THE POWER OF ASYMMETRY IN 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: $f(x) \equiv \operatorname{sgn} p(x)$. Introduced in the seminal work of Minsky and Papert (1969), this notion is central to some of the strongest algorithmic and complexity-theoretic results for constant-depth circuits. One problem that has remained open for several decades, with applications to computational learning and communication complexity, is to determine the maximum threshold degree of a polynomial-size constant-depth circuit in n variables. The best lower bound prior to our work was $\Omega(n^{(d-1)/(2d-1)})$ for circuits of depth d. We obtain a polynomial improvement for every depth d, with a lower bound of $\Omega(n^{3/7})$ for depth 3 and $\Omega(\sqrt{n})$ for depth $d \geqslant 4$. The proof contributes an approximation-theoretic technique of independent interest, which exploits asymmetry in circuits to prove their hardness for polynomials.

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

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

1. Introduction.

Representations of Boolean functions by polynomials have played an important role in theoretical computer science. The idea of representing a Boolean function by the *sign* of a polynomial has been particularly fruitful. Formally, a real polynomial p is said to represent a given Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ 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 *threshold degree* of f, denoted $\deg_{\pm}(f)$, is the minimum degree of a sign-representing polynomial for f. Equivalent terms in the literature include *strong degree* [5], *voting polynomial degree* [20], *PTF degree* [24], and *sign degree* [11]. The formal study of this complexity measure began in 1969 with the pioneering work of Minsky and Papert [22]. Motivated by applications to neural networks, the authors of [22] proved that the parity function on n variables has the maximum possible threshold degree, n. They obtained lower bounds on the threshold degree of several other functions, including DNF formulas and intersections of halfspaces. Minsky and Papert's work has found applications far beyond artificial intelligence. In theoretical computer science, applications of sign-representing polynomials range from circuit lower bounds [20, 21] and size-depth trade-offs [26, 38] to computational learning [18, 17, 25, 4, 32, 34, 13, 35, 39] and the closure properties of PP [9].

The notion of threshold degree has been especially influential in the study of AC^0, the class of constant-depth polynomial-size circuits with $\wedge, \vee, \neg$ gates of unbounded fan-in. Aspnes et al. [5] used sign-representing polynomials to give an entirely different proof of classic lower bounds for AC^0. In communication complexity, the notion of

[*] Submitted to the editors March 4, 2016. An extended abstract of this paper appeared in *Proceedings of the Fifty-Sixth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 431–450, 2015.

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, http://web.cs.ucla.edu/~sherstov).

Table 1.1

Known bounds on the maximum threshold degree of $\wedge, \vee, \neg$-circuits of polynomial size and constant depth. In all bounds, n denotes the number of variables.

Depth	**Threshold degree**	**Reference**
2	$\Omega(n^{1/3})$	Minsky and Papert [22]
2	$O(n^{1/3}\log n)$	Klivans and Servedio [18]
d	$\Omega(n^{1/3}\log^{2(d-2)/3} n)$	O'Donnell and Servedio [25]
d	$\Omega(n^{\frac{d-1}{2d-1}})$	Sherstov [35]
3	$\Omega(n^{3/7})$	This paper
4	$\Omega(\sqrt{n})$	This paper

threshold degree was critical in constructing [29, 30] the first AC^0 circuit with exponentially small discrepancy and hence maximum communication complexity in nearly every model. That discrepancy result was used in [29] to show the optimality of Allender's classic simulation of AC^0 by majority circuits, solving the open problem [20] on the relation between the two circuit classes. A more sophisticated application of threshold degree gave the first exponential lower bound on the sign-rank of AC^0 circuits [27], twenty-two years after the problem was posed by Babai et al. [6]. Subsequent work [14, 7, 37, 36] resolved other questions in communication complexity and circuit complexity related to constant-depth circuits by generalizing the threshold degree method of [29, 30].

Sign-representing polynomials have also enabled *algorithmic* breakthroughs in the study of constant-depth circuits. An illustrative example is the fastest known algorithm for learning polynomial-size DNF formulas, due to Klivans and Servedio [18], with running time $\exp(\tilde{O}(n^{1/3}))$. The authors of [18] obtained their algorithm by proving an essentially tight upper bound of $O(n^{1/3}\log n)$ on the threshold degree of that concept class. Another such learning-theoretic breakthrough is the fastest algorithm for learning Boolean formulas, obtained by O'Donnell and Servedio [25] for formulas of constant depth and by Ambainis et al. [4] for arbitrary depth. The algorithm runs in time $\exp(\tilde{O}(n^{(2^{d-1}-1)/(2^d-1)}))$ for formulas of size n and constant depth d, and in time $\exp(\tilde{O}(\sqrt{n}))$ for formulas of unbounded depth. In both cases, the bound on the running time follows from the corresponding upper bound on the threshold degree.

1.1. Our results. A longstanding open problem in the area is to determine the maximum threshold degree of an AC^0 circuit in n variables. This problem is motivated by algorithmic and complexity-theoretic applications [18, 25, 19, 27, 13], in addition to being a natural question in its own right. Table 1.1 summarizes the progress to date. In their seminal monograph, Minsky and Papert [22] proved a lower bound of $\Omega(n^{1/3})$ on the threshold degree of the following DNF formula in n variables:

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

Three decades later, Klivans and Servedio [18] obtained a bound of $O(n^{1/3}\log n)$ on the threshold degree of any polynomial-size DNF formula in n variables, matching Minsky and Papert's result and resolving the problem for depth 2. Attempts to determine the threshold degree for depth $d \geqslant 3$ have been met with limited success. The only upper bound known to date is $O(n)$, which follows trivially from the definition of threshold degree. In particular, it is consistent with our knowledge that there are AC^0 circuits with linear threshold degree. On the lower bounds side, the only progress until recently was due to O'Donnell and Servedio [25], who constructed circuits of depth d with threshold degree $\Omega(n^{1/3}\log^{2(d-2)/3} n)$. The authors of [25] formally posed the challenge of obtaining a polynomial improvement on Minsky and Papert's lower bound. Such an improvement was obtained in [35], with a threshold degree lower bound of $\Omega(n^{(d-1)/(2d-1)})$ for circuits of depth d. In particular, the result in [35] subsumes all previous lower bounds, with a strict improvement starting at depth $d = 3$. The main contribution of our paper is a polynomially stronger lower bound for every depth $d \geqslant 3$. For depth 3, we obtain:

Theorem 1.1 (Main result, $d = 3$). *Let $f\colon \{0,1\}^{n+n^{6/7}} \to \{0,1\}$ be the depth-3 read-once formula given by*

$$f(x,y) = \bigvee_{i=1}^{n^{1/7}} \left(\bigwedge_{j=1}^{n^{2/7}} \bigvee_{k=1}^{n^{4/7}} x_{i,j,k} \right) \wedge \left(\bigwedge_{j=1}^{n^{3/7}} \bigvee_{k=1}^{n^{2/7}} y_{i,j,k} \right).$$

Then

$$\deg_{\pm}(f) = \Omega(n^{3/7}).$$

Apart from improving on the previous lower bound for depth-3 *circuits*, this theorem is of interest in the study of *formulas*. Specifically, it matches a known upper bound of $\tilde{O}(n^{(2^{d-1}-1)/(2^d-1)})$ on the threshold degree of formulas of depth d and size n, due to O'Donnell and Servedio [25]. Prior to our work, that upper bound was known to be tight only for $d = 1$ and $d = 2$, by the classic result of Minsky and Papert [22] mentioned above. Theorem 1.1 suggests that O'Donnell and Servedio's upper bound may be tight for all d, a fascinating possibility.

A comment is in order on the admittedly unusual formula f in Theorem 1.1. A traditional approach to lower bounds for constant-depth circuits would certainly favor a more symmetric construction. Surprisingly, asymmetry turns out to be essential to the tight lower bound in Theorem 1.1. In particular, the previous lower bound of $\Omega(n^{2/5})$ for depth-3 formulas, obtained in [35], was shown in that paper to be tight for all formulas of the form $f(x) = \bigvee_{i=1}^{n_1} \bigwedge_{j=1}^{n_2} \bigvee_{k=1}^{n_3} x_{i,j,k}$ with $n_1 n_2 n_3 = n$. Asymmetry plays a critical role in all of our constructions, as we will shortly explain in detail.

For circuits of depth $d \geqslant 4$, we obtain a lower bound of $\Omega(\sqrt{n})$, improving polynomially on previous work and Theorem 1.1.

Theorem 1.2 (Main result, $d \geqslant 4$). *There is an explicitly given $\wedge,\vee$-circuit $f\colon \{0,1\}^n \to \{0,1\}$ of depth 4 and polynomial size such that*

$$\deg_{\pm}(f) = \Omega(\sqrt{n}).$$

As we discuss in Remark 9.6 at the end of the paper, one can use O'Donnell and Servedio's technique [25] to improve the lower bound of Theorem 1.2 by an arbitrary polylogarithmic factor, at the expense of increasing the circuit depth by a constant. Theorem 1.2 solves a recent open problem due to Bun and Thaler [13] and Thaler [39], who discussed the challenge of proving an $\Omega(\sqrt{n})$ lower bound for constant-depth circuits and proposed several candidate functions. Intriguingly, the construction in Theorem 1.2 seems unrelated to Bun and Thaler's candidate functions, whose status remains open.

Finally, we note that the threshold degree lower bounds in this paper imply improved lower bounds in communication complexity and learning theory. Our main result, stated as Theorem 1.2 above and restated in technical detail as Theorem 9.4 at the end of the paper, gives an $\wedge,\vee$-circuit $f\colon \{0,1\}^n \to \{0,1\}$ of polynomial size and depth 4 with discrepancy $\exp(-\Omega(\sqrt{n}))$ and threshold weight and threshold density $\exp(\Omega(\sqrt{n}))$. The best previous bounds [35] were $\exp(-\Omega(n^{\frac{1}{2}-\frac{1}{4d-6}}))$ for discrepancy and $\exp(\Omega(n^{\frac{1}{2}-\frac{1}{4d-6}}))$ for threshold weight and density, where $d \geqslant 2$ stands for the depth of the circuit. The passage from threshold degree to these other complexity measures uses by-now standard reductions [30, 20]. We refer the interested reader to [35, section 8] for details, including all definitions.

1.2. Proof overview. At first glance, Theorems 1.1 and 1.2 seem unrelated. In reality, they are corollaries to a more general result that we prove. The key notion here is that of *one-sided approximate degree*, defined for a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ as the least degree of a real polynomial p that is close to zero on $f^{-1}(0)$ and far from zero on $f^{-1}(1)$:

$$p(x) \in \begin{cases} [-\epsilon,\epsilon] & \text{if } f(x)=0,\\ [1-\epsilon,+\infty) & \text{if } f(x)=1. \end{cases}$$

The error parameter ϵ in this definition is typically a small constant, with $\epsilon = 1/3$ being the default setting. One-sided approximate degree has played an important role in the area [18, 15, 12, 31, 13, 35], with applications to both complexity theory and algorithms. Its relation to threshold degree is straightforward: if p is a one-sided approximant for f, then $p - \frac{1}{2}$ is a sign-representing polynomial for f. One-sided approximate degree is therefore always at least as large as threshold degree, and the gap between them can be arbitrary.

The central technical contribution of this paper is a *hardness amplification* result that transforms any Boolean function with high one-sided approximate degree into a function with proportionately high threshold degree. Quantitatively, our hardness amplification theorem transforms any given circuit $f\colon \{0,1\}^n \to \{0,1\}$ with one-sided approximate degree n^α in a black-box manner into a circuit $F\colon \{0,1\}^N \to \{0,1\}$ with threshold degree $\Omega(N^\beta)$, where $\beta = \beta(\alpha)$ is the monotonically increasing function given by

$$\beta = \begin{cases} 3/7 & \text{if } \alpha < 1/2,\\ 3\alpha/(3\alpha+2) & \text{if } 1/2 \leqslant \alpha < 2/3,\\ 1/2 & \text{otherwise.} \end{cases} \tag{1.1}$$

A graph of this function is shown in Figure 1.1 (right). The technical statement of our result is as follows.

Theorem 1.3 (Hardness amplification). *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a given function with one-sided approximate degree n^α, where $0 \leqslant \alpha \leqslant 1$. Then there is an explicitly given function $F\colon \{0,1\}^N \to \{0,1\}$ of the form $F = \mathrm{OR}_{m_1} \circ ((\mathrm{AND}_{m_2} \circ f') \wedge (\mathrm{AND}_{m_3} \circ \mathrm{OR}_{m_4}))$ with threshold degree*

$$\deg_\pm(F) = \Omega(N^\beta),$$

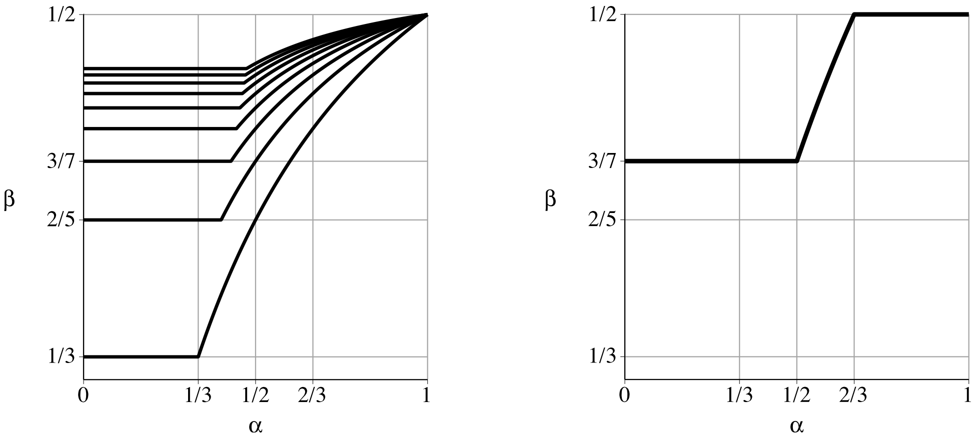

FIG. 1.1. *Transforming a circuit $f\colon \{0,1\}^n \to \{0,1\}$ with one-sided approximate degree n^α into a polynomially larger circuit $F\colon \{0,1\}^N \to \{0,1\}$ with threshold degree $\Omega(N^\beta)$. The graphs plot the dependence $\beta = \beta(\alpha)$ in previous work and this paper.* Left: *the best previous construction [35], with the distinct curves corresponding to an increase of $2, 3, \ldots 10$, respectively, in circuit depth in going from f to F.* Right: *the construction in this paper, corresponding to a depth increase of 2.*

where $f' \in \{\neg f, \mathrm{OR}_n\}$ and $\beta = \beta(\alpha)$ is given by (1.1).

In this theorem, the composition operator $\circ$ denotes componentwise composition on disjoint sets of variables, and similarly $\wedge$ denotes a conjunction on disjoint sets of variables. Thus, F is a function on $N = m_1(m_2 n + m_3 m_4)$ variables. Observe that the transformation $f \mapsto F$ preserves polynomial size and increases the circuit depth only by 2, which is essential for our applications. Our main results follow immediately from Theorem 1.3 and known lower bounds on the one-sided approximate degree of constant-depth circuits. Specifically, we obtain Theorem 1.1 by applying the hardness amplification to the function $f = \neg\mathrm{OR}_n$ with one-sided approximate degree $\Theta(\sqrt{n})$. To obtain Theorem 1.2, we instead use a certain polynomial-size CNF formula f with one-sided approximate degree $\Omega(n^{2/3})$.

We find Theorem 1.3 of interest in its own right, independent of its role in proving the main results of this paper. It is helpful to contrast it with the best previous hardness amplification result for threshold degree, obtained in [35, Theorem 1.6]. In that work, we showed how to transform any given circuit $f\colon \{0,1\}^n \to \{0,1\}$ with one-sided approximate degree n^α into a polynomially larger circuit $F\colon \{0,1\}^N \to \{0,1\}$ with threshold degree $\Omega(N^\beta)$, where

$$\beta = \max\left\{\frac{d-1}{2d-1}, \frac{d\alpha}{(2d-1)\alpha+1}\right\}.$$

The integer parameter d refers to the increase in circuit depth in going from f to F. The dependence $\beta = \beta(\alpha)$ improves monotonically with the depth parameter d, approaching $\beta = 1/2$ in the limit as $d \to \infty$. In contrast, the construction in our paper features no depth parameter; the passage $f \mapsto F$ in Theorem 1.3 always corresponds to a depth increase of 2. Figure 1.1 compares the previous hardness amplification result from [35] with Theorem 1.3 in this paper, plotting the dependence $\beta = \beta(\alpha)$ in the two cases. As the figure shows, we improve on previous work for $d = 2$ and $d = 3$, as well as for all d starting at $\alpha \geqslant 2/3$. These improvements directly translate in the polynomially stronger lower bounds in our main results.

Our proof of Theorem 1.3 departs significantly from previous work. After all, we need to somehow amplify one-sided approximate degree $n^{2/3}$ to threshold degree $\Omega(\sqrt{N})$ using only two levels of gates, as opposed to infinitely many in previous work. We achieve these efficiency gains as follows, using asymmetry along with new intermediate notions of approximation.

(i) By hypothesis, the original function f cannot be approximated in a one-sided manner by a polynomial of low degree. In the notation of Theorem 1.3, our first step is to show that the composition $\mathrm{AND}_{m_2} \wedge \neg f$ cannot be approximated in a one-sided manner to within a small constant by any low-degree rational function with ℓ_∞ norm $2^{O(m_2)}$. This passage from polynomials to rational functions is the first stage in the hardness amplification process.

(ii) In parallel, we show that the composition $\mathrm{AND}_{m_3} \circ \mathrm{OR}_{m_4}$ cannot be approximated in a one-sided manner to exponentially small error by any low-degree rational function.

(iii) Using the conclusions of the previous two steps, we prove that the conjunction $(\mathrm{AND}_{m_2} \wedge \neg f) \wedge (\mathrm{AND}_{m_3} \wedge \mathrm{OR}_{m_4})$ cannot be approximated in a one-sided manner to within a small constant by any low-degree rational function. This step is the centerpiece of our paper, and it holds in considerable generality. Specifically, we are able to prove a general "composition theorem" that characterizes the one-sided rational approximation of any composition $g \wedge h$ in terms of approximation-theoretic properties of the individual functions g and h. Note the critical role of asymmetry in this step.

(iv) Finally, we invoke a result from previous work [35] that characterizes the threshold degree of a disjunction of functions in terms of the one-sided rational approximation of the individual functions.

Steps (i), (ii), and (iii) in this program correspond to sections 7, 8, and 6, respectively. These components are put together in section 9, completing the proof. We provide additional details and intuition at each stage of the proof, and conclude the paper by discussing the potential of our approach to give stronger bounds.

2. Preliminaries.

2.1. Basic notation. Given the key role of rational functions in this work, it will be convenient to use the extended real number system $\mathbb{R} \cup \{-\infty, +\infty\}$ for all calculations. We additionally adopt the conventions that $0^0 = 1$ and $x/0 = +\infty$ for $x > 0$, where the former is justified by continuity. As usual, $\log x$ refers to the logarithm of x to base 2. 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 sign function is given by

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

For a logical condition C, we use Iverson bracket notation

$$\mathbf{I}[C] = \begin{cases} 1 & \text{if } C \text{ is true}, \\ 0 & \text{otherwise}. \end{cases}$$

We use the term *Euclidean space* to refer to $\mathbb{R}^n$ for some positive integer n. We let e_i denote the vector whose ith component is 1 and the others are 0. Thus, the vectors

$e_1, e_2, \ldots, e_n$ form the standard basis for $\mathbb{R}^n$. Generalizing this notation somewhat, we let $\mathbf{1}_S$ denote the characteristic vector of the set S, so that $\mathbf{1}_S = \sum_{i \in S} e_i$. For a linear subspace L, we let $L^\perp$ denote its orthogonal complement.

Set membership notation, when used in the subscript of an expectation operator, indicates that the expectation is taken with respect to a uniformly random element of the indicated set. A generic instance of this notation is $\mathbf{E}_{x \in S} f(x)$, which we will often shorten further to $\mathbf{E}_S f$. We will often omit the argument in equations and inequalities involving functions, as in $\operatorname{sgn} p = (-1)^f$. Relational and arithmetic operators for functions are to be interpreted pointwise. For example, the statement "$f \geqslant 2|g|$ on X" means that $f(x) \geqslant 2|g(x)|$ for every $x \in X$.

2.2. Boolean functions, formulas, and circuits. Throughout this paper, Boolean functions are mappings $X \to \{0,1\}$ for some finite subset X of Euclidean space, most often $X = \{0,1\}^n$. The functions $\mathrm{AND}_n, \mathrm{OR}_n, \mathrm{XOR}_n$ on the Boolean hypercube $\{0,1\}^n$ have their standard definitions: $\mathrm{AND}_n(x) = \bigwedge_{i=1}^n x_i$, $\mathrm{OR}_n(x) = \bigvee_{i=1}^n x_i$, and $\mathrm{XOR}_n(x) = \bigoplus_{i=1}^n x_i$. For a Boolean function f, we let $\neg f$ denote the negation of f. We use the common shorthands $\mathrm{NAND}_n = \neg\mathrm{AND}_n$ and $\mathrm{NOR}_n = \neg\mathrm{OR}_n$. To avoid clutter, we will often omit the floor and ceiling operators when indicating the input length of Boolean functions. For example, $\mathrm{OR}_{\sqrt{n}}$ stands for $\mathrm{OR}_{\lceil\sqrt{n}\rceil}$ or $\mathrm{OR}_{\lfloor\sqrt{n}\rfloor}$, depending on context. A key function in this paper is the *element distinctness* function $\mathrm{ED}_{n,m} \colon \{e_1, e_2, \ldots, e_m\}^n \to \{0,1\}$, defined for $m \geqslant n$ by

$$\mathrm{ED}_{n,m}(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if } x_1, x_2, \ldots, x_n \text{ are pairwise distinct,} \\ 0 & \text{otherwise.} \end{cases}$$

The input to $\mathrm{ED}_{n,m}$ can be viewed as an $m \times n$ Boolean matrix in which every column contains *exactly* one nonzero entry. In that representation, the function evaluates to true if and only if every row contains *at most* one nonzero entry. Observe that $\mathrm{ED}_{n,m}$ is defined on a small subset of the ambient hypercube $\{0,1\}^{nm}$, unlike $\mathrm{AND}_n, \mathrm{OR}_n$, and XOR_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 functions $f \colon X \to \{0,1\}$ and $g \colon Y \to \{0,1\}$, we let $f \wedge g$ stand for the function on $X \times Y$ given by $(f \wedge g)(x,y) = f(x) \wedge g(y)$. The shorthand $f \vee g$ is defined analogously. We often use this notation along with the composition operator, as in $\mathrm{OR}_\ell \circ ((\mathrm{AND}_k \circ \neg f) \wedge g)$. Observe that in our notation, f and $f \wedge f$ are completely different functions, with domain X and $X \times X$, respectively.

For our purposes, a *Boolean circuit*, or equivalently an $\wedge, \vee$*-circuit*, is a circuit with gates $\wedge$ and $\vee$ of unbounded fan-in, with negations allowed at the input gates. In this terminology, the circuit class AC^0 consists of function families $\{f_n\}_{n=1}^\infty$ such that each $f_n \colon \{0,1\}^n \to \{0,1\}$ can be represented by an $\wedge, \vee$-circuit with cn^c gates and depth c, for some constant $c \geqslant 1$ and all n. A *Boolean formula*, or equivalently an $\wedge, \vee$*-formula*, is an $\wedge, \vee$-circuit in which every gate has fan-out 1. Common examples of $\wedge, \vee$-formulas are DNF and CNF formulas. We define *size* somewhat differently for circuits vs. formulas, as the number of gates in the former case and the number of leaf nodes in the latter case. An $\wedge, \vee$-formula is called *read-once* if its leaf nodes correspond to pairwise distinct input variables. In particular, the size of a read-once $\wedge, \vee$-formula never exceeds the number of input variables. We refer to an $\wedge, \vee$-circuit or -formula

$f\colon \{0,1\}^n \to \{0,1\}$ as *explicitly given* if our manuscript provides an algorithm that runs in time $n^{O(1)}$ and produces the representation of f as a circuit or formula.

2.3. Combinatorial identities. For an integer $k \geqslant 0$ and an arbitrary real number α, recall the generalized binomial coefficient

$$\binom{\alpha}{k} = \frac{\alpha}{k} \cdot \frac{\alpha-1}{k-1} \cdot \dots \cdot \frac{\alpha-k+1}{1}.$$

This notation specifically allows $\alpha = 0$ and $\alpha < 0$, both of which arise frequently in this paper. For example,

$$\binom{-1}{k} = (-1)^k \qquad\qquad (k = 1, 2, 3, \dots)$$

and more generally

$$\binom{-n}{k} = (-1)^k \binom{n+k-1}{k} \qquad\qquad (n, k = 1, 2, 3, \dots).$$

We will need the following combinatorial identities.

FACT 2.1.

(i) *For any integer $n \geqslant 1$ and any real polynomial p of degree less than n,*

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} p(i) = 0;$$

(ii) *for any integers $n \geqslant 0$ and $k \geqslant 1$,*

$$\sum_{i=0}^{n} \frac{(-1)^i}{k+i} \binom{n}{i} = \frac{1}{k} \binom{n+k}{k}^{-1};$$

(iii) *for any integers $n \geqslant k \geqslant 0$,*

$$\binom{n}{k} \int_0^1 x^k (1-x)^{n-k}\, dx = \frac{1}{n+1};$$

(iv) *for any integers $n \geqslant k \geqslant 0$,*

$$\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \dots + \binom{n}{k} = \binom{n+1}{k+1}.$$

The first identity, (i), is well-known [16]. The other three are less so, and we provide their short proofs for the reader's convenience.

Proof. (ii) We have

$$\begin{aligned}
\sum_{i=0}^{n} \frac{(-1)^i}{i+k} \binom{n}{i} &= \sum_{i=0}^{n} \frac{(-1)^i (i+k-1)(i+k-2)\cdots(i+1)}{(n+k)(n+k-1)\cdots(n+1)} \binom{n+k}{i+k} \\
&= \sum_{i=-k+1}^{n} \frac{(-1)^i (i+k-1)(i+k-2)\cdots(i+1)}{(n+k)(n+k-1)\cdots(n+1)} \binom{n+k}{i+k} \\
&= \sum_{i=-k}^{n} \frac{(-1)^i (i+k-1)(i+k-2)\cdots(i+1)}{(n+k)(n+k-1)\cdots(n+1)} \binom{n+k}{i+k} + \frac{1}{k} \binom{n+k}{k}^{-1} \\
&= \frac{1}{k} \binom{n+k}{k}^{-1},
\end{aligned}$$

where the final step uses (i).

(iii) Applying the binomial theorem,

$$\begin{aligned}\binom{n}{k}\int_0^1 x^k(1-x)^{n-k}\,dx &= \binom{n}{k}\sum_{i=0}^{n-k}(-1)^i\binom{n-k}{i}\int_0^1 x^{k+i}\,dx\\ &= \binom{n}{k}\sum_{i=0}^{n-k}(-1)^i\binom{n-k}{i}\frac{1}{k+i+1}\\ &= \binom{n}{k}\cdot\frac{1}{k+1}\binom{n-k+(k+1)}{k+1}^{-1}\\ &= \frac{1}{n+1},\end{aligned}$$

where the third step uses (ii).

(iv) This equality has a simple combinatorial interpretation: to choose a size-$(k+1)$ subset $S\subseteq\{1,2,\ldots,n+1\}$, one can first choose an integer m and then choose one of the size-$(k+1)$ subsets $S\subseteq\{1,2,\ldots,n+1\}$ with $\max S=m$. □

2.4. 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:

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

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 *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 $\lambda_1f_1+\lambda_2f_2+\cdots+\lambda_kf_k$, 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.

Throughout this manuscript, we view probability distributions as real functions. This convention makes available the shorthands 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.

2.5. Symmetrization. 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. We let S_n stand for the symmetric group of order n, and define $\sigma x=x_{\sigma(1)}x_{\sigma(2)}\ldots x_{\sigma(n)}$ for any $\sigma\in S_n$ and $x\in\{0,1\}^n$. The following simple but fundamental fact, due to Minsky and Papert [22], allows one to transform a multivariate real polynomial on $\{0,1\}^n$ into a related univariate polynomial on $\{0,1,2,\ldots,n\}$ without an increase in degree.

Proposition 2.2 (Minsky and Papert). *Let $p\colon \{0,1\}^n \to \mathbb{R}$ be an arbitrary polynomial. Then the mapping*

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

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

Minsky and Papert's result has the following multivariate generalization.

Corollary 2.3 (cf. Minsky and Papert). *Let $p\colon (\{0,1\}^n)^m \to \mathbb{R}$ be an arbitrary polynomial. Then there is a polynomial $q\colon \mathbb{R}^m \to \mathbb{R}$ of degree at most $\deg p$ such that*

$$\mathop{\mathbf{E}}_{\sigma_1\in S_n} \mathop{\mathbf{E}}_{\sigma_2\in S_n} \cdots \mathop{\mathbf{E}}_{\sigma_m\in S_n} p(\sigma_1 x_1, \dots, \sigma_m x_m) = q(|x_1|, |x_2|, \dots, |x_m|)$$

for all $x_1, x_2, \dots, x_m \in \{0,1\}^n$.

This generalization follows by induction on m, where the base case $m = 1$ corresponds to Proposition 2.2. For details, see, e.g., [27].

2.6. 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* (equivalently, *ℓ_∞-norm approximant*) *for f with error ϵ*. In the study of Boolean functions, the standard setting of the error parameter is $\epsilon = 1/3$. A related notion is that of *threshold degree*, denoted $\deg_\pm(f)$ and defined as the least degree of a real polynomial p that represents f in sign:

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

It is intuitively clear that sign-representation is a weaker notion than uniform approximation. Formally, we have

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

In particular,

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

The term "threshold degree" appears to be due to Saks [28]. Equivalent terms in the literature include "strong degree" [5], "voting polynomial degree" [20], "polynomial threshold function degree" [24], and "sign degree" [11].

Key to our work is a hybrid of uniform approximation and sign-representation, 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 as a *one-sided approximant for f with error ϵ*. Again, the canonical setting of the error parameter is $\epsilon = 1/3$. The gap between the one-sided approximate degree of a Boolean function $f\colon\{0,1\}^n\to\{0,1\}$ and that of its negation $\neg f$ can be as large as 1 versus $\Omega(\sqrt{n})$, achieved for $f=\mathrm{OR}_n$. In contrast, threshold degree and approximate degree are invariant under 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.

Basic approximation theory allows one to efficiently reduce the error in a uniform or one-sided approximation of a Boolean function. We will only need error reduction in the setting of one-sided approximation, where the analysis is particularly simple.

