, \psi_1\colon X \to \mathbb{R}$ such that

(v) $\psi_0 \geqslant \epsilon|\psi_1|$ *on* $f^{-1}(0)$,
(vi) $\psi_1 \geqslant \epsilon|\psi_0|$ *on* $f^{-1}(1)$,
(vii) $\deg p \leqslant d_0 \implies \langle \psi_0, p\rangle = 0$,
(viii) $\deg p \leqslant d_1 \implies \langle \psi_1, p\rangle = 0$,
(ix) $\psi_0 \not\equiv 0$,
(x) $\psi_1 \not\equiv 0$.

Proof. Let P_0 and P_1 denote the linear subspaces of real polynomials on X of degree at most d_0 and d_1, respectively. Conditions (i) and (ii) can be rewritten as

$$\epsilon^{1-f} p_0 + \epsilon^f p_1 > 0,$$
$$(-\epsilon)^{1-f} p_0 + (-\epsilon)^f p_1 < 0$$

on X. By linear programming duality, this system of inequalities in $p_0 \in P_0, p_1 \in P_1$ is infeasible if and only if there exist nonnegative functions μ, λ on X, not both identically zero, such that

$$\epsilon^{1-f}\mu - (-\epsilon)^{1-f}\lambda \in P_0^{\perp}, \tag{6.5}$$
$$\epsilon^{f}\mu - (-\epsilon)^{f}\lambda \in P_1^{\perp}. \tag{6.6}$$

The existence of such μ and λ is in turn equivalent to the existence of $\psi_0, \psi_1\colon X \to \mathbb{R}$, not both identically zero, that obey (v)–(viii), where we identify ψ_0 and ψ_1 with the left-hand sides of (6.5) and (6.6), respectively.

Finally, the requirement that at least one of ψ_0, ψ_1 be not identically zero is logically equivalent to the requirement that $\psi_0 \not\equiv 0$ and $\psi_1 \not\equiv 0$ simultaneously. Indeed, if *exactly* one of ψ_0, ψ_1 were identically zero, then by (v)–(vi) the other would have to be a nonnegative function, contradicting $\langle \psi_0, 1\rangle = \langle \psi_1, 1\rangle = 0$. ◻

COROLLARY 6.7. *Let $f\colon X \to \{0,1\}$ be a given function, $R(f, d_0, d_1) > 0$. Then $R(f, d_0, d_1)$ is the supremum over all $\epsilon > 0$ for which there exist $\psi_0, \psi_1\colon X \to \mathbb{R}$ with*

(i) $\psi_0 \geqslant \epsilon|\psi_1|$ *on* $f^{-1}(0)$,
(ii) $\psi_1 \geqslant \epsilon|\psi_0|$ *on* $f^{-1}(1)$,
(iii) $\deg p \leqslant d_0 \implies \langle \psi_0, p\rangle = 0$,
(iv) $\deg p \leqslant d_1 \implies \langle \psi_1, p\rangle = 0$,
(v) $\psi_0 \not\equiv 0$,
(vi) $\psi_1 \not\equiv 0$.

6.3. Lower bound on the threshold degree. We are now in a position to prove a lower bound on the threshold degree of any composition $\mathrm{OR}_n \circ f$. The following first-principles construction plays an important role in the proof.

LEMMA 6.8. *For integers n, d with $n \geqslant 1$ and $0 \leqslant d \leqslant n$, let*

$$p_{n,d}(t) = \prod_{i=n-d+1}^{n} \frac{i-t}{i}.$$

Then

$$p_{n,d}(0) = 1,$$
$$|p_{n,d}(t)| \leqslant \left(1 - \frac{d}{n}\right)^t, \qquad t = 1, 2, \ldots, n.$$

Proof. The cases $t = 0$ and $t > n - d$ are straightforward, with $p_{n,d}$ evaluating to 1 in the former case and vanishing in the latter. For $t = 1, 2, \ldots, n - d$, we have the closed form

$$p_{n,d}(t) = \binom{n-t}{d}\binom{n}{d}^{-1},$$

whence

$$|p_{n,d}(t)| = \frac{n-d}{n} \cdot \frac{n-d-1}{n-1} \cdot \cdots \cdot \frac{n-d-t+1}{n-t+1} \leqslant \left(\frac{n-d}{n}\right)^t. \qquad \square$$

We have reached the main technical result of this section.

THEOREM 6.9. *Let $d_0, d_1 \geqslant 0$ be integers, $f\colon X \to \{0, 1\}$ a given Boolean function. If $R(f, d_0, d_1) > \epsilon$, then*

$$\deg_{\pm}(\mathrm{OR}_n \circ f) \geqslant \min\{\lfloor \epsilon^2 n \rfloor (d_0 + 1), d_1 + 1\}, \qquad n = 1, 2, 3, \ldots.$$

Proof. Abbreviate $F = \mathrm{OR}_n \circ f$. We need only consider the case $\epsilon > 0$, the theorem being trivial otherwise. Since $R(f, d_0, d_1) > \delta$ for sufficiently small $\delta > \epsilon$, Corollary 6.7 guarantees the existence of $\psi_0, \psi_1\colon X \to \mathbb{R}$ such that

$$f(x) = 0 \implies \psi_0(x) \geqslant \delta|\psi_1(x)|, \tag{6.7}$$
$$f(x) = 1 \implies \psi_1(x) \geqslant \delta|\psi_0(x)|, \tag{6.8}$$
$$\deg p < d_0 + 1 \implies \langle \psi_0, p \rangle = 0, \tag{6.9}$$
$$\deg p < d_1 + 1 \implies \langle \psi_1, p \rangle = 0, \tag{6.10}$$
$$\psi_0 \not\equiv 0. \tag{6.11}$$

For integers n, d, let $p_{n,d}$ denote the degree-d polynomial constructed in Lemma 6.8. Define $A, B\colon X^n \to \mathbb{R}$ by

$$A(x) = p_{n,n-\lfloor \epsilon^2 n \rfloor}\left(\sum_{i=1}^{n} f(x_i)\right) \prod_{i=1}^{n} \psi_0(x_i),$$
$$B(x) = \prod_{i:f(x_i)=0} |\psi_0(x_i)| \cdot \prod_{i:f(x_i)=1} \delta\psi_1(x_i)$$
$$- \prod_{i=1}^{n}(1 - f(x_i)) \cdot \prod_{i=1}^{n}(|\psi_0(x_i)| - \delta\psi_1(x_i)).$$

We have

$$F(x) = 0 \implies A(x) = \prod_{i=1}^{n} |\psi_0(x_i)|, \tag{6.12}$$

$$F(x) = 1 \implies |A(x)| \leqslant \epsilon^{2\sum f(x_i)} \prod_{i=1}^{n} |\psi_0(x_i)|, \tag{6.13}$$

where the first item follows from Lemma 6.8 and the nonnegativity of ψ_0 on $f^{-1}(0)$, and the second is immediate from Lemma 6.8. Continuing, (6.7) and (6.8) imply

$$F(x) = 0 \implies B(x) \leqslant \prod_{i=1}^{n} |\psi_0(x_i)|, \tag{6.14}$$

$$F(x) = 1 \implies B(x) \geqslant \delta^{2\sum f(x_i)} \prod_{i=1}^{n} |\psi_0(x_i)|, \tag{6.15}$$

respectively. Finally, we claim that

$$\deg P < \lfloor \epsilon^2 n \rfloor (d_0 + 1) \implies \langle A, P \rangle = 0, \tag{6.16}$$

$$\deg P < d_1 + 1 \implies \langle B, P \rangle = 0. \tag{6.17}$$

The first claim follows directly from (6.9) and Proposition 2.8, whereas the second follows from (6.10) once one rewrites

$$\begin{aligned} B(x) &= \prod_{i=1}^{n} \{\delta\psi_1(x_i) + (1 - f(x_i))(|\psi_0(x_i)| - \delta\psi_1(x_i))\} \\ &\qquad - \prod_{i=1}^{n} (1 - f(x_i))(|\psi_0(x_i)| - \delta\psi_1(x_i)) \\ &= \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ S \neq \varnothing}} \prod_{i \in S} \delta\psi_1(x_i) \cdot \prod_{i \notin S} (1 - f(x_i))(|\psi_0(x_i)| - \delta\psi_1(x_i)). \end{aligned}$$

By (6.12)–(6.15), the function $\Psi = \frac{1}{\delta} B - \frac{1}{\epsilon} A$ satisfies

$$(-1)^{1-F(x)} \Psi(x) \geqslant (\delta - \epsilon)^{2n} \prod_{i=1}^{n} |\psi_0(x_i)|.$$

Recalling (6.11), we obtain $(-1)^{1-F}\Psi \geqslant 0$ and $\Psi \not\equiv 0$. Moreover, equations (6.16) and (6.17) ensure that Ψ is orthogonal to every polynomial of degree less than $\min\{\lfloor \epsilon^2 n \rfloor (d_0 + 1), d_1 + 1\}$. By the dual characterization of threshold degree (Theorem 2.4), the proof is complete. ∎

We now reword the previous theorem in terms of one-sided approximate degree.

Corollary 6.10. *Let $f \colon X \to \{0, 1\}$ be given. Then for all $\epsilon \geqslant 0$ and all integers $n \geqslant 1$ and $d \geqslant 0$,*

$$\deg_{\pm}(\mathrm{OR}_n \circ f) \geqslant \min \left\{ \left\lfloor n \sqrt{\epsilon(1+\epsilon)} \right\rfloor \left\lceil \frac{d+1}{2} \right\rceil, \left\lceil \frac{\deg^{+}_{\epsilon}(f, d)}{2} \right\rceil \right\}. \tag{6.18}$$

Proof. If $\epsilon = 0$ or $\deg_\epsilon^+(f,d) = 0$, then the right-hand side of (6.18) vanishes, and the claim is trivially true. As a result, we may assume that $\epsilon > 0$ and $\deg_\epsilon^+(f,d) \geqslant 1$. Consider the nonnegative integers $d_0 = \lfloor \frac{1}{2}d \rfloor$ and $d_1 = \lfloor \frac{1}{2}\deg_\epsilon^+(f,d) - \frac{1}{2} \rfloor$. Then $\deg_\epsilon^+(f, 2d_0) > 2d_1$ by definition, whence

$$R(f, d_0, d_1) \geqslant \sqrt[4]{\frac{\epsilon}{1-\epsilon}} > \sqrt[4]{\epsilon(1+\epsilon)}$$

by Proposition 6.5. Therefore, Theorem 6.9 implies that

$$\begin{aligned}\deg_\pm(\mathrm{OR}_n \circ f) &\geqslant \min\left\{ \left\lfloor n\sqrt{\epsilon(1+\epsilon)} \right\rfloor (d_0+1), d_1+1 \right\} \\ &= \min\left\{ \left\lfloor n\sqrt{\epsilon(1+\epsilon)} \right\rfloor \left\lceil \frac{d+1}{2} \right\rceil, \left\lceil \frac{\deg_\epsilon^+(f,d)}{2} \right\rceil \right\}. \qquad \square\end{aligned}$$

As a special case, we recover Corollary 4.7 with an entirely new proof:

COROLLARY 6.11. *Let $f\colon X \to \{0,1\}$ be given. Then*

$$\deg_\pm(\mathrm{OR}_n \circ f) \geqslant \frac{1}{2}\min\{n, \deg_{1/3}^+(f)\}.$$

In particular,

$$\deg_\pm(\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}) = \Omega(n^{2/5}).$$

Proof. The first assertion is trivial for $n = 1$, whereas for $n \geqslant 2$ it follows by taking $d = 0$ and $\epsilon = 1/3$ in Corollary 6.10. The second assertion follows from the first by Theorem 2.3. $\square$

6.4. Upper bound on the threshold degree. We now recall a matching upper bound on the threshold degree of any composition $\mathrm{OR}_n \circ f$. This result was already implicit in the original paper of Beigel et al. [8], with various related statements obtained in subsequent work [24, 46, 48].

THEOREM 6.12 (cf. Beigel et al.). *Let $f\colon X \to \{0,1\}$ be given. Then for all integers $n \geqslant 1$,*

$$\deg_\pm(\mathrm{OR}_n \circ f) \leqslant \min_{0 \leqslant \epsilon < \frac{1}{2n}} \min_{d=0,1,2,\ldots} \left\{2nd + \deg_\epsilon^+(f,d)\right\} \tag{6.19}$$

$$\leqslant \lceil \log 2n \rceil \min_{d=0,1,2,\ldots} \{2nd + \deg_{1/3}^+(f,d)\}. \tag{6.20}$$

Proof (cf. [8, 24]). Abbreviate $F = \mathrm{OR}_n \circ f$, and fix an integer $d \geqslant 0$ and a real number $0 \leqslant \epsilon < \frac{1}{2n}$. By definition, there are polynomials p_0, p_1 of degree at most d and $\deg_\epsilon^+(f,d)$, respectively, such that

$$\begin{aligned}&\left|\frac{p_1}{p_0}\right| < \frac{1}{2n} && \text{on } f^{-1}(0), \\ &\frac{p_1}{p_0} > 1 - \frac{1}{2n} && \text{on } f^{-1}(1).\end{aligned}$$

Then

$$\mathrm{sgn}\left(\sum_{i=1}^n \frac{p_1(x_i)}{p_0(x_i)} - \frac{1}{2}\right) = \begin{cases} -1 & \text{if } F(x_1, x_2, \ldots, x_n) = 0, \\ 1 & \text{otherwise.}\end{cases}$$

Multiplying the expression in parentheses by the positive quantity $\prod p_0(x_i)^2$ gives a sign-representing polynomial for F of degree at most $2nd + \deg^+_\epsilon(f,d)$, namely,

$$\sum_{i=1}^n p_0(x_i)p_1(x_i)\prod_{\substack{j=1\\ j\neq i}}^n p_0(x_j)^2 - \frac{1}{2}\prod_{j=1}^n p_0(x_j)^2.$$

This completes the proof of (6.19). Now (6.20) can be verified as follows:

$$\begin{aligned}\deg_\pm(\mathrm{OR}_n\circ f) &\leqslant \min_{d=0,1,2,\dots}\left\{2n\cdot\lceil\log 2n\rceil d + \deg^+_{\frac{1}{2n+1}}(f,\lceil\log 2n\rceil d)\right\}\\ &\leqslant \lceil\log 2n\rceil \min_{d=0,1,2,\dots}\{2nd+\deg^+_{1/3}(f,d)\},\end{aligned}$$

where the first inequality follows by taking $\epsilon = \frac{1}{2n+1}$ in (6.19), and the second follows by taking $\epsilon=\frac{1}{3}$ and $k=\lceil\log 2n\rceil$ in Proposition 6.3. ☐

6.5. The final characterization. It remains to show that our lower and upper bounds on the threshold degree of $\mathrm{OR}_n\circ f$ essentially coincide. We start with a technical observation.