Fact 2.4. *For any Boolean function $f\colon X\to\{0,1\}$ and any $0\leqslant\epsilon\leqslant 1/2$,*

$$\deg^+_{\frac{\epsilon^k}{\epsilon^k+(1-\epsilon)^k}}(f)\leqslant k\deg^+_\epsilon(f)\qquad\qquad (k=1,2,3,\ldots).$$

Proof. If p is a one-sided approximant for f with error ϵ, then $p^k/(\epsilon^k+(1-\epsilon)^k)$ is a one-sided approximant for f with error $\epsilon^k/(\epsilon^k+(1-\epsilon)^k)$. ∎

Fact 2.4 makes it clear, among other things, that the canonical constant $\epsilon=1/3$ in the definition of one-sided approximate degree can be replaced by any other number in $(0,1/2)$ without changing the model in any significant way.

A natural approach to approximating a composed function $f\circ g$ is to approximate f and g individually 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 [33] gives an optimal procedure for making any polynomial robust to noise in the inputs.

Theorem 2.5 (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 $\epsilon\in[-1/3,1/3]^n$.

2.7. Approximate degree of concrete functions. The most studied Boolean functions in the context of polynomial approximation are OR_n and AND_n. The following seminal theorem, due to Nisan and Szegedy [23], was one of the first results in this line of work.

Theorem 2.6 (Nisan and Szegedy).

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

Buhrman et al. [10] and de Wolf [41] generalized Nisan and Szegedy's theorem to an arbitrary error parameter ϵ. For our purposes, only the upper bound is needed.

Theorem 2.7 (Buhrman et al.; de Wolf). *For $\epsilon\leqslant 1/3$,*

$$\deg_\epsilon(\mathrm{AND}_n)=\deg_\epsilon(\mathrm{OR}_n)=O\left(\sqrt{n\log\frac{1}{\epsilon}}\right).$$

Another extensively studied function in the context of polynomial approximation is element distinctness, $\mathrm{ED}_{n,m}$. It has played an important role in quantum query complexity [1, 2] and more recently in the study of constant-depth circuits [13, 35]. The following lower bound is due to Ambainis [2].

Theorem 2.8 (Ambainis).

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

We note that Theorem 2.8 is tight, with a matching upper bound obtained also by Ambainis [3]. Bun and Thaler [13] recently showed, with a short and elegant proof, that Ambainis's lower bound on the approximate degree of element distinctness carries over to the one-sided setting:

Theorem 2.9 (Bun and Thaler).

$$\deg^{+}_{1/3}(\mathrm{ED}_{n,n}) = \Omega(n^{2/3}).$$

For the reader's convenience, we include the proof.

Proof (adapted from [13]). Bun and Thaler define $\mathrm{ED}_{n,n}$ somewhat differently, with domain $(\{0,1\}^{\log n})^n$ rather than $\{e_1, e_2, \ldots, e_n\}^n$. However, their proof works equally well in both settings. Details follow.

Let $0 \leqslant \epsilon < 1/2$ be arbitrary, and let $p\colon (\mathbb{R}^n)^n \to \mathbb{R}$ be a one-sided approximant for $\mathrm{ED}_{n,n}$ with error ϵ. Symmetrize p to obtain a new polynomial of the same or smaller degree,

$$q(x_1, x_2, \ldots, x_n) = \mathop{\mathbf{E}}_{\sigma \in S_n} p(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}).$$

Recall that $\mathrm{ED}_{n,n}(x_1, x_2, \ldots, x_n) = \mathrm{ED}_{n,n}(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$ for all $\sigma \in S_n$. Hence, the symmetrized polynomial q is also a valid one-sided approximant for this function, with $|q| \leqslant \epsilon$ on $\mathrm{ED}_{n,n}^{-1}(0)$ and $q \geqslant 1 - \epsilon$ on $\mathrm{ED}_{n,n}^{-1}(1)$. The key observation is that $\mathrm{ED}_{n,n}^{-1}(1) = \{(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}) : \sigma \in S_n\}$ and therefore the symmetrized polynomial q is constant on $\mathrm{ED}_{n,n}^{-1}(1)$. By normalizing q, we obtain a uniform approximant for $\mathrm{ED}_{n,n}$:

$$\left\| \mathrm{ED}_{n,n} - \frac{1-\epsilon}{q(e_1, e_2, \ldots, e_n)} \cdot q \right\|_\infty \leqslant \epsilon.$$

We conclude that $\deg_\epsilon(\mathrm{ED}_{n,n}) = \deg^{+}_\epsilon(\mathrm{ED}_{n,n})$ for all ϵ, which completes the proof in view of Theorem 2.8. ∎

2.8. Dual characterizations. Approximate degree, one-sided approximate degree, and threshold degree each have an exact dual characterization, obtained by an appeal to linear programming duality. We will only need the dual characterization for one-sided approximate degree, due to Bun and Thaler [13].

Theorem 2.10 (Bun and Thaler). *Let $f\colon X \to \{0,1\}$ be given, $d > 0$. Then $\deg^{+}_\epsilon(f) \geqslant d$ if and only if there exists $\psi\colon X \to \mathbb{R}$ 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 $f(x) = 1$.*

Recent papers refer to the function ψ in Theorem 2.10 as a *dual object* or *dual polynomial.* Its role is to serve as an explicit witness to the fact that f has one-sided approximate degree larger than a given number. It will be convenient to specialize this result to the negated function $\neg f$:

Corollary 2.11. *Let $f\colon X \to \{0,1\}$ be given, $d > 0$. Then $\deg_\epsilon^+(\neg f) \geqslant d$ if and only if there exists $\psi\colon X \to \mathbb{R}$ 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) $\psi(x) \leqslant 0$ *whenever $f(x) = 0$.*

Proof. Theorem 2.10 ensures the existence of $\psi'\colon X \to \mathbb{R}$ with

(i) $\langle 1 - f, \psi'\rangle > \epsilon\|\psi'\|_1$,
(ii) $\langle \psi', p\rangle = 0$ for every polynomial p of degree less than d, and
(iii) $\psi'(x) \geqslant 0$ whenever $f(x) = 0$.

Property (ii) implies in particular that $\langle 1, \psi'\rangle = 0$, whence $-\langle f, \psi'\rangle > \epsilon\|\psi'\|_1$ by (i). The proof is now complete by letting $\psi = -\psi'$. ◻

The dual objects that arise in Theorem 2.10 and its corollary share the following metric properties [35].

Proposition 2.12. *Let $\psi\colon X \to \mathbb{R}$ 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 $f\colon X \to \{0,1\}$.*

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 $\{0,1\}$. ◻

It is often necessary to have a dual object with additional properties, beyond what linear programming duality guarantees. In such cases the dual object must be constructed from first principles. In this paper, we will need such a specially constructed dual object for the OR function. Previous constructions due to Špalek [40], Bun and Thaler [12], and the author [35] do not provide all the properties that we require.

Theorem 2.13. *Let ϵ be given, $0 < \epsilon < 1$. Then for every $n \geqslant 2$ and every probability distribution κ on $\{1,2,\ldots,n\}$, there is an (explicitly given) function $\omega\colon \{0,1,2,\ldots,n\} \to \mathbb{R}$ such that*

$$\begin{aligned}
&\omega(0) > \frac{1-\epsilon}{2} \cdot \|\omega\|_1, \\
&(-1)^{n+t}\omega(t) \geqslant \frac{\epsilon\kappa(t)}{3} \cdot \|\omega\|_1 \qquad\qquad (t = 1,2,\ldots,n), \\
&\deg p < \sqrt{\delta n} \implies \langle \omega, p\rangle = 0,
\end{aligned}$$

where $\delta = \delta(\epsilon) > 0$ is a constant independent of n.

Theorem 2.13 generalizes a result from previous work [35, Theorem 2.8], which in our notation corresponds to the special case

$$\kappa = \left(\frac{c}{1}, \frac{c}{2^2}, \frac{c}{3^2}, \ldots, \frac{c}{n^2}\right)$$

for a normalizing factor $c \sim 6/\pi^2$. We provide the proof of Theorem 2.13 in Appendix B.

We will also need a special kind of dual object for the high-accuracy approximation of OR_n, in contrast to the bounded-error regime of the previous theorem.

Theorem 2.14. *Let $0 < c < 1$ be a sufficiently small absolute constant. Then for all integers n, r with $1 \leqslant r \leqslant n/2$, there exists a function $\nu \colon \{0, 1, 2, \ldots, n\} \to \mathbb{R}$ such that*

$$\begin{aligned}
&\deg p \leqslant c\sqrt{nr} \implies \langle \nu, p\rangle = 0, && \\
&\nu(t) = 0 && (t = 1, 2, \ldots, r-1), \\
&(-1)^{n+t}\nu(t) \geqslant 0 && (t = 1, 2, \ldots, n), \\
&\nu(0) > c^r \|\nu\|_1, && \\
&|\nu(t)| \geqslant c^n \|\nu\|_1 && (t > n/2), \\
&|\nu(t)| \leqslant \left(\frac{r}{ct}\right)^r \|\nu\|_1 && (t = 1, 2, \ldots, n).
\end{aligned}$$

This theorem is a modification of an earlier result due to Bun and Thaler [12], who constructed a dual object for the *bounded-error* approximation of any given symmetric function. It is straightforward to verify that their construction is also a dual object for OR in the high-accuracy regime of interest to us, and we need only adapt it to ensure the additional metric properties and sign behavior. We provide a complete proof of Theorem 2.14 in Appendix C.

3. A hard CNF formula.

The technical centerpiece of this paper, developed in sections 4 to 8, is a technique that transforms any given constant-depth $\wedge, \vee$-circuit with high one-sided approximate degree into a constant-depth $\wedge, \vee$-circuit with proportionately high threshold degree. This section focuses on constructing the former object, a circuit of polynomial size and small depth (in fact, a CNF formula) with high one-sided approximate degree. Prior to our work, the strongest lower bound on the one-sided approximate degree of a polynomial-size CNF formula in n variables was $\Omega(n/\log n)^{2/3}$, due to Bun and Thaler [13]. Here, we obtain a logarithmically stronger bound of $\omega(n^{2/3})$. We pursue this quantitative improvement solely for aesthetic reasons, although the technique in question seems quite general and is likely relevant to other problems in polynomial approximation and quantum query complexity (see Remark 3.4 for details). The reader can skip this section on a first reading and proceed with the rest of the development in section 4 without loss of continuity.

3.1. The role of the input encoding.

When studying the approximation of Boolean functions by polynomials, one most often considers functions defined on the entire hypercube. A notable departure from this convention is a line of research in quantum query complexity that studies Boolean functions on the set $\{e_1, e_2, \ldots, e_m\}^n$ for appropriate integers n and m, which is a tiny subset of the ambient hypercube $\{0,1\}^{nm}$. For example, the function $\mathrm{SURJ}_{n,m} \colon \{e_1, e_2, \ldots, e_m\}^n \to \{0,1\}$ is defined for $m \leqslant n$ by

$$\mathrm{SURJ}_{n,m}(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if } \{x_1, x_2, \ldots, x_n\} = \{e_1, e_2, \ldots, e_m\}, \\ 0 & \text{otherwise.} \end{cases}$$

If we interpret the input as encoding a mapping $i \mapsto x_i$, then $\mathrm{SURJ}_{n,m}$ evaluates to true precisely when the input represents a surjection. For this reason, $\mathrm{SURJ}_{n,m}$

is known as the *surjection problem* [8]. It would be logical to call its counterpart the "injection problem," but instead it is better known as *element distinctness*. As the reader will recall from section 2, the element distinctness function $\mathrm{ED}_{n,m}\colon \{e_1, e_2, \ldots, e_m\}^n \to \{0,1\}$ is defined for $m \geqslant n$ by

$$\mathrm{ED}_{n,m}(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if } x_1, x_2, \ldots, x_n \text{ are pairwise distinct,} \\ 0 & \text{otherwise.} \end{cases}$$

The choice of $\{e_1, e_2, \ldots, e_m\}^n$ as the domain for these functions has to do with the manner in which a quantum query algorithm accesses the input bits. Our applications are concerned with the parameter setting $m = n$, in which case these two functions are the same: $\mathrm{SURJ}_{n,n} = \mathrm{ED}_{n,n}$. Their study in this paper and previous work [8, 13, 35] is motivated by the fact that they are efficiently representable as constant-depth $\wedge, \vee$-circuits.

From the standpoint of applications, a technical obstacle arises due to the inefficient encoding of the input in the definition of these functions. To illustrate, consider the element distinctness function $\mathrm{ED}_{n,n}$. Recall from Theorem 2.8 due to Ambainis [2] that $\deg_{1/3}(\mathrm{ED}_{n,n}) = \Omega(n^{2/3})$. At first glance, this lower bound on the approximate degree seems clearly superior to the $\Omega(\sqrt{n})$ lower bound for the OR_n function. One quickly realizes, however, that the input to the former is much larger in terms of bit length, with n^2 bits for $\mathrm{ED}_{n,n}$ versus n bits for OR_n. As a result, the lower bound for $\mathrm{ED}_{n,n}$ is only a cube root of the input length, versus a square root for OR_n. It is the wasteful encoding of the input to $\mathrm{ED}_{n,n}$ that artificially weakens the otherwise strong lower bound on its approximate degree.

To overcome this obstacle, it is necessary to encode the input to $\mathrm{ED}_{n,n}$ more efficiently. The folklore approach [8, 13, 35] is to work with the composition $\mathrm{ED}_{n,n} \circ \iota$, where the gadget $\iota\colon \{0,1\}^{\log n} \to \{e_1, e_2, \ldots, e_n\}$ is the canonical one-to-one correspondence. The input to $\mathrm{ED}_{n,n} \circ \iota$ has bit length $n \log n$, whereas its approximate degree is easily seen to satisfy $\deg_{1/3}(\mathrm{ED}_{n,n} \circ \iota) \geqslant \deg_{1/3}(\mathrm{ED}_{n,n}) = \Omega(n^{2/3})$. Thus, $\mathrm{ED}_{n,n} \circ \iota$ is a function on $N = n \log n$ variables with approximate degree $\Omega(N/\log N)^{2/3}$, a significant improvement. In fact, prior to our paper, this was the strongest lower bound on the approximate degree of a polynomial-size CNF formula [2, 13].

3.2. The new gadget. The folklore construction just described is still unsatisfying in that the gadget ι counts toward the bit length of the input but does not contribute to the function's approximate degree, which weakens the resulting lower bound. Here, we present an alternate construction that entirely eliminates this inefficiency. Specifically, we construct a gadget $\phi\colon \{0,1\}^{6\lceil \log m \rceil} \to \{e_1, e_2, \ldots, e_m\}$ such that

$$\deg_{1/3}(f \circ \phi) \geqslant \deg_{1/3}(f) \lceil 1 + \log m \rceil \tag{3.1}$$

for every function $f\colon \{e_1, e_2, \ldots, e_m\}^n \to \{0,1\}$. Thus, the new gadget actually increases the approximate degree by a factor proportional to the gadget's size. This contrasts with the folklore gadget ι, which only guarantees $\deg_{1/3}(f \circ \iota) \geqslant \deg_{1/3}(f)$. The lower bound (3.1) carries over to other approximation-theoretic measures such as one-sided approximate degree and threshold degree, as well as to partial functions on $\{e_1, e_2, \ldots, e_m\}^n$.

Our construction of ϕ is based exclusively on elementary linear algebra. Its crux is the following first-principles lemma.

Lemma 3.1. *For every integer $m \geqslant 2$, there is a surjection $\phi\colon \{0,1\}^{6\lceil \log m\rceil} \to \{e_1, e_2, \ldots, e_m\}$ such that*

$$\mathop{\mathbf{E}}_{\phi^{-1}(e_i)} p = \mathop{\mathbf{E}}_{\{0,1\}^{6\lceil \log m\rceil}} p \qquad (i = 1, 2, \ldots, m) \tag{3.2}$$

for every polynomial $p\colon \{0,1\}^{6\lceil \log m\rceil} \to \mathbb{R}$ of degree at most $\lceil \log m\rceil$. Moreover, ϕ can be constructed deterministically in time $m^{O(1)}$.

Proof. We will abbreviate $k = \lceil \log m\rceil$ and identify $\{0,1\}$ throughout the proof with the two-element field $\mathbb{F}_2$. We start by constructing vectors $v_1, v_2, \ldots, v_{6k} \in \mathbb{F}_2^{5k}$ among which any k are linearly independent, using a greedy algorithm. Assume that $v_1, v_2, \ldots, v_{i-1}$ have already been constructed, and we need to find a vector v_i that is not contained in the span of fewer than k vectors from among $v_1, v_2, \ldots, v_{i-1}$. The union of all such spans has size at most $\binom{6k}{k-1}2^{k-1} < 2^{5k}$. Thus, it suffices to deterministically enumerate these offending possibilities in time $2^{O(k)}$ and take v_i to be one of the remaining vectors in $\mathbb{F}_2^{5k}$. This completes the construction.

Now, let $L \subset \mathbb{F}_2^{6k}$ be the row span of the matrix with columns $v_1, v_2, \ldots, v_{6k}$. By linear algebra, the linear independence of any k vectors from among $v_1, v_2, \ldots, v_{6k}$ implies that the coordinates of a uniformly random vector in L are k-wise independent and distributed uniformly in $\mathbb{F}_2$. Let $X_1, X_2, X_3, \ldots$ denote the distinct cosets of L in the ambient linear space $\mathbb{F}_2^{6k}$. There are $2^{6k-\dim L} \geqslant 2^k$ such cosets. We let ϕ be the surjection that sends $X_i \mapsto e_{\min\{i,m\}}$ for $i = 1, 2, 3, \ldots$. The k-wise independence property for a uniformly random vector of L is inherited by any translate of L, whence (3.2) for every real polynomial of degree at most k. ∎

The construction of ϕ in the previous lemma can be made more efficient with regard to running time using coding theory. However, this efficiency improvement is irrelevant for our purposes because the input length in our applications will be polynomial in m, making the running time in Lemma 3.1 efficient to start with. We have reached the main technical result of this section.

Theorem 3.2. *Let $m \geqslant 2$ be a given integer. Then there exists a function $\phi\colon \{0,1\}^{6\lceil \log m\rceil} \to \{e_1, e_2, \ldots, e_m\}$, constructible in time $m^{O(1)}$, such that*

$$\deg_\epsilon^+(f \circ \phi) \geqslant \deg_\epsilon^+(f)\lceil 1 + \log m\rceil, \tag{3.3}$$

$$\deg_\epsilon(f \circ \phi) \geqslant \deg_\epsilon(f)\lceil 1 + \log m\rceil \tag{3.4}$$

for every ϵ and every (possibly partial) Boolean function f on $\{e_1, e_2, \ldots, e_m\}^n$.

Passing to the limit in (3.4) as $\epsilon \nearrow 1/2$ gives an analogous conclusion for threshold degree: $\deg_\pm(f \circ \phi) \geqslant \deg_\pm(f)\lceil 1 + \log m\rceil$. We will not need this additional bound, however.

Proof of Theorem 3.2. Let ϕ be the surjection constructed in Lemma 3.1. Fix n arbitrarily and consider the following averaging operator A that linearly sends every function $p\colon (\{0,1\}^{6\lceil \log m\rceil})^n \to \mathbb{R}$ to a function $Ap\colon \{e_1, e_2, \ldots, e_m\}^n \to \mathbb{R}$, according to

$$(Ap)(x_1, x_2, \ldots, x_n) = \mathop{\mathbf{E}}_{\phi^{-1}(x_1)\times\phi^{-1}(x_2)\times\cdots\times\phi^{-1}(x_n)} p.$$

The intuition behind this definition is straightforward: if p is a one-sided approximant for $f \circ \phi$ with error ϵ, then Ap is a one-sided approximant for f with the same error ϵ (and likewise for ℓ_∞-norm approximation). Therefore, the proof will be complete once we show that

$$\deg Ap \leqslant \frac{\deg p}{\lceil \log m \rceil + 1} \tag{3.5}$$

for every real polynomial p.

By the linearity of A, it suffices to prove (3.5) for factored polynomials of the form $p(x_1, x_2, \ldots, x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n)$, where $p_1, p_2, \ldots, p_n$ are real polynomials on $\{0,1\}^{6\lceil \log m \rceil}$. Then the defining equation simplifies to

$$(Ap)(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \mathop{\mathbf{E}}_{\phi^{-1}(x_i)} p_i.$$

We now examine the individual contributions of $p_1, p_2, \ldots, p_n$ to the degree of Ap as a real polynomial. For any polynomial p_i of degree at most $\lceil \log m \rceil$, Lemma 3.1 ensures that the corresponding expectation $\mathbf{E}_{\phi^{-1}(x_i)}\, p_i$ is a constant independent of the input x_i. Thus, polynomials p_i of degree at most $\lceil \log m \rceil$ do not contribute to the degree of Ap. For the other polynomials p_i, the expectation $\mathbf{E}_{\phi^{-1}(x_i)}\, p_i$ is a linear polynomial in x_i, namely,

$$\mathop{\mathbf{E}}_{\phi^{-1}(x_i)} p_i = x_{i,1} \mathop{\mathbf{E}}_{\phi^{-1}(e_1)} p_i + x_{i,2} \mathop{\mathbf{E}}_{\phi^{-1}(e_2)} p_i + \cdots + x_{i,m} \mathop{\mathbf{E}}_{\phi^{-1}(e_m)} p_i,$$

where we are crucially exploiting the fact that the input x_i is a vector in the restricted set $\{e_1, e_2, \ldots, e_m\}$ rather than an arbitrary vector in $\{0,1\}^m$. Thus, polynomials p_i of degree greater than $\lceil \log m \rceil$ contribute at most 1 each to the degree of Ap. In summary, we have shown that $\deg Ap \leqslant |\{i : \deg p_i \geqslant \lceil \log m \rceil + 1\}|$, which immediately implies (3.5). ◻

3.3. The CNF construction.

By applying Theorem 3.2 to the element distinctness function, we will now obtain an explicit CNF formula $F\colon \{0,1\}^N \to \{0,1\}$ of polynomial size with one-sided approximate degree $\deg^+_{1/3}(F) = \omega(N^{2/3})$. This lower bound improves on all previous lower bounds for CNF formulas [13] and matches up to a polylogarithmic factor all known lower bounds for $\wedge, \vee$-circuits of arbitrary constant depth [25]. Details follow.

THEOREM 3.3. *Consider the function $F\colon \{0,1\}^N \to \{0,1\}$ on $N = 6\lceil \log n \rceil n$ variables given by*

$$F = \mathrm{ED}_{n,n} \circ \phi,$$

where $\phi\colon \{0,1\}^{6\lceil \log n \rceil} \to \{e_1, e_2, \ldots, e_n\}$ is as constructed in Theorem 3.2. Then F is computable by a CNF formula of polynomial size and satisfies

$$\deg^+_{1/3}(F) = \Omega(N^{2/3} \log^{1/3} N).$$

Proof. Recall from Theorem 2.9 that $\deg^+_{1/3}(\mathrm{ED}_{n,n}) = \Omega(n^{2/3})$. As a result, Theorem 3.2 shows that

$$\begin{aligned} \deg^+_{1/3}(F) &= \Omega(n^{2/3} \log n) \\ &= \Omega(N^{2/3} \log^{1/3} N). \end{aligned}$$

It remains to verify that F is computable by a polynomial-size CNF formula. By definition,

$$\begin{aligned} F(x_1, x_2, \ldots, x_n) &= \mathrm{ED}_{n,n}(\phi(x_1), \phi(x_2), \ldots, \phi(x_n)) \\ &= \bigwedge_{i \neq j} (\phi(x_i) \neq \phi(x_j)). \end{aligned}$$

Each of the predicates $\phi(x_i) \neq \phi(x_j)$ in this expression features only $12\lceil \log n \rceil$ Boolean variables and can therefore be computed by a CNF formula with $O(n^{12})$ clauses. The conjunction of these CNF formulas for all pairs of distinct i and j gives the desired polynomial-size CNF formula for F. ◻

Remark 3.4. The technique of Theorem 3.2 seems quite general and is likely relevant to other problems where a logarithmic gap arises due to the input encoding. In this paper, we have focused on the application to polynomial approximation (Theorem 3.3). Another application is to quantum query complexity, as follows. Recently, Beame and Machmouchi [8] proved a lower bound of $\Omega(n/\log n)$ on the quantum query complexity of a function in AC^0, improving on the previous bound of $\Omega(n/\log n)^{2/3}$ due to Ambainis [2]. The logarithmic factor in both cases is due to the use of the folklore gadget ι, which counts toward the function's input length but does not contribute to its query complexity. We are confident that by using our gadget ϕ instead of ι, one can eliminate the logarithmic factors in previous work [2, 8] and obtain a tight lower bound of $\Omega(n)$ on the quantum query complexity of AC^0, answering Beame and Machmouchi's question.

4. One-sided rational approximation.

We now review one-sided rational approximation of Boolean functions, studied recently in [35]. Let $f\colon X \to \{0,1\}$ be a Boolean function of interest, $d_0, d_1 \geqslant 0$ given reals. Following [35], we define $R(f, d_0, d_1)$ as the infimum over all $\epsilon > 0$ for which there exist 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) $\deg p_1 \leqslant d_1$.

Observe that $R(f, d_0, d_1)$ is always well-defined and ranges in $[0,1]$. This quantity formalizes one-sided approximation of f by rational functions in that the quotient p_1/p_0 is close to zero on $f^{-1}(0)$ and far from zero on $f^{-1}(1)$:

$$\left|\frac{p_1}{p_0}\right| \in \begin{cases} [0,\epsilon) & \text{on } f^{-1}(0), \\ (\frac{1}{\epsilon}, +\infty] & \text{on } f^{-1}(1). \end{cases} \tag{4.1}$$

4.1. Examples and key facts.

To illustrate, consider the familiar functions OR_n and AND_n with domain $X = \{0,1\}^n$. For any $\epsilon > 0$, we have $R(\mathrm{OR}_n, 0, 1) < \epsilon$ by taking $p_1(x) = x_1 + x_2 + \cdots + x_n$ and $p_0(x) = \epsilon/2$ in the definition above. Passing to the limit, we conclude that

$$R(\mathrm{OR}_n, 0, 1) = 0. \tag{4.2}$$

An analogous argument shows that

$$R(\mathrm{AND}_n, 1, 0) = 0. \tag{4.3}$$

Furthermore, it is straightforward to see that OR_n and AND_n have ℓ_∞-norm rational approximants of degree 1 with error arbitrarily close to 0. Indeed,

$$\lim_{\epsilon \searrow 0} \left\| \mathrm{AND}_n - \frac{\epsilon}{\epsilon + \sum(1 - x_i)} \right\|_\infty = 0, \tag{4.4}$$

$$\lim_{\epsilon \searrow 0} \left\| \mathrm{OR}_n - \frac{\sum x_i}{\epsilon + \sum x_i} \right\|_\infty = 0. \tag{4.5}$$

These results on rational approximation should be contrasted with Theorem 2.6, which states that approximating the AND_n function even in the one-sided sense requires a polynomial of degree $\Omega(\sqrt{n})$.

As one might expect, the constructions in (4.4) and (4.5) are helpful in analyzing formulas of greater depth as well. Specifically, it turns out that any read-once formula $f\colon \{0,1\}^n \to \{0,1\}$ of depth 2 satisfies

$$R\left(f, t, \frac{n}{t}\right) = 0 \qquad\qquad (0 < t < \infty).$$

We give a detailed proof of this fact, for use in later sections.

Lemma 4.1. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be an arbitrary read-once formula of depth 2. Then for any $\epsilon > 0$ and $t > 0$, there is a one-sided rational approximant $R_{\epsilon,t}$ for f with a positive denominator of degree at most t, a nonnegative numerator of degree at most n/t, and error ϵ.*

Proof. There are two cases to consider, depending on the top gate of f. If $f = \mathrm{OR}_{n_1} \wedge \mathrm{OR}_{n_2} \wedge \cdots \wedge \mathrm{OR}_{n_r}$ for some integers $n_1, n_2, \ldots, n_r$ with $\sum n_i \leqslant n$, then the desired approximant is

$$R_{\epsilon,t}(x_1, x_2, \ldots, x_r) = \frac{\epsilon \prod_{i:n_i>t} \sum_{j=1}^{n_i} x_{i,j}}{\epsilon + n^n \sum_{i:n_i \leqslant t} \mathrm{NOR}_{n_i}(x_i)},$$

where we view NOR_{n_i} as a degree-n_i real polynomial. In the complementary case $f = \mathrm{AND}_{n_1} \vee \mathrm{AND}_{n_2} \vee \cdots \vee \mathrm{AND}_{n_r}$, the approximant is

$$R_{\epsilon,t}(x_1, x_2, \ldots, x_r) = \frac{\epsilon + n^n \sum_{i:n_i \leqslant n/t} \mathrm{AND}_{n_i}(x_i)}{\epsilon + \prod_{i:n_i>n/t} \sum_{j=1}^{n_i}(1 - x_{i,j})},$$

where we similarly view AND_{n_i} as a degree-n_i real polynomial. □

Analogous to polynomial approximation, we have the following efficient procedure for reducing the error in a one-sided rational approximant.

Proposition 4.2. *For all $d_0, d_1 \geqslant 0$ and every Boolean function $f\colon X \to \{0,1\}$,*

$$R(f, kd_0, kd_1) \leqslant R(f, d_0, d_1)^k \qquad\qquad (k = 1, 2, 3, \ldots).$$

Proof. Let p_0, p_1 be any polynomials of degree at most d_0, d_1, respectively, such that $|p_1| < \epsilon p_0$ on $f^{-1}(0)$ and $|p_0| < \epsilon p_1$ on $f^{-1}(1)$. Then clearly $|p_1^k| < \epsilon^k p_0^k$ on $f^{-1}(0)$ and $|p_0^k| < \epsilon^k p_1^k$ on $f^{-1}(1)$. □

Another aesthetically pleasing feature of the above formalism of rational approximation is the ease of switching between a function and its negation:

Proposition 4.3. *For all $d_0, d_1 \geqslant 0$ and every $f\colon X \to \{0,1\}$,*

$$R(\neg f, d_0, d_1) = R(f, d_1, d_0).$$

Proof. Immediate from the definition. □

4.2. Relation to sign-representation. Our interest in rational approximation is motivated by its central role in constructing sign-representing polynomials. In particular, rational approximation allows for a complete and elegant characterization of the threshold degree of every composition of the form $\mathrm{OR}_\ell \circ f$. The upper bound on the threshold degree of $\mathrm{OR}_\ell \circ f$ in terms of rational approximation was discovered by Beigel et al. [9] in their breakthrough paper on the closure of PP under intersection. Several variations and reformulations of that upper bound have been obtained in subsequent work [17, 32, 34]. The tightest and most recent version is as follows [35], stated in the terminology of this paper.