PROPOSITION 6.13. *Let $G\colon\mathbb{N}\to\mathbb{R}$ and $g\colon\mathbb{N}\to\mathbb{R}$ be given functions, where*
(i) *G is nondecreasing and unbounded,*
(ii) *g is nonincreasing,*
(iii) *$G(0)\leqslant g(0)$.*
Then

$$\max_{i=0,1,2,\dots}\min\{G(i+1),g(i)\}\geqslant\frac{1}{2}\min_{i=0,1,2\dots}\{G(i)+g(i)\}.$$

Proof. We will prove the claimed result under much weaker assumptions on G and g. Specifically, the only consequence of (i)–(iii) that we will use is the existence of $i^*\geqslant 0$ such that $G(i^*)\leqslant g(i^*)$ and $G(i^*+1)\geqslant g(i^*+1)$. We have:

$$\begin{aligned}2\max_{i\geqslant 0}\min\{G(i+1),g(i)\} &\geqslant \min\{2G(i^*+1),2g(i^*)\}\\ &\geqslant \min\{2G(i^*+1),G(i^*)+g(i^*)\}\\ &\geqslant \min\{G(i^*+1)+g(i^*+1),G(i^*)+g(i^*)\}\\ &\geqslant \min_{i\geqslant 0}\{G(i)+g(i)\}.\end{aligned}$$

☐

The desired characterization of the threshold degree of $\mathrm{OR}_n\circ f$ is as follows.

THEOREM 6.14. *For every function $f\colon X\to\{0,1\}$ and every $n\geqslant 2$,*

$$\frac{D}{8}\leqslant\deg_\pm(\mathrm{OR}_n\circ f)\leqslant D\cdot 2\lceil\log 2n\rceil,$$

where

$$D=\min_{d=0,1,2,\dots}\left\{nd+\deg^+_{1/3}(f,d)\right\}.$$

Proof. The upper bound on the threshold degree follows directly from Theorem 6.12. The lower bound can be verified as follows:

$$\begin{aligned}\deg_\pm(\mathrm{OR}_n\circ f) &\geqslant \max_{d\geqslant 0}\min\left\{\frac{n(d+1)}{4},\frac{\deg^+_{1/3}(f,d)}{2}\right\}\\ &\geqslant \min_{d\geqslant 0}\left\{\frac{nd}{8}+\frac{\deg^+_{1/3}(f,d)}{4}\right\},\end{aligned}$$

where the first inequality holds by taking $\epsilon = 1/3$ in Corollary 6.10, and the second follows by Proposition 6.13. ∎

Prior to our work, the characterization in Theorem 6.14 was only known for $n = 2$, with the upper and lower bounds for that case obtained by Beigel et al. [8] and Sherstov [46], respectively. Specifically, those authors showed that up to a small multiplicative constant, the threshold degree of $\mathrm{OR}_2 \circ f$ equals the smallest degree of a rational function that approximates f pointwise within $1/3$:

$$\deg_\pm(\mathrm{OR}_2 \circ f) = \Theta\left(\min_{p,q}\left\{\deg p + \deg q : \left\|f - \frac{p}{q}\right\|_\infty \leqslant \frac{1}{3}\right\}\right).$$

By Proposition 6.2, this characterization is equivalent to Theorem 6.14 for $n = 2$.

It is instructive to examine the behavior of the threshold degree as $n \to \infty$:

Theorem 6.15. *Let $f\colon X \to \{0,1\}$ be given. Then for all n large enough,*

$$\deg_\pm(\mathrm{OR}_n \circ f) \leqslant \deg^+_{1/3}(f) \cdot \lceil \log 2n \rceil,$$

$$\deg_\pm(\mathrm{OR}_n \circ f) \geqslant \frac{\deg^+_{1/3}(f)}{2}.$$

Proof. The upper bound holds for all n by taking $d = 0$ in Theorem 6.12. Taking $\epsilon = 1/3$ and $d = 0$ in Corollary 6.10 shows that the lower bound holds for n large enough. ∎

In other words, for n sufficiently large the threshold degree of $\mathrm{OR}_n \circ f$ essentially equals the one-sided *polynomial* approximate degree of f. This conclusion is intuitively satisfying in light of the construction of Theorem 6.12, in which rational approximants with nonconstant denominators become inefficient for large n.

7. A simpler proof for depth 2. In Corollary 4.7, we proved that

$$\deg_\pm(\mathrm{OR}_n \circ f) = c \min\{n, \deg^+_{1/3}(f)\}$$

for some absolute constant $c > 0$ and every function $f\colon X \to \{0,1\}$, with the following important special case:

$$\deg_\pm(\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}) = \Omega(n^{2/5}).$$

We gave an alternate proof of these results in the previous section, using our characterization of the threshold degree for compositions $\mathrm{OR}_n \circ f$. We will now present a third and simpler yet proof, which combines the techniques of this paper with a construction due to Bun and Thaler [13]. Unfortunately, this proof does not generalize to compositions of greater depth and does not allow us to recover the general result of Theorem 4.5 nor the main result of this paper, Theorem 1.1.

Theorem 7.1. *Let $f\colon X \to \{0,1\}$ be given. Suppose that there exist $\psi_0, \psi_1\colon X \to \mathbb{R}$ such that*

(i) $\psi_1 \geqslant |\psi_0|$ *on* $f^{-1}(1)$,
(ii) $\psi_0 = \max\{\psi_1, 0\}$ *on* $f^{-1}(0)$,
(iii) $\deg p < d_0 \implies \langle \psi_0, p\rangle = 0$,
(iv) $\deg p < d_1 \implies \langle \psi_1, p\rangle = 0$,
(v) $\psi_1 \not\equiv 0$.

Then

$$\deg_\pm(\mathrm{OR}_n \circ f) \geqslant \min\{nd_0, d_1\} \qquad (n = 1, 2, 3, \dots).$$

Proof. We may assume that $d_1 > 0$, the claimed lower bound being trivial otherwise. Write $\psi_1 = \mu_+ - \mu_-$, where $\mu_+ = \max\{\psi_1, 0\}$ and $\mu_- = \max\{-\psi_1, 0\}$ are the positive and negative parts of ψ_1, respectively. Observe that $\mu_+ \not\equiv 0$ and $\mu_- \not\equiv 0$ by (iv), (v). As usual, we need to decide what dual object to use for the function in question, $F = \mathrm{OR}_n \circ f$. Bun and Thaler [13] used $\mu_+^{\otimes n} - \mu_-^{\otimes n}$ for this purpose, an elegant choice that works well in the setting of pointwise approximation. Since our interest is in sign-representation instead, we must additionally ensure agreement in sign with F. To this end, we define our dual object to be

$$\Psi = \mu_+^{\otimes n} - \mu_-^{\otimes n} - {\psi_0}^{\otimes n}.$$

By (i) and (ii),

$$|\psi_0| \leqslant \mu_+ \qquad \text{on } f^{-1}(1), \tag{7.1}$$

$$\psi_0 = \mu_+ \qquad \text{on } f^{-1}(0), \tag{7.2}$$

$$\operatorname{supp} \mu_- \subseteq f^{-1}(0). \tag{7.3}$$

On $F^{-1}(1)$,

$$\begin{aligned}
\Psi(x) &= \prod_{i=1}^n \mu_+(x_i) - \prod_{i=1}^n \mu_-(x_i) - \prod_{i=1}^n \psi_0(x_i) && \text{by definition} \\
&= \prod_{i=1}^n \mu_+(x_i) - \prod_{i=1}^n \psi_0(x_i) && \text{by (7.3)} \\
&\geqslant 0 && \text{by (7.1) and (7.2).}
\end{aligned}$$

On $F^{-1}(0)$,

$$\begin{aligned}
\Psi(x) &= \prod_{i=1}^n \mu_+(x_i) - \prod_{i=1}^n \mu_-(x_i) - \prod_{i=1}^n \psi_0(x_i) && \text{by definition} \\
&= -\prod_{i=1}^n \mu_-(x_i) && \text{by (7.2).}
\end{aligned}$$

Since $\mu_- \not\equiv 0$, the last equation additionally shows that $\Psi \not\equiv 0$.

Summarizing, we have shown that $(-1)^{1-F}\Psi \geqslant 0$ and $\Psi \not\equiv 0$. In light of Theorem 2.4, the claimed lower bound on the threshold degree of F will follow once we show that Ψ is orthogonal to every polynomial P of degree less than $\min\{nd_0, d_1\}$. By linearity, it suffices to consider factored polynomials

$$P(x_1, x_2, \dots, x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n).$$

Use a telescoping sum to write

$$\mu_+^{\otimes n} - \mu_-^{\otimes n} = \sum_{j=1}^n \underbrace{\mu_+ \otimes \cdots \otimes \mu_+}_{j-1} \otimes (\mu_+ - \mu_-) \otimes \underbrace{\mu_- \otimes \cdots \otimes \mu_-}_{n-j}.$$

Then

$$\langle \Psi, P\rangle = \sum_{j=1}^{n} \langle \mu_+, p_1\rangle \cdots \langle \mu_+, p_{j-1}\rangle \underbrace{\langle \psi_1, p_j\rangle}_{=0} \langle \mu_-, p_{j+1}\rangle \cdots \langle \mu_-, p_n\rangle$$
$$- \underbrace{\langle {\psi_0}^{\otimes n}, P\rangle}_{=0},$$

where the marked inner products are zero by (iii) and (iv). ∎

COROLLARY 7.2. *For every* $f\colon X \to \{0,1\}$,

$$\deg_{\pm}(\mathrm{OR}_n \circ f) \geqslant \min\{n, \deg^{+}_{1/4}(f)\}.$$

In particular,

$$\deg_{\pm}(\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}) = \Omega(n^{2/5}).$$

Proof. We may assume that $\deg^{+}_{1/4}(f) > 0$, the claim being trivial otherwise. Then Lemma 4.2 guarantees that f has a $(1, \deg^{+}_{1/4}(f), 1)$-dual pair, which means in particular that the hypothesis of Theorem 7.1 holds with $d_0 = 1$ and $d_1 = \deg^{+}_{1/4}(f)$. This proves the first claim. The second claim follows from the first in view of Theorem 2.3. ∎

8. Additional applications. In this concluding section, we examine additional applications of our main result and in particular prove Theorems 1.4 and 1.5 from the Introduction. We assume basic familiarity with communication complexity theory and computational learning. For a concise introduction to these research areas, we refer the reader to the monographs by Kushilevitz and Nisan [31] and Kearns and Vazirani [23].

8.1. Communication complexity. Let $f\colon X \times Y \to \{0,1\}$ be a given two-party communication problem. The *ϵ-error randomized communication complexity* of f, denoted $R_\epsilon(f)$, is the minimum cost of a communication protocol with public randomness that computes f with error at most ϵ on every input. For a probability distribution μ on $X \times Y$, the *discrepancy of f with respect to μ* is given by

$$\mathrm{disc}_\mu(f) = \max_{\substack{X' \subseteq X \\ Y' \subseteq Y}} \left| \sum_{x \in X'} \sum_{y \in Y'} (-1)^{f(x,y)} \mu(x,y) \right|.$$

The minimum discrepancy of f over all probability distributions is denoted

$$\mathrm{disc}(f) = \min_{\mu} \mathrm{disc}_\mu(f).$$

Discrepancy plays a central role in communication complexity theory because it implies communication lower bounds in almost every model, with low discrepancy corresponding to high communication complexity. In particular, the randomized communication complexity of every function f obeys

$$R_\epsilon(f) \geqslant \log \frac{1-2\epsilon}{\mathrm{disc}(f)}, \tag{8.1}$$

a fundamental inequality known as the *discrepancy method* [31, section 3.5].

Discrepancy is difficult to analyze, except in a handful of canonical cases. A useful technique in this context is the *pattern matrix method* [42, 43], which among other things translates lower bounds on approximate degree into upper bounds on discrepancy. We will use the following version [43, Theorem 7.3] of the pattern matrix method.

THEOREM 8.1 (Sherstov). *Let* $f\colon \{0,1\}^n \to \{0,1\}$ *be a given Boolean function. Define* $F\colon \{0,1\}^{4n} \times \{0,1\}^{4n} \to \{0,1\}$ *by*

$$F(x,y) = f\left(\dots, \bigvee_{j=1}^{4} (x_{i,j} \wedge y_{i,j}), \dots\right).$$

Then

$$\operatorname{disc}(F) \leqslant 2^{-\deg_{\pm}(f)/2}.$$

Combining this theorem with our main result, we obtain:

THEOREM 8.2. *Fix an arbitrary constant* $k \geqslant 1$ *and define* $f_n\colon \{0,1\}^n \to \{0,1\}$ *by*

$$f_n = \mathrm{NOR}_{n^{\frac{1}{2k-1}}} \circ \underbrace{\mathrm{NOR}_{n^{\frac{2}{2k-1}}} \circ \cdots \circ \mathrm{NOR}_{n^{\frac{2}{2k-1}}}}_{k-1}.$$

Consider the two-party communication problem $F_n\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *given by* $F_n(x,y) = f_n(x \wedge y)$. *Then for some constant* $c = c(k) > 0$ *and all* n,

$$\begin{aligned} &\operatorname{disc}(F_n) \leqslant \exp\left(-cn^{\frac{k-1}{2k-1}}\right), \\ &R_{\frac{1}{2}-\exp\left(-cn^{\frac{k-1}{2k-1}}\right)}(F_n) \geqslant cn^{\frac{k-1}{2k-1}}. \end{aligned}$$

Proof. By the discrepancy method (8.1), it suffices to prove the discrepancy upper bound. The identity $\mathrm{NOR}_s \circ \mathrm{OR}_t = \mathrm{NOR}_{st}$ implies that $f_n \circ \mathrm{OR}_4 \circ \mathrm{AND}_2$ is a subfunction of $F_{4^{2k-1}n}$. Therefore,

$$\begin{aligned} \operatorname{disc}(F_{4^{2k-1}n}) &\leqslant \operatorname{disc}(f_n \circ \mathrm{OR}_4 \circ \mathrm{AND}_2) \\ &\leqslant 2^{-\deg_{\pm}(f_n)/2} \\ &\leqslant \exp(-\Omega(n^{\frac{k-1}{2k-1}})), \end{aligned}$$

where the last two inequalities use Theorem 8.1 and Theorem 1.1, respectively. ◻

This settles Theorem 1.4 from the Introduction. For any $d \geqslant 3$, Theorem 8.2 gives an explicit two-party communication problem $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, computable by a read-once $\{\wedge, \vee\}$-formula of depth d, with discrepancy

$$\exp\left(-\Omega\left(n^{\frac{1}{2}-\frac{1}{4d-6}}\right)\right).$$

This result matches all previous lower bounds for $\{\wedge, \vee\}$-circuits of polynomial size and depth $d = 3$, and strictly improves on previous work for depth $d > 3$. Table 8.1 gives a quantitative and bibliographic summary of this line of research. Finally, we remark that Theorem 8.2 generalizes to three or more parties, by the multiparty version of the pattern matrix method [49].

TABLE 8.1
Discrepancy of $\{\wedge,\vee\}$-circuits of constant depth and polynomial size.