Theorem 4.4 (cf. Beigel et al.). *Let $f\colon X \to \{0,1\}$ be given. Then for all ℓ,*

$$\deg_\pm(\mathrm{OR}_\ell \circ f) \leqslant 2 \min_{d_0,d_1} \left\{ \ell d_0 + d_1 : R(f,d_0,d_1) < \frac{1}{\ell^{1/4}} \right\}. \tag{4.6}$$

In particular,

$$\deg_\pm(\mathrm{OR}_\ell \circ f) \leqslant 2 \min_{d_0,d_1} \left\{ \ell d_0 + d_1 : R(f,d_0,d_1) < \frac{1}{2} \right\} \left\lceil \frac{\log \ell}{4} \right\rceil. \tag{4.7}$$

Proof (cf. [9, 17]). Fix arbitrary polynomials p_0, p_1 of degree at most d_0, d_1, respectively, such that $p_0 > \ell^{1/4}\,|p_1|$ on $f^{-1}(0)$ and $p_1 > \ell^{1/4}\,|p_0|$ on $f^{-1}(1)$. By perturbing p_0 if necessary, we may assume that p_0 does not vanish on the domain of f. As a result,

$$\frac{p_1^2}{\sqrt{\ell}\,p_0^2} < \frac{1}{\ell} \qquad \text{on } f^{-1}(0),$$

$$\frac{p_1^2}{\sqrt{\ell}\,p_0^2} > 1 \qquad \text{on } f^{-1}(1).$$

Then

$$\operatorname{sgn}\left(\sum_{i=1}^{\ell} \frac{p_1(x_i)^2}{\sqrt{\ell}\,p_0(x_i)^2} - 1 \right) = \begin{cases} -1 & \text{on } (\mathrm{OR}_\ell \circ f)^{-1}(0), \\ 1 & \text{on } (\mathrm{OR}_\ell \circ f)^{-1}(1). \end{cases}$$

Multiplying the expression in parentheses by the positive quantity $\prod p_0(x_i)^2$ gives a sign-representing polynomial for $\mathrm{OR}_\ell \circ f$ of degree at most $2\ell d_0 + 2d_1$, namely,

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

This settles (4.6), which in turn implies (4.7) in light of Proposition 4.2. ∎

It was recently shown in [35] that Theorem 4.4 is optimal up to a logarithmic factor, an unexpected finding given the construction's simplicity. Specifically, we have the following matching lower bound on the threshold degree of $\mathrm{OR}_\ell \circ f$ in terms of one-sided rational approximation [35, Theorem 6.7].

Theorem 4.5 (Sherstov). *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*

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

4.3. A dual characterization. An essential property of $R(f,d_0,d_1)$ for our purposes is that it admits an exact and intuitive dual characterization. To start with, an appeal to linear programming duality reveals the following fact [35, Theorem 6.4].

Theorem 4.6 (Sherstov). *Let $f\colon X \to \{0,1\}$ be a given 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 $\phi_0, \phi_1\colon X \to \mathbb{R}$ such that

(v) $\phi_0 \geqslant \epsilon|\phi_1|$ *on* $f^{-1}(0)$,
(vi) $\phi_1 \geqslant \epsilon|\phi_0|$ *on* $f^{-1}(1)$,
(vii) $\deg p \leqslant d_0 \implies \langle \phi_0, p\rangle = 0$,
(viii) $\deg p \leqslant d_1 \implies \langle \phi_1, p\rangle = 0$,
(ix) $\phi_0 \not\equiv 0$,
(x) $\phi_1 \not\equiv 0$.

As an immediate corollary, we have the following dual characterization of one-sided rational approximation [35, Corollary 6.5].

Corollary 4.7 (Sherstov). *Let $f\colon X \to \{0,1\}$ be given with $R(f, d_0, d_1) > 0$. Then $R(f, d_0, d_1)$ is the supremum over all $\epsilon > 0$ for which there exist $\phi_0, \phi_1\colon X \to \mathbb{R}$ with*

(i) $\phi_0 \geqslant \epsilon|\phi_1|$ *on* $f^{-1}(0)$,
(ii) $\phi_1 \geqslant \epsilon|\phi_0|$ *on* $f^{-1}(1)$,
(iii) $\deg p \leqslant d_0 \implies \langle \phi_0, p\rangle = 0$,
(iv) $\deg p \leqslant d_1 \implies \langle \phi_1, p\rangle = 0$,
(v) $\phi_0 \not\equiv 0$,
(vi) $\phi_1 \not\equiv 0$.

5. Hybrid rational approximation.

We now introduce a hybrid notion of approximation by rational functions, which seamlessly interpolates between ℓ_∞-norm and one-sided approximation and plays a key role in this paper. Fix $d_0, d_1 \geqslant 0$ and a Boolean function $f\colon X \to \{0,1\}$. For $\Delta > 1$, we define $R_\Delta(f, d_0, d_1)$ as the infimum over all $\epsilon > 0$ for which there exist polynomials p_0, p_1 such that

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

It is instructive to compare this formalism with one-sided rational approximation, embodied by the quantity $R(f, d_0, d_1)$ from section 4. What sets them apart is item (ii), which in the case of one-sided rational approximation reads "$|p_0| < \epsilon p_1$ on $f^{-1}(1)$." A moment's reflection shows that the feasibility of (i)–(iv) for hybrid rational approximation is monotonic in ϵ, in the sense that increasing ϵ can only make it easier to satisfy (i)–(iv). As a result, $R_\Delta(f, d_0, d_1)$ is always well-defined and ranges in $[0, 1]$.

5.1. Relation to one-sided approximation.

The quotient of the polynomials in the above definition obeys

$$\left|\frac{p_1}{p_0}\right| \in \begin{cases} [0, \epsilon) & \text{on } f^{-1}(0), \\ (\frac{1}{\epsilon}, \frac{\Delta}{\epsilon}) & \text{on } f^{-1}(1). \end{cases} \tag{5.1}$$

It is helpful to contrast (5.1) with its counterpart (4.1) for one-sided rational approximation. Simply put, $R_\Delta(f, d_0, d_1)$ formalizes the approximation of f by rational functions whereby the approximant is close to zero on $f^{-1}(0)$ and is "large but not too large" on $f^{-1}(1)$. As Δ ranges in $(1,+\infty)$, this new formalism monotonically interpolates between ℓ_∞-norm approximation ($\Delta\approx 1$) and one-sided approximation ($\Delta\to+\infty$). In particular, we have:

Theorem 5.1. *Let $f\colon X\to\{0,1\}$ be given. Then for all $d_0,d_1\geqslant 0$,*

$$R(f,d_0,d_1)\leqslant \lim_{\Delta\to+\infty} R_\Delta(f,d_0,d_1), \tag{5.2}$$

$$R(f,d_0,d_1)\geqslant R(f,d_0,d_1)^2\geqslant \lim_{\Delta\to+\infty} R_\Delta(f,2d_0,2d_1). \tag{5.3}$$

Proof. For any pair of polynomials p_0,p_1 and any $\Delta>1$, the conditions

$$\begin{aligned}&|p_1|<\epsilon p_0 \text{ on } f^{-1}(0),\\ &p_0\in(\tfrac{\epsilon}{\Delta}p_1,\epsilon p_1) \text{ on } f^{-1}(1)\end{aligned}$$

trivially imply

$$\begin{aligned}&|p_1|<\epsilon p_0 \text{ on } f^{-1}(0),\\ &|p_0|<\epsilon p_1 \text{ on } f^{-1}(1),\end{aligned}$$

which proves (5.2). Conversely, fix $\epsilon>0$ and polynomials p_0,p_1 that obey the last two equations. By perturbing p_0 if necessary, we may assume that p_0 does not vanish on the domain of f. Taking

$$M>\epsilon^2\max_{x\in X}\frac{p_1(x)^2}{p_0(x)^2},$$

we arrive at

$$\begin{aligned}&p_1^2<\epsilon^2p_0^2 \text{ on } f^{-1}(0),\\ &p_0^2\in\left(\frac{\epsilon^2}{M}p_1^2,\epsilon^2p_1^2\right) \text{ on } f^{-1}(1),\end{aligned}$$

which yields $\lim_{\Delta\to+\infty}R_\Delta(f,2d_0,2d_1)\leqslant R(f,d_0,d_1)^2$. This directly implies (5.3) since $R(f,d_0,d_1)\in[0,1]$. □

5.2. A dual characterization. Hybrid rational approximation admits an exact dual characterization. It is helpful to compare the theorem that follows with its earlier counterpart for one-sided rational approximation (Theorem 4.6).

Theorem 5.2. *Let $f\colon X\to\{0,1\}$ be a given Boolean function, $\epsilon>0$, and $\Delta>1$. Then for all $d_0,d_1\geqslant 0$, the nonexistence of polynomials p_0,p_1 such that*

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

is equivalent to the existence of $\phi_0,\phi_1\colon X\to\mathbb{R}$ such that

(v) $\phi_0\geqslant\epsilon|\phi_1|$ *on* $f^{-1}(0)$,
(vi) $\phi_1\geqslant\epsilon\max\{-\phi_0,-\frac{1}{\Delta}\phi_0\}$ *on* $f^{-1}(1)$,
(vii) $\deg p\leqslant d_0 \implies \langle\phi_0,p\rangle=0$,
(viii) $\deg p\leqslant d_1 \implies \langle\phi_1,p\rangle=0$,
(ix) $\phi_0\not\equiv 0$,
(x) $\phi_1\not\equiv 0$.

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

$$\begin{aligned} &\epsilon^{1-f} p_0 + (-\tfrac{\epsilon}{\Delta})^f p_1 > 0, \\ &(-\epsilon)^{1-f} p_0 + (-\epsilon)^f p_1 < 0 \end{aligned}$$

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

$$\epsilon^{1-f}\mu - (-\epsilon)^{1-f}\lambda \in P_0^{\perp}, \tag{5.4}$$

$$(-\tfrac{\epsilon}{\Delta})^f \mu - (-\epsilon)^f \lambda \in P_1^{\perp}. \tag{5.5}$$

By basic arithmetic, the existence of such μ and λ is in turn equivalent to the existence of $\phi_0, \phi_1 \colon X \to \mathbb{R}$, not both identically zero, that obey (v)–(viii), where we identify ϕ_0 and ϕ_1 with the left-hand side of (5.4) and (5.5), respectively.

Finally, the requirement that at least one of ϕ_0, ϕ_1 be not identically zero is logically equivalent to the requirement that $\phi_0 \not\equiv 0$ and $\phi_1 \not\equiv 0$ simultaneously. Indeed, if *exactly* one of ϕ_0, ϕ_1 were identically zero, then by (v)–(vi) the other would have to be a nonnegative function, contradicting $\langle \phi_0, 1\rangle = \langle \phi_1, 1\rangle = 0$. ∎

As a corollary, we obtain a complete dual characterization of $R_\Delta(f, d_0, d_1)$.

COROLLARY 5.3. *Let $f \colon X \to \{0,1\}$ be a given Boolean function, $\Delta > 1$, and $d_0, d_1 \geqslant 0$. If $R_\Delta(f, d_0, d_1) > 0$, then $R_\Delta(f, d_0, d_1)$ is the supremum over all $\epsilon > 0$ for which there exist $\phi_0, \phi_1 \colon X \to \mathbb{R}$ with*

(i) $\phi_0 \geqslant \epsilon|\phi_1|$ *on* $f^{-1}(0)$,
(ii) $\phi_1 \geqslant \epsilon \max\{-\phi_0, -\frac{1}{\Delta}\phi_0\}$ *on* $f^{-1}(1)$,
(iii) $\deg p \leqslant d_0 \implies \langle \phi_0, p\rangle = 0$,
(iv) $\deg p \leqslant d_1 \implies \langle \phi_1, p\rangle = 0$,
(v) $\phi_0 \not\equiv 0$,
(vi) $\phi_1 \not\equiv 0$.

6. The composition theorem.

Recall that our goal is to construct an $\wedge, \vee$-circuit of constant depth and polynomial size with high threshold degree. We focus in our search on circuits of the form $\mathrm{OR}_\ell \circ F$ for some F and $\ell \geqslant 2$. Our starting point is Theorem 4.5, which characterizes the threshold degree of every such composition. Specifically, that theorem shows that the threshold degree of $\mathrm{OR}_\ell \circ F$ is large if F does not have a low-degree one-sided rational approximant with constant error. Quantitatively,

$$\deg_\pm(\mathrm{OR}_\ell \circ F) = \Omega(\min\{\ell(d+1), D+1\})$$

whenever F does not have a one-sided rational approximant with numerator degree D, denominator degree d, and error $1/3$. The theorem holds for all D and d, but clearly it is only meaningful to work with $D \geqslant d$. To summarize, the project of this paper reduces to proving lower bounds for the one-sided rational approximation of small-depth circuits F.

Rational approximation is, however, itself a challenging model for which to prove lower bounds. After exploring various lines of attack, we discovered an approach that is at once intuitive and sufficiently powerful to give optimal lower bounds for the rational approximation of all functions of interest to us. Specifically, we study functions of the form $F = f \wedge g$ for arbitrary nonconstant f and g, and characterize the one-sided rational approximation of any such composition F in terms of natural analytic properties of f and g. Approximating F in a one-sided manner is of course at least as hard as approximating f or g individually; what our results in this section show is that approximating F is *much* harder in general, and we are able to characterize by how much. This "composition theorem" is the technical centerpiece of our paper.

6.1. The upper bound. Before we state our lower bound for the one-sided rational approximation of $f \wedge g$, it is helpful to pause and think about upper bounds first. To use the notation of the opening paragraph, suppose that we would like to construct an ϵ-error one-sided approximant for $f \wedge g$ with numerator and denominator degree on the order of D and d, respectively, where $0 < \epsilon \leqslant 1/3$ and $D \geqslant d$. The simplest approach is to take one-sided rational approximants $\tilde{f}$ and $\tilde{g}$ for the corresponding functions and approximate $f \wedge g$ in a one-sided manner by

$$\tilde{f} \cdot \tilde{g}. \tag{6.1}$$

For this construction to work, $\tilde{f}$ and $\tilde{g}$ must have error sufficiently small relative to $\|\tilde{g}\|_\infty$ and $\|\tilde{f}\|_\infty$, respectively, as well as numerator degree $O(D)$ and denominator degree $O(d)$. Another, incomparable approach is to appeal to DeMorgan's law and approximate $f \wedge g$ by

$$\frac{1}{\dfrac{1}{\tilde{f}^2} + \dfrac{1}{\tilde{g}^2}}, \tag{6.2}$$

where $\tilde{f}$ and $\tilde{g}$ again stand for one-sided rational approximants of f and g, respectively. In this alternate construction, it suffices for $\tilde{f}$ and $\tilde{g}$ to have error ϵ, but now both the numerator and denominator in these approximants must have degree $O(d)$.

These two constructions can be succinctly described using our notation for one-sided and hybrid rational approximation. The first construction shows that for any $\Delta, \Delta' > 1$, the conditions

$$R_\Delta(f, d, D) \leqslant \frac{\epsilon}{\sqrt{\Delta'}},$$
$$R_{\Delta'}(g, d, D) \leqslant \frac{\epsilon}{\sqrt{\Delta}}$$

are sufficient to conclude that

$$R(f \wedge g, O(d), O(D)) \leqslant \epsilon.$$

The second construction allows one to reach the same conclusion whenever

$$R(f, d, d) \leqslant \epsilon,$$
$$R(g, d, d) \leqslant \epsilon.$$

These equations make it clear that in both constructions, the individual approximants for f and g must in general have significantly stronger parameters—error or degree—than the target parameters for the composed function $f \wedge g$.

6.2. The lower bound. Given the restricted form of (6.1) and (6.2), there is no reason a priori to expect these constructions to give an optimal approximant. We are nevertheless able to show quite generally that they do, a result to which we refer in this paper as the "composition theorem":

THEOREM 6.1. *Let $f\colon X \to \{0,1\}$ and $g\colon Y \to \{0,1\}$ be given functions. Let $0 < \epsilon \leqslant 1$ and $\Delta > 1$. Assume that there exist $\phi_0, \phi_1', \phi_1''\colon X \to \mathbb{R}$ such that*

$$\phi_0 \geqslant \epsilon|\phi_1'| \text{ on } f^{-1}(0), \tag{6.3}$$

$$\phi_0 \geqslant \epsilon|\phi_1''| \text{ on } f^{-1}(0), \tag{6.4}$$

$$\phi_1' \geqslant \epsilon \max\{-\phi_0, -\tfrac{1}{\Delta}\phi_0\} \text{ on } f^{-1}(1), \tag{6.5}$$

$$\phi_1'' \geqslant \epsilon|\phi_0| \text{ on } f^{-1}(1), \tag{6.6}$$

$$\deg p \leqslant d \implies \langle \phi_0, p\rangle = 0, \tag{6.7}$$

$$\deg p \leqslant D \implies \langle \phi_1', p\rangle = 0, \tag{6.8}$$

$$\deg p \leqslant d \implies \langle \phi_1'', p\rangle = 0, \tag{6.9}$$

$$\phi_0 \not\equiv 0, \tag{6.10}$$

$$\phi_1' \not\equiv 0, \tag{6.11}$$

$$\phi_1'' \not\equiv 0. \tag{6.12}$$

Assume furthermore that

$$R(g, d, D) > \frac{\epsilon}{\sqrt{\Delta}}. \tag{6.13}$$

Then

$$R\left(f \wedge g, \frac{d}{2}, \frac{D}{2}\right) \geqslant \frac{\epsilon}{\sqrt{2}}. \tag{6.14}$$

The statement of Theorem 6.1 is admittedly technical but its intuitive content is satisfying and easy to explain. Conditions (6.3), (6.5), (6.7), (6.8), (6.10), (6.11), (6.13) can be summarized as

$$\left.\begin{aligned} R_\Delta(f, d, D) &> \frac{\epsilon}{\sqrt{\Delta'}} \\ R_{\Delta'}(g, d, D) &> \frac{\epsilon}{\sqrt{\Delta}} \end{aligned}\right\} \text{for every } \Delta' > 1, \tag{6.15}$$

by the dual characterization of hybrid rational approximation (Theorem 5.2). The remaining conditions (6.4), (6.6), (6.7), (6.9), (6.10), (6.12) correspond to

$$R(f, d, d) \geqslant \epsilon, \tag{6.16}$$

by the dual characterization of one-sided rational approximation (Theorem 4.6). Of note here is the use of a dual object, ϕ_0, for both (6.15) and (6.16). Hypothesis (6.15) rules out an approximant for $f \wedge g$ of the form (6.1), whereas hypothesis (6.16) rules out an approximant of the form (6.2). The theorem states, roughly, that ruling out these two constructions is enough to rule out all possibilities.

The reader may have expected to see the conclusion of the composition theorem arrived at under the following weaker hypotheses:

$$\begin{aligned} R_\Delta(f, d, D) &> \frac{\epsilon}{\sqrt{\Delta'}}, \\ R_{\Delta'}(g, d, D) &> \frac{\epsilon}{\sqrt{\Delta}}, \\ R(f, d, d) &\geqslant \epsilon \end{aligned}$$

for some *fixed* values $\Delta, \Delta' > 1$. A moment's thought shows, however, that this expectation is misplaced. Indeed, under these weaker assumptions it may well turn out that f and g have one-sided approximants with error 0 and degree $d+1$, in which case $f \wedge g$ would have an efficient approximant of the form (6.1).

Proof of Theorem 6.1. Applying the dual characterization of one-sided rational approximation (Theorem 4.6), we infer from (6.13) the existence of $\psi_0, \psi_1 \colon Y \to \mathbb{R}$ such that

$$\psi_0 \geqslant \epsilon|\psi_1|/\sqrt{\Delta} \text{ on } g^{-1}(0), \tag{6.17}$$

$$\psi_1 \geqslant \epsilon|\psi_0|/\sqrt{\Delta} \text{ on } g^{-1}(1), \tag{6.18}$$

$$\deg p \leqslant d \implies \langle \psi_0, p \rangle = 0, \tag{6.19}$$

$$\deg p \leqslant D \implies \langle \psi_1, p \rangle = 0, \tag{6.20}$$

$$\psi_0 \not\equiv 0, \tag{6.21}$$

$$\psi_1 \not\equiv 1. \tag{6.22}$$

Define $\zeta_0, \zeta_1 \colon X \times Y \to \mathbb{R}$ by

$$\zeta_0 = \left(\frac{1}{\epsilon}\phi_0 + \frac{1}{2}\phi_1''\right) \cdot |\psi_0| \cdot g + \frac{\epsilon}{2}|\phi_0| \cdot f \cdot \psi_0, \tag{6.23}$$

$$\zeta_1 = \phi_1' \cdot |\psi_0| \cdot g + \frac{1}{\sqrt{\Delta}}|\phi_0| \cdot f \cdot \psi_1, \tag{6.24}$$

where $f, \phi_0, \phi_1', \phi_1'', g, \psi_0, \psi_1, \zeta_0, \zeta_1$ above and in the rest of the proof are shorthands for $f(x)$, $\phi_0(x)$, $\phi_1'(x)$, $\phi_1''(x)$, $g(y)$, $\psi_0(y)$, $\psi_1(y)$, $\zeta_0(x, y)$, $\zeta_1(x, y)$, respectively. Then by (6.7), (6.9), and (6.19), and linearity,

$$\deg p \leqslant d \implies \langle \zeta_0, p \rangle = 0. \tag{6.25}$$

Analogously, by (6.8), (6.20), and linearity,

$$\deg p \leqslant D \implies \langle \zeta_1, p \rangle = 0. \tag{6.26}$$

We now establish relevant metric properties of ζ_0 and ζ_1.

Claim 6.2. *$\zeta_1 \geqslant \epsilon^2 \max\{-\zeta_0, 0\}$ whenever $f \wedge g = 1$.*

Proof. For $f = g = 1$,

$$\begin{aligned} -\zeta_0 &= -\left(\frac{1}{\epsilon}\phi_0 + \frac{1}{2}\phi_1''\right) \cdot |\psi_0| - \frac{\epsilon}{2}|\phi_0| \cdot \psi_0 && \text{by (6.23)} \\ &\leqslant -\left(\frac{1}{\epsilon}\phi_0 + \frac{\epsilon}{2}|\phi_0|\right) \cdot |\psi_0| - \frac{\epsilon}{2}|\phi_0| \cdot \psi_0 && \text{by (6.6)} \\ &\leqslant -\frac{1}{\epsilon}\phi_0 \cdot |\psi_0| \end{aligned}$$

and

$$\begin{aligned} \zeta_1 &= \phi_1' \cdot |\psi_0| + \frac{1}{\sqrt{\Delta}}|\phi_0| \cdot \psi_1 && \text{by (6.24)} \\ &\geqslant \epsilon \max\left\{-\phi_0, -\frac{1}{\Delta}\phi_0\right\} \cdot |\psi_0| + \frac{\epsilon}{\Delta}|\phi_0| \cdot |\psi_0| && \text{by (6.5) and (6.18)} \\ &\geqslant \epsilon \max\{-\phi_0 \cdot |\psi_0|, 0\}. && (6.27) \end{aligned}$$

Comparing these bounds for ζ_0 and ζ_1 immediately gives $\zeta_1 \geqslant \epsilon^2 \max\{-\zeta_0, 0\}$. □

CLAIM 6.3. $\zeta_0 \geqslant \frac{\epsilon^2}{2}|\zeta_1|$ *whenever* $f \wedge g = 0$.

Proof. We examine the three possibilities, depending on the values of f and g. If $f = 0$ and $g = 1$,

$$\begin{aligned}
\zeta_0 &= \left(\frac{1}{\epsilon}\phi_0 + \frac{1}{2}\phi_1''\right) \cdot |\psi_0| && \text{by (6.23)} \\
&\geqslant \left(\frac{1}{2}|\phi_1'| + \frac{1}{2}|\phi_1''| + \frac{1}{2}\phi_1''\right) \cdot |\psi_0| && \text{by (6.3) and (6.4)} \\
&\geqslant \frac{1}{2}|\phi_1'| \cdot |\psi_0| \\
&= \frac{1}{2}|\zeta_1| && \text{by (6.24).}
\end{aligned}$$

If $f = 1$ and $g = 0$,

$$\begin{aligned}
\zeta_0 &= \frac{\epsilon}{2}|\phi_0| \cdot \psi_0 && \text{by (6.23)} \\
&\geqslant \frac{\epsilon}{2}|\phi_0| \cdot \frac{\epsilon}{\sqrt{\Delta}}|\psi_1| && \text{by (6.17)} \\
&= \frac{\epsilon^2}{2}|\zeta_1| && \text{by (6.24).}
\end{aligned}$$

In the remaining case when $f = g = 0$, we immediately have $\zeta_0 = \zeta_1 = 0$ from the defining equations for ζ_0 and ζ_1. □

CLAIM 6.4. $\zeta_0 \not\equiv 0$ *and* $\zeta_1 \not\equiv 1$.

Proof. Recall from (6.3), (6.7), and (6.10) that ϕ_0 is orthogonal to the constant function 1, is nonnegative on $f^{-1}(0)$, and is not identically zero. It follows that

$$\min_{f^{-1}(1)} \phi_0 < 0.$$

Similarly, recall from (6.17), (6.19), and (6.21) that ψ_0 is orthogonal to the constant function 1, is nonnegative on $g^{-1}(0)$, and is not identically zero, whence

$$\min_{g^{-1}(1)} \psi_0 < 0.$$

But (6.27) guarantees that $\zeta_1 \geqslant -\epsilon\phi_0 \cdot |\psi_0|$ whenever $f = g = 1$. We conclude that ζ_1 is strictly positive at some point where $f = g = 1$. In particular, $\zeta_1 \not\equiv 0$ as desired. Since by (6.26) the constant function 1 is orthogonal to ζ_1, we also conclude that ζ_1 must take on a negative value at some point:

$$\min_{X \times Y} \zeta_1 < 0. \tag{6.28}$$

It remains to show that $\zeta_0 \not\equiv 0$. Claims 6.2 and 6.3 ensure that ζ_1 is nonnegative when $f \wedge g = 1$ and is bounded in absolute value by $\frac{2}{\epsilon^2}\zeta_0$ when $f \wedge g = 0$. Therefore, $\zeta_0 \equiv 0$ would force $\zeta_1 \geqslant 0$ everywhere, in contradiction to (6.28). □

The newly established properties of ζ_0 and ζ_1 in (6.25), (6.26), and Claims 6.2 to 6.4 imply, in light of the dual characterization of hybrid rational approximation (Theorem 5.2), that

$$R_{\Delta'}(f \wedge g, d, D) \geqslant \frac{\epsilon^2}{2}$$

for every $\Delta' > 1$. Passing to the limit as $\Delta' \to \infty$ and applying Theorem 5.1, we arrive at the desired lower bound (6.14). □

7. From polynomial to hybrid rational approximation. The composition theorem of the previous section allows us to obtain lower bounds for one-sided constant-error rational approximation from lower bounds for two substantially more restricted models, namely, hybrid rational approximation with constant error and one-sided rational approximation with exponentially small error. We tackle these two restricted models in this section and the next, respectively. Our focus here, Theorem 7.1, is a hardness amplification result that gives lower bounds for the hybrid rational approximation of a large class of functions. This theorem translates lower bounds for one-sided *polynomial* approximation, of which there is an abundant supply in the literature, into lower bounds for hybrid *rational* approximation.

THEOREM 7.1. *Let* $f\colon X \to \{0,1\}$ *be a nonconstant Boolean function,* $0 < \epsilon \leqslant 1/2$. *For* $c = c(\epsilon) > 0$ *sufficiently large, define*

$$F = \mathrm{AND}_{cn} \circ f.$$

Then there exist functions $\Phi_0, \Phi_1', \Phi_1''\colon X^{cn} \to \mathbb{R}$ *such that:*

(i) $\Phi_1' \geqslant (1-\epsilon)\max\{-\Phi_0, -2^{-n}\Phi_0\}$ *on* $F^{-1}(1)$,

(ii) $\Phi_1'' \geqslant (1-\epsilon)|\Phi_0|$ *on* $F^{-1}(1)$,

(iii) $\Phi_0 \geqslant (1-\epsilon)\max\{|\Phi_1'|, |\Phi_1''|\}$ *on* $F^{-1}(0)$,

(iv) $\langle \Phi_1', P\rangle = 0$ *whenever* $\deg P \leqslant \frac{1}{c}\deg^+_{1/3}(\neg f)\sqrt{n}$,

(v) $\langle \Phi_0, P\rangle = \langle \Phi_1'', P\rangle = 0$ *whenever* $\deg P \leqslant \min\{\frac{1}{c}\deg^+_{1/3}(\neg f), \frac{1}{2}n\}$,

(vi) $\Phi_0 \not\equiv 0$,

(vii) $\Phi_1' \not\equiv 0$,

(viii) $\Phi_1'' \not\equiv 0$.

The conclusion of Theorem 7.1 is easiest to understand in terms of the dual characterization of one-sided and hybrid rational approximation (Theorems 4.6 and 5.2, respectively). Specifically, properties (i)–(viii) correspond to the following two lower bounds for rational approximation:

$$R_{2^n}\left(\mathrm{AND}_{cn} \circ f, \min\left\{\frac{1}{c}\deg^+_{1/3}(\neg f), \frac{n}{2}\right\}, \frac{1}{c}\deg^+(\neg f)\sqrt{n}\right) \geqslant 1-\epsilon,$$

$$R\left(\mathrm{AND}_{cn} \circ f, \min\left\{\frac{1}{c}\deg^+_{1/3}(\neg f), \frac{n}{2}\right\}, \min\left\{\frac{1}{c}\deg^+_{1/3}(\neg f), \frac{n}{2}\right\}\right) \geqslant 1-\epsilon,$$

where $c = c(\epsilon) > 0$ is a constant. These two inequalities are incomparable: the former gives a stronger lower bound on the numerator degree, whereas the latter applies to a more general model (one-sided vs. hybrid approximation). The only property needed to reach these conclusions is the one-sided approximate degree of $\neg f$. Thus, Theorem 7.1 transforms a function that is hard to approximate by polynomials into a related function that is hard to approximate by rational functions.

Our proof of Theorem 7.1 is an adaptation of a recent hardness amplification result in [35, section 5], used in that paper to obtain the strongest lower bound on the threshold degree of AC^0 prior to our work. That earlier result is logically incomparable with ours but requires a more complex proof. Both proofs start with a dual object for the original function f and build from it a sequence of dual objects of increasing complexity, culminating in one that witnesses the claimed approximation-theoretic property of the composition $\mathrm{AND}_{cn} \circ f$. In our case, the starting point is a dual object that witnesses the one-sided approximate degree of $\neg f$, and the end result is the triple of dual objects $\Phi_0, \Phi_1', \Phi_1''$ in the theorem statement. The intermediate building blocks are all borrowed from [35], but we are able to combine them in a way that is considerably simpler and more intuitive. We have structured our proof of Theorem 7.1 to emphasize both the similarities and differences with the earlier work. Specifically, the preparatory subsections 7.1 to 7.4 below are in close correspondence with [35], whereas the heart of our argument, in subsection 7.5, is different and simpler. We provide additional details and intuition at each stage of the proof.