Depth	**Discrepancy**	**Reference**
3	$\exp\{-\Omega(n^{1/3})\}$	[11, 42, 43]
4	$\exp\{-\Omega(n/\log n)^{2/5}\}$	[13]
$d \geqslant 3$	$\exp\{-\Omega(n^{\frac{1}{2}-\frac{1}{4d-6}})\}$	this paper

8.2. Computational learning. Apart from threshold degree, several other complexity measures are of interest when representing a function $f\colon \{0,1\}^n \to \{0,1\}$ by the sign of a real polynomial. Two such are the *density* and *weight* of the sign-representing polynomial. Unlike threshold degree, these measures depend on the exact choice of basis for the subspace of real polynomials of a given degree. The canonical choice is the *parity basis* χ_S for $S \subseteq \{1,2,\ldots,n\}$, where $\chi_S\colon \{0,1\}^n \to \{-1,+1\}$ is given by

$$\chi_S(x) = (-1)^{\sum_{i\in S} x_i}.$$

This basis derives its name from the fact that χ_S computes the parity of the bits in S, with output values -1 and $+1$ corresponding to odd and even parity, respectively. The *threshold density* of f, denoted $\operatorname{dns}(f)$, is the minimum size of a set family $\mathscr{S}$ such that

$$\operatorname{sgn}\left(\sum_{S\in\mathscr{S}} \lambda_S\chi_S\right) \equiv \begin{cases} -1 & \text{if } f(x)=0, \\ +1 & \text{if } f(x)=1 \end{cases}$$

for some reals λ_S. A more subtle complexity measure is the *threshold weight* of f, denoted $W(f)$ and defined as the minimum sum $\sum_{S\subseteq\{1,2,\ldots,n\}} |\lambda_S|$ over all integers λ_S such that

$$\operatorname{sgn}\left(\sum_{S\subseteq\{1,2,\ldots,n\}} \lambda_S\chi_S\right) \equiv \begin{cases} -1 & \text{if } f(x)=0, \\ +1 & \text{if } f(x)=1. \end{cases}$$

In other words, $\operatorname{dns}(f)$ is the minimum *number* of functions χ_S in any linear combination that sign-represents f, whereas $W(f)$ is the minimum *sum of coefficients* in any integer linear combination of χ_S that sign-represents f. In circuit complexity terms, the threshold density and threshold weight of f exactly correspond to the minimum size of a threshold-of-parity and threshold-of-majority circuit for f, respectively. It is clear that $\operatorname{dns}(f) \leqslant W(f)$ for every f, and a little more thought reveals that $1 \leqslant \operatorname{dns}(f) \leqslant 2^n$ and $1 \leqslant W(f) \leqslant (2\sqrt{2})^n$. These complexity measures have been extensively studied [9, 10, 17, 20, 7, 22, 29, 24, 28, 26, 27, 11, 39], motivated by applications to computational learning and circuit complexity.

The following ingenious theorem, due to Krause and Pudlák [29, Proposition 2.1], translates lower bounds on threshold degree into lower bounds on threshold density.

TABLE 8.2
Threshold weight and threshold density of $\{\wedge, \vee\}$-circuits of constant depth and polynomial size.

Depth	**Threshold weight**	**Threshold density**	**Reference**
3	$\exp\{\Omega(n^{1/3})\}$	$\exp\{\Omega(n^{1/3})\}$	[29]
4	$\exp\{\Omega(n/\log n)^{2/5}\}$	no bound	[13]
$d \geqslant 3$	$\exp\{\Omega(n^{\frac{1}{2}-\frac{1}{4d-6}})\}$	$\exp\{\Omega(n^{\frac{1}{2}-\frac{1}{4d-6}})\}$	this paper

THEOREM 8.3 (Krause and Pudlák). *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a given Boolean function. Define $F\colon (\{0,1\}^n)^3 \to \{0,1\}$ by $F(x,y,z) = f(\dots, (\overline{z_i} \wedge x_i) \vee (z_i \wedge y_i), \dots)$. Then*

$$\operatorname{dns}(F) \geqslant 2^{\deg_\pm(f)}.$$

Combining Krause and Pudlák's technique with the main result of this paper, we obtain the desired lower bound on the threshold density of constant-depth circuits.

THEOREM 8.4. *Fix an arbitrary constant $k \geqslant 1$ and define $F_n\colon \{0,1\}^{2n} \to \{0,1\}$ by*

$$F_n = \mathrm{NOR}_{n^{\frac{1}{2k-1}}} \circ \underbrace{\mathrm{NOR}_{n^{\frac{2}{2k-1}}} \circ \cdots \circ \mathrm{NOR}_{n^{\frac{2}{2k-1}}}}_{k-1} \circ \mathrm{NOR}_2.$$

Then

$$W(F_n) \geqslant \operatorname{dns}(F_n) \geqslant \exp\left(\Omega\left(n^{\frac{k-1}{2k-1}}\right)\right).$$

Proof. Define $f_n\colon \{0,1\}^n \to \{0,1\}$ by

$$f_n = \mathrm{NOR}_{n^{\frac{1}{2k-1}}} \circ \underbrace{\mathrm{NOR}_{n^{\frac{2}{2k-1}}} \circ \cdots \circ \mathrm{NOR}_{n^{\frac{2}{2k-1}}}}_{k-1}.$$

The identity $\mathrm{NOR}_s \circ \mathrm{OR}_t = \mathrm{NOR}_{st}$ implies that $f_n \circ \mathrm{OR}_2 \circ \mathrm{AND}_2$ is a subfunction of $F_{2^{2k-1}n}$. The claimed lower bound for F_n now follows from

$$\begin{aligned}\operatorname{dns}(F_{2^{2k-1}n}) &\geqslant \operatorname{dns}(f_n \circ \mathrm{OR}_2 \circ \mathrm{AND}_2)\\ &\geqslant 2^{\deg_\pm(f_n)}\\ &\geqslant \exp(\Omega(n^{\frac{k-1}{2k-1}})),\end{aligned}$$

where the last two inequalities use Theorem 8.3 and Theorem 1.1, respectively. ∎

This establishes Theorem 1.5 from the Introduction. For any $d \geqslant 3$, Theorem 8.4 gives a read-once $\{\wedge, \vee\}$-formula $F\colon \{0,1\}^n \to \{0,1\}$ of depth d with threshold weight and threshold density

$$\exp\left(\Omega\left(n^{\frac{1}{2}-\frac{1}{4d-6}}\right)\right).$$

This result matches all previous lower bounds for $\{\wedge,\vee\}$-circuits of polynomial size and depth $d=3$, and strictly improves on previous work for depth $d>3$. The reader will find a quantitative and bibliographic summary of this line of research in Table 8.2.

Remark 8.5. Threshold weight and threshold density are sometimes defined in terms of a different monomial basis, whose elements are the conjunction functions $x\mapsto\prod_{i\in S}x_i$ for $S\subseteq\{1,2,\ldots,n\}$. Krause and Pudlák's theorem easily generalizes to that setting, as does Theorem 8.4.

8.3. Approximate degree of AND-OR trees. The approximate degree of AND-OR trees has been the focus of much work over the past two decades. A series of breakthroughs [19, 21, 15, 3, 41] in quantum query complexity have culminated in an $O(\sqrt{n})$ upper bound on the 1/3-approximate degree of every AND-OR tree on n variables. Obtaining a matching lower bound has turned out to be surprisingly difficult, even for trees as simple as $\mathrm{NOR}_{n_1}\circ\mathrm{NOR}_{n_2}$ on $n=n_1n_2$ variables. This depth-2 tree was finally shown to have approximate degree $\Omega(\sqrt{n})$ by Bun and Thaler [12] and independently by the author [45], closing a long line of incremental improvements [35, 51, 2, 46, 12, 45]. In follow-up work, Bun and Thaler [13] generalized this lower bound to arbitrary constant depth $k\geqslant 2$, showing that the tree $\mathrm{NOR}_{n_1}\circ\mathrm{NOR}_{n_2}\circ\cdots\circ\mathrm{NOR}_{n_k}$ on $n=n_1n_2\cdots n_k$ variables has approximate degree $\Omega(\sqrt{n}/\log^{(k-2)/2}n)$. Our amplification theorem sharpens this lower bound to a tight $\Omega(\sqrt{n})$ for the depth-k tree with all fan-ins $n^{1/k}$:

Theorem 8.6. *Let $k\geqslant 1$ be an arbitrary integer constant. Then for all $n\geqslant 1$, the composition*

$$f=\underbrace{\mathrm{NOR}_{n^{1/k}}\circ\mathrm{NOR}_{n^{1/k}}\circ\cdots\circ\mathrm{NOR}_{n^{1/k}}}_{k}$$

obeys

$$\deg_{1/3}(f)\geqslant\deg^{+}_{1/3}(f)=\Omega(\sqrt{n}).$$

Proof. We will prove that

$$\deg^{+}_{1/4}(\underbrace{\mathrm{NOR}_{cn}\circ\mathrm{NOR}_{cn}\circ\cdots\circ\mathrm{NOR}_{cn}}_{k})\geqslant n^{k/2}\tag{8.2}$$

for a sufficiently large absolute constant $c\geqslant 1$ and all positive integers k and n. This suffices to prove the theorem because the error in a one-sided approximation of any given Boolean function can be reduced from 1/3 to 1/4 at the expense of a constant-factor increase in the degree of the approximant.

We claim that for all n and k, the function in (8.2) has a $(n^{(k-1)/2},n^{k/2},1/2)$-dual pair. Indeed, the base case $k=1$ of this claim is immediate from Theorem 2.2 and Lemma 4.2, whereas the inductive step follows from Theorem 4.3. Now (8.2) follows directly from the dual characterization of one-sided approximate degree (Theorem 2.6) and the definition of a dual pair. ∎

9. Open problems. There are several natural directions for future work. The most obvious open problem is to obtain improved lower and upper bounds on the maximum threshold degree of constant-depth circuits. In particular, is there a constant-depth circuit $f\colon\{0,1\}^n\to\{0,1\}$ with threshold degree $\Omega(n)$? An analogous question for approximate degree is also of great interest. A related open problem, discussed in the previous section, concerns the approximate degree of arbitrary AND-OR trees on n variables. As of this writing, matching lower and upper bounds are known only for AND-OR trees of constant depth. Finally, it is a fascinating open problem to determine whether approximate degree is multiplicative in the sense that $\deg_{1/3}(f\circ g) = \Theta(\deg_{1/3}(f)\deg_{1/3}(g))$ for all Boolean functions $f, g\colon \{0,1\}^n \to \{0,1\}$. Currently, only the upper bound is known to hold (Corollary 2.11).

Appendix A. Constructing a dual object for NOR. The purpose of this appendix is to prove Theorem 2.9, which gives a dual object for the NOR function with a number of additional properties. The development here closely follows earlier work by Špalek [54] and Bun and Thaler [12]. The main points of departure are a more careful choice of roots for the dual object and the use of shifts, to induce the desired sign behavior and metric properties. We start with a well-known binomial identity [18].

FACT A.1. *For every polynomial p of degree less than n,*

$$\sum_{t=0}^{n}(-1)^t\binom{n}{t}p(t) = 0.$$

The next lemma constructs a dual object for NOR that has the sign behavior claimed in Theorem 2.9 but may lack the corresponding metric properties.

LEMMA A.2. *Let ϵ be given, $0<\epsilon<1$. Then for some $\delta=\delta(\epsilon)>0$ and every $n\geqslant 2$, there exists an (explicitly given) function $\omega\colon\{0,1,2,\dots,n\}\to\mathbb{R}$ such that*

$$\omega(0) > \frac{1-\epsilon}{2}\cdot\|\omega\|_1, \tag{A.1}$$

$$(-1)^{n+t}\omega(t)\geqslant 0 \qquad (t=1,2,\dots,n), \tag{A.2}$$

$$\deg p < \sqrt{\delta n} \implies \langle \omega, p\rangle = 0. \tag{A.3}$$

Proof. We first consider the case of n odd. Let $m = 2\lceil 4/\epsilon\rceil + 1$ and $d=\lfloor\sqrt{n/m}\rfloor$. Define $S=\{2\}\cup\{i^2m : i=0,1,2,\dots,d\}$, so that $S\subseteq\{0,1,2,\dots,n\}$. Consider the function $\omega\colon\{0,1,2,\dots,n\}\to\mathbb{R}$ given by

$$\omega(t) = \frac{(-1)^{n+t+|S|+1}}{n!}\binom{n}{t}\prod_{\substack{i=0,1,2,\dots,n:\\ i\notin S}}(t-i).$$

Fact A.1 implies that ω is orthogonal to every polynomial of degree at most d, so that (A.3) holds with

$$\delta = \frac{1}{2\lceil 4/\epsilon\rceil+1}.$$

A routine calculation reveals that

$$\omega(t) = \begin{cases} (-1)^{|\{i\in S : i<t\}|}\displaystyle\prod_{i\in S\setminus\{t\}}\frac{1}{|t-i|} & \text{if } t\in S,\\[2ex] 0 & \text{otherwise.}\end{cases} \tag{A.4}$$

In particular,

$$\frac{\omega(0)}{|\omega(2)|} = \prod_{i=1}^{d}\frac{i^2m-2}{i^2m} \geqslant 1-\sum_{i=1}^{d}\frac{2}{i^2m} > 1-\frac{2}{m}\sum_{i=1}^{\infty}\frac{1}{i^2} = 1-\frac{\pi^2}{3m}$$

and

$$\frac{\omega(0)}{|\omega(i^2m)|} = \frac{i^2m-2}{4} \cdot \frac{(d-i)!\,(d+i)!}{d!\,d!} \geqslant \frac{i^2m-2}{4} \qquad (i=1,2,\dots,d).$$

Hence,

$$\begin{aligned}
\frac{\|\omega\|_1}{\omega(0)} &= 1 + \frac{|\omega(2)|}{\omega(0)} + \sum_{i=1}^{d} \frac{|\omega(i^2m)|}{\omega(0)} \\
&\leqslant 2 + \frac{\pi^2}{3m-\pi^2} + \sum_{i=1}^{d} \frac{4}{i^2m-2} \\
&\leqslant 2 + \frac{\pi^2}{3m-\pi^2} + \frac{4}{m-2}\sum_{i=1}^{\infty} \frac{1}{i^2} \\
&= 2 + \frac{\pi^2}{3m-\pi^2} + \frac{2}{m-2}\cdot\frac{\pi^2}{3} \\
&\leqslant 2 + 2\epsilon,
\end{aligned}$$

where the last step holds because $m \geqslant 8/\epsilon$. Now (A.1) is immediate.

It remains to examine the sign behavior of ω. Since ω vanishes outside S, the requirement (A.2) holds trivially at those points. For $t \in S$, it follows from (A.4) that

$$\begin{aligned}
&\operatorname{sgn}\omega(2) = -1, \\
&\operatorname{sgn}\omega(i^2m) = (-1)^{i+1} \qquad (i=1,2,\dots,d).
\end{aligned}$$

Since m is odd, these equations yield $\operatorname{sgn}\omega(t) = (-1)^{t+1}$ for positive $t \in S$. This settles (A.2) and completes the proof for n odd. The proof for n even is closely analogous, with the difference that one works with the set $S = \{0\} \cup \{i^2m+1 : i = 0,1,2,\dots\}$ for an odd integer $m = \Theta(1/\epsilon)$. ∎

We have reached the main result of this section, stated earlier as Theorem 2.9.