7.1. Outer dual object. Analogous to [35], the starting point in our proof is what we call the "outer" dual object. It is derived from the dual polynomial for the OR function in Theorem 2.13 and represents the linear combination that we use to combine the various building blocks of our construction. Without loss of generality, we may assume that $c = c(\epsilon)$ is a sufficiently large even integer. Then for each $k = 0, 1, 2, \ldots, n$, Theorem 2.13 gives an explicit function $\omega_k\colon \{0, 1, 2, \ldots, cn-k\} \to \mathbb{R}$ such that

$$\|\omega_k\|_1 = 1, \tag{7.1}$$

$$\omega_k(0) > \frac{1}{2} - \frac{\epsilon}{8}, \tag{7.2}$$

$$|\omega_k(t)| \geqslant \frac{\epsilon}{12 \cdot 2^t} \qquad (t \geqslant 1), \tag{7.3}$$

$$\deg p < \sqrt{n} \implies \langle \omega_k, p \rangle = 0. \tag{7.4}$$

By Proposition 2.12(ii),

$$\|\omega_k\|_\infty \leqslant \frac{1}{2}. \tag{7.5}$$

7.2. Inner dual objects. We now turn to the innermost part of the construction, namely, the dual object that witnesses the one-sided approximate degree of $\neg f$ and the probability distributions that it induces on X. This step, too, is closely analogous to [35]. Define

$$d = \deg^+_{\frac{1-(\epsilon/30)^2}{2}}(\neg f).$$

Then for a sufficiently large constant $c = c(\epsilon) > 0$, we have

$$d > \frac{1}{c} \deg^+_{1/3}(\neg f) \tag{7.6}$$

by the error-reduction property of one-sided approximate degree (Fact 2.4). By the dual characterization of one-sided approximate degree (Corollary 2.11), there exists a function $\phi\colon X \to \mathbb{R}$ with

$$\deg p < d \implies \langle \phi, p \rangle = 0, \tag{7.7}$$

$$f(x) = 0 \implies \phi(x) \leqslant 0, \tag{7.8}$$

$$\langle f, \phi \rangle > \frac{1-(\epsilon/30)^2}{2} \|\phi\|_1. \tag{7.9}$$

Then

$$\langle \phi, 1 \rangle = 0, \tag{7.10}$$

$$\phi \not\equiv 0 \tag{7.11}$$

by (7.7) and (7.9), respectively. By homogeneity, we may assume that

$$\|\phi\|_1 = 1. \tag{7.12}$$

Define α by

$$\langle f, \phi \rangle = \frac{1-\alpha}{2}. \tag{7.13}$$

Then

$$0 \leqslant \alpha < \left(\frac{\epsilon}{30}\right)^2, \tag{7.14}$$

where the upper bound is immediate from (7.9) and (7.12), whereas the lower bound holds by (7.10), (7.12), and Proposition 2.12(iii).

We now consider several probability distributions that ϕ induces on X. By (7.12), the function $|\phi|$ itself is a probability distribution on X. We further define μ_0 and μ_1 to be the probability distributions induced by $|\phi|$ on the sets $\{x \in X : \phi(x) < 0\}$ and $\{x \in X : \phi(x) > 0\}$, respectively. Equations (7.10) and (7.11) guarantee that these two sets are nonempty, so that μ_0 and μ_1 are well-defined. By (7.10) and (7.12),

$$\phi = \frac{1}{2}\mu_1 - \frac{1}{2}\mu_0. \tag{7.15}$$

Multiplying on both sides by f and applying (7.8), we find that $\langle f, \phi\rangle = \frac{1}{2}\langle f, \mu_1\rangle - \frac{1}{2}\langle f, \mu_0\rangle = \frac{1}{2} - \frac{1}{2}\langle f, \mu_0\rangle$, which in view of (7.13) gives $\langle f, \mu_0\rangle = \alpha$. In particular,

$$\begin{aligned} \langle f \cdot \mu_0 - \alpha\mu_0, 1\rangle &= \langle f, \mu_0\rangle - \alpha\langle \mu_0, 1\rangle \\ &= \langle f, \mu_0\rangle - \alpha \\ &= 0. \end{aligned} \tag{7.16}$$

We will need the following technical result from [31, 35].

Lemma 7.2 (Sherstov). *Let $\xi\colon X \to \mathbb{R}$ be an arbitrary function. Then for every polynomial $P\colon X^N \to \mathbb{R}$ and every $k = 0, 1, 2, \ldots, N$, the mapping*

$$z \mapsto \left\langle \xi^{\otimes k} \otimes \bigotimes_{i=1}^{N-k} \mu_{z_i}, P \right\rangle, \qquad z \in \{0,1\}^{N-k}, \tag{7.17}$$

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

Proof (adapted from [31]). By linearity, it suffices to consider factored polynomials of the form $P(x_1, \ldots, x_N) = p_1(x_1) \cdots p_N(x_N)$. In this case, (7.17) simplifies to

$$z \mapsto \prod_{i=1}^{k} \langle \xi, p_i\rangle \cdot \prod_{i=1}^{N-k} \langle \mu_{z_i}, p_{k+i}\rangle, \qquad z \in \{0,1\}^{N-k}. \tag{7.18}$$

By (7.7) and (7.15), polynomials p_i of degree less than d satisfy $\langle \mu_0, p_i\rangle = \langle \mu_1, p_i\rangle$ and therefore do not contribute to the degree of (7.18) as a real function on $\{0,1\}^{N-k}$. It follows that the degree of (7.18) is at most $|\{i : \deg p_i \geqslant d\}| \leqslant (\deg P)/d$. ∎

7.3. Auxiliary distributions in tensor space. Following [35], we will now use μ_0 and μ_1 to construct auxiliary functions $\Lambda^N_{k,m}$ on the tensor space X^N. For nonnegative integers k, m, N with $k + m \leqslant N$, define

$$\Lambda^N_{k,m}(x_1, x_2, \ldots, x_N) = \mathop{\mathbf{E}}_{S,T}\left[\prod_{i\in S} f(x_i)\mu_0(x_i) \cdot \prod_{i\in T} \mu_0(x_i) \cdot \prod_{i\notin S\cup T} \mu_1(x_i)\right], \tag{7.19}$$

where the expectation is taken over a uniformly random pair of disjoint sets $S, T \subseteq \{1, 2, \ldots, N\}$ of size $|S| = k$ and $|T| = m$. Observe that $\Lambda^N_{k,m}$ is a nonnegative function, a fact that we will use frequently without further mention. The following lemma from [35] collects basic properties of $\Lambda^N_{k,m}$. For the reader's convenience, we include its short proof.

Lemma 7.3 (Sherstov).

(i) $\operatorname{supp} \Lambda^N_{k,0} \subseteq f^{-1}(1)^N$,

(ii) $\Lambda^N_{k,m} = \Lambda^N_{k+m,0}$ *on* $f^{-1}(1)^N$,

(iii) $\Lambda^N_{k,m}(x) \neq 0$ *only if* $|\{i : \phi(x_i) < 0\}| = k + m$,

(iv) *for an arbitrary real polynomial* $P\colon X^N \to \mathbb{R}$, *the mapping* $m \mapsto \langle \Lambda^N_{k,m}, P\rangle$ $(m = 0, 1, 2, \ldots, N-k)$ *is a univariate polynomial of degree at most* $(\deg P)/d$.

Proof (adapted from [35]).

(i) Immediate from the fact that $\operatorname{supp} \mu_1 \subseteq f^{-1}(1)$.

(ii) Immediate from the defining equation for $\Lambda^N_{k,m}$.

(iii) Immediate from the fact that μ_0 and μ_1 are supported on $\{x \in X : \phi(x) < 0\}$ and $\{x \in X : \phi(x) > 0\}$, respectively.

(iv) For $S \subseteq \{1, 2, \ldots, N\}$ with $|S| = k$, define

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

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 the mapping in (iv) is a convex combination of mappings

$$m \mapsto \langle \Lambda^N_{S,m}, P\rangle \qquad (m = 0, 1, 2, \ldots, N-k) \tag{7.20}$$

as S ranges over k-element subsets. As a result, the proof will be complete once we show that (7.20) is a polynomial of degree at most $(\deg P)/d$.

By symmetry, we may assume that $S = \{1, 2, \ldots, k\}$. By Lemma 7.2, the mapping

$$z \mapsto \left\langle (f\cdot\mu_0)^{\otimes k} \otimes \bigotimes_{i=1}^{N-k} \mu_{z_i}, P\right\rangle, \qquad z \in \{0,1\}^{N-k},$$

has degree at most $(\deg P)/d$. Therefore by Proposition 2.2, the mapping

$$m \mapsto \langle \Lambda^N_{S,m}, P\rangle = \mathop{\mathbf{E}}_{\substack{z\in\{0,1\}^{N-k}\\ |z|=m}} \left\langle (f\cdot\mu_0)^{\otimes k} \otimes \bigotimes_{i=1}^{N-k} \mu_{z_i}, P\right\rangle \tag{7.21}$$

is a univariate polynomial on $\{0, 1, 2, \ldots, N-k\}$ of degree at most $(\deg P)/d$. □

7.4. Corrector for false negatives. Recall from the statement of Theorem 7.1 that the sought dual object Φ_1'' must be nonnegative on $F^{-1}(1)$. We ensure this sign behavior by means of a "corrector" object, whose role is to correct the sign on any offending inputs in $F^{-1}(1)$ without disturbing the signs elsewhere on the domain. For integers k and N with $1 \leqslant k \leqslant N$, let $\tilde{\Lambda}_k^N \colon X^N \to \mathbb{R}$ be given by

$$\tilde{\Lambda}_k^N(x) = \frac{1}{(\lceil k/2 \rceil - 1)!} \times \mathop{\mathbf{E}}_{|S|=k} \left[\prod_{i \in S} \frac{(f(x_i) - \alpha)\mu_0(x_i)}{1-\alpha} \cdot \prod_{i=\lfloor k/2 \rfloor + 1}^{k-1} \left(\sum_{j \in S} f(x_j) - i \right) \cdot \prod_{i \notin S} \mu_1(x_i) \right],$$

where the expectation is over a uniformly random set $S \subseteq \{1, 2, \ldots, N\}$ of size $|S| = k$. This definition of $\tilde{\Lambda}_k^N$ is borrowed with minor changes from [35], and the following lemma is an adaptation of the corresponding result from [35].

Lemma 7.4 (cf. Sherstov).

(i) $\langle \tilde{\Lambda}_k^N, P \rangle = 0$ for every polynomial P of degree at most $k/2$,

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

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

(iv) $|\tilde{\Lambda}_k^N| \leqslant \frac{1}{2} \left(\frac{2\alpha}{1-\alpha} \right)^{k/2} \Lambda_{0,k}^N$ outside $f^{-1}(1)^N$.

Proof (adapted from [35]).

(i) Recall from (7.16) that $\langle f \cdot \mu_0 - \alpha\mu_0, 1 \rangle = 0$. For $t = 0, 1, 2, \ldots,$ it follows that $(f \cdot \mu_0 - \alpha\mu_0)^{\otimes t}$ is orthogonal to every polynomial of degree less than t. In particular, the function

$$x \mapsto \prod_{i \in S} (f(x_i) - \alpha)\mu_0(x_i) \cdot \prod_{i=\lfloor k/2 \rfloor + 1}^{k-1} \left(\sum_{j \in S} f(x_j) - i \right) \cdot \prod_{i \notin S} \mu_1(x_i),$$

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

(ii) Immediate from the fact that μ_0 and μ_1 are supported on $\{x \in X : \phi(x) < 0\}$ and $\{x \in X : \phi(x) > 0\}$, respectively.

(iii) Substituting $f(x_i) = 1$ in the defining equation for $\tilde{\Lambda}_k^N$, we obtain

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

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

$$\frac{1}{(\lceil k/2 \rceil - 1)!} \prod_{i \in S} \frac{|f(x_i) - \alpha|\, \mu_0(x_i)}{1-\alpha} \cdot \prod_{i=\lfloor k/2 \rfloor + 1}^{k-1} \left| \sum_{j \in S} f(x_j) - i \right| \cdot \prod_{i \notin S} \mu_1(x_i) \leqslant \frac{1}{2} \left(\frac{2\alpha}{1-\alpha} \right)^{k/2} \prod_{i \in S} \mu_0(x_i) \cdot \prod_{i \notin S} \mu_1(x_i). \tag{7.22}$$

To see this, consider the nonempty set $Z = \{i : f(x_i) = 0\}$. There are three possibilities. If $Z \nsubseteq S$, then both sides of (7.22) vanish because μ_1 has support inside $f^{-1}(1)$. If $Z \subseteq S$ and $1 \leqslant |Z| \leqslant \lceil k/2 \rceil - 1$, then $\prod_{i=\lfloor k/2 \rfloor+1}^{k-1} |\sum_{j \in S} f(x_j) - i| = 0$ and the left-hand side of (7.22) vanishes. In the remaining case that $Z \subseteq S$ and $\lceil k/2 \rceil \leqslant |Z| \leqslant k$, the left-hand side of (7.22) simplifies to

$$\begin{aligned}
&\binom{|Z|-1}{\lceil k/2 \rceil - 1} \prod_{i \in S} \frac{|f(x_i) - \alpha|\, \mu_0(x_i)}{1-\alpha} \cdot \prod_{i \notin S} \mu_1(x_i) \\
&\qquad \leqslant 2^{|Z|-1} \prod_{i \in S} \frac{|f(x_i) - \alpha|\, \mu_0(x_i)}{1-\alpha} \cdot \prod_{i \notin S} \mu_1(x_i) \\
&\qquad = 2^{|Z|-1} \prod_{i \in Z} \frac{|f(x_i) - \alpha|\, \mu_0(x_i)}{1-\alpha} \cdot \prod_{i \in S \setminus Z} \frac{|f(x_i) - \alpha|\, \mu_0(x_i)}{1-\alpha} \cdot \prod_{i \notin S} \mu_1(x_i) \\
&\qquad = 2^{|Z|-1} \left(\frac{\alpha}{1-\alpha}\right)^{|Z|} \prod_{i \in Z} \mu_0(x_i) \cdot \prod_{i \in S \setminus Z} \mu_0(x_i) \cdot \prod_{i \notin S} \mu_1(x_i) \\
&\qquad \leqslant \frac{1}{2} \left(\frac{2\alpha}{1-\alpha}\right)^{k/2} \prod_{i \in S} \mu_0(x_i) \cdot \prod_{i \notin S} \mu_1(x_i),
\end{aligned}$$

where the final step is valid because $\alpha \leqslant 1/3$ by (7.14). This completes the proof of (7.22). Passing to expectations on both sides of (7.22) with respect to a uniformly random subset S of cardinality k, we arrive at the claimed conclusion:

$$|\tilde{\Lambda}_k^N(x)| \leqslant \frac{1}{2} \left(\frac{2\alpha}{1-\alpha}\right)^{k/2} \Lambda_{0,k}^N(x). \qquad \square$$

7.5. Final construction. With the preparatory work now complete, we are in a position to define the desired dual objects $\Phi_0, \Phi_1', \Phi_1''$ and verify their properties. This concluding part of the proof of Theorem 7.1 departs from [35] and seems considerably simpler. Let

$$\begin{aligned}
\Phi_0 &= \sum_{k=0}^{n} 3^k \left(\sum_{m=0}^{cn-k} |\omega_k(m)| \Lambda_{k,m}^{cn} + (2^{n+1} - 1) \sum_{m=n-k+1}^{cn-k} |\omega_k(m)| \tilde{\Lambda}_{k+m}^{cn} - \Lambda_{k,0}^{cn} \right), \\
\Phi_1' &= \sum_{k=0}^{n} 3^k \sum_{m=0}^{cn-k} \omega_k(m) \Lambda_{k,m}^{cn}, \\
\Phi_1'' &= \sum_{k=0}^{n} 3^k \left(\sum_{m=0}^{cn-k} \omega_k(m) \Lambda_{k,m}^{cn} + (2^{n+1} + 1) \sum_{m=n-k+1}^{cn-k} |\omega_k(m)| \tilde{\Lambda}_{k+m}^{cn} \right).
\end{aligned}$$

We proceed to verify one by one the properties of these functions required by Theorem 7.1.

LEMMA 7.5. *On $F^{-1}(1)$, one has*

$$\begin{aligned}
\Phi_1' &\geqslant (1-\epsilon) \max\{-\Phi_0, -2^{-n}\Phi_0\}, \\
\Phi_1'' &\geqslant (1-\epsilon) |\Phi_0|.
\end{aligned}$$

Proof. Take an arbitrary point $x \in F^{-1}(1) = f^{-1}(1)^{cn}$ and let $\ell = |\{i : \phi(x_i) < 0\}|$. There are two cases to consider, depending on the value of ℓ.

Case $0 \leqslant \ell \leqslant n$. Using (7.2) and (7.5), one easily verifies that

$$\sum_{k=0}^{\ell} 3^k \omega_k(\ell - k) \geqslant (1-\epsilon)\left(3^\ell - \sum_{k=0}^{\ell} 3^k |\omega_k(\ell-k)|\right). \tag{7.23}$$

We have

$$\begin{aligned}
|\Phi_0(x)| &= \left|3^\ell \Lambda^{cn}_{\ell,0}(x) - \sum_{k=0}^{\ell} 3^k |\omega_k(\ell-k)| \Lambda^{cn}_{k,\ell-k}(x)\right| && \text{by Lemmas 7.3(iii), 7.4(ii)}\\
&= \left|3^\ell - \sum_{k=0}^{\ell} 3^k |\omega_k(\ell-k)|\right| \Lambda^{cn}_{\ell,0}(x) && \text{by Lemma 7.3(ii)}\\
&= \left(3^\ell - \sum_{k=0}^{\ell} 3^k |\omega_k(\ell-k)|\right) \Lambda^{cn}_{\ell,0}(x) && \text{by (7.5)}
\end{aligned}$$

and

$$\begin{aligned}
\Phi_1''(x) &= \Phi_1'(x) && \text{by Lemma 7.4(ii)}\\
&= \sum_{k=0}^{\ell} 3^k \omega_k(\ell-k) \Lambda^{cn}_{k,\ell-k}(x) && \text{by Lemma 7.3(iii)}\\
&= \left(\sum_{k=0}^{\ell} 3^k \omega_k(\ell-k)\right) \Lambda^{cn}_{\ell,0}(x) && \text{by Lemma 7.3(ii),}
\end{aligned}$$

which gives $\Phi_1''(x) = \Phi_1'(x) \geqslant (1-\epsilon)|\Phi_0(x)|$ in light of (7.23).

Case $\ell \geqslant n+1$. We have

$$\begin{aligned}
\Phi_0(x) &= \sum_{k=0}^{n} 3^k \left(|\omega_k(\ell-k)| \Lambda^{cn}_{k,\ell-k}(x) + (2^{n+1}-1)|\omega_k(\ell-k)| \tilde{\Lambda}^{cn}_{\ell}(x)\right)\\
&\qquad\qquad \text{by Lemmas 7.3(iii), 7.4(ii)}\\
&= \left(\sum_{k=0}^{n} 3^k (|\omega_k(\ell-k)| + (2^{n+1}-1)|\omega_k(\ell-k)|)\right) \Lambda^{cn}_{\ell,0}(x)\\
&\qquad\qquad \text{by Lemmas 7.3(ii), 7.4(iii)}\\
&= 2^{n+1} \left(\sum_{k=0}^{n} 3^k |\omega_k(\ell-k)|\right) \Lambda^{cn}_{\ell,0}(x).
\end{aligned}$$

Moreover,

$$\begin{aligned} |\Phi_1'(x)| &= \left|\sum_{k=0}^{n} 3^k \omega_k(\ell-k)\Lambda_{k,\ell-k}^{cn}(x)\right| && \text{by Lemma 7.3(iii)} \\ &= \left|\sum_{k=0}^{n} 3^k \omega_k(\ell-k)\right| \Lambda_{\ell,0}^{cn}(x) && \text{by Lemma 7.3(ii)} \\ &\leqslant \left(\sum_{k=0}^{n} 3^k |\omega_k(\ell-k)|\right) \Lambda_{\ell,0}^{cn}(x) \end{aligned}$$

and

$$\begin{aligned} \Phi_1''(x) &= \sum_{k=0}^{n} 3^k \left(\omega_k(\ell-k)\Lambda_{k,\ell-k}^{cn}(x) + (2^{n+1}+1)|\omega_k(\ell-k)|\tilde{\Lambda}_\ell^{cn}(x)\right) \\ & \qquad\qquad\qquad\qquad\qquad\qquad \text{by Lemmas 7.3(iii), 7.4(ii)} \\ &= \left(\sum_{k=0}^{n} 3^k (\omega_k(\ell-k) + (2^{n+1}+1)|\omega_k(\ell-k)|)\right) \Lambda_{\ell,0}^{cn}(x) \\ & \qquad\qquad\qquad\qquad\qquad\qquad \text{by Lemmas 7.3(ii), 7.4(iii)} \\ &\geqslant 2^{n+1} \left(\sum_{k=0}^{n} 3^k |\omega_k(\ell-k)|\right) \Lambda_{\ell,0}^{cn}(x). \end{aligned}$$

The estimates for Φ_0 and Φ_1' show that $|\Phi_1'| \leqslant 2^{-n-1}\Phi_0$, which yields the desired inequality $\Phi_1' \geqslant (1-\epsilon)\max\{-2^{-n}\Phi_0, -\Phi_0\}$ in view of $0 \leqslant \epsilon \leqslant 1/2$. The estimates for Φ_0 and Φ_1'' immediately give $\Phi_1'' \geqslant |\Phi_0|$. □

Lemma 7.6. *$\Phi_0 \geqslant (1-\epsilon)\max\{|\Phi_1'|, |\Phi_1''|\}$ on $F^{-1}(0)$.*

Proof. Take an arbitrary point $x \in F^{-1}(0)$ and let $\ell = |\{i : \phi(x_i) < 0\}|$. Analogous to the previous lemma, there are two cases to consider, depending on the value of ℓ.

Case $0 \leqslant \ell \leqslant n$. We have

$$\begin{aligned} \Phi_0(x) &= \sum_{k=0}^{\ell} 3^k |\omega_k(\ell-k)|\Lambda_{k,\ell-k}^{cn}(x) - 3^\ell \Lambda_{\ell,0}^{cn}(x) && \text{by Lemmas 7.3(iii), 7.4(ii)} \\ &= \sum_{k=0}^{\ell} 3^k |\omega_k(\ell-k)|\Lambda_{k,\ell-k}^{cn}(x) && \text{by Lemma 7.3(i)} \end{aligned}$$

and

$$\begin{aligned} |\Phi_1''(x)| &= |\Phi_1'(x)| && \text{by Lemma 7.4(ii)} \\ &= \left|\sum_{k=0}^{\ell} 3^k \omega_k(\ell-k)\Lambda_{k,\ell-k}^{cn}(x)\right| && \text{by Lemma 7.3(iii).} \end{aligned}$$

Therefore, $\Phi_0(x) \geqslant |\Phi_1'(x)| = |\Phi_1''(x)|$ in this case.

Case $\ell \geqslant n+1$. By Lemmas 7.3(i), 7.3(iii), and 7.4(ii), the defining equations for $\Phi_0, \Phi_1', \Phi_1''$ simplify to

$$\begin{aligned}
\Phi_0(x) &= \sum_{k=0}^{n} 3^k \left(|\omega_k(\ell-k)|\Lambda^{cn}_{k,\ell-k}(x) + (2^{n+1}-1)|\omega_k(\ell-k)|\tilde\Lambda^{cn}_{\ell}(x) \right), \\
\Phi_1'(x) &= \sum_{k=0}^{n} 3^k \omega_k(\ell-k)\Lambda^{cn}_{k,\ell-k}(x), \\
\Phi_1''(x) &= \sum_{k=0}^{n} 3^k \left(\omega_k(\ell-k)\Lambda^{cn}_{k,\ell-k}(x) + (2^{n+1}+1)|\omega_k(\ell-k)|\tilde\Lambda^{cn}_{\ell}(x) \right).
\end{aligned}$$

Therefore, the proof will be complete once we show that

$$\begin{aligned}
\sum_{k=0}^{n} 3^k & \left(|\omega_k(\ell-k)|\Lambda^{cn}_{k,\ell-k}(x) - (2^{n+1}-1)|\omega_k(\ell-k)|\,|\tilde\Lambda^{cn}_{\ell}(x)| \right) \\
& \geqslant (1-\epsilon) \sum_{k=0}^{n} 3^k \left(|\omega_k(\ell-k)|\Lambda^{cn}_{k,\ell-k}(x) + (2^{n+1}+1)|\omega_k(\ell-k)|\,|\tilde\Lambda^{cn}_{\ell}(x)| \right).
\end{aligned}$$

Rearranging, it suffices to show that

$$\epsilon \sum_{k=0}^{n} 3^k |\omega_k(\ell-k)|\Lambda^{cn}_{k,\ell-k}(x) \geqslant 2^{n+2} \left(\sum_{k=0}^{n} 3^k |\omega_k(\ell-k)| \right) |\tilde\Lambda^{cn}_{\ell}(x)|.$$

Dropping all but the first term on the left-hand side, one arrives at the stronger inequality

$$\epsilon |\omega_0(\ell)|\Lambda^{cn}_{0,\ell}(x) \geqslant 2^{n+2} \left(\sum_{k=0}^{n} 3^k |\omega_k(\ell-k)| \right) |\tilde\Lambda^{cn}_{\ell}(x)|.$$

This final inequality follows immediately from the estimates in (7.3), (7.5), (7.14), and Lemma 7.4(iv). ∎

Lemma 7.7. *Let $P, Q\colon X^{cn} \to \mathbb{R}$ be polynomials with*

$$\deg P \leqslant \frac{1}{c} \deg^+_{1/3}(\neg f)\sqrt{n}, \tag{7.24}$$

$$\deg Q \leqslant \min\left\{ \frac{1}{c} \deg^+_{1/3}(\neg f), \frac{n}{2} \right\}. \tag{7.25}$$

Then

$$\langle \Phi_1', P \rangle = \langle \Phi_1'', Q \rangle = \langle \Phi_0, Q \rangle = 0.$$

Proof. By (7.6) and Lemma 7.3(iv), there are polynomials $p_0, p_1, \ldots, p_n$ such that

$$\langle \Lambda^{cn}_{k,m}, P \rangle = p_k(m) \qquad (k = 0, 1, \ldots, n; \quad m = 0, 1, \ldots, cn-k), \tag{7.26}$$

$$\deg p_k < \sqrt{n} \qquad (k = 0, 1, \ldots, n). \tag{7.27}$$

Therefore,

$$\begin{aligned}
\langle \Phi_1', P\rangle &= \sum_{k=0}^{n} 3^k \sum_{m=0}^{cn-k} \omega_k(m)\langle \Lambda_{k,m}^{cn}, P\rangle \\
&= \sum_{k=0}^{n} 3^k \langle \omega_k, p_k\rangle && \text{by (7.26)} \\
&= \sum_{k=0}^{n} 3^k \cdot 0 && \text{by (7.4) and (7.27)} \\
&= 0.
\end{aligned}$$

We now prove analogous claims for Φ_1'' and Φ_0. By (7.6) and Lemma 7.3(iv), there are reals (degree-zero univariate polynomials) $q_0, q_1, \ldots, q_n$ such that

$$\langle \Lambda_{k,m}^{cn}, Q\rangle = q_k \qquad (k = 0, 1, \ldots, n; \quad m = 0, 1, \ldots, cn - k). \tag{7.28}$$

Thus,

$$\begin{aligned}
\langle \Phi_1'', Q\rangle &= \sum_{k=0}^{n} 3^k \sum_{m=0}^{cn-k} \omega_k(m)\langle \Lambda_{k,m}^{cn}, Q\rangle && \text{by (7.25) and Lemma 7.4(i)} \\
&= \sum_{k=0}^{n} 3^k q_k \cdot \sum_{m=0}^{cn-k} \omega_k(m) && \text{by (7.28)} \\
&= \sum_{k=0}^{n} 3^k q_k \cdot 0 && \text{by (7.4)} \\
&= 0.
\end{aligned}$$

An analogous argument shows that Φ_1'' is orthogonal to polynomials of degree up to $\min\left\{\frac{1}{c}\deg_{1/3}^{+}(\neg f)\sqrt{n}, \frac{n}{2}\right\}$, but we will not need this improved bound.

Finally,

$$\begin{aligned}
\langle \Phi_0, Q\rangle &= \sum_{k=0}^{n} 3^k \left(\sum_{m=0}^{cn-k} |\omega_k(m)|\langle \Lambda_{k,m}^{cn}, Q\rangle - \langle \Lambda_{k,0}^{cn}, Q\rangle\right) \\
& && \text{by (7.25) and Lemma 7.4(i)} \\
&= \sum_{k=0}^{n} 3^k q_k \cdot \left(\sum_{m=0}^{cn-k} |\omega_k(m)| - 1\right) && \text{by (7.28)} \\
&= \sum_{k=0}^{n} 3^k q_k \cdot 0 && \text{by (7.1)} \\
&= 0.
\end{aligned}$$

□

Lemma 7.8. $\Phi_0, \Phi_1', \Phi_1'' \not\equiv 0$.

Proof. Let $x^* \in X$ be an arbitrary point with $\phi(x^*) > 0$, which exists by (7.10) and (7.11). In light of Lemmas 7.3(iii) and 7.4(ii), the defining equations for $\Phi_0, \Phi_1', \Phi_1''$

show that

$$\Phi_0(x^*,x^*,\ldots,x^*) = (|\omega_0(0)|-1)\Lambda_{0,0}^{cn}(x^*,x^*,\ldots,x^*), \tag{7.29}$$

$$\Phi_1'(x^*,x^*,\ldots,x^*) = \omega_0(0)\Lambda_{0,0}^{cn}(x^*,x^*,\ldots,x^*), \tag{7.30}$$

$$\Phi_1''(x^*,x^*,\ldots,x^*) = \omega_0(0)\Lambda_{0,0}^{cn}(x^*,x^*,\ldots,x^*). \tag{7.31}$$

Recall that $\Lambda_{0,0}^{cn}(x^*,x^*,\ldots,x^*) = \mu_1(x^*)^{cn} = 2^{cn}|\phi(x^*)|^{cn} > 0$ by definition, whereas $\omega_0(0) \in (1/4, 1/2]$ by (7.2) and (7.5). Therefore, the right-hand sides of (7.29)–(7.31) are nonzero. □

Lemmas 7.5 to 7.8 settle the required properties (i)–(viii) in Theorem 7.1, completing the proof.

8. High-accuracy approximation of the AND-OR tree.

As a final building block of our main result, we will now study the one-sided rational approximation of $\mathrm{AND}_n \circ \mathrm{OR}_r$ for arbitrary parameters n and r. To be more specific, we are interested in the numerator and denominator degree required for one-sided approximation with error 2^{-r}. We give a complete solution to this problem, with matching upper and lower bounds. We start with the upper bound, which is significantly simpler and is actually achieved for polynomials.

Theorem 8.1 (Upper bound). *There exists an absolute constant $c>0$ such that*

$$\deg^+_{2^{-r}}(\mathrm{AND}_n \circ \mathrm{OR}_r) \leqslant c \min\left\{r\sqrt{n}, n\right\}.$$

Proof. We consider two cases, depending on the value of r.

Case $1 \leqslant r \leqslant \sqrt{n}$. By Theorem 2.7, there is a polynomial $p\colon \{0,1\}^n \to [0,1]$ of degree $O(\sqrt{nr})$ with

$$|\mathrm{AND}_n(x) - p(x)| \leqslant 2^{-r-1}, \qquad x \in \{0,1\}^n. \tag{8.1}$$

Theorem 2.5 ensures that this approximating polynomial can be made highly robust to noise in the inputs with only a constant-factor increase in degree. More precisely, there exists a polynomial $p_{\mathrm{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(\sqrt{nr})$ with

$$|p(x) - p_{\mathrm{robust}}(x+\epsilon)| < 2^{-\sqrt{nr}-1}, \qquad x \in \{0,1\}^n,\ \epsilon \in [-1/3,1/3]^n. \tag{8.2}$$

Again by Theorem 2.7, there is a degree-$O(\sqrt{r})$ polynomial q with $\|\mathrm{OR}_r - q\|_\infty \leqslant 1/3$. By (8.1) and (8.2), the composed polynomial $p_{\mathrm{robust}} \circ q$ satisfies

$$\begin{aligned}\|\mathrm{AND}_n \circ \mathrm{OR}_r - p_{\mathrm{robust}} \circ q\|_\infty &\leqslant 2^{-r-1} + 2^{-\sqrt{nr}-1}\\ &\leqslant 2^{-r}.\end{aligned}$$

In particular, $p_{\mathrm{robust}} \circ q$ is a one-sided approximant for $\mathrm{AND}_n \circ \mathrm{OR}_r$ with error 2^{-r}. This completes the proof since $p_{\mathrm{robust}} \circ q$ has degree $\deg(p_{\mathrm{robust}})\deg(q) = O(r\sqrt{n}) = O(\min\{r\sqrt{n}, n\})$.

Case $r \geqslant \sqrt{n}$. Consider the polynomial $p(x) = \prod_{i=1}^n \sum_{j=1}^r x_{i,j}$. We have $p = 0$ whenever $\mathrm{AND}_n \circ \mathrm{OR}_r = 0$, and $p \geqslant 1$ whenever $\mathrm{AND}_n \circ \mathrm{OR}_r = 1$. Thus, p is a one-sided approximant for $\mathrm{AND}_n \circ \mathrm{OR}_r$ with error 0 and degree n. This completes the proof since $n \leqslant \min\{r\sqrt{n}, n\}$. □

To rephrase Theorem 8.1, $\mathrm{AND}_n \circ \mathrm{OR}_r$ can be approximated in a one-sided manner to within 2^{-r} by a rational function with denominator degree 0 and numerator degree $\Theta(\min\{r\sqrt{n}, n\})$. This construction turns out to be optimal in a strong sense: the numerator degree $\Theta(\min\{r\sqrt{n}, n\})$ is best possible even if one allows denominator degree as large as $\Theta(r)$. The formal statement follows.

Theorem 8.2 (Lower bound). *There is an absolute constant $c > 0$ such that*

$$R(\mathrm{AND}_n \circ \mathrm{OR}_r, cr, c\min\{r\sqrt{n}, n\}) > 2^{-r}. \tag{8.3}$$

The rest of section 8 is devoted to the proof of Theorem 8.2, which unlike the upper bound is quite lengthy and technical. We start by settling the degenerate cases (subsection 8.1), which facilitates the exposition of the general proof. The remainder of the argument (subsections 8.2 to 8.7) is structured to emphasize similarities with Theorem 7.1, whose proof was the subject of the previous section. In particular, we are able to reuse several key results from that earlier development. The principal point of departure is the new and challenging subsection 8.6, which constructs a corrector object for false positives. No such object was needed in the previous section. A lesser difference, in subsection 8.2, is the use of a high-accuracy dual polynomial for OR, in contrast to the bounded-error dual polynomial in the previous section.

8.1. Degenerate cases. Before proving Theorem 8.2 for general n and r, we first take care of the degenerate cases when either n or r is small.

Theorem 8.3. *For all integers $r \geqslant 1$,*

$$R(\mathrm{OR}_r, r-1, 0) = \frac{1}{\sqrt{2^r - 1}}. \tag{8.4}$$

Proof. Define $\phi_0, \phi_1 \colon \{0,1\}^r \to \mathbb{R}$ by $\phi_0(x) = (-1)^{x_1+x_2+\cdots+x_r}$ and

$$\phi_1(x) = \begin{cases} -\sqrt{2^r-1} & \text{if } x = 0, \\ 1/\sqrt{2^r-1} & \text{otherwise.} \end{cases}$$

It is straightforward to verify the following:

(i) $\phi_0 \geqslant |\phi_1|/\sqrt{2^r-1}$ on $\mathrm{OR}_r^{-1}(0)$,
(ii) $\phi_1 \geqslant |\phi_0|/\sqrt{2^r-1}$ on $\mathrm{OR}_r^{-1}(1)$,
(iii) $\langle \phi_0, p\rangle = 0$ whenever $\deg p \leqslant r-1$,
(iv) $\langle \phi_1, p\rangle = 0$ whenever $\deg p = 0$,
(v) $\phi_0 \not\equiv 0$ and $\phi_1 \not\equiv 0$.

As a result, $R(\mathrm{OR}_r, r-1, 0) \geqslant 1/\sqrt{2^r-1}$ by the dual characterization of one-sided rational approximation (Theorem 4.6).

For the matching upper bound, consider polynomials $p_0, p_1 \colon \{0,1\}^r \to \mathbb{R}$ given by $p_0(x) = 2^r(1 - \mathrm{OR}_r(x)) - (-1)^{x_1+x_2+\cdots+x_r}$ and $p_1(x) = \sqrt{2^r-1}$. The proof will be complete once we establish that:

(i) $\deg p_0 \leqslant r-1$,
(ii) $\deg p_1 = 0$,
(iii) $p_0 = 2^r - 1 = \sqrt{2^r-1}|p_1|$ on $\mathrm{OR}_r^{-1}(0)$,
(iv) $p_1 = \sqrt{2^r-1} \geqslant \sqrt{2^r-1}|p_0|$ on $\mathrm{OR}_r^{-1}(1)$.

The last three properties are obvious, whereas the first property follows from the representation

$$\mathrm{OR}_r(x) = 1 - \prod_{i=1}^{r} \frac{1+(-1)^{x_i}}{2}. \qquad \square$$

THEOREM 8.4. *There exist constants $c_1, c_2 > 0$ such that*

$$R(\mathrm{AND}_n, 0, c_1\sqrt{n}) \geqslant \frac{1}{\sqrt{2}},$$
$$R(\mathrm{AND}_n, 0, c_2\sqrt{n}) \leqslant \frac{1}{\sqrt{2}}.$$

Proof. By definition, $R(\mathrm{AND}_n, 0, d)$ is the infimum over $\epsilon > 0$ for which there exists a polynomial p of degree at most d with $p > 1/\epsilon$ on $\mathrm{AND}_n^{-1}(1)$ and $|p| < \epsilon$ on $\mathrm{AND}_n^{-1}(0)$. Now the claim is immediate from Theorem 2.6, which asserts that the minimum degree of a polynomial p with $p \geqslant 2/3$ on $\mathrm{AND}_n^{-1}(1)$ and $|p| \leqslant 1/3$ on $\mathrm{AND}_n^{-1}(0)$ equals $\Theta(\sqrt{n})$. □

The last two theorems settle the special case of Theorem 8.2 when either n or r is bounded by a constant. For example, Theorem 8.3 shows that (8.3) holds with $c = 1/100$ for all $n < 100$ and all r. Similarly, Theorem 8.4 shows that (8.3) holds with $c = \min\{c_1, 1\}/100$ for all $r < 100$ and all n. In particular, we may assume henceforth that

$$n \geqslant 12, \tag{8.5}$$
$$r \geqslant 11. \tag{8.6}$$

We may further assume without loss of generality that

$$n \equiv 0 \pmod 4, \tag{8.7}$$
$$r \equiv 1 \pmod 2. \tag{8.8}$$

These divisibility assumptions can be ensured in the usual manner, by working with a subfunction $\mathrm{AND}_{n'} \circ \mathrm{OR}_{r'}$ if necessary.

8.2. Outer dual object. Analogous to section 7, we start by constructing the "outer" dual object, so called because it serves as the glue that holds together the remaining building blocks of the construction. Recall from (8.5) and (8.7) that $n \geqslant 12$ is an integer divisible by 4. As a result, for each $k = 0, 1, 2, \ldots, n/2$, Theorem 2.14 ensures the existence of a function $\nu_k \colon \{0, 1, 2, \ldots, n-k\} \to \mathbb{R}$ such that

$$\deg p \leqslant c_{\mathrm{out}} \min\{\sqrt{nr}, n\} \implies \langle \nu_k, p\rangle = 0, \tag{8.9}$$
$$\|\nu_k\|_1 = 1, \tag{8.10}$$
$$\nu_k(t) = 0 \qquad (t = 1, 2, \ldots, \min\{r, n/4\} - 1), \tag{8.11}$$
$$(-1)^{k+t}\nu_k(t) \geqslant 0 \qquad (t \geqslant 1), \tag{8.12}$$
$$\nu_k(0) > c_{\mathrm{out}}^{\min\{r,n/4\}}, \tag{8.13}$$
$$|\nu_k(t)| \geqslant c_{\mathrm{out}}^{n} \qquad (t > n/2), \tag{8.14}$$
$$|\nu_k(t)| \leqslant \left(\frac{\min\{r, n/4\}}{c_{\mathrm{out}}\, t}\right)^{\min\{r,n/4\}} \qquad (t \geqslant 1), \tag{8.15}$$

for a sufficiently small constant

$$0 < c_{\mathrm{out}} < \frac{1}{4}. \tag{8.16}$$

It follows from (8.9) that $\langle \nu_k, 1\rangle = 0$ for all k, so that (8.10) and Proposition 2.12(ii) imply that

$$\|\nu_k\|_\infty \leqslant \frac{1}{2} \qquad (k = 0, 1, 2, \ldots, n/2). \tag{8.17}$$

Finally, we claim that

$$\left(\frac{\min\{r, n/4\}}{c_{\text{out}}\, t}\right)^{\min\{r,n/4\}} \leqslant \left(\frac{\min\{r, n/4\}}{c_{\text{out}}\, t}\right)^{r} \qquad (t = 1, 2, \ldots, n).$$

Indeed, the inequality is trivial when $r \leqslant n/4$ and follows from (8.16) otherwise. In particular, (8.15) gives

$$|\nu_k(t)| \leqslant \left(\frac{\min\{r, n/4\}}{c_{\text{out}}\, t}\right)^{r} \qquad (t \geqslant 1). \tag{8.18}$$

8.3. Inner dual objects. We now turn our attention to the "inner" dual object, whose domain $\{0,1\}^r$ is the domain of the inner function in the composition $\text{AND}_n \circ \text{OR}_r$. Recall from (8.8) that $r \geqslant 11$ is an odd integer. Let

$$\epsilon = \frac{c_{\text{out}}^6}{2 + c_{\text{out}}^6}. \tag{8.19}$$

Taking κ in Theorem 2.13 to be the uniform distribution over $\{2, 4, 6, \ldots, r-1\}$, we infer the existence of a function $\omega\colon \{0, 1, 2, \ldots, r\} \to \mathbb{R}$ such that

$$\|\omega\|_1 = 1, \tag{8.20}$$

$$\omega(0) > \frac{1-\epsilon}{2}, \tag{8.21}$$

$$(-1)^{t+1}\omega(t) \geqslant 0 \qquad (t = 1, 2, \ldots, r), \tag{8.22}$$

$$|\omega(t)| \geqslant \frac{2\epsilon}{3(r-1)} \qquad (t = 2, 4, 6, \ldots, r-1), \tag{8.23}$$

$$\deg p < c_{\text{in}}\sqrt{r} \implies \langle \omega, p\rangle = 0, \tag{8.24}$$

for some constant $c_{\text{in}} = c_{\text{in}}(\epsilon)$ with $0 < c_{\text{in}} < 1$. Now define $\phi\colon \{0,1\}^r \to \mathbb{R}$ by

$$\phi(x) = -\binom{r}{|x|}^{-1} \omega(|x|).$$

Then

$$\|\phi\|_1 = 1, \tag{8.25}$$

$$\phi(0^r) < -\frac{1-\epsilon}{2}, \tag{8.26}$$

$$(-1)^{|x|}\phi(x) \geqslant 0 \qquad (x \neq 0^r), \tag{8.27}$$

$$|\phi(x)| \geqslant \frac{2\epsilon}{3(r-1)}\binom{r}{|x|}^{-1} \qquad (|x| = 2, 4, 6, \ldots, r-1) \tag{8.28}$$

by (8.20)–(8.23), respectively. In addition, it is straightforward to deduce from (8.24) that for any polynomial $p\colon \{0,1\}^r \to \mathbb{R}$,

$$\deg p < c_{\text{in}}\sqrt{r} \implies \langle \phi, p\rangle = 0, \tag{8.29}$$

with the notable special case

$$\langle \phi, 1\rangle = 0. \tag{8.30}$$

This can be seen by writing $\langle \phi, p\rangle = -\sum_{t=0}^{r}\omega(t)\,\mathbf{E}_{|x|=t}\,p(x)$ and recalling from Proposition 2.2 that $t \mapsto \mathbf{E}_{|x|=t}\,p(x)$ is a univariate polynomial on $\{0,1,2,\ldots,r\}$ of degree at most $\deg p$. Finally, we note that

$$\begin{aligned}\langle \phi, \mathrm{OR}_r\rangle &= \langle \phi, \mathrm{OR}_r - 1\rangle\\ &= -\phi(0^r)\\ &> \frac{1-\epsilon}{2},\end{aligned} \tag{8.31}$$

where the first step uses (8.30) and the third step uses (8.26). We define α by

$$\frac{1-\alpha}{2} = \langle \phi, \mathrm{OR}_r\rangle. \tag{8.32}$$

Then

$$0 \leqslant \alpha < \epsilon, \tag{8.33}$$

where the upper bound is immediate from (8.31), and the lower bound holds by (8.25), (8.30), and Proposition 2.12(iii).

Analogous to the development in section 7, we now consider several probability distributions that ϕ induces on $\{0,1\}^r$. By (8.25), the function $|\phi|$ itself is a probability distribution on $\{0,1\}^r$. We further define μ_0 and μ_1 to be the probability distributions induced by $|\phi|$ on the sets $\{x : \phi(x) < 0\}$ and $\{x : \phi(x) > 0\}$, respectively. In particular, (8.26) shows that

$$\mu_0(0^r) > 0, \tag{8.34}$$

$$\mu_1(0^r) = 0. \tag{8.35}$$

Equations (8.25) and (8.30) imply that

$$\phi = \frac{1}{2}\mu_1 - \frac{1}{2}\mu_0, \tag{8.36}$$

with μ_0 and μ_1 well-defined. For every nonzero input $x \in \{0,1\}^r$ of even Hamming weight, we have $\phi(x) \geqslant \frac{2\epsilon}{3r}\binom{r}{|x|}^{-1}$ from (8.27) and (8.28), whence

$$\begin{aligned}\mu_1(x) &= 2|\phi|\\ &\geqslant \frac{4\epsilon}{3r}\binom{r}{|x|}^{-1} \qquad (|x| = 2,4,6,\ldots,r-1).\end{aligned} \tag{8.37}$$

Multiplying (8.36) on both sides by OR_r and applying (8.35), we find that $\langle \mathrm{OR}_r, \phi\rangle = \frac{1}{2}\langle \mathrm{OR}_r, \mu_1\rangle - \frac{1}{2}\langle \mathrm{OR}_r, \mu_0\rangle = \frac{1}{2} - \frac{1}{2}\langle \mathrm{OR}_r, \mu_0\rangle$, which in view of (8.32) gives

$$\langle \mathrm{OR}_r, \mu_0\rangle = \alpha, \tag{8.38}$$

$$\langle 1 - \mathrm{OR}_r, \mu_0\rangle = 1 - \alpha. \tag{8.39}$$

In particular,

$$\begin{aligned}\langle \mathrm{OR}_r \cdot \mu_0 - \alpha\mu_0, 1\rangle &= \langle \mathrm{OR}_r, \mu_0\rangle - \alpha\langle \mu_0, 1\rangle \\ &= \langle \mathrm{OR}_r, \mu_0\rangle - \alpha \\ &= 0. \end{aligned} \tag{8.40}$$

Since we defined μ_0 and μ_1 in terms of ϕ exactly as in section 7, an analogue of Lemma 7.2 applies here as well, with the same proof as before. We restate it here, in the notation of this section.

LEMMA 8.5. *Let $\xi\colon \{0,1\}^r \to \mathbb{R}$ be an arbitrary function. Then for every polynomial $P\colon (\{0,1\}^r)^n \to \mathbb{R}$ and every $k = 0,1,2,\ldots,n$, the mapping*

$$z \mapsto \left\langle \xi^{\otimes k} \otimes \bigotimes_{i=1}^{n-k} \mu_{z_i}, P \right\rangle, \qquad z \in \{0,1\}^{n-k}, \tag{8.41}$$

is a polynomial of degree at most $(\deg P)/(c_{\mathrm{in}}\sqrt{r})$.

Proof. Analogous to Lemma 7.2. □

8.4. Auxiliary distributions. Analogous to section 7, we will now use μ_0 and μ_1 to construct auxiliary functions $\Lambda^n_{k,m}$ on the tensor space $(\{0,1\}^r)^n$. For nonnegative integers k, m with $k + m \leqslant n$, we define a nonnegative function

$$\Lambda^n_{k,m}(x_1, x_2, \ldots, x_n) = \mathop{\mathbf{E}}_{S,T}\left[\prod_{i\in S} \mathrm{OR}_r(x_i)\mu_0(x_i) \cdot \prod_{i\in T} \mu_0(x_i) \cdot \prod_{i\notin S\cup T} \mu_1(x_i)\right],$$

where the expectation is taken over a uniformly random pair of disjoint sets $S, T \subseteq \{1,2,\ldots,n\}$ of size $|S| = k$ and $|T| = m$. Observe that this definition is identical to the one in section 7 with $f = \mathrm{OR}_r$. In particular, Lemma 7.3 applies in its entirety, with the same proof. For convenience, we restate the lemma here in the notation of this section.

LEMMA 8.6.
(i) $\operatorname{supp} \Lambda^n_{k,0} \subseteq (\{0,1\}^r \setminus \{0^r\})^n$,
(ii) $\Lambda^n_{k,m} = \Lambda^n_{k+m,0}$ *on* $(\{0,1\}^r \setminus \{0^r\})^n$,
(iii) $\Lambda^n_{k,m}(x) \neq 0$ *only if* $|\{i : \phi(x_i) < 0\}| = k + m$,
(iv) *for an arbitrary real polynomial* $P\colon (\{0,1\}^r)^n \to \mathbb{R}$, *the map* $m \mapsto \langle \Lambda^n_{k,m}, P\rangle$ $(m = 0,1,2,\ldots,n-k)$ *is a polynomial of degree at most* $(\deg P)/(c_{\mathrm{in}}\sqrt{r})$.

Proof. Analogous to the proof of Lemma 7.3, with the obvious difference that the appeal to Lemma 7.2 in part (iv) should be replaced with its counterpart from this section (Lemma 8.5). □

8.5. Corrector for false negatives. As one can see from the dual characterization of one-sided rational approximation (Theorem 4.6), the dual objects that we are to construct must exhibit very specific sign behavior. To that end, we will use "corrector" objects to force the correct sign on the relevant portions on the domain. Analogous to the development in section 7, these corrector objects are orthogonal to low-degree polynomials and are close to zero on all but a handful of relevant inputs. Here, we build a corrector object for $(\{0,1\}^r \setminus \{0^r\})^n$, corresponding to the inputs where $\mathrm{AND}_n \circ \mathrm{OR}_r$ evaluates to true.

For an integer k with $1 \leqslant k \leqslant n$, let $\tilde{\Lambda}^n_k \colon (\{0,1\}^r)^n \to \mathbb{R}$ be given by

$$\begin{aligned}\tilde{\Lambda}^n_k(x) &= \frac{1}{(\lceil k/2 \rceil - 1)!}\\ &\times \mathop{\mathbf{E}}_{|S|=k}\left[\prod_{i\in S}\frac{(\mathrm{OR}_r(x_i)-\alpha)\mu_0(x_i)}{1-\alpha}\cdot\prod_{i=\lfloor k/2\rfloor+1}^{k-1}\left(\sum_{j\in S}\mathrm{OR}_r(x_j)-i\right)\cdot\prod_{i\notin S}\mu_1(x_i)\right],\end{aligned}$$

where the expectation is over a uniformly random set $S \subseteq \{1,2,\ldots,n\}$ of size $|S| = k$. Observe that this definition is identical to that in section 7, with $f = \mathrm{OR}_r$. In particular, Lemma 7.4 from that section carries over in its entirety, with the same proof as before. We restate it here in the notation of this section.

Lemma 8.7.

(i) $\langle \tilde{\Lambda}^n_k, P\rangle = 0$ for every polynomial P of degree at most $k/2$,

(ii) $\tilde{\Lambda}^n_k(x) \neq 0$ only if $|\{i : \phi(x_i) < 0\}| = k$,

(iii) $\tilde{\Lambda}^n_k = \Lambda^n_{0,k}$ on $(\{0,1\}^r \setminus \{0^r\})^n$,

(iv) $|\tilde{\Lambda}^n_k| \leqslant \frac{1}{2}\left(\frac{2\alpha}{1-\alpha}\right)^{k/2}\Lambda^n_{0,k}$ outside $(\{0,1\}^r \setminus \{0^r\})^n$.

Proof. Analogous to Lemma 7.4. □

8.6. Corrector for false positives. We will now build a corrector object for the complementary portion of the domain, where $\mathrm{AND}_n \circ \mathrm{OR}_r$ evaluates to false. This part of the proof is the most challenging and has no analogue in section 7. The crux of our argument is the following technical result.

Theorem 8.8. *Let N, R, d be positive integers, $d \leqslant R/5$. Then there exists a function $Z\colon (\{0,1\}^R)^N \to \mathbb{R}$ such that*

$$\langle Z, P\rangle = 0 \text{ whenever } \deg P \leqslant d, \tag{8.42}$$

$$\operatorname{supp} Z \subseteq \{0^{RN}\} \cup (\{0,1\}^R \setminus \{0^R\})^N, \tag{8.43}$$

$$Z(0^{RN}) = 1, \tag{8.44}$$

$$|Z(x_1,x_2,\ldots,x_N)| \leqslant 60^d\left(\frac{6}{R+1}\right)^N\prod_{i=1}^N\binom{R}{|x_i|}^{-1} \quad \text{for } x \neq 0^{RN}. \tag{8.45}$$

Proof. For subsets $S_1, S_2, \ldots, S_{2d} \subseteq \{1,2,\ldots,N\} \times \{1,2,\ldots,R\}$, consider the function $Z_{S_1,S_2,\ldots,S_{2d}}\colon (\{0,1\}^R)^N \to \mathbb{R}$ given by

$$Z_{S_1,S_2,\ldots,S_{2d}}(x) = \sum_{T\subseteq\{1,2,\ldots,2d\}}\binom{d-1-|T|}{d-1}(-1)^{|T|}\,\mathbf{I}[x = \mathbf{1}_{\bigcup_{j\in T}S_j}].$$

We will refer to a family of sets $S_1, S_2, \ldots, S_{2d} \subseteq \{1,2,\ldots,N\}\times\{1,2,\ldots,R\}$ as *good* if they meet the following three criteria:

$$\begin{aligned}&S_j \cap S_{j'} = \varnothing && (\forall\, j \neq j'),\\ &S_j \cap \{(1,1),(1,2),\ldots,(1,R)\} \neq \varnothing && (j = 1,2,\ldots,2d),\\ &\left(\bigcup_{j\in T}S_j\right)\cap\{(i,1),(i,2),\ldots,(i,R)\}\neq\varnothing && (i = 1,2,\ldots,N,\quad |T| \geqslant d).\end{aligned}$$

Claim 8.9. *Let $S_1, S_2, \ldots, S_{2d} \subseteq \{1, 2, \ldots, N\} \times \{1, 2, \ldots, R\}$ be a good set family. Then*

(i) $\langle Z_{S_1,S_2,\ldots,S_{2d}}, P\rangle = 0$ *whenever* $\deg P \leqslant d$,
(ii) $\operatorname{supp} Z_{S_1,S_2,\ldots,S_{2d}} \subseteq \{0^{RN}\} \cup (\{0,1\}^R \setminus \{0^R\})^N$,
(iii) $Z_{S_1,S_2,\ldots,S_{2d}}(0^{RN}) = 1$.

For clarity of exposition, we provide the proof of this and other claims after the proof of the theorem. Now let $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2d} \subseteq \{1, 2, \ldots, N\} \times \{1, 2, \ldots, R\}$ be a random set family generated by Algorithm 8.1.

Algorithm 8.1 Procedure for generating $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2d}$.

```
Choose $p_1, p_2, \ldots, p_N \in [1/2, 1]$ uniformly at random.
Let $\iota\colon \{1,2,\ldots,N\} \times \{1,2,\ldots,R\} \to \{0,1,2,\ldots,2d\}$ be a random function
    whose values at each point of its domain are independent random variables,
    distributed according to
        $\iota(i,j) = \begin{cases} 0 & \text{with probability } 1 - p_i, \\ 1 & \text{with probability } p_i/2d, \\ 2 & \text{with probability } p_i/2d, \\ \vdots & \\ 2d & \text{with probability } p_i/2d. \end{cases}$
Define $\mathbf{S}_j = \iota^{-1}(j)$ for $j = 1, 2, \ldots, 2d$.
```

Claim 8.10. $\mathbf{P}[\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2d}$ *are good*$] > \mathrm{e}^{-2d}(3/4)^{N-1}$.

Claim 8.11. *For $x \neq 0^{RN}$,*

$$\mathop{\mathbf{E}}_{\mathbf{S}_1,\mathbf{S}_2,\ldots,\mathbf{S}_{2d}} |Z_{\mathbf{S}_1,\mathbf{S}_2,\ldots,\mathbf{S}_{2d}}(x)| \leqslant 8^d \left(\frac{4}{R+1}\right)^N \prod_{i=1}^N \left(\frac{R}{x_{i,1} + x_{i,2} + \cdots + x_{i,R}}\right)^{-1}.$$

We now complete the proof of the theorem by combining Claims 8.9 to 8.11. Define

$$Z(x) = \mathop{\mathbf{E}}_{\mathbf{S}_1,\mathbf{S}_2,\ldots,\mathbf{S}_{2d}} [Z_{\mathbf{S}_1,\mathbf{S}_2,\ldots,\mathbf{S}_{2d}}(x) \mid \mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2d} \text{ are good}].$$

Then properties (8.42)–(8.44) are immediate by Claim 8.9, whereas the remaining property (8.45) follows from Claims 8.10 and 8.11. □

Proof of Claim 8.9. (i) Recall from the first two properties of a good set family that $S_1, S_2, \ldots, S_{2d}$ are nonempty and pairwise disjoint. As a result,

$$Z_{S_1,S_2,\ldots,S_{2d}}(x) = \binom{d - 1 - \sum_{j=1}^{2d} \mathbf{I}[x|_{S_j} = 11\ldots1]}{d-1} \times \mathbf{I}[x|_{\overline{S_1 \cup S_2 \cup \cdots \cup S_{2d}}} = 00\ldots0] \prod_{j=1}^{2d} (\mathbf{I}[x|_{S_j} = 00\ldots0] - \mathbf{I}[x|_{S_j} = 11\ldots1]).$$

Multiplying out the binomial coefficient, we find that $Z_{S_1,S_2,\ldots,S_{2d}}$ is a linear combination of functions of the form

$$\begin{aligned}\mathbf{I}[x|_{\overline{S_1 \cup S_2 \cup \cdots \cup S_{2d}}} = 00\ldots0] \prod_{j\in A} \mathbf{I}[x|_{S_j} = 11\ldots1] \\ \times \prod_{j\notin A} (\mathbf{I}[x|_{S_j} = 00\ldots0] - \mathbf{I}[x|_{S_j} = 11\ldots1]),\end{aligned}$$

with $A \subseteq \{1,2,\ldots,2d\}$ ranging over sets of size at most $d-1$. This is a product of $2d+1$ functions on disjoint sets of variables, where the final $2d-|A| \geqslant d+1$ functions are orthogonal to polynomials of degree less than 1. As a result, the entire product is orthogonal to polynomials of degree less than $(2d-|A|)\cdot 1 \geqslant d+1$.

(ii) Fix an arbitrary input $x \neq 0^{RN}$ in the support of $Z_{S_1,S_2,\ldots,S_{2d}}$. Then

$$\begin{aligned} Z_{S_1,S_2,\ldots,S_{2d}}(x) &= \sum_{T\subseteq\{1,2,\ldots,2d\}} \binom{d-1-|T|}{d-1} (-1)^{|T|}\, \mathbf{I}[x = \mathbf{1}_{\bigcup_{j\in T} S_j}] \\ &= \sum_{\substack{T\subseteq\{1,2,\ldots,2d\}\\ T\neq\varnothing}} \binom{d-1-|T|}{d-1} (-1)^{|T|}\, \mathbf{I}[x = \mathbf{1}_{\bigcup_{j\in T} S_j}] \\ &= \sum_{\substack{T\subseteq\{1,2,\ldots,2d\}\\ |T|\geqslant d}} \binom{d-1-|T|}{d-1} (-1)^{|T|}\, \mathbf{I}[x = \mathbf{1}_{\bigcup_{j\in T} S_j}], \end{aligned}$$

where the first equality holds by definition, the second uses the fact that $x \neq 0^{RN}$, and the third follows from the definition of a binomial coefficient. We conclude that $x = \mathbf{1}_{\bigcup_{j\in T} S_j}$ for some set T of cardinality at least d. But by the third property of a good set family, the union of any d sets from among $S_1, S_2, \ldots, S_{2d}$ intersects each of the sets

$$\begin{gathered} \{1\}\times\{1,2,\ldots,R\}, \\ \{2\}\times\{1,2,\ldots,R\}, \\ \vdots \\ \{N\}\times\{1,2,\ldots,R\}. \end{gathered}$$

As a result, $\bigwedge_{i=1}^{N} \bigvee_{j=1}^{R} x_{ij} = 1$.