Theorem A.3. *Let ϵ be given, $0<\epsilon<1$. Then for some $\delta=\delta(\epsilon)>0$ and every $n \geqslant 2$, there exists an (explicitly given) function $\omega\colon \{0,1,2,\dots,n\} \to \mathbb{R}$ such that*

$$\omega(0) > \frac{1-\epsilon}{2}\cdot\|\omega\|_1, \tag{A.5}$$

$$(-1)^{n+t}\omega(t) \geqslant \frac{\epsilon}{4t^2}\cdot\|\omega\|_1 \qquad (t=1,2,\dots,n), \tag{A.6}$$

$$\deg p < \sqrt{\delta n} \implies \langle \omega, p\rangle = 0. \tag{A.7}$$

Proof. The cases $n=2$ and $n=3$ can be handled directly by taking $\delta=\delta(\epsilon)=1/4$ and defining

$$\begin{aligned}
\omega\colon (0,1,2) &\mapsto \left(\frac{1}{2}-\frac{\epsilon}{3}, -\frac{1}{2}, \frac{\epsilon}{3}\right), \\
\omega\colon (0,1,2,3) &\mapsto \left(\frac{1}{2}-\frac{\epsilon}{3}, \frac{\epsilon}{4}, -\frac{1}{2}, \frac{\epsilon}{12}\right),
\end{aligned}$$

respectively. In the rest of the proof, we treat the complementary case $n \geqslant 4$.

For some $\delta = \delta(\epsilon) > 0$ and all $n \geqslant 4$, Lemma A.2 ensures the existence of functions $\omega_0 \colon \{0, 1, 2, \ldots, 2\lfloor n/4 \rfloor\} \to \mathbb{R}$ and $\omega_1 \colon \{0, 1, 2, \ldots, n\} \to \mathbb{R}$ such that

$$\|\omega_0\|_1 = \|\omega_1\|_1 = 1, \tag{A.8}$$

$$\omega_0(0) > \frac{6-\epsilon}{12}, \tag{A.9}$$

$$\omega_1(0) > \frac{6-\epsilon}{12}, \tag{A.10}$$

$$(-1)^t \omega_0(t) \geqslant 0, \qquad t \geqslant 0, \tag{A.11}$$

$$(-1)^{n+t} \omega_1(t) \geqslant 0, \qquad t \geqslant 1, \tag{A.12}$$

$$\deg p < \sqrt{\delta n} \implies \langle \omega_0, p \rangle = \langle \omega_1, p \rangle = 0. \tag{A.13}$$

For convenience, extend ω_0 and ω_1 to all of $\mathbb{Z}$ by defining these functions to be zero outside their original domain. Define $\omega \colon \{0, 1, 2, \ldots, n\} \to \mathbb{R}$ by

$$\omega(t) = \omega_1(t) + \rho \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{(-1)^{i+n}}{i^2} \omega_0(t-i) + \rho \sum_{i=\lfloor n/2 \rfloor + 1}^{n} \frac{(-1)^{i+n}}{i^2} \omega_0(-t+i),$$

where

$$\rho = \frac{5\epsilon}{\pi^2(1-\epsilon)}.$$

We proceed to verify the three properties of ω claimed in the theorem statement. To begin with,

$$\begin{aligned}\|\omega\|_1 &\leqslant \|\omega_1\|_1 + \rho \sum_{i=1}^{n} \frac{1}{i^2} \|\omega_0\|_1 \leqslant 1 + \rho \sum_{i=1}^{\infty} \frac{1}{i^2} = 1 + \rho \cdot \frac{\pi^2}{6} \\ &= \frac{6-\epsilon}{6(1-\epsilon)}, \end{aligned} \tag{A.14}$$

where the second inequality uses (A.8). Now (A.5) is immediate because $\omega(0) = \omega_1(0) > (6-\epsilon)/12$ by (A.10).

Property (A.6) for $t \geqslant 1$ can be verified as follows:

$$\begin{aligned}(-1)^{n+t}\omega(t) &= |\omega_1(t)| + \rho \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{|\omega_0(t-i)|}{i^2} + \rho \sum_{i=\lfloor n/2 \rfloor + 1}^{n} \frac{|\omega_0(-t+i)|}{i^2} \\ &\geqslant \rho \cdot \frac{|\omega_0(0)|}{t^2} \\ &\geqslant \frac{5\epsilon}{\pi^2(1-\epsilon)} \cdot \frac{6-\epsilon}{12t^2} \\ &\geqslant \frac{\epsilon}{4t^2} \cdot \|\omega\|_1, \end{aligned}$$

where the first step follows from (A.11) and (A.12), the third from (A.9), and the fourth from (A.14).

The remaining property (A.7) is immediate from (A.13). ∎

Acknowledgments. The author is thankful to Mark Bun, Ryan O'Donnell, Rocco Servedio, Justin Thaler, and the anonymous reviewers for their valuable feedback on an earlier version of this manuscript.

REFERENCES

[1] S. Aaronson and Y. Shi, *Quantum lower bounds for the collision and the element distinctness problems*, J. ACM, 51 (2004), pp. 595–605, doi:10.1145/1008731.1008735.

[2] A. Ambainis, *Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range*, Theory of Computing, 1 (2005), pp. 37–46, doi:10.4086/toc.2005.v001a003.

[3] A. Ambainis, A. M. Childs, B. Reichardt, R. Špalek, and S. 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 (2010), pp. 2513–2530, doi:10.1137/080712167.

[4] J. Aspnes, R. Beigel, M. L. Furst, and S. Rudich, *The expressive power of voting polynomials*, Combinatorica, 14 (1994), pp. 135–148, doi:10.1007/BF01215346.

[5] L. Babai, P. Frankl, and J. Simon, *Complexity classes in communication complexity theory*, in *Proceedings of the Twenty-Seventh Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 1986, pp. 337–347, doi:10.1109/SFCS.1986.15.

[6] P. Beame and T. Huynh, *Multiparty communication complexity and threshold circuit size of* AC^0, SIAM J. Comput., 41 (2012), pp. 484–518, doi:10.1137/100792779.

[7] R. Beigel, *Perceptrons,* PP*, and the polynomial hierarchy*, Computational Complexity, 4 (1994), pp. 339–349, doi:10.1007/BF01263422.

[8] R. Beigel, N. Reingold, and D. A. Spielman, PP *is closed under intersection*, J. Comput. Syst. Sci., 50 (1995), pp. 191–202, doi:10.1006/jcss.1995.1017.

[9] J. Bruck, *Harmonic analysis of polynomial threshold functions*, SIAM J. Discrete Math., 3 (1990), pp. 168–177, doi:10.1137/0403015.

[10] J. Bruck and R. Smolensky, *Polynomial threshold functions,* AC^0 *functions, and spectral norms*, SIAM J. Comput., 21 (1992), pp. 33–42, doi:10.1137/0221003.

[11] H. Buhrman, N. K. Vereshchagin, and R. de Wolf, *On computation and communication with small bias*, in *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity* (CCC), 2007, pp. 24–32, doi:10.1109/CCC.2007.18.

[12] M. Bun and J. Thaler, *Dual lower bounds for approximate degree and Markov–Bernstein inequalities*, Inf. Comput., 243 (2015), pp. 2–25, doi:10.1016/j.ic.2014.12.003.

[13] M. Bun and J. Thaler, *Hardness amplification and the approximate degree of constant-depth circuits*, in *Proceedings of the Forty-Second International Colloquium on Automata, Languages and Programming* (ICALP), 2015, pp. 268–280, doi:10.1007/978-3-662-47672-7_22.

[14] A. Chattopadhyay and A. Ada, *Multiparty communication complexity of disjointness*, in Electronic Colloquium on Computational Complexity (ECCC), January 2008. Report TR08-002.

[15] E. Farhi, J. Goldstone, and S. Gutmann, *A quantum algorithm for the Hamiltonian NAND tree*, Theory of Computing, 4 (2008), pp. 169–190, doi:10.4086/toc.2008.v004a008.