(iii) Recall from the second property of a good set family that $S_1, S_2, \ldots, S_{2d}$ are nonempty. As a result, the claim is immediate from the definition of $Z_{S_1,S_2,\ldots,S_{2d}}$. □

Proof of Claim 8.10. By construction, $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2d}$ are pairwise disjoint. As a result,

$$\mathbf{P}[\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2d} \text{ are good}] = \mathbf{P}[E_1 \wedge E_2 \wedge \cdots \wedge E_N],$$

where E_1 denotes the event that

$$S_j \cap \{(1,1),(1,2),\ldots,(1,R)\} \neq \varnothing, \qquad j = 1,2,\ldots,2d,$$

and E_i $(i = 2,3,\ldots,N)$ denotes the event that

$$\left(\bigcup_{j\in T} S_j\right) \cap \{(i,1),(i,2),\ldots,(i,R)\} \neq \varnothing, \qquad T \in \binom{\{1,2,\ldots,2d\}}{d}.$$

Since $E_1, E_2, \ldots, E_N$ are independent, we further obtain

$$\begin{aligned}\mathbf{P}[\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2d} \text{ are good}] &= \prod_{i=1}^{N} \mathbf{P}[E_i] \\ &= \mathbf{P}[E_1]\,\mathbf{P}[E_2]^{N-1}, \qquad (8.46)\end{aligned}$$

where the second step holds by symmetry. In what follows, $p_1, p_2, \ldots, p_N \in [1/2, 1]$ and $\iota\colon \{1, 2, \ldots, N\} \times \{1, 2, \ldots, R\} \to \{0, 1, 2, \ldots, 2d\}$ refer to the random variables in Algorithm 8.1.

Rephrasing, E_1 is the event that the sequence

$$\iota(1,1), \iota(1,2), \iota(1,3), \ldots, \iota(1,R) \qquad (8.47)$$

contains each of the numbers $1, 2, 3, \ldots, 2d$, in some order. Conditioned on p_1, the R random variables in (8.47) are independent and identically distributed, each taking on $0, 1, 2, \ldots, 2d$ with probability $1 - p_1, p_1/2d, p_1/2d, \ldots, p_1/2d$, respectively. Since $p_1 \geqslant 1/2$ and $R \geqslant 4d$, with probability at least $1/2$ the sequence contains at least $2d$ nonzeroes. Conditioned on this event, the probability that the sequence features each of the numbers $1, 2, \ldots, 2d$ is at least $(2d)!/(2d)^{2d}$. In summary,

$$\mathbf{P}[E_1] \geqslant \frac{(2d)!}{2(2d)^{2d}} > \exp(-2d)\,. \qquad (8.48)$$

We now analyze E_2. Fix an arbitrary set $T \subseteq \{1, 2, \ldots, 2d\}$ of cardinality d. Observe from the definition of $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{2d}$ that $\bigcup_{j \in T} \mathbf{S}_j$ is a random subset of $\{1, 2, \ldots, N\} \times \{1, 2, \ldots, R\}$ obtained by choosing $p_1, p_2, \ldots, p_N \in [1/2, 1]$ independently and uniformly at random and including each element $(i', j') \in \{1, 2, \ldots, N\} \times \{1, 2, \ldots, R\}$ independently with probability $|T| \cdot p_{i'}/2d = p_{i'}/2$. In particular,

$$\left(\bigcup_{j \in T} \mathbf{S}_j\right) \cap \{(2,1), (2,2), (2,3), \ldots, (2,R)\} \qquad (8.49)$$

is a random set obtained by including each of $(2,1), (2,2), (2,3), \ldots, (2,R)$ independently with probability $p_2/2$. Therefore,

$$\begin{aligned}\mathbf{P}\left[\left(\bigcup_{j \in T} \mathbf{S}_j\right) \cap \{(2,1), (2,2), (2,3), \ldots, (2,R)\} = \varnothing\right] &= 2\int_{1/2}^{1} \left(1 - \frac{p_2}{2}\right)^R dp_2 \\ &\leqslant 2\int_{1/2}^{2} \left(1 - \frac{p_2}{2}\right)^R dp_2 \\ &= \frac{3}{R+1} \cdot \left(\frac{3}{4}\right)^R.\end{aligned}$$

Recall that E_2 is the event that (8.49) is nonempty for each T of cardinality d. Therefore,

$$\begin{aligned}\mathbf{P}[E_2] &\geqslant 1 - \binom{2d}{d} \cdot \frac{3}{R+1} \left(\frac{3}{4}\right)^R \\ &\geqslant 1 - \binom{2d}{d} \cdot \frac{3}{5d+1} \cdot \left(\frac{3}{4}\right)^{5d} \\ &\geqslant \frac{3}{4}, \qquad (8.50)\end{aligned}$$

where the first step uses the union bound and second step uses $d \leqslant R/5$. By (8.46), (8.48), and (8.50), the proof is complete. $\square$

Proof of Claim 8.11. Let $T \subseteq \{1, 2, \ldots, 2d\}$ be an arbitrary nonempty set. Recall that $\bigcup_{j\in T} \mathbf{S}_j$ is a random set obtained by choosing $p_1, p_2, \ldots, p_N \in [1/2, 1]$ independently and uniformly at random and including each element $(i', j') \in \{1, 2, \ldots, N\} \times \{1, 2, \ldots, R\}$ independently with probability $|T|p_{i'}/2d$. Abbreviating

$$x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,R}),$$

we obtain:

$$\begin{aligned}
\mathbf{P}\left[\mathbf{1}_{\bigcup_{j\in T}\mathbf{S}_j} = x\right] &= \prod_{i=1}^{N} 2\int_{1/2}^{1} \left(\frac{|T|p_i}{2d}\right)^{|x_i|} \left(1-\frac{|T|p_i}{2d}\right)^{R-|x_i|} dp_i \\
&\leqslant \prod_{i=1}^{N} 2\int_{0}^{2d/|T|} \left(\frac{|T|p_i}{2d}\right)^{|x_i|} \left(1-\frac{|T|p_i}{2d}\right)^{R-|x_i|} dp_i \\
&= \left(\frac{4d}{|T|}\right)^{N} \prod_{i=1}^{N} \int_0^1 p^{|x_i|}\,(1-p)^{R-|x_i|}\; dp \\
&= \left(\frac{4d}{|T|(R+1)}\right)^{N} \prod_{i=1}^{N} \binom{R}{|x_i|}^{-1}, \qquad (8.51)
\end{aligned}$$

where the final step follows by Fact 2.1(iii). Therefore,

$$\begin{aligned}
&\mathbf{E}\,|Z_{\mathbf{S}_1,\mathbf{S}_2,\ldots,\mathbf{S}_{2d}}(x)| \\
&\quad\leqslant \sum_{T\subseteq\{1,2,\ldots,2d\}} \left|\binom{d-1-|T|}{d-1}\right| \mathbf{P}[x = \mathbf{1}_{\bigcup_{j\in T}\mathbf{S}_j}] \\
&\quad= \sum_{\substack{T\subseteq\{1,2,\ldots,2d\}\\ |T|\geqslant d}} \binom{|T|-1}{d-1} \mathbf{P}[x = \mathbf{1}_{\bigcup_{j\in T}\mathbf{S}_j}] + \mathbf{I}[x = 0^{RN}] \\
&\quad= \sum_{\substack{T\subseteq\{1,2,\ldots,2d\}\\ |T|\geqslant d}} \binom{|T|-1}{d-1} \mathbf{P}[x = \mathbf{1}_{\bigcup_{j\in T}\mathbf{S}_j}] && \text{since } x \neq 0^{RN} \\
&\quad\leqslant \sum_{\substack{T\subseteq\{1,2,\ldots,2d\}\\ |T|\geqslant d}} \binom{|T|}{d} \left(\frac{4}{R+1}\right)^{N} \prod_{i=1}^{N} \binom{R}{|x_i|}^{-1} && \text{by (8.51)} \\
&\quad= \sum_{t=d}^{2d} \binom{2d}{t}\binom{t}{d} \left(\frac{4}{R+1}\right)^{N} \prod_{i=1}^{N} \binom{R}{|x_i|}^{-1} \\
&\quad= 2^d \binom{2d}{d} \left(\frac{4}{R+1}\right)^{N} \prod_{i=1}^{N} \binom{R}{|x_i|}^{-1}. \qquad \square
\end{aligned}$$

This completes the proof of Theorem 8.8. We will now reinterpret it in our setting of interest, relating it among other things to the probability distribution μ_1.

COROLLARY 8.12. *For every integer $t \geqslant 1$, there exists $Z_t\colon (\{0,1\}^r)^t \to \mathbb{R}$ such* *that*

$$Z_t(0^{rt}) = 1, \tag{8.52}$$

$$|Z_t| \leqslant 60^{r/10}\,(9/\epsilon)^t \mu_1^{\otimes t} \textit{ on } (\{0,1\}^r)^t \setminus \{0^{rt}\}, \tag{8.53}$$

$$\langle Z_t, P\rangle = 0 \textit{ whenever } \deg P \leqslant r/10. \tag{8.54}$$

Proof. Recall from (8.6) that $r \geqslant 11$. Taking $R = \lfloor r/2 \rfloor$ and $d = \lfloor r/10 \rfloor$ in Theorem 8.8, we infer the existence of $Z\colon (\{0,1\}^R)^t \to \mathbb{R}$ such that

$$Z(0^{Rt}) = 1, \tag{8.55}$$

$$\operatorname{supp} Z \subseteq \{0^{Rt}\} \cup (\{0,1\}^R \setminus 0^R)^t, \tag{8.56}$$

$$|Z(y_1,\ldots,y_t)| \leqslant 60^{\frac{r}{10}} \left(\frac{6}{R+1}\right)^t \prod_{i=1}^t \binom{R}{|y_i|}^{-1} \text{ for } (y_1,\ldots,y_t) \neq 0^{Rt}, \tag{8.57}$$

$$\deg P \leqslant r/10 \implies \langle Z, P\rangle = 0. \tag{8.58}$$

Define $Z_t\colon (\{0,1\}^r)^t \to \mathbb{R}$ by

$$Z_t(x_1,x_2,\ldots,x_t) = \sum_{\substack{y_1\in\{0,1\}^R\\ |y_1|=|x_1|/2}} \sum_{\substack{y_2\in\{0,1\}^R\\ |y_2|=|x_2|/2}} \cdots \sum_{\substack{y_t\in\{0,1\}^R\\ |y_t|=|x_t|/2}} Z(y_1,y_2,\ldots,y_t) \prod_{i=1}^t \binom{r}{|x_i|}^{-1}$$

if $|x_1|, |x_2|, \ldots, |x_t|$ are all even, and $Z_t(x_1,x_2,\ldots,x_t) = 0$ otherwise.

We proceed to verify the three properties required of Z_t. The first property, (8.52), is immediate from (8.55). To verify (8.53), fix an arbitrary input $(x_1,x_2,\ldots,x_t) \neq 0^{rt}$. There are three cases to examine. If at least one of $|x_1|, |x_2|, \ldots, |x_t|$ is odd, we have $Z_t(x_1,x_2,\ldots,x_t) = 0$ by definition and thus (8.53) holds trivially. If $|x_1|, |x_2|, \ldots, |x_t|$ are all even but not all positive, then by (8.56) we again have $Z_t(x_1,x_2,\ldots,x_t) = 0$. In the remaining case that $|x_1|, |x_2|, \ldots, |x_t|$ are even and positive, (8.37) and (8.57) imply that

$$\prod_{i=1}^t \mu_1(x_i) \geqslant \left(\frac{4\epsilon}{3r}\right)^t \prod_{i=1}^t \binom{r}{|x_i|}^{-1},$$

$$|Z_t(x_1,x_2,\ldots,x_t)| \leqslant 60^{r/10} \left(\frac{6}{R+1}\right)^t \prod_{i=1}^t \binom{r}{|x_i|}^{-1},$$

respectively, again forcing (8.53).

It remains to verify (8.54). Let $P\colon (\{0,1\}^r)^t \to \mathbb{R}$ be an arbitrary polynomial of degree at most $r/10$. By Corollary 2.3, there is a polynomial $Q\colon \mathbb{R}^t \to \mathbb{R}$ of degree at most $r/10$ such that

$$\mathop{\mathbf{E}}_{\sigma_1\in S_r} \cdots \mathop{\mathbf{E}}_{\sigma_t\in S_r} P(\sigma_1 x_1,\ldots,\sigma_t x_t) = Q(|x_1|,\ldots,|x_t|) \tag{8.59}$$

for all $x \in (\{0,1\}^r)^t$. Since $Z_t(x_1,\ldots,x_t)$ is uniquely determined by $|x_1|,\ldots,|x_t|$, we have $Z_t(x_1,\ldots,x_t) = Z_t(\sigma_1 x_1,\ldots,\sigma_t x_t)$ for all permutations $\sigma_1,\ldots,\sigma_t \in S_r$. In particular,

$$\begin{aligned}\langle Z_t, P\rangle &= \sum_{x\in(\{0,1\}^r)^t} Z_t(x_1,\ldots,x_t)P(x_1,\ldots,x_t)\\ &= \sum_{x\in(\{0,1\}^r)^t} Z_t(x_1,\ldots,x_t) \mathop{\mathbf{E}}_{\sigma_1\in S_1}\cdots\mathop{\mathbf{E}}_{\sigma_t\in S_t} P(\sigma_1 x_1,\ldots,\sigma_t x_t)\\ &= \sum_{x\in(\{0,1\}^r)^t} Z_t(x_1,\ldots,x_t)Q(|x_1|,\ldots,|x_t|) && \text{by (8.59)}\\ &= \sum_{y\in(\{0,1\}^R)^t} Z(y_1,\ldots,y_t)Q(2|y_1|,\ldots,2|y_t|) && \text{by definition of } Z_t\\ &= 0 && \text{by (8.58).} \qquad \square\end{aligned}$$

At last, the following theorem constructs our desired corrector object, which allows us to force the correct sign on all inputs outside $(\{0,1\}^r\setminus\{0^r\})^n$.

Theorem 8.13. *Let $k\geqslant 0$ and $m\geqslant 1$ be integers, where $m+k\leqslant n$. Then there is a function $\tilde\Lambda^n_{k,m}\colon(\{0,1\}^r)^n\to\mathbb{R}$ such that:*

(i) *$\tilde\Lambda^n_{k,m}=\Lambda^n_{k,m}$ outside $(\{0,1\}^r\setminus\{0^r\})^n$;*

(ii) *$|\tilde\Lambda^n_{k,m}|\leqslant(\max\{1,\sqrt{m/r}\}/c_{\rm cor})^r\sum_{i=0}^m\Lambda^n_{k,i}$ on $(\{0,1\}^r\setminus\{0^r\})^n$ for some constant $c_{\rm cor}=c_{\rm cor}(\epsilon)$ with $0<c_{\rm cor}<1$;*

(iii) *$\langle\tilde\Lambda^n_{k,m},P\rangle=0$ whenever $\deg P\leqslant\min\{r/10, c_{\rm in}r\sqrt{r}/2\}$.*

Proof. The proof has two parts to it, corresponding to m small and m large. The constructions in these two cases are quite different.

Case $m\leqslant r$. By the hypothesis of the theorem, we have $1\leqslant m\leqslant\min\{r,n-k\}$ in this case. For each $t=1,2,\ldots,m$, Corollary 8.12 provides an explicit function $Z_t\colon(\{0,1\}^r)^t\to\mathbb{R}$ that obeys (8.52)–(8.54). Define

$$\begin{aligned}\tilde\Lambda^n_{k,m}(x)=\sum_{t=1}^m\binom{m}{t}(1-\alpha)^t \mathop{\mathbf{E}}_{i_1<i_2<\cdots<i_t}\ \mathop{\mathbf{E}}_{\substack{S:|S|=k+m-t\\ i_1,i_2,\ldots,i_t\notin S}}\Bigg[&Z_t(x_{i_1},x_{i_2},\ldots,x_{i_t})\\ &\times\prod_{i\in S}\mathrm{OR}_r(x_i)\mu_0(x_i)\cdot\prod_{i\notin S\cup\{i_1,i_2,\ldots,i_t\}}\mu_1(x_i)\Bigg],\end{aligned}$$

where the first expectation is over uniformly random $i_1,i_2,\ldots,i_t\in\{1,2,\ldots,n\}$ with $i_1<i_2<\cdots<i_t$, and the second expectation is over a uniformly random subset $S\subseteq\{1,2,\ldots,n\}\setminus\{i_1,i_2,\ldots,i_t\}$ of cardinality $|S|=k+m-t$.

We proceed to verify the properties required of $\tilde\Lambda^n_{k,m}$. The orthogonality property follows immediately from (8.54):

$$\deg P\leqslant\frac{r}{10}\implies\langle\tilde\Lambda^n_{k,m},P\rangle=0. \tag{8.60}$$

Continuing, fix an arbitrary input $x\in(\{0,1\}^r\setminus\{0^r\})^n$ and substitute (8.53) in the defining equation for $\tilde{\Lambda}^n_{k,m}$ to obtain

$$\begin{aligned}|\tilde{\Lambda}^n_{k,m}(x)| &\leqslant 60^{\frac{r}{10}} \sum_{t=1}^{m} \binom{m}{t} \left(\frac{9}{\epsilon}\right)^t \mathop{\mathbf{E}}_{|S|=k+m-t} \left[\prod_{i\in S} \mathrm{OR}_r(x_i)\mu_0(x_i) \cdot \prod_{i\notin S} \mu_1(x_i)\right] \\ &= 60^{\frac{r}{10}} \sum_{t=1}^{m} \binom{m}{t} \left(\frac{9}{\epsilon}\right)^t \Lambda^n_{k+m-t,0}(x) \\ &\leqslant 60^{\frac{r}{10}} \left(1+\frac{9}{\epsilon}\right)^m \sum_{t=1}^{m} \Lambda^n_{k+m-t,0}(x).\end{aligned}$$

Using Lemma 8.6(ii) and the assumption that $m \leqslant r$, we arrive at

$$|\tilde{\Lambda}^n_{k,m}(x)| \leqslant \left(\frac{15}{\epsilon}\right)^r \sum_{i=0}^{m-1} \Lambda^n_{k,i}(x), \qquad x \in (\{0,1\}^r \setminus \{0^r\})^n. \tag{8.61}$$

It remains to verify that $\tilde{\Lambda}^n_{k,m} = \Lambda^n_{k,m}$ outside $(\{0,1\}^r \setminus \{0^r\})^n$. To start with,

$$\mu_0 = \mathrm{OR}_r \cdot \mu_0 + \mathrm{NOR}_r \cdot \mu_0 = \mathrm{OR}_r \cdot \mu_0 + (1-\alpha)\mathrm{NOR}_r,$$

where the second step uses (8.39). As a result,

$$\begin{aligned}\Lambda^n_{k,m}(x) &= \mathop{\mathbf{E}}_{T,S} \prod_{i\in T} \mu_0(x_i) \cdot \prod_{i\in S} \mathrm{OR}_r(x_i)\mu_0(x_i) \cdot \prod_{i\notin S\cup T} \mu_1(x_i) \\ &= \mathop{\mathbf{E}}_{T,S} \prod_{i\in T} (\mathrm{OR}_r(x_i)\mu_0(x_i) + (1-\alpha)\mathrm{NOR}_r(x_i)) \cdot \prod_{i\in S} \mathrm{OR}_r(x_i)\mu_0(x_i) \\ &\qquad\qquad \times \prod_{i\notin S\cup T} \mu_1(x_i),\end{aligned}$$

where the expectation is over a uniformly random choice of sets $T \subseteq \{1,2,\ldots,n\}$ and $S \subseteq \{1,2,\ldots,n\} \setminus T$ of cardinalities $|T| = m$ and $|S| = k$. Multiplying out,

$$\begin{aligned}\Lambda^n_{k,m}(x) = \sum_{t=0}^{m} \binom{m}{t} (1-\alpha)^t \mathop{\mathbf{E}}_{i_1<i_2<\cdots<i_t} \mathop{\mathbf{E}}_{\substack{S:|S|=k+m-t\\ i_1,i_2,\ldots,i_t\notin S}} \Bigg[&\mathrm{NOR}_{rt}(x_{i_1}x_{i_2}\ldots x_{i_t}) \\ &\times \prod_{i\in S} \mathrm{OR}_r(x_i)\mu_0(x_i) \cdot \prod_{i\notin S\cup\{i_1,i_2,\ldots,i_t\}} \mu_1(x_i)\Bigg],\end{aligned} \tag{8.62}$$

where the first expectation is over uniformly random $i_1, i_2, \ldots, i_t \in \{1,2,\ldots,n\}$ with $i_1 < i_2 < \cdots < i_t$, and the second expectation is over a uniformly random subset $S \subseteq \{1,2,\ldots,n\} \setminus \{i_1,i_2,\ldots,i_t\}$ of cardinality $|S| = k+m-t$. For the remainder of the proof, fix an arbitrary input $x = (x_1, x_2, \ldots, x_n)$ with $x_{i^*} = 0^r$ for at least one coordinate i^*. We claim that for any two disjoint sets S and $\{i_1, i_2, \ldots, i_t\}$,

$$\begin{aligned}\mathrm{NOR}_{rt}(x_{i_1}x_{i_2}\ldots x_{i_t}) &\prod_{i\in S} \mathrm{OR}_r(x_i)\mu_0(x_i) \cdot \prod_{i\notin S\cup\{i_1,i_2,\ldots,i_t\}} \mu_1(x_i) \\ &= Z_t(x_{i_1}, x_{i_2}, \ldots, x_{i_t}) \prod_{i\in S} \mathrm{OR}_r(x_i)\mu_0(x_i) \cdot \prod_{i\notin S\cup\{i_1,i_2,\ldots,i_t\}} \mu_1(x_i).\end{aligned} \tag{8.63}$$

Indeed, if $i^* \notin \{i_1, i_2, \ldots, i_t\}$ then the left- and right-hand sides of (8.63) both vanish because $\mathrm{OR}_r(x_{i^*})\mu_0(x_{i^*}) = \mu_1(x_{i^*}) = 0$. In the complementary case when $i^* \in \{i_1, i_2, \ldots, i_t\}$, we have from (8.52) and (8.53) that

$$\begin{aligned} Z_t(x_{i_1}, x_{i_2}, \ldots, x_{i_t}) &= \begin{cases} 1 & \text{if } x_{i_1} = x_{i_2} = \cdots = x_{i_t} = 0^r, \\ 0 & \text{otherwise} \end{cases} \\ &= \mathrm{NOR}_{rt}(x_{i_1}, x_{i_2}, \ldots, x_{i_t}), \end{aligned}$$

settling (8.63). Substituting (8.63) in (8.62),

$$\begin{aligned} \Lambda^n_{k,m}(x) = \sum_{t=0}^{m} \binom{m}{t} (1-\alpha)^t \mathop{\mathbf{E}}_{i_1 < i_2 < \cdots < i_t} \mathop{\mathbf{E}}_{\substack{S : |S| = k+m-t \\ i_1, i_2, \ldots, i_t \notin S}} \Bigg[& Z_t(x_{i_1}, x_{i_2}, \ldots, x_{i_t}) \\ & \times \prod_{i \in S} \mathrm{OR}_r(x_i)\mu_0(x_i) \cdot \prod_{i \notin S \cup \{i_1, i_2, \ldots, i_t\}} \mu_1(x_i) \Bigg]. \end{aligned}$$

In this sum, the term corresponding to $t = 0$ vanishes because $\mathrm{OR}_r(x_{i^*})\mu_0(x_{i^*}) = \mu_1(x_{i^*}) = 0$. As a result, we arrive at the desired conclusion:

$$\tilde{\Lambda}^n_{k,m}(x) = \Lambda^n_{k,m}(x), \qquad x \notin (\{0,1\}^r \setminus \{0^r\})^n. \tag{8.64}$$

The newly established properties (8.60), (8.61), and (8.64) complete the proof for the case $m \leqslant r$.

Case $m \geqslant r + 1$. By the theorem hypothesis, we have $r + 1 \leqslant m \leqslant n - k$ in this case. Let $\tilde{\Lambda}^n_{k,1}, \tilde{\Lambda}^n_{k,2}, \ldots, \tilde{\Lambda}^n_{k,\lfloor r/2 \rfloor}$ be the functions constructed in the first half of the proof, with properties (8.60), (8.61), and (8.64). Define

$$\begin{aligned} \tilde{\Lambda}^n_{k,m} = \Lambda^n_{k,m} &- (-1)^{\lfloor r/2 \rfloor} \binom{m-1}{\lfloor r/2 \rfloor} \Lambda^n_{k,0} \\ &- \sum_{i=1}^{\lfloor r/2 \rfloor} (-1)^{\lfloor r/2 \rfloor - i} \binom{m}{i} \binom{m-i-1}{\lfloor r/2 \rfloor - i} (\Lambda^n_{k,i} - \tilde{\Lambda}^n_{k,i}). \end{aligned} \tag{8.65}$$

We proceed to establish properties (i)–(iii) in the theorem statement. On inputs outside $(\{0,1\}^r \setminus \{0^r\})^n$, we have $\Lambda^n_{k,0} = 0$ by Lemma 8.6(i) and $\Lambda^n_{k,i} - \tilde{\Lambda}^n_{k,i} = 0$ by (8.64). Making these substitutions in (8.65) gives $\tilde{\Lambda}^n_{k,m} = \Lambda^n_{k,m}$ outside $(\{0,1\}^r \setminus \{0^r\})^n$, establishing (i). On $(\{0,1\}^r \setminus \{0^r\})^n$, we have

$$\begin{aligned} |\tilde{\Lambda}^n_{k,m}| &\leqslant \Lambda^n_{k,m} + \binom{m-1}{\lfloor r/2 \rfloor} \Lambda^n_{k,0} + \sum_{i=1}^{\lfloor r/2 \rfloor} \binom{m}{i} \binom{m-i-1}{\lfloor r/2 \rfloor - i} (\Lambda^n_{k,i} + |\tilde{\Lambda}^n_{k,i}|) \\ &\leqslant \Lambda^n_{k,m} + 2^{\lfloor r/2 \rfloor} \binom{m}{\lfloor r/2 \rfloor} \sum_{i=0}^{\lfloor r/2 \rfloor} \Lambda^n_{k,i} + 2^{\lfloor r/2 \rfloor} \binom{m}{\lfloor r/2 \rfloor} \sum_{i=1}^{\lfloor r/2 \rfloor} |\tilde{\Lambda}^n_{k,i}| \\ &\leqslant \Lambda^n_{k,m} + 2^{\lfloor r/2 \rfloor} \binom{m}{\lfloor r/2 \rfloor} \sum_{i=0}^{\lfloor r/2 \rfloor} \Lambda^n_{k,i} + 2^{\lfloor r/2 \rfloor} \binom{m}{\lfloor r/2 \rfloor} \cdot \left\lfloor \frac{r}{2} \right\rfloor \left(\frac{15}{\epsilon} \right)^r \sum_{i=0}^{\lfloor r/2 \rfloor} \Lambda^n_{k,i} \\ &\leqslant \Lambda^n_{k,m} + 2^{\lfloor r/2 \rfloor} \binom{m}{\lfloor r/2 \rfloor} \left(1 + \left\lfloor \frac{r}{2} \right\rfloor \left(\frac{15}{\epsilon} \right)^r \right) \sum_{i=0}^{\lfloor r/2 \rfloor} \Lambda^n_{k,i}, \end{aligned}$$

where the third step uses (8.61). This settles (ii).

It remains to prove (iii). Fix a polynomial P with $\deg P \leqslant \min\{r/10, c_{\mathrm{in}} r\sqrt{r}/2\}$. By Lemma 8.6(iv), there exists a univariate polynomial p with

$$\langle \Lambda^n_{k,i}, P\rangle = p(i) \qquad (i = 0, 1, 2, \ldots, m), \tag{8.66}$$

$$\deg p \leqslant \frac{r}{2}. \tag{8.67}$$

By definition,

$$\begin{aligned}\langle \tilde{\Lambda}^n_{k,m}, P\rangle = \langle \Lambda^n_{k,m}, P\rangle - (-1)^{\lfloor r/2\rfloor}\binom{m-1}{\lfloor r/2\rfloor}\langle \Lambda^n_{k,0}, P\rangle \\ - \sum_{i=1}^{\lfloor r/2\rfloor}(-1)^{\lfloor r/2\rfloor - i}\binom{m}{i}\binom{m-i-1}{\lfloor r/2\rfloor - i}(\langle \Lambda^n_{k,i}, P\rangle - \langle \tilde{\Lambda}^n_{k,i}, P\rangle).\end{aligned}$$

Applying (8.60) and (8.66),

$$\begin{aligned}\langle \tilde{\Lambda}^n_{k,m}, P\rangle &= p(m) - \sum_{i=0}^{\lfloor r/2\rfloor}(-1)^{\lfloor r/2\rfloor - i}\binom{m}{i}\binom{m-i-1}{\lfloor r/2\rfloor - i}p(i) \\ &= p(m) - \sum_{i=0}^{\lfloor r/2\rfloor}(-1)^{\lfloor r/2\rfloor - i}\binom{m}{i}\binom{m-i-1}{m-\lfloor r/2\rfloor - 1}p(i) \\ &= -\sum_{i=0}^{m}(-1)^{\lfloor r/2\rfloor - i}\binom{m}{i}\binom{m-i-1}{m-\lfloor r/2\rfloor - 1}p(i) \\ &= 0,\end{aligned}$$

where last step is valid by (8.67) and Fact 2.1(i) because the degree of $\binom{m-i-1}{m-\lfloor r/2\rfloor-1}p(i)$ as a univariate polynomial in i is at most $m - \lfloor r/2\rfloor - 1 + \deg p \leqslant m-1$. This establishes (iii), completing the proof. ∎

8.7. Final construction. We are finally in a position to define the dual objects required to prove Theorem 8.2. Let $\Phi_0, \Phi_1 \colon (\{0,1\}^r)^n \to \mathbb{R}$ be given by

$$\Phi_0 = \sum_{k=0}^{n/2}\sum_{m=1}^{n-k}\frac{|\nu_k(m)|}{c_{\mathrm{out}}^k}\tilde{\Lambda}^n_{k,m},$$

$$\Phi_1 = \sum_{k=0}^{n/2}\sum_{m=0}^{n-k}\frac{\nu_k(m)}{c_{\mathrm{out}}^k}\Lambda^n_{k,m} + \sum_{k=\frac{n}{2}+1}^{n}\frac{1}{c_{\mathrm{out}}^k}\tilde{\Lambda}^n_{k},$$

where $\tilde{\Lambda}^n_{k,m}$ is as constructed in Theorem 8.13 and ν_k is as defined in subsection 8.2. The four lemmas that follow establish the properties required of Φ_0 and Φ_1 by the dual characterization of one-sided rational approximation (Theorem 4.6).

Lemma 8.14. *$\Phi_0 \geqslant |\Phi_1|/2$ outside $(\{0,1\}^r \setminus \{0^r\})^n$.*

Proof. For any integer m with $n/2 < m \leqslant n$, we have the following bounds outside

$(\{0,1\}^r \setminus \{0^r\})^n$:

$$\begin{aligned}
|\tilde{\Lambda}^n_m| &\leqslant \left(\frac{2\alpha}{1-\alpha}\right)^{m/2} \Lambda^n_{0,m} && \text{by Lemma 8.7(iv)}\\
&\leqslant \left(\frac{2\epsilon}{1-\epsilon}\right)^{m/2} \Lambda^n_{0,m} && \text{by (8.33)}\\
&\leqslant c_{\text{out}}^{3m}\Lambda^n_{0,m} && \text{by (8.19)}\\
&\leqslant c_{\text{out}}^{n+m}\Lambda^n_{0,m} && \text{by (8.16)}\\
&\leqslant |\nu_0(m)|\, c_{\text{out}}^{m}\Lambda^n_{0,m} && \text{by (8.14).} \qquad (8.68)
\end{aligned}$$

It follows that outside $(\{0,1\}^r \setminus \{0^r\})^n$,

$$\begin{aligned}
|\Phi_1| &\leqslant \sum_{k=0}^{n/2}\sum_{m=0}^{n-k} \frac{|\nu_k(m)|}{c_{\text{out}}^k}\Lambda^n_{k,m} + \sum_{m=\frac{n}{2}+1}^{n} \frac{1}{c_{\text{out}}^m}|\tilde{\Lambda}^n_m|\\
&\leqslant \sum_{k=0}^{n/2}\sum_{m=0}^{n-k} \frac{|\nu_k(m)|}{c_{\text{out}}^k}\Lambda^n_{k,m} + \sum_{m=\frac{n}{2}+1}^{n} |\nu_0(m)|\Lambda^n_{0,m} && \text{by (8.68)}\\
&\leqslant 2\sum_{k=0}^{n/2}\sum_{m=0}^{n-k} \frac{|\nu_k(m)|}{c_{\text{out}}^k}\Lambda^n_{k,m}\\
&= 2\sum_{k=0}^{n/2}\sum_{m=1}^{n-k} \frac{|\nu_k(m)|}{c_{\text{out}}^k}\Lambda^n_{k,m} && \text{by Lemma 8.6(i)}\\
&= 2\sum_{k=0}^{n/2}\sum_{m=1}^{n-k} \frac{|\nu_k(m)|}{c_{\text{out}}^k}\tilde{\Lambda}^n_{k,m} && \text{by Theorem 8.13(i)}\\
&= 2\Phi_0.
\end{aligned}$$

□

Lemma 8.15. *$\Phi_1 \geqslant (c_{\text{cor}}\, c_{\text{out}}^2)^r\, |\Phi_0|/14$ on $(\{0,1\}^r \setminus \{0^r\})^n$, where $0 < c_{\text{cor}} < 1$ is the constant from Theorem 8.13.*

Proof. On $(\{0,1\}^r \setminus \{0^r\})^n$, we have

$$\begin{aligned}
\Phi_1 &= \sum_{k=0}^{n/2}\sum_{m=0}^{n-k} \frac{\nu_k(m)}{c_{\text{out}}^k}\Lambda^n_{k,m} + \sum_{\ell=\frac{n}{2}+1}^{n} \frac{1}{c_{\text{out}}^\ell}\tilde{\Lambda}^n_\ell\\
&= \sum_{k=0}^{n/2}\sum_{m=0}^{n-k} \frac{\nu_k(m)}{c_{\text{out}}^k}\Lambda^n_{k+m,0} + \sum_{\ell=\frac{n}{2}+1}^{n} \frac{1}{c_{\text{out}}^\ell}\Lambda^n_{\ell,0} && \text{by Lemmas 8.6(ii), 8.7(iii)}\\
&= \sum_{\ell=0}^{n/2}\left(\frac{\nu_\ell(0)}{c_{\text{out}}^\ell} + \sum_{i=1}^{\ell}\frac{\nu_{\ell-i}(i)}{c_{\text{out}}^{\ell-i}}\right)\Lambda^n_{\ell,0}\\
&\qquad + \sum_{\ell=\frac{n}{2}+1}^{n}\left(\frac{1}{c_{\text{out}}^\ell} + \sum_{i=\ell-\frac{n}{2}}^{\ell}\frac{\nu_{\ell-i}(i)}{c_{\text{out}}^{\ell-i}}\right)\Lambda^n_{\ell,0} && \text{by algebra}\\
&= \sum_{\ell=0}^{n/2}\left(\frac{\nu_\ell(0)}{c_{\text{out}}^\ell} + \sum_{i=\min\{r,n/4\}}^{\ell}\frac{\nu_{\ell-i}(i)}{c_{\text{out}}^{\ell-i}}\right)\Lambda^n_{\ell,0}
\end{aligned}$$

$$
\begin{aligned}
&\quad + \sum_{\ell=\frac{n}{2}+1}^{n} \left(\frac{1}{c_{\text{out}}^{\ell}} + \sum_{i=\ell-\frac{n}{2}}^{\ell} \frac{\nu_{\ell-i}(i)}{c_{\text{out}}^{\ell-i}} \right) \Lambda_{\ell,0}^{n} && \text{by (8.11)} \\
&\geqslant \sum_{\ell=0}^{n/2} \left(\frac{1}{c_{\text{out}}^{\ell-\min\{r,n/4\}}} - \frac{1}{2} \sum_{i=\min\{r,n/4\}}^{\infty} \frac{1}{c_{\text{out}}^{\ell-i}} \right) \Lambda_{\ell,0}^{n} \\
&\quad + \sum_{\ell=\frac{n}{2}+1}^{n} \left(\frac{1}{c_{\text{out}}^{\ell}} - \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{c_{\text{out}}^{\ell-i}} \right) \Lambda_{\ell,0}^{n} && \text{by (8.13), (8.17)} \\
&\geqslant \frac{1}{3} \sum_{\ell=0}^{n} \frac{1}{c_{\text{out}}^{\ell-\min\{r,n/4\}}} \Lambda_{\ell,0}^{n} && \text{by (8.16)} \\
&\geqslant \frac{c_{\text{out}}^{r}}{3} \sum_{\ell=0}^{n} \frac{1}{c_{\text{out}}^{\ell}} \Lambda_{\ell,0}^{n}. && (8.69)
\end{aligned}
$$

We now turn to Φ_0. For any $k \geqslant 0$,

$$
\begin{aligned}
&\sum_{m=1}^{n-k} |\nu_k(m)| \max\left\{1, \sqrt{\frac{m}{r}}\right\}^{r} \\
&\qquad = \sum_{m=\min\{r,n/4\}}^{n-k} |\nu_k(m)| \max\left\{1, \sqrt{\frac{m}{r}}\right\}^{r} && \text{by (8.11)} \\
&\qquad \leqslant \sum_{m=\min\{r,n/4\}}^{n-k} |\nu_k(m)| \left(\frac{m}{\min\{r,n/4\}} \right)^{r/2} \\
&\qquad \leqslant \sum_{m=\min\{r,n/4\}}^{n-k} \left(\frac{\min\{r,n/4\}}{c_{\text{out}}\, m} \cdot \sqrt{\frac{m}{\min\{r,n/4\}}} \right)^{r} && \text{by (8.18)} \\
&\qquad = \sum_{m=\min\{r,n/4\}}^{n-k} \left(\frac{1}{c_{\text{out}}} \sqrt{\frac{\min\{r,n/4\}}{m}} \right)^{r} \\
&\qquad \leqslant \frac{1}{c_{\text{out}}^{r}} \left(1 + \int_{\min\{r,n/4\}}^{\infty} \left(\frac{\min\{r,n/4\}}{m} \right)^{r/2} \, dm \right) \\
&\qquad = \frac{1}{c_{\text{out}}^{r}} \left(1 + \frac{2\min\{r,n/4\}}{r-2} \right) \\
&\qquad \leqslant \frac{1}{c_{\text{out}}^{r}} \left(1 + \frac{2r}{r-2} \right).
\end{aligned}
$$

In view of (8.6), we arrive at

$$
\sum_{m=1}^{n-k} |\nu_k(m)| \max\left\{1, \sqrt{\frac{m}{r}}\right\}^{r} \leqslant \frac{31}{9 c_{\text{out}}^{r}}. \tag{8.70}
$$

It remains to piece the above calculations together. On $(\{0,1\}^r \setminus \{0^r\})^n$,

$$
|\Phi_0| \leqslant \sum_{k=0}^{n/2} \sum_{m=1}^{n-k} \frac{|\nu_k(m)|}{c_{\text{out}}^{k}} \, |\tilde{\Lambda}_{k,m}^{n}|
$$

$$\begin{aligned}
&\leqslant \sum_{k=0}^{n/2} \sum_{m=1}^{n-k} \frac{|\nu_k(m)|}{c_{\text{out}}^k} \left(\frac{1}{c_{\text{cor}}} \max\left\{1, \sqrt{\frac{m}{r}}\right\} \right)^r \sum_{i=0}^{n-k} \Lambda_{k,i}^n && \text{by Theorem 8.13(ii)} \\
&\leqslant \frac{31}{9(c_{\text{cor}}\, c_{\text{out}})^r} \sum_{k=0}^{n/2} \frac{1}{c_{\text{out}}^k} \sum_{i=0}^{n-k} \Lambda_{k,i}^n && \text{by (8.70)} \\
&= \frac{31}{9(c_{\text{cor}}\, c_{\text{out}})^r} \sum_{k=0}^{n/2} \frac{1}{c_{\text{out}}^k} \sum_{i=0}^{n-k} \Lambda_{k+i,0}^n && \text{by Lemma 8.6(ii)} \\
&\leqslant \frac{31}{9(c_{\text{cor}}\, c_{\text{out}})^r} \sum_{\ell=0}^{n} \left(\sum_{k=0}^{\ell} \frac{1}{c_{\text{out}}^k} \right) \Lambda_{\ell,0}^n && \text{by algebra} \\
&\leqslant \frac{31}{9(c_{\text{cor}}\, c_{\text{out}})^r} \cdot \frac{4}{3} \sum_{\ell=0}^{n} \frac{1}{c_{\text{out}}^\ell} \Lambda_{\ell,0}^n && \text{by (8.16)} \\
&\leqslant \frac{14}{(c_{\text{cor}}\, c_{\text{out}})^r} \cdot \frac{1}{3} \sum_{\ell=0}^{n} \frac{1}{c_{\text{out}}^\ell} \Lambda_{\ell,0}^n \\
&\leqslant \frac{14}{(c_{\text{cor}}\, c_{\text{out}}^2)^r} \cdot \Phi_1 && \text{by (8.69).} \qquad \square
\end{aligned}$$

Lemma 8.16. *Let $P, Q\colon (\{0,1\}^r)^n \to \mathbb{R}$ be polynomials with*

$$\deg P \leqslant \min\{r/10, c_{\text{in}} r\sqrt{r}/2\}, \tag{8.71}$$

$$\deg Q \leqslant \min\{c_{\text{in}}\, c_{\text{out}}\, r\sqrt{n}, c_{\text{in}}\, c_{\text{out}}\, n\sqrt{r}, n/4\}. \tag{8.72}$$

Then

$$\langle \Phi_0, P\rangle = \langle \Phi_1, Q\rangle = 0.$$

Proof. The claim for Φ_0 is immediate by Theorem 8.13(iii). To prove it for Φ_1, recall from Lemma 8.6(iv) that there exist univariate polynomials $q_0, q_1, \ldots, q_{n/2}$ with

$$\langle \Lambda_{k,m}^n, Q\rangle = q_k(m) \qquad (k = 0, 1, \ldots, n/2;\ m = 0, 1, \ldots, n-k), \tag{8.73}$$

$$\deg q_k \leqslant c_{\text{out}} \min\{\sqrt{rn}, n\} \qquad (k = 0, 1, \ldots, n/2). \tag{8.74}$$

Then

$$\begin{aligned}
\langle \Phi_1, Q\rangle &= \sum_{k=0}^{n/2} \sum_{m=0}^{n-k} \frac{\nu_k(m)}{c_{\text{out}}^k} \langle \Lambda_{k,m}^n, Q\rangle + \sum_{k=\frac{n}{2}+1}^{n} \frac{1}{c_{\text{out}}^k} \langle \tilde{\Lambda}_k^n, Q\rangle \\
&= \sum_{k=0}^{n/2} \sum_{m=0}^{n-k} \frac{\nu_k(m)}{c_{\text{out}}^k} \langle \Lambda_{k,m}^n, Q\rangle && \text{by (8.72) and Lemma 8.7(i)} \\
&= \sum_{k=0}^{n/2} \frac{\langle \nu_k, q_k\rangle}{c_{\text{out}}^k} && \text{by (8.73)} \\
&= 0. && \text{by (8.9) and (8.74).} \qquad \square
\end{aligned}$$

Lemma 8.17. $\Phi_0, \Phi_1 \not\equiv 0$.

Proof. We have

$$\begin{aligned}
\Phi_0(0^r,\dots,0^r) &= \sum_{k=0}^{n/2}\sum_{m=1}^{n-k}\frac{|\nu_k(m)|}{c_{\text{out}}^k}\tilde{\Lambda}^n_{k,m}(0^r,\dots,0^r) && \text{by definition}\\
&= \sum_{k=0}^{n/2}\sum_{m=1}^{n-k}\frac{|\nu_k(m)|}{c_{\text{out}}^k}\Lambda^n_{k,m}(0^r,\dots,0^r) && \text{by Theorem 8.13(i)}\\
&\geqslant |\nu_0(n)|\Lambda^n_{0,n}(0^r,\dots,0^r) && \text{since } \Lambda^n_{k,m}\geqslant 0\\
&= |\nu_0(n)|\mu_0(0^r)^n && \text{by definition}\\
&> 0 && \text{by (8.14), (8.16), and (8.34).}
\end{aligned}$$

The analysis for Φ_1 is based on a similar argument. On the nonempty set $(\operatorname{supp}\mu_1)^n = \{x : \phi(x) > 0\}^n$,

$$\begin{aligned}
\Phi_1 &= \sum_{k=0}^{n/2}\sum_{m=0}^{n-k}\frac{\nu_k(m)}{c_{\text{out}}^k}\Lambda^n_{k,m} + \sum_{k=\frac{n}{2}+1}^{n}\frac{1}{c_{\text{out}}^k}\tilde{\Lambda}^n_k && \text{by definition}\\
&= \nu_0(0)\Lambda^n_{0,0} && \text{by Lemmas 8.6(iii), 8.7(ii)}\\
&= \nu_0(0)\mu_1^{\otimes n} && \text{by definition}\\
&> 0 && \text{by (8.13) and (8.16).}
\end{aligned}$$

□

By the dual characterization of one-sided rational approximation (Theorem 4.6), the newly established Lemmas 8.14 to 8.17 imply that

$$\begin{aligned}
&R\big(\text{AND}_n \circ \text{OR}_r,\\
&\quad \min\{r/10, c_{\text{in}} r\sqrt{r}/2\},\\
&\quad \min\{c_{\text{in}}\, c_{\text{out}}\, r\sqrt{n}, c_{\text{in}}\, c_{\text{out}}\, n\sqrt{r}, n/4\}\big) \geqslant \frac{(c_{\text{cor}}\, c_{\text{out}}^2)^r}{14},
\end{aligned}$$

where $c_{\text{in}}, c_{\text{out}}, c_{\text{cor}} \in (0,1)$ are sufficiently small absolute constants. We conclude that

$$R(\text{AND}_n \circ \text{OR}_r, cr, c\min\{r\sqrt{n}, n\}) \geqslant c^{-r}$$

for a sufficiently small absolute constant $c > 0$. This conclusion is logically equivalent to Theorem 8.2, in view of the error reduction procedure for rational approximation (Proposition 4.2).

9. Main results. The main technical contribution of this paper is a hardness amplification result that transforms any Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ with high one-sided approximate degree into a related Boolean function $F\colon \{0,1\}^N \to \{0,1\}$ with proportionately high threshold degree. The transformed function is of the form $F = \text{OR}_\ell \circ ((\text{AND}_k \circ \neg f) \wedge g)$, where g is an auxiliary function to which we refer as the *amplifier*. If the original function has one-sided approximate degree n^α, then the transformed function has threshold degree $\Omega(N^\beta)$ for some monotonically growing exponent $\beta = \beta(\alpha)$ that depends on g. We formalize our technique in this generality in subsection 9.1. In subsection 9.2, we specialize the amplifier g to be a read-once formula of depth 2 and prove that the resulting construction achieves

$$\beta = \begin{cases} 3/7 & \text{if } \alpha < 1/2,\\ 3\alpha/(3\alpha+2) & \text{if } 1/2 \leqslant \alpha \leqslant 2/3,\\ 1/2 & \text{if } \alpha > 2/3.\end{cases}$$

As corollaries, we obtain our main lower bounds on the threshold degree of constant-depth circuits and read-once formulas, by choosing f accordingly in each case. In subsection 9.3, we prove matching upper bounds. In the concluding subsection 9.4, we discuss the limitations of our technique and propose directions for future work.

9.1. The general theorem. We start with a general statement of our hardness amplification technique. This result brings together the dual view of one-sided and hybrid rational approximation from sections 4 and 5, the composition theorem from section 6, and the lower bound on hybrid rational approximation from section 7.

THEOREM 9.1. *Let f and g be nonconstant Boolean functions. Fix integers $k \geqslant 1$ and $d, D \geqslant 0$ such that*

$$R(g, d, D) > 2^{-k}. \tag{9.1}$$

Then

$$R((\mathrm{AND}_k \circ \neg f) \wedge g, c\min\{k, d, \deg^+_{1/3}(f)\}, c\min\{\sqrt{k}\deg^+_{1/3}(f), D\}) > \frac{1}{\sqrt{2}} \tag{9.2}$$

and

$$\begin{aligned}\deg_\pm(\mathrm{OR}_\ell \circ ((\mathrm{AND}_k \circ \neg f) \wedge g))&\\ \geqslant c\min\{\ell k, \ell(d+1), \ell\deg^+_{1/3}(f), \sqrt{k}\deg^+_{1/3}(f), D\}&\end{aligned} \tag{9.3}$$

for all $\ell \geqslant 2$, where $c > 0$ is an absolute constant, independent of f, g, D, d, k, ℓ.

Proof. The lower bound (9.3) on the threshold degree is a direct consequence of (9.2) and Theorem 4.5. Thus, it suffices to prove (9.2).

For some absolute integer constant $c_0 \geqslant 1$ and all $k \geqslant c_0$, Theorem 7.1 ensures the existence of functions $\Phi_0, \Phi'_1, \Phi''_1 \colon (\operatorname{dom} f)^k \to \mathbb{R}$ such that

$$\Phi_0 \geqslant \frac{1}{2}|\Phi'_1| \text{ on } (\mathrm{AND}_k \circ \neg f)^{-1}(0), \tag{9.4}$$

$$\Phi_0 \geqslant \frac{1}{2}|\Phi''_1| \text{ on } (\mathrm{AND}_k \circ \neg f)^{-1}(0), \tag{9.5}$$

$$\Phi'_1 \geqslant \frac{1}{2}\max\{-\Phi_0, -2^{-k/c_0}\Phi_0\} \text{ on } (\mathrm{AND}_k \circ \neg f)^{-1}(1), \tag{9.6}$$

$$\Phi''_1 \geqslant \frac{1}{2}|\Phi_0| \text{ on } (\mathrm{AND}_k \circ \neg f)^{-1}(1), \tag{9.7}$$

$$\langle \Phi_0, P\rangle = 0 \text{ whenever } \deg P \leqslant \frac{1}{2c_0}\min\{\deg^+_{1/3}(f), k\}, \tag{9.8}$$

$$\langle \Phi'_1, P\rangle = 0 \text{ whenever } \deg P \leqslant \frac{1}{c_0\sqrt{c_0}}\deg^+_{1/3}(f)\sqrt{k}, \tag{9.9}$$

$$\langle \Phi''_1, P\rangle = 0 \text{ whenever } \deg P \leqslant \frac{1}{2c_0}\min\{\deg^+_{1/3}(f), k\}, \tag{9.10}$$

$$\Phi_0, \Phi'_1, \Phi''_1 \not\equiv 0. \tag{9.11}$$

On the other hand, it follows from (9.1) and the error-reduction property of rational approximation (Proposition 4.2) that

$$R\left(g, \frac{d}{2c_0}, \frac{D}{2c_0}\right) > 2^{-k/2c_0}. \tag{9.12}$$

Applying Theorem 6.1 to (9.4)–(9.12) with parameters $\epsilon = 1/2$ and $\Delta = 2^{k/c_0}$, we infer that

$$R\left((\mathrm{AND}_k \circ \neg f) \wedge g, \frac{1}{4c_0} \min\{k, d, \deg^+_{1/3}(f)\}, \right. \\ \left. \frac{1}{4c_0\sqrt{c_0}} \min\{\deg^+_{1/3}(f)\sqrt{k}, D\}\right) \geqslant \frac{1}{2\sqrt{2}} \tag{9.13}$$

for all $k \geqslant c_0$.

We now claim that (9.2) holds with $c = 1/(16c_0\sqrt{c_0})$. For $k \geqslant c_0$, the bound follows directly from (9.13) and the error-reduction property (Proposition 4.2). To prove validity in the complementary case $k < c_0$, observe that the left-hand side of (9.2) is trivially bounded from below by $R((\mathrm{AND}_k \circ \neg f) \wedge g, cd, cD)$, where

$$\begin{aligned} R((\mathrm{AND}_k \circ \neg f) \wedge g, cd, cD) &\geqslant R(g, cd, cD) && \text{since } f \text{ is nonconstant} \\ &\geqslant R\left(g, \frac{d}{2k}, \frac{D}{2k}\right) && \text{since } c \leqslant \frac{1}{2c_0} < \frac{1}{2k} \\ &\geqslant R(g, d, D)^{1/(2k)} && \text{by Proposition 4.2} \\ &> \frac{1}{\sqrt{2}} && \text{by (9.1).} \end{aligned}$$

□

9.2. Results using depth-2 amplifiers. We now establish the main results of our paper by invoking Theorem 9.1 with appropriate functions f and g. In all of our applications, the amplifier g will be a read-once formula of depth 2.

Our first application concerns the threshold degree of constant-depth formulas. In their inspiring work twelve years ago, O'Donnell and Servedio [25, Theorem 5.2] obtained an upper bound of $\tilde{O}(N^{(2^{d-1}-1)/(2^d-1)})$ on the threshold degree of any $\wedge, \vee$-formula of depth d and size N. This bound was known to be tight only for $d = 1$ and $d = 2$, by the classic results of Minsky and Papert [22]. We are able to show that O'Donnell and Servedio's bound is tight for depth $d = 3$ as well by constructing a depth-3 formula of size N and threshold degree $\Omega(N^{3/7})$. The best previous lower bound, obtained in [35], was polynomially weaker: $\Omega(N^{2/5})$.

Theorem 9.2. *Let $F\colon \{0,1\}^{N+N^{6/7}} \to \{0,1\}$ be the read-once formula given by*

$$F = \mathrm{OR}_{N^{1/7}} \circ ((\mathrm{AND}_{N^{2/7}} \circ \mathrm{OR}_{N^{4/7}}) \wedge (\mathrm{AND}_{N^{3/7}} \circ \mathrm{OR}_{N^{2/7}})).$$

Then

$$\deg_\pm(F) = \Omega(N^{3/7}).$$

This result settles Theorem 1.1 from the Introduction. The reader will note that our constructed formula is highly asymmetric. It turns out that asymmetry is crucial to the optimal lower bound in Theorem 9.2. Specifically, we showed in [35] that all formulas of the form $\mathrm{OR}_{N_1} \circ \mathrm{AND}_{N_2} \circ \mathrm{OR}_{N_3}$ on $N = N_1N_2N_3$ variables have threshold degree $O(N^{2/5} \log N)$.

Proof of Theorem 9.2. Theorem 8.2 implies that the function $g = \mathrm{AND}_{n^{3/4}} \circ \mathrm{OR}_{\sqrt{n}}$ has one-sided rational approximation error

$$R(g, c\sqrt{n}, cn^{3/4}) > 2^{-\sqrt{n}}$$

for some constant $c > 0$. On the other hand, Theorem 2.6 states that

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

for $f = \mathrm{NOR}_n$. Appealing to Theorem 9.1 with $d = c\sqrt{n}$, $D = cn^{3/4}$, $k = \sqrt{n}$, and $\ell = n^{1/4}$, we obtain a lower bound of $\Omega(n^{3/4})$ on the threshold degree of the composition

$$\mathrm{OR}_{n^{1/4}} \circ ((\mathrm{AND}_{\sqrt{n}} \circ \mathrm{OR}_n) \wedge (\mathrm{AND}_{n^{3/4}} \circ \mathrm{OR}_{\sqrt{n}})).$$

Setting $n = N^{4/7}$ completes the proof. □

We now obtain a general hardness amplification result for polynomial approximation, which transforms any Boolean function with given one-sided approximate degree into a related Boolean function with proportionately high threshold degree. This result extends Theorem 9.2 on the threshold degree of constant-depth formulas and settles Theorem 1.3 from the Introduction.

Theorem 9.3. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be given with $\deg^+_{1/3}(f) \geqslant n^\alpha$, where $\alpha \in [0,1]$. Consider the function $F\colon \{0,1\}^N \to \{0,1\}$ on $N = \max\{n^{7/4}+n^{3/2}, n\cdot n^\alpha\sqrt{n^\alpha}+n^{3\alpha}\}$ variables, given by*

$$F = \begin{cases} \mathrm{OR}_{n^{1/4}} \circ ((\mathrm{AND}_{\sqrt{n}} \circ \mathrm{OR}_n) \wedge (\mathrm{AND}_{n^{3/4}} \circ \mathrm{OR}_{\sqrt{n}})) & \text{if } \alpha < 1/2, \\ \mathrm{OR}_{\sqrt{n^\alpha}} \circ ((\mathrm{AND}_{n^\alpha} \circ \neg f) \wedge (\mathrm{AND}_{n^\alpha\sqrt{n^\alpha}} \circ \mathrm{OR}_{n^\alpha})) & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \deg_\pm(F) &= \Omega(\max\{n^{3/4}, n^\alpha\sqrt{n^\alpha}\}) \\ &\geqslant \begin{cases} cN^{3/7} & \text{if } \alpha < 1/2, \\ cN^{3\alpha/(3\alpha+2)} & \text{if } 1/2 \leqslant \alpha < 2/3, \\ c\sqrt{N} & \text{otherwise,} \end{cases} \end{aligned}$$

where $c > 0$ is an absolute constant, independent of f, α, n.

Proof. The claim for $\alpha < 1/2$ is a restatement of Theorem 9.2, and we focus on the complementary case $\alpha \geqslant 1/2$. Theorem 8.2 implies that for some absolute constant $c > 0$, the function $g = \mathrm{AND}_{n^\alpha\sqrt{n^\alpha}} \circ \mathrm{OR}_{n^\alpha}$ obeys

$$R(g, cn^\alpha, cn^\alpha\sqrt{n^\alpha}) > 2^{-n^\alpha}.$$

Invoking Theorem 9.1 with parameters $d = cn^\alpha$, $D = cn^\alpha\sqrt{n^\alpha}$, $k = n^\alpha$, and $\ell = \sqrt{n^\alpha}$, we obtain $\deg_\pm(F) = \Omega(n^\alpha\sqrt{n^\alpha})$. □

As a corollary, we now obtain a lower bound of $\Omega(\sqrt{N})$ on the threshold degree of an $\wedge,\vee$-circuit $F\colon \{0,1\}^N \to \{0,1\}$ of constant depth and polynomial size. This lower bound is the main result of our paper, stated earlier as Theorem 1.2.

Theorem 9.4. *Consider the function $F\colon \{0,1\}^N \to \{0,1\}$ on $N = \Theta(n\log n)^2$ variables given by*

$$F = \mathrm{OR}_{(n\log n)^{1/3}} \circ ((\mathrm{AND}_{(n\log n)^{2/3}} \circ \neg(\mathrm{ED}_{n,n} \circ \phi)) \wedge (\mathrm{AND}_{n\log n} \circ \mathrm{OR}_{(n\log n)^{2/3}})),$$

where $\phi\colon \{0,1\}^{6\lceil\log n\rceil} \to \{e_1, e_2, \ldots, e_n\}$ is as constructed in Theorem 3.2. Then

$$\deg_\pm(F) = \Omega(\sqrt{N}). \tag{9.14}$$

Moreover, F is computable by an $\wedge,\vee$-circuit of depth 4 and polynomial size.

Proof. Recall from Theorem 3.3 that the composition $\mathrm{ED}_{n,n} \circ \phi$ on $6n\lceil \log n \rceil$ variables has one-sided approximate degree $\deg^+_{1/3}(\mathrm{ED}_{n,n} \circ \phi) = \Omega(n \log n)^{2/3}$. As a result, (9.14) follows directly from Theorem 9.3. Theorem 3.3 further states that $\mathrm{ED}_{n,n} \circ \phi$ is computable by a CNF formula of polynomial size, which settles the claim regarding the circuit complexity of F. ◻

The tight lower bound in Theorem 9.4 crucially depends on the newly developed gadget ϕ from section 3. For completeness, we also include a simpler version that only uses the folklore gadget, achieving a logarithmically weaker lower bound.

THEOREM 9.5. *Consider the function $F\colon \{0,1\}^N \to \{0,1\}$ on $N = \Theta(n^2 \log n)$ variables given by*

$$F = \mathrm{OR}_{n^{1/3}} \circ ((\mathrm{AND}_{n^{2/3}} \circ \neg f) \wedge (\mathrm{AND}_n \circ \mathrm{OR}_{n^{2/3}})),$$

where $f\colon (\{0,1\}^{\lceil \log n \rceil})^n \to \{0,1\}$ is defined by

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

Then

$$\deg_\pm(F) = \Omega\left(\sqrt{\frac{N}{\log N}}\right). \tag{9.15}$$

Moreover, F is computable by an $\wedge, \vee$-circuit of depth 4 and polynomial size.

Proof. Without loss of generality, we may assume that n is a power of 2. Observe that $f = \mathrm{ED}_{n,n} \circ \iota$, where $\iota\colon \{0,1\}^{\log n} \to \{e_1, e_2, \ldots, e_n\}$ is the lexicographic bijection referred to as the "folklore gadget" in section 3. Bun and Thaler [13] show that this composition has one-sided approximate degree $\deg^+_{1/3}(f) = \Omega(n^{2/3})$. As a result, (9.15) follows directly from Theorem 9.3. Since f is a polynomial-size CNF formula, the claim regarding the circuit complexity of F is immediate. ◻

Remark 9.6. O'Donnell and Servedio [25] proved the equality $\deg_\pm(F \circ \mathrm{XOR}_k) = k \deg_\pm(F)$ for every Boolean function F. As a result, our lower bounds on the threshold degree of AC^0 can be strengthened by an arbitrary polylogarithmic factor by composing the functions in Theorems 9.4 and 9.5 with $\mathrm{XOR}_{\log^{O(1)} N}$.

9.3. Tightness for degree-2 amplifiers. In Theorem 9.3 on hardness amplification, the lower bound on the threshold degree of the transformed function $F\colon \{0,1\}^N \to \{0,1\}$ never exceeds $\Omega(\sqrt{N})$, no matter how large the one-sided approximate degree of the original function $f\colon \{0,1\}^n \to \{0,1\}$. We now show that this square root barrier is inherent rather than an artifact of our analysis. Along the way, we will prove that the lower bound in our main result, Theorem 9.4, is tight up to a logarithmic factor.

THEOREM 9.7. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be given. Then for all k and ℓ, and all depth-2 read-once formulas $g\colon \{0,1\}^{kn} \to \{0,1\}$, the composition*

$$F = \mathrm{OR}_\ell \circ ((\mathrm{AND}_k \circ \neg f) \wedge g)$$

on $N = 2\ell kn$ variables obeys

$$\deg_\pm(F) \leqslant 3\sqrt{2N \deg^+_0(\neg f)}.$$

To see the relevance of this result to our work, observe that $\deg_0^+(\neg\mathrm{ED}_{n,n}) \leqslant 2$ and therefore $\deg_0^+(\neg\mathrm{ED}_{n,n} \circ \phi) = O(\log n)$ in Theorem 9.4. In particular, Theorem 9.7 shows that the threshold degree lower bound in our main result (Theorem 9.4) is tight up to a factor of $O(\sqrt{\log N})$ and cannot be improved by adjusting the fan-ins or using a different depth-2 amplifier g. More generally, Theorem 9.7 shows that the square root barrier in our hardness amplification technique (Theorem 9.3) is inherent due to the possibility of a large gap between the one-sided approximate degree of f and that of $\neg f$.

Proof of Theorem 9.7. Define $d = \deg_0^+(\neg f)$ and fix a polynomial $p\colon \{0,1\}^n \to \mathbb{R}$ of degree d that vanishes on $f^{-1}(1)$ and ranges in $[1,+\infty)$ on $f^{-1}(0)$. For every $\epsilon > 0$ and $t > 0$, Lemma 4.1 gives a one-sided rational approximant $R_{\epsilon,t}$ for g with a positive denominator of degree at most t, a nonnegative numerator of degree at most kn/t, and error ϵ. Then for $\delta > 0$ small enough, the rational functions

$$\frac{\delta}{\delta + \sum_{i=1}^k f(x_i)} \cdot R_{\epsilon,t}(y), \qquad \prod_{i=1}^k p(x_i) \cdot R_{\epsilon,t}(y)$$

are one-sided approximants for $\neg f(x_1) \wedge \cdots \wedge \neg f(x_k) \wedge g(y)$ with error ϵ and $\epsilon\|p\|_\infty^k$, respectively. Passing to the limit as $\epsilon \searrow 0$, we conclude that

$$R\left((\mathrm{AND}_k \circ \neg f) \wedge g, n+t, \frac{kn}{t}\right) = 0,$$
$$R\left((\mathrm{AND}_k \circ \neg f) \wedge g, t, \frac{kn}{t} + kd\right) = 0,$$

where $t > 0$ is arbitrary. Theorem 4.4 now gives

$$\deg_\pm(F) \leqslant 2 \min_{t>0} \left\{ \frac{kn}{t} + \ell(n+t) \right\} \leqslant 4\sqrt{\ell kn} + 2\ell n$$

and

$$\deg_\pm(F) \leqslant 2 \min_{t>0} \left\{ \frac{kn}{t} + kd + \ell t \right\} \leqslant 4\sqrt{\ell kn} + 2kd.$$

Taking the minimum of these two upper bounds on the threshold degree of F completes the proof:

$$\begin{aligned}\deg_\pm(F) &\leqslant 4\sqrt{\ell kn} + 2\min\{\ell n, kd\}\\ &\leqslant 4\sqrt{\ell kn} + 2\sqrt{\ell knd}\\ &\leqslant 6\sqrt{\ell knd}.\end{aligned}$$

□

9.4. Tightness for arbitrary amplifiers. In this final section, we explore the limitations of Theorem 9.1 as a technique for hardness amplification and propose directions for future work. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a given Boolean function, with $\deg_{1/3}^+(f) \geqslant n^\alpha$. The theorem is concerned with the composition

$$F = \mathrm{OR}_\ell \circ ((\mathrm{AND}_k \circ \neg f) \wedge g)$$

for a suitably chosen Boolean function g and integer parameters ℓ and k. This composition, when viewed as a Boolean function $F\colon \{0,1\}^N \to \{0,1\}$, is defined on $N \geqslant \ell kn$

variables. It is clear from the statement of Theorem 9.1 that it cannot give a threshold degree lower bound for F better than $\Omega(\min\{\ell k, \ell n^\alpha, \sqrt{k}n^\alpha\})$. Passing to a judiciously chosen geometric mean,

$$\begin{aligned}\min\{\ell k, \ell n^\alpha, \sqrt{k}n^\alpha\} &\leqslant \begin{cases} (\ell n^\alpha)^{\frac{1}{3}}(\sqrt{k}n^\alpha)^{\frac{2}{3}} & \text{if } \alpha < 1/3,\\ (\ell k)^{\frac{3\alpha-1}{3\alpha+2}}(\ell n^\alpha)^{\frac{1}{3\alpha+2}}(\sqrt{k}n^\alpha)^{\frac{2}{3\alpha+2}} & \text{otherwise}\end{cases}\\ &\leqslant \max\{(\ell k n)^{\frac{1}{3}}, (\ell k n)^{\frac{3\alpha}{3\alpha+2}}\}\\ &\leqslant \max\{N^{\frac{1}{3}}, N^{\frac{3\alpha}{3\alpha+2}}\}.\end{aligned}$$

Thus, Theorem 9.1 by itself cannot give a threshold degree lower bound asymptotically superior to

$$\max\{N^{1/3}, N^{3\alpha/(3\alpha+2)}\} \tag{9.16}$$

for any composition $F\colon \{0,1\}^N \to \{0,1\}$.

Recall that Theorem 9.3 in this paper actually achieves (9.16) for any $0 \leqslant \alpha \leqslant 2/3$, with g taken to be a suitable read-once formula of depth 2. We are confident that it is possible to achieve (9.16) for $\alpha > 2/3$ as well by using read-once formulas g of somewhat larger depth—in fact, depth 3 may well suffice. In particular, we believe that the approach of this paper paves the way to lower bounds as large as $\Omega(N^{3/5})$ on the threshold degree of constant-depth $\wedge,\vee$-circuits $F\colon \{0,1\}^N \to \{0,1\}$, provided of course that strong enough lower bounds for one-sided polynomial approximation are discovered soon.

Apart from matching the hardness amplification in (9.16) for all α, it is natural to wonder how to go *beyond* it. In other words, given a constant-depth polynomial-size $\wedge,\vee$-circuit $f\colon \{0,1\}^n \to \{0,1\}$ with one-sided approximate degree n^α, we would like to construct a related constant-depth polynomial-size $\wedge,\vee$-circuit $F\colon \{0,1\}^N \to \{0,1\}$ with threshold degree $\Omega(N^\beta)$ for some $\beta > 3\alpha/(3\alpha+2)$. We are optimistic on this front as well and believe that the ideas of this paper provide a good starting point. Specifically, a promising construction is to take $F = \mathrm{OR}_\ell \circ ((h \circ \neg f) \wedge g)$ for some parameter ℓ and some read-once formulas h and g of constant depth. In this paper, we have only instantiated this approach for h and g of depth 1 and 2, respectively. Higher-depth constructions will likely give stronger results.

Acknowledgments. The author is thankful to Mark Bun, Adam Klivans, Ryan O'Donnell, Rocco Servedio, Justin Thaler, and the anonymous reviewers for their valuable feedback on an earlier version of this manuscript. I am additionally thankful to Justin for sending me an early preprint of [39].

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, https://doi.org/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, https://doi.org/10.4086/toc.2005.v001a003.

[3] A. Ambainis, *Quantum walk algorithm for element distinctness*, SIAM J. Comput., 37 (2007), pp. 210–239, https://doi.org/10.1137/S0097539705447311.

[4] 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, https://doi.org/10.1137/080712167.

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

[6] 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, https://doi.org/10.1109/SFCS.1986.15.

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

[8] P. Beame and W. Machmouchi, *The quantum query complexity of* AC^0, Quantum Information & Computation, 12 (2012), pp. 670–676.

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

[10] H. Buhrman, R. Cleve, R. de Wolf, and C. Zalka, *Bounds for small-error and zero-error quantum algorithms*, in *Proceedings of the Fortieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 1999, pp. 358–368, https://doi.org/10.1109/SFFCS.1999.814607.

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

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

[17] 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, https://doi.org/10.1016/j.jcss.2003.11.002.

[18] 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, https://doi.org/10.1016/j.jcss.2003.07.007.

[19] 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.

[20] 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, https://doi.org/10.1016/S0304-3975(96)00019-9.

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

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

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

[24] R. O'Donnell and R. A. Servedio, *Extremal properties of polynomial threshold functions*, J. Comput. Syst. Sci., 74 (2008), pp. 298–312, https://doi.org/10.1016/j.jcss.2007.06.021.

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

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

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

[28] M. E. Saks, *Slicing the hypercube*, Surveys in Combinatorics, (1993), pp. 211–255, https://doi.org/10.1017/CBO9780511662089.009.

[29] A. A. Sherstov, *Separating* AC^0 *from depth-2 majority circuits*, SIAM J. Comput., 38 (2009), pp. 2113–2129, https://doi.org/10.1137/08071421X. Preliminary version in *Proceedings of*

the Thirty-Ninth Annual ACM Symposium on Theory of Computing (STOC), 2007.

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

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

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

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

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

[35] A. A. Sherstov, *Breaking the Minsky–Papert barrier for constant-depth circuits*, in *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* (STOC), 2014, pp. 223–232, https://doi.org/10.1145/2591796.2591871. Full version available as ECCC Report TR14-009, January 2014.

[36] A. A. Sherstov, *Communication lower bounds using directional derivatives*, J. ACM, 61 (2014), pp. 1–71, https://doi.org/10.1145/2629334. Preliminary version in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing* (STOC), 2013.

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

[38] 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, https://doi.org/10.1109/18.312168.

[39] J. Thaler, *Lower bounds for the approximate degree of block-composed functions*, in *Proceedings of the Forty-Third International Colloquium on Automata, Languages and Programming* (ICALP), 2016, pp. 17:1–17:15, https://doi.org/10.4230/LIPIcs.ICALP.2016.17.

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

[41] R. de Wolf, *A note on quantum algorithms and the minimal degree of ϵ-error polynomials for symmetric functions*, Quantum Information and Computation, 8 (2008), pp. 943–950.

Appendix A. Useful bounds. In this appendix, we collect various bounds that are useful in the construction of dual objects for the OR function (Theorems 2.13 and 2.14). We start with two facts that involve factorials.

Fact A.1. *For any integers $n \geqslant k \geqslant 0$,*

$$k!\,(n-k)! \geqslant \left\lfloor \frac{n}{2} \right\rfloor ! \left\lceil \frac{n}{2} \right\rceil !.$$

Proof. Immediate from the inequality $\binom{n}{k} \leqslant \binom{n}{\lfloor n/2 \rfloor}$, after dividing through by $n!$. ◻

Fact A.2. *There are constants $c_1 > 0$ and $c_2 > 0$ such that for every integer $n \geqslant 1$,*

$$c_1\sqrt{n}\left(\frac{n}{\mathrm{e}}\right)^n \leqslant n! \leqslant c_2\sqrt{n}\left(\frac{n}{\mathrm{e}}\right)^n,$$
$$\frac{c_1 4^n}{\sqrt{n}} \leqslant \binom{2n}{n} \leqslant \frac{c_2 4^n}{\sqrt{n}}.$$

Proof. Immediate from Stirling's factorial approximation. ◻

We move on to somewhat specialized bounds that pertain to products of differences of squares.

Fact A.3. *For any integer $i \geqslant 1$,*

$$\prod_{\substack{j=1\\ j\neq i}}^{\infty}\left(1-\frac{1}{|i^2-j^2|}\right) \geqslant \frac{4\sin(\sqrt{5}\pi)}{\sqrt{5}\pi} = 0.3846\ldots.$$

This lower bound is tight, as one can verify by setting $i=2$.

Proof. To restate the claim, we are interested in the minimum value of $f(i)g(i)$ over $i=1,2,3,\ldots$, where the functions $f,g\colon \{1,2,3,\ldots\}\to\mathbb{R}$ are given by

$$\begin{aligned} f(i) &= \prod_{j=1}^{i-1}\left(1-\frac{1}{i^2-j^2}\right),\\ g(i) &= \prod_{j=i+1}^{\infty}\left(1-\frac{1}{j^2-i^2}\right). \end{aligned}$$

The first function satisfies

$$\begin{aligned} &f(1)=1,\\ &\min_{i\geqslant 2} f(i) \geqslant \min_{i\geqslant 2}\left(1-\frac{1}{i^2-(i-1)^2}\right)\left(1-\frac{i-2}{i^2-(i-2)^2}\right) = \frac{2}{3}. \end{aligned}$$

A factor-by-factor comparison shows that g is monotonically increasing, with the first two values

$$\begin{aligned} g(1) &= \frac{2\sin(-\sqrt{2}\pi)}{\sqrt{2}\pi},\\ g(2) &= \frac{6\sin(\sqrt{5}\pi)}{\sqrt{5}\pi}. \end{aligned}$$

As a result,

$$\min_{i\geqslant 1} f(i)g(i) \geqslant \min\left\{f(1)g(1), \frac{2}{3}g(2)\right\} = \frac{4\sin(\sqrt{5}\pi)}{\sqrt{5}\pi}. \qquad \square$$

Fact A.4. *For any integers $d \geqslant i \geqslant 1$,*

$$\prod_{\substack{j=1\\ j\neq i}}^{d} |i^2-j^2| \geqslant \frac{d!\,d!}{2i^2}.$$

Proof. We have

$$\begin{aligned} \prod_{\substack{j=1\\ j\neq i}}^{d} |i^2-j^2| &= \prod_{\substack{j=1\\ j\neq i}}^{d} |i-j| \cdot \prod_{\substack{j=1\\ j\neq i}}^{d} (i+j)\\ &= (i-1)!\,(d-i)! \cdot \frac{(d+i)!}{i!\,2i}\\ &= \frac{(d-i)!\,(d+i)!}{2i^2}\\ &\geqslant \frac{d!\,d!}{2i^2}, \end{aligned}$$

where the final step uses Fact A.1. $\square$

Appendix B. Constant-error approximation of OR. The purpose of this appendix is to prove Theorem 2.13, which gives a bounded-error dual object for the OR function with a number of additional properties. Our construction is a minor modification of the corresponding dual object in [35], which has almost all of the properties that we need. We start with a technical lemma from that work [35, Lemma A.2].

LEMMA B.1. *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*

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

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

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

We have reached the main result of this section.

THEOREM (restatement of Theorem 2.13). *Let ϵ be given, $0 < \epsilon < 1$. Then for every $n \geqslant 2$ and every probability distribution κ on $\{1, 2, \ldots, n\}$, there is an (explicitly given) function $\omega\colon \{0, 1, 2, \ldots, n\} \to \mathbb{R}$ such that*

$$\omega(0) > \frac{1-\epsilon}{2} \cdot \|\omega\|_1, \tag{B.4}$$

$$(-1)^{n+t}\omega(t) \geqslant \frac{\epsilon\kappa(t)}{3} \cdot \|\omega\|_1 \qquad (t = 1, 2, \ldots, n), \tag{B.5}$$

$$\deg p < \sqrt{\delta n} \implies \langle \omega, p\rangle = 0, \tag{B.6}$$

where $\delta = \delta(\epsilon) > 0$ is a constant independent of κ and n.

Our proof closely follows [35]. Specifically, we obtain the desired sign behavior and metric properties by defining ω as the convex combination of several shifted copies of the dual object in Lemma B.1.

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

$$\omega\colon (0, 1, 2) \mapsto \left(\frac{1}{2} - \frac{\epsilon}{3}, -\frac{1}{2}, \frac{\epsilon}{3}\right),$$
$$\omega\colon (0, 1, 2, 3) \mapsto \left(\frac{1}{2} - \frac{\epsilon}{3}, \frac{\epsilon\kappa(1)}{3}, -\frac{1}{2}, \frac{\epsilon(1-\kappa(1))}{3}\right),$$

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

For some $\delta = \delta(\epsilon) > 0$ and all $n \geqslant 4$, Lemma B.1 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

$$\|\omega_0\|_1 = \|\omega_1\|_1 = 1, \tag{B.7}$$

$$\omega_0(0) > \frac{1}{2} - \frac{\epsilon}{6}, \tag{B.8}$$

$$\omega_1(0) > \frac{1}{2} - \frac{\epsilon}{6}, \tag{B.9}$$

$$(-1)^t\omega_0(t) \geqslant 0, \qquad t \geqslant 0, \tag{B.10}$$

$$(-1)^{n+t}\omega_1(t) \geqslant 0, \qquad t \geqslant 1, \tag{B.11}$$

$$\deg p < \sqrt{\delta n} \implies \langle \omega_0, p\rangle = \langle \omega_1, p\rangle = 0. \tag{B.12}$$

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

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

where

$$\rho = \frac{2}{3}\cdot\frac{\epsilon}{1-\epsilon}.$$

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

$$\begin{aligned} \|\omega\|_1 &\leqslant \|\omega_1\|_1 + \rho\sum_{i=1}^{n}\kappa(i)\|\omega_0\|_1 \\ &= 1+\rho \\ &= \frac{3-\epsilon}{3(1-\epsilon)}, \end{aligned} \tag{B.13}$$

where the second step uses (B.7). Now (B.4) is immediate because $\omega(0) = \omega_1(0) > (3-\epsilon)/6$ by (B.9).

Property (B.5) for $t \geqslant 1$ can be verified as follows:

$$\begin{aligned} (-1)^{n+t}\omega(t) &= |\omega_1(t)| + \rho\sum_{i=1}^{\lfloor n/2\rfloor}\kappa(i)|\omega_0(t-i)| + \rho\sum_{i=\lfloor n/2\rfloor+1}^{n}\kappa(i)|\omega_0(-t+i)| \\ &\geqslant \rho\cdot\kappa(t)\,|\omega_0(0)| \\ &\geqslant \rho\cdot\kappa(t)\cdot\frac{3-\epsilon}{6} \\ &\geqslant \frac{\epsilon\kappa(t)}{3}\cdot\|\omega\|_1, \end{aligned}$$

where the first step follows from (B.10) and (B.11), the third from (B.8), and the fourth from (B.13).

The remaining property (B.6) is immediate from (B.12). □

Appendix C. High-accuracy approximation of OR. In Appendix B, we constructed a dual object for the bounded-error approximation of OR. Here, we obtain its counterpart for *high-accuracy* approximation. Analogous to the bounded-error case, the new dual object is tailored to the needs of this paper and has special metric properties and sign behavior. Our construction is a modification of an earlier result due to Bun and Thaler [12], who studied the bounded-error approximation of arbitrary symmetric functions. A glance at the dual object in that paper reveals that it doubles as a high-accuracy dual object for OR, and we need only adapt it somewhat to ensure the additional properties that we need. Overall, the analysis below seems less demanding than in [12] because we do not need to ensure the large inner products that the bounded-error case requires.

THEOREM (restatement of Theorem 2.14). *Let $0 < c < 1$ be a sufficiently small absolute constant. Then for all integers n and r with $1 \leqslant r \leqslant n/2$, there exists a* *function* $\nu\colon \{0,1,2,\ldots,n\} \to \mathbb{R}$ *such that*

$$\begin{aligned}
&\deg p \leqslant c\sqrt{nr} \implies \langle \nu, p\rangle = 0, && \\
&\nu(t) = 0 && (t = 1,2,\ldots,r-1),\\
&(-1)^{n+t}\nu(t) \geqslant 0 && (t = 1,2,\ldots,n),\\
&\nu(0) > c^r \|\nu\|_1, && \\
&|\nu(t)| \geqslant c^n \|\nu\|_1 && (t > n/2),\\
&|\nu(t)| \leqslant \left(\frac{r}{ct}\right)^r \|\nu\|_1 && (t = 1,2,\ldots,n).
\end{aligned}$$

Proof. We first consider the case $\lfloor n/5\rfloor \leqslant r \leqslant n/2$. Define $\nu'\colon \{0,1,2,\ldots,n\} \to \mathbb{R}$ by

$$\nu'(t) = \frac{(-1)^t}{m!}\binom{n}{t}\prod_{i=1}^{m}(i-t),$$

where $m \in \{\lfloor n/2\rfloor - 1, \lfloor n/2\rfloor\}$ is chosen such that $m \equiv n \pmod 2$. Then

$$\deg p < n/2 \implies \langle \nu', p\rangle = 0, \tag{C.1}$$

$$\nu'(t) = 0 \qquad (t = 1,2,\ldots,\lfloor n/2\rfloor - 1), \tag{C.2}$$

$$(-1)^{n+t}\nu'(t) \geqslant 0 \qquad (t = 1,2,\ldots,n), \tag{C.3}$$

where the first assertion follows from Fact 2.1(i), and the other two are immediate. Continuing, we have

$$\begin{aligned}
&\nu'(0) = 1, && \\
&|\nu'(t)| = \binom{n}{t}\binom{t-1}{m} \geqslant 1 && (t = m+1, m+2, \ldots, n),\\
&\|\nu'\|_1 \leqslant 4^n, &&
\end{aligned}$$

where the first two bounds are trivial, and the third follows from the observation that $|\nu'(t)| \leqslant 2^n\binom{n}{t}$. As a result,

$$\nu'(0) \geqslant 4^{-n}\|\nu'\|_1, \tag{C.4}$$

$$|\nu'(t)| \geqslant 4^{-n}\|\nu'\|_1 \qquad (t > n/2). \tag{C.5}$$

Finally, the assumption that $r \geqslant \max\{\lfloor n/5\rfloor, 1\}$ trivially implies that

$$\begin{aligned}
|\nu'(t)| &\leqslant \|\nu'\|_1\\
&\leqslant \left(\frac{10r}{t}\right)^r \|\nu'\|_1 \qquad (t = 1,2,\ldots,n).
\end{aligned} \tag{C.6}$$

In view of (C.1)–(C.6), the theorem holds in this case for $\nu = \nu'$ and small enough $c > 0$.

We now examine the complementary case $1 \leqslant r < \lfloor n/5\rfloor$. The construction here uses the function ν' defined above as well as an additional function ν'', to be introduced shortly. Let

$$\begin{aligned} R &= \begin{cases} r & \text{if } r \text{ is odd,} \\ r+1 & \text{otherwise,} \end{cases} \\ d &= \left\lfloor \sqrt{\frac{n}{R} - 1} \right\rfloor, \\ S_1 &= \{R + (n \bmod 2) + j : j = 0, 1, 2, \ldots, 3R-1\}, \\ S_i &= \{i^2R + (n \bmod 2) + j : j = 0, 1, 2, \ldots, R-1\} \qquad (i = 2, 3, \ldots, d). \end{aligned}$$

Note that the sets $S_1, S_2, \ldots, S_d$ are pairwise disjoint. Define

$$S = \{0\} \cup S_1 \cup S_2 \cup \cdots \cup S_d,$$

so that $S \subseteq \{0, 1, 2, \ldots, n\}$. Define $\nu'' \colon \{0, 1, 2, \ldots, n\} \to \mathbb{R}$ by

$$\nu''(t) = \frac{(-1)^t}{n!} \binom{n}{t} \prod_{\substack{i=0,1,2,\ldots,n: \\ i \notin S}} (i - t).$$

It follows from Fact 2.1(i) that ν'' is orthogonal to every polynomial of degree less than $n - (n + 1 - |S|) = R(d + 2)$. Thus,

$$\deg p < \sqrt{rn} \implies \langle \nu'', p \rangle = 0. \tag{C.7}$$

A routine calculation reveals that

$$\nu''(t) = \begin{cases} (-1)^{|\{i \in S : i < t\}|} \displaystyle\prod_{i \in S \setminus \{t\}} \frac{1}{|t - i|} & \text{if } t \in S, \\ 0 & \text{otherwise.} \end{cases} \tag{C.8}$$

In particular,

$$\nu''(t) = 0 \qquad (t = 1, 2, \ldots, r-1), \tag{C.9}$$

$$(-1)^{n+t} \nu''(t) \geqslant 0 \qquad (t = 1, 2, \ldots, n). \tag{C.10}$$

We now proceed to examine metric properties of ν'':

$$\begin{aligned} \frac{1}{\nu''(0)} &\leqslant (2R+1)(2R+2) \cdots (4R) \prod_{i=1}^{d} \prod_{j=1}^{R} (i^2R + j) \\ &= \frac{(4R)!}{(2R)!} (d!\, d!\, R^d)^R \prod_{i=1}^{d} \prod_{j=1}^{R} \left(1 + \frac{j}{i^2R}\right) \\ &< \frac{(4R)!}{(2R)!} (d!\, d!\, R^d)^R \exp\left(\sum_{i=1}^{\infty} \sum_{j=1}^{R} \frac{j}{i^2R} \right) \\ &= \frac{(4R)!}{(2R)!} (d!\, d!\, R^d)^R \exp\left(\frac{\pi^2(R+1)}{12} \right) \\ &\leqslant \left(c_1 \cdot d!\, d!\, R^{d+2}\right)^R \end{aligned} \tag{C.11}$$

for some constant $c_1 > 0$, where the final step uses Fact A.2. For $t \in S_1 \cup S_2$, we have

$$\begin{aligned}
\frac{1}{|\nu''(t)|} &= t \cdot (t - R - (n \bmod 2))!\,(5R + (n \bmod 2) - 1 - t)! \\
&\qquad\qquad \times \prod_{i=3}^{d} \prod_{j=0}^{R-1} (i^2 R + (n \bmod 2) + j - t) \\
&\geqslant (2R)!\,(2R)! \prod_{i=3}^{d} (i^2 R - 5R)^R \\
&\geqslant (2R)!\,(2R)! \left(\frac{d!\,d!\,R^{d-2}}{4} \right)^R \prod_{i=3}^{\infty} \left(1 - \frac{5}{i^2} \right)^R \\
&\geqslant \left(c_2 \cdot d!\,d!\,R^{d+2} \right)^R
\end{aligned} \tag{C.12}$$

for some constant $c_2 > 0$, where the second and fourth steps use Fact A.1 and Fact A.2, respectively. Finally, for $t \in S_{i'}$ with $i' \geqslant 3$,

$$\begin{aligned}
\frac{1}{|\nu''(t)|} &= t \cdot \binom{t - 2R - (n \bmod 2)}{2R} \cdot (2R)!\,(t - i'^2 R - (n \bmod 2))! \\
&\qquad \times (i'^2 R + (n \bmod 2) + R - 1 - t)! \\
&\qquad \times \prod_{\substack{i=1 \\ i \neq i'}}^{d} \prod_{j=0}^{R-1} |i^2 R + (n \bmod 2) + j - t| \\
&\geqslant \binom{t - 2R - (n \bmod 2)}{2R} \cdot (2R)! \left\lfloor \frac{R}{2} \right\rfloor ! \left\lceil \frac{R}{2} \right\rceil ! \prod_{\substack{i=1 \\ i \neq i'}}^{d} \prod_{j=0}^{R-1} (|i^2 - i'^2| - 1) R \\
&\geqslant \binom{t - 2R - (n \bmod 2)}{2R} \cdot (2R)! \left\lfloor \frac{R}{2} \right\rfloor ! \left\lceil \frac{R}{2} \right\rceil ! \\
&\qquad \times R^{R(d-1)} \prod_{\substack{i=1 \\ i \neq i'}}^{d} |i^2 - i'^2|^R \cdot \prod_{\substack{i=1 \\ i \neq i'}}^{\infty} \left(1 - \frac{1}{|i^2 - i'^2|} \right)^R \\
&\geqslant \binom{t - 2R - (n \bmod 2)}{2R} \cdot (2R)! \left\lfloor \frac{R}{2} \right\rfloor ! \left\lceil \frac{R}{2} \right\rceil ! \, R^{R(d-1)} \left(\frac{d!d!}{2i'^2} \right)^R \cdot \frac{1}{3^R} \\
&\geqslant \left(\frac{t}{R} \right)^R \left(c_3 \cdot d!\,d!\,R^{d+2} \right)^R
\end{aligned} \tag{C.13}$$

for some constant $c_3 > 0$, where the second step uses Fact A.1, and the final two steps use Facts A.2 to A.4. As a result,

$$\begin{aligned}
\|\nu''\|_1 - |\nu''(0)| &= \sum_{t \in S_1 \cup S_2} |\nu''(t)| + \sum_{i=3}^{d} \sum_{t \in S_i} |\nu''(t)| \\
&\leqslant \frac{4R}{(c_2 \cdot d!\,d!\,R^{d+2})^R} + \frac{R}{(c_3 \cdot d!\,d!\,R^{d+2})^R} \sum_{i=3}^{\infty} \left(\frac{1}{i^2} \right)^R \\
&\leqslant \frac{5R}{(\min\{c_2, c_3\} \cdot d!\,d!\,R^{d+2})^R},
\end{aligned} \tag{C.14}$$

where the second step uses the estimates in (C.12) and (C.13). Now the bounds

$$\nu''(0) > c_4^r \|\nu''\|_1, \tag{C.15}$$

$$|\nu''(t)| \leqslant \left(\frac{r}{c_4 t}\right)^r \|\nu''\|_1 \qquad (t = 1, 2, \ldots, n) \tag{C.16}$$

are immediate from (C.8) and (C.12)–(C.14), where $c_4 > 0$ is a small enough constant.

We now claim that the theorem holds for

$$\nu = \frac{1}{2^n \|\nu'\|_1} \nu' + \frac{1}{\|\nu''\|_1} \nu''.$$

To start with, properties (C.1)–(C.3) of ν' and properties (C.7)–(C.10) of ν'' directly imply

$$\deg p < \sqrt{nr/2} \implies \langle \nu, p \rangle = 0, \tag{C.17}$$

$$\nu(t) = 0 \qquad (t = 1, 2, \ldots, r-1), \tag{C.18}$$

$$(-1)^{n+t} \nu(t) \geqslant 0 \qquad (t = 1, 2, \ldots, n). \tag{C.19}$$

By (C.4) and (C.15),

$$\nu(0) > c^r \|\nu\|_1 \tag{C.20}$$

for a small enough constant $c > 0$. Similarly, properties (C.3) and (C.5) of ν' and property (C.10) of ν'' give

$$|\nu(t)| \geqslant c^n \|\nu\|_1 \qquad (t > n/2) \tag{C.21}$$

for a small enough constant $c > 0$. Finally, (C.16) forces

$$|\nu(t)| \leqslant \left(\frac{r}{ct}\right)^r \|\nu\|_1 \qquad (t = 1, 2, \ldots, n), \tag{C.22}$$

again for $c > 0$ small enough. By (C.17)–(C.22), the proof is complete. □