[16] D. Gavinsky and A. A. Sherstov, *A separation of* NP *and* coNP *in multiparty communication complexity*, Theory of Computing, 6 (2010), pp. 227–245, doi:10.4086/toc.2010.v006a010.

[17] M. Goldmann, J. Håstad, and A. A. Razborov, *Majority gates vs. general weighted threshold gates*, Computational Complexity, 2 (1992), pp. 277–300, doi:10.1007/BF01200426.

[18] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, 2nd ed., 1994.

[19] L. K. Grover, *A fast quantum mechanical algorithm for database search*, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* (STOC), 1996, pp. 212–219, doi:10.1145/237814.237866.

[20] J. Håstad, *On the size of weights for threshold gates*, SIAM J. Discret. Math., 7 (1994), pp. 484–492, doi:10.1137/S0895480192235878.

[21] P. Høyer, M. Mosca, and R. de Wolf, *Quantum search on bounded-error inputs*, in *Proceedings of the Thirtieth International Colloquium on Automata, Languages and Programming* (ICALP), 2003, pp. 291–299, doi:10.1007/3-540-45061-0_25.

[22] J. C. Jackson, *An efficient membership-query algorithm for learning DNF with respect to the uniform distribution*, J. Comput. Syst. Sci., 55 (1997), pp. 414–440, doi:10.1006/jcss.1997.1533.

[23] M. J. Kearns and U. V. Vazirani, *An Introduction to Computational Learning Theory*, MIT Press, Cambridge, 1994.

[24] A. R. Klivans, R. O'Donnell, and R. A. Servedio, *Learning intersections and thresholds of halfspaces*, J. Comput. Syst. Sci., 68 (2004), pp. 808–840, doi:10.1016/j.jcss.2003.11.002.

[25] A. R. Klivans and R. A. Servedio, *Learning DNF in time* $2^{\tilde{O}(n^{1/3})}$, J. Comput. Syst. Sci., 68 (2004), pp. 303–318, doi:10.1016/j.jcss.2003.07.007.

[26] A. R. Klivans and R. A. Servedio, *Toward attribute efficient learning of decision lists and*

parities, J. Machine Learning Research, 7 (2006), pp. 587–602.

[27] A. R. Klivans and A. A. Sherstov, *Unconditional lower bounds for learning intersections of halfspaces*, Machine Learning, 69 (2007), pp. 97–114, doi:10.1007/s10994-007-5010-1. Preliminary version in *Proceedings of the Nineteenth Annual Conference on Computational Learning Theory* (COLT), 2006.

[28] M. Krause, *On the computational power of Boolean decision lists*, Computational Complexity, 14 (2006), pp. 362–375, doi:10.1007/s00037-005-0203-0.

[29] M. Krause and P. Pudlák, *On the computational power of depth-2 circuits with threshold and modulo gates*, Theor. Comput. Sci., 174 (1997), pp. 137–156, doi:10.1016/S0304-3975(96)00019-9.

[30] M. Krause and P. Pudlák, *Computing Boolean functions by polynomials and threshold circuits*, Comput. Complex., 7 (1998), pp. 346–370, doi:10.1007/s000370050015.

[31] E. Kushilevitz and N. Nisan, *Communication complexity*, Cambridge University Press, 1997.

[32] T. Lee, *A note on the sign degree of formulas*, 2009. Available at http://arxiv.org/abs/0909.4607.

[33] T. Lee and A. Shraibman, *Disjointness is hard in the multiparty number-on-the-forehead model*, Computational Complexity, 18 (2009), pp. 309–336, doi:10.1007/s00037-009-0276-2.

[34] M. L. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge, Mass., 1969.

[35] N. Nisan and M. Szegedy, *On the degree of Boolean functions as real polynomials*, Computational Complexity, 4 (1994), pp. 301–313, doi:10.1007/BF01263419.

[36] R. O'Donnell and R. A. Servedio, *New degree bounds for polynomial threshold functions*, Combinatorica, 30 (2010), pp. 327–358, doi:10.1007/s00493-010-2173-3.

[37] R. Paturi and M. E. Saks, *Approximating threshold circuits by rational functions*, Inf. Comput., 112 (1994), pp. 257–272, doi:10.1006/inco.1994.1059.

[38] V. V. Podolskii, *Perceptrons of large weight*, in *Proceedings of the Second International Computer Science Symposium in Russia* (CSR), 2007, pp. 328–336, doi:10.1007/978-3-540-74510-5_33.

[39] V. V. Podolskii, *A uniform lower bound on weights of perceptrons*, in *Proceedings of the Third International Computer Science Symposium in Russia* (CSR), 2008, pp. 261–272, doi:10.1007/978-3-540-79709-8_27.

[40] A. A. Razborov and A. A. Sherstov, *The sign-rank of* AC^0, SIAM J. Comput., 39 (2010), pp. 1833–1855, doi:10.1137/080744037. Preliminary version in *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2008.

[41] B. Reichardt, *Reflections for quantum query algorithms*, in *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA), 2011, pp. 560–569, doi:10.1137/1.9781611973082.44.

[42] A. A. Sherstov, *Separating* AC^0 *from depth-2 majority circuits*, SIAM J. Comput., 38 (2009), pp. 2113–2129, doi:10.1137/08071421X. Preliminary version in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* (STOC), 2007.

[43] A. A. Sherstov, *The pattern matrix method*, SIAM J. Comput., 40 (2011), pp. 1969–2000, doi:10.1137/080733644. Preliminary version in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* (STOC), 2008.

[44] A. A. Sherstov, *Strong direct product theorems for quantum communication and query complexity*, SIAM J. Comput., 41 (2012), pp. 1122–1165, doi:10.1137/110842661. Preliminary version in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (STOC), 2011.

[45] A. A. Sherstov, *Approximating the AND-OR tree*, Theory of Computing, 9 (2013), pp. 653–663, doi:10.4086/toc.2013.v009a020.

[46] A. A. Sherstov, *The intersection of two halfspaces has high threshold degree*, SIAM J. Comput., 42 (2013), pp. 2329–2374, doi:10.1137/100785260. Preliminary version in *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2009.

[47] A. A. Sherstov, *Making polynomials robust to noise*, Theory of Computing, 9 (2013), pp. 593–615, doi:10.4086/toc.2013.v009a018. Preliminary version in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2012.

[48] A. A. Sherstov, *Optimal bounds for sign-representing the intersection of two halfspaces by polynomials*, Combinatorica, 33 (2013), pp. 73–96, doi:10.1007/s00493-013-2759-7. Preliminary version in *Proceedings of the Forty-Second Annual ACM Symposium on Theory of Computing* (STOC), 2010.

[49] A. A. Sherstov, *Communication lower bounds using directional derivatives*, J. ACM, 61

(2014), pp. 1–71, doi:10.1145/2629334. Preliminary version in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing* (STOC), 2013.

[50] A. A. Sherstov, *The multiparty communication complexity of set disjointness*, SIAM J. Comput., 45 (2016), pp. 1450–1489, doi:10.1137/120891587. Preliminary version in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2009.

[51] Y. Shi, *Approximating linear restrictions of Boolean functions*. Manuscript, 2002.

[52] Y. Shi and Y. Zhu, *Quantum communication complexity of block-composed functions*, Quantum Information & Computation, 9 (2009), pp. 444–460.

[53] K.-Y. Siu, V. P. Roychowdhury, and T. Kailath, *Rational approximation techniques for analysis of neural networks*, IEEE Transactions on Information Theory, 40 (1994), pp. 455–466, doi:10.1109/18.312168.

[54] R. Špalek, *A dual polynomial for OR*. Available at http://arxiv.org/abs/0803.4516, 2008